Focusing Techniques

Murali Subbarao Tae Choi Arman Nikzad
Dept. of Electrical Engg., State University of New York, Stony Brook, New York 11794-2350

Abstract

We use the paraxial geometric optics model of image formation to derive a set of camera focusing techniques. These techniques do not require calibration of cameras but involve a search of the camera parameter space. The techniques are proved to be theoretically sound. They include energy maximization of unfiltered, low-pass filtered, high-pass filtered, and band-pass filtered images. It is shown that in the presence of high spatial frequencies, noise, and aliasing, focusing techniques based on band-pass filters perform well. The focusing techniques are implemented on a a prototype camera system named SPARCS. The architecture of SPARCS is described briefly. The performance of the different techniques are compared experimentally. All techniques are found to perform well. One of them which has better overall characteristics is recommended for practical applications.

1 Introduction

Focusing cameras is an important problem in computer vision and microscopy. In this paper we consider only those passive focusing techniques which do not require calibration of the cameras. These techniques necessarily involve a search of the camera parameter space. Therefore we will call these *search focusing techniques*. The camera parameters include the distance between the lens and the image detector, and the focal length of the lens. There are some focusing methods which do not search the camera parameter space [8, 16, 15], but they require accurate calibration of the camera parameters and the corresponding optical transfer function or the point spread function. We first consider the case of focusing a camera to a target object by moving the lens along the optical axis of the lens. The results of this case can be easily extended to focusing a camera by adjusting its focal length, or by moving the target object along the optical axis.

Many focusing techniques have been investigated and compared in the literature (e.g. Krotkov 1987, Ligthart and Groen 1982, Schlag et al). In these techniques, a focus measure is defined which is a maximum for the best focused image and it generally decreases as the defocus increases. Therefore, when the lens of a camera is moved from one end to the other, typically, the focus measure of the image of the target object gradually increases, reaches a maximum at the focused lens position, and then decreases gradually thereafter. The problem then is to find the lens position at which the focus measure is a maximum. This is essentially a search of the lens position space to find the maximum of focus measure.

In the previous literature, most definitions of focus measure have been based on heuristics. No proof has been provided about the theoretical soundness of these focus measures, i.e. it has not been proved theoretically that the focus measures have a global maximum for the best focused images. Some definitions of focus measures involve non-linear operators and it is hard to prove or disprove their correctness. Some other definitions are useful only for a very limited class of images. We have not found any systematic and explicit analysis of the image defocusing phenomenon and derivation of a focus measure based on such an analysis. Further, previous literature has ignored an important aspect we call the "side-lobe effect" (explained later) which may give rise to local maxima and cause the global maximum position to be shifted. This makes searching for the position of global maximum difficult and introcuces error.

In this paper, first we describe a model of image defocusing derived from paraxial geometric optics (Gaskill 1978, Hecht and Zajac 1979, Born and Wolf 1975). Based on this model we derive a class of focus measures which we prove to be theoretically sound, i.e. the focus measures have global maximum for the best focused images. We then analyze the side-lobe effect and show how it can be reduced through low pass filtering. Next we discuss a few other focusing techniques from previous literature.

The focus measures derived in this paper have been implemented on a prototype camera system named Stonybrook Passive Autofocusing and Ranging Camera System (SPARCS). A brief description of SPARCS architecture is included. A number of experiments were carried out using SPARCS to evaluate the focus measures derived here. The experiments and their results are described. The experimental results show that the focus measures derived here perform well. Based on these results, we find that, one of the focus measures named *energy of low pass filtered image gradient* has the best performance overall. We conclude this paper by recommending it for practical applications.

2 Camera Model

Image formation in a simple camera is shown in Figure 1. We have shown a thin lens model for the optical system, but the analysis here can be easily extended to a thick lens model (Gaskill 1978). Let P be a point on a visible surface in the scene and p' be its focused image. The relation between the positions of P and p' is given by the lens formula,

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

where u is the distance between the lens plane and the object plane and v is the distance between the lens plane and the image plane. In this figure, ID is the image detector (CCD array), D is the aperture diameter, and s is the distance between the lens plane and the image detector. The distance s, focal length f, and aperture diameter D, will be referred together as *camera parameters* and denoted by $\mathbf{e}$, i.e.

$$\mathbf{e} \equiv (s, f, D). \tag{2}$$

In order to illustrate the theoretical basis of focus measures, we take the optical system to be circularly symmetric around the optical axis and we use a paraxial geometric optics model [1] for image formation. This is a good approximation in practice to actual image formation process modeled by physical optics (Subbarao and Lu 1992, Subbarao 90.02.07). However, our analysis itself is applicable to physical optics model also.

3 Point Spread Function

In Figure 1, if the object point P is not in focus, then it gives rise to a blurred image p on the image detector ID. According to geometric optics, the blurred image of P has the same shape as the lens aperture but scaled by a factor. This holds irrespective of the position of P on the object plane. Since we have taken the aperture to be circular, the blurred image of P is also a circle with uniform brightness inside the circle and zero outside. This is called a *blur circle*.

Let the light energy incident on the lens from the point P during one exposure period of the camera be one unit. Then, the blurred image of P is the response of the camera to a unit point source and hence it is the Point Spread Function (PSF) of the camera system. We will denote this PSF by $h_a(x, y)$.

Let R be the radius of the blur circle and q be the scaling factor defined as $q = 2R/D$. In Figure 1, from similar triangles, we have

$$q = \frac{2R}{D} = \frac{s-v}{v} = s\left[\frac{1}{v} - \frac{1}{s}\right] \tag{3}$$

Substituting for $1/v$ from Eq. (1) in the above equation, we obtain

$$q = s\left[\frac{1}{f} - \frac{1}{u} - \frac{1}{s}\right] \tag{4}$$

Therefore

$$R = q\frac{D}{2} = s\frac{D}{2}\left[\frac{1}{f} - \frac{1}{u} - \frac{1}{s}\right] \tag{5}$$

Note that q and therefore R can be either positive or negative depending on whether $s \geq v$ or $s < v$. In the former case the image detector plane is behind the focused image of P and in the latter case it is in front of the focused image of P.

If we assume the camera to be a lossless system (i.e., no light energy is absorbed by the camera system) then

$$\int\int h_a(x, y)\, dx\, dy = 1 \tag{6}$$

because the light energy incident on the lens was taken to be one unit. Using this and the fact that the blur circle has uniform brightness inside a circle of radius R and zero outside, we get

$$h_a(x, y) = \begin{cases} \frac{1}{\pi R^2} & \text{if } x^2 + y^2 \leq R^2 \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

In a practical camera system, if a sequence of images $g_i(x, y)$ are taken at camera parameter settings of $\mathbf{e}_i$ for $i = 1, 2, 3, \cdots$, then image magnification and mean image brightness may change even though nothing has changed in the scene. For example, moving the lens away from the image detector will increase image magnification and changing the aperture diameter changes mean image brightness. In order to define and compare focus measures for this sequence of images in a correct and consistent manner, the sequence of images must be first normalized with respect to these factors. Normalization with respect to image brightness is carried out by dividing the image brightness at every point by the mean brightness of the image.

Normalization with respect to image magnification is more complicated. It can be done by image interpolation and resampling such that all images correspond to the same field of view [12]. The relation between an original image $g(x, y)$ taken with $s = s_0$ and the corresponding magnification normalized image $g_n(x, y)$ is given by $g_n(x/s_0, y/s_0) = g(x, y)$. However, in most practical applications, the magnification change is less than 3% and can be ignored. It is probably for this reason that most previous literature fails to mention the magnification correction. But this cannot be overlooked from a theoretical point of view.

In the following discussion we assume that the images have been normalized. Without loss of generality, we assume that both the mean brightness and magnification have been normalized to be 1. After magnification normalization, the radius of the blur circle can be expressed as a function of the camera parameter setting $\mathbf{e}$ and object distance u as

$$R(\mathbf{e}; u) = \frac{D}{2}\left(\frac{1}{f} - \frac{1}{u} - \frac{1}{s}\right). \tag{8}$$

The Optical Transfer Function (OTF) corresponding to the above PSF (Eq. 7) is

$$H_a(\omega, \nu; \mathbf{e}, u) = 2\frac{J_1(R(\mathbf{e}; u)\, \rho(\omega, \nu))}{R(\mathbf{e}; u)\, \rho(\omega, \nu)} \tag{9}$$

where ω, ν, and ρ are spatial frequencies specified in radians/unit distance, J_1 is the first order Bessel function, and ρ is the radial spatial frequency

$$\rho(\omega, \nu) = \sqrt{\omega^2 + \nu^2}. \tag{10}$$

Eq. (9) explicitly represents the dependence of the OTF on the camera parameter setting $\mathbf{e}$ and the object distance u.

In practice, the image of a point object is not a crisp circular patch of constant brightness as suggested by geometric optics. Instead, due to diffraction, poly-chromatic illumination, lens aberrations, etc., it will be a roughly circular blob with the brightness falling off gradually at the border rather than sharply. Therefore, as an alternative to the above PSF model, often [3, ?, 8, 10] a two-dimensional Gaussian is suggested which is defined by

$$h_b(x,y) \;=\; \frac{1}{2\pi r^2} e^{-\frac{x^2+y^2}{2r^2}} \tag{11}$$

where r is a spread parameter corresponding to the *standard deviation* of the distribution of the PSF. In practice, it is found that [11, 13] r is proportional to R, i.e.

$$r \;=\; c\,R \;\text{ for }\; c > 0 \tag{12}$$

where c is a constant. It is approximately equal to $1/\sqrt{2}$ in practice [13]. Since blur circle radius R is a function of $\mathbf{e}$ and u, r can be written as $r(\mathbf{e}, u)$. (However, the image of an actual point light source for our camera was quite close to that predicted by geometric optics and was far from a Gaussian.)

The OTF corresponding to the above PSF is (ω, ν in radians/unit dist.)

$$H_b(\omega,\nu;\mathbf{e},u) \;=\; e^{-\frac{1}{2}\rho^2(\omega,\nu)\; r^2(\mathbf{e};u)} \tag{13}$$

where

$$r(\mathbf{e};u) \;=\; c\,\frac{D}{2}\left(\frac{1}{f}-\frac{1}{u}-\frac{1}{s}\right). \tag{14}$$

Once again, Eq. (13) explicitly represents the dependence of the OTF on the camera parameter setting $\mathbf{e}$ and the object distance u.

4 Focus Measures

In this section we develop a theoretical basis for focus measures. Let $f(x,y)$ be the focused image of a planar object at distance u. The *focused image* $f(x,y)$ at a point (x,y) of a scene is defined as the total light energy incident on the camera aperture (entrance pupil) during one exposure period from the object point along the direction corresponding to (x,y) (Subbarao and Nikzad, 1990). We do not know of any previous literature on focusing techniques which gives a precise and correct (we believe) definition of the focused image as we have done here. This definition is essential for a sound analysis of the focusing techniques.

Let $g_i(x,y)$ be a sequence of images of the object recorded for a sequence of camera parameter settings $\mathbf{e}_i$ where

$$\mathbf{e_i} \;=\; (s_i, f, D) \qquad \text{or} \qquad \mathbf{e_i} \;=\; (s, f_i, D) \tag{15}$$

for $i = 1, 2, 3, \cdots$. Here we are considering the variation of only one camera parameter at a time– either the lens position or the focal length because this is the usual mode of operation of almost all cameras. Simultaneous variation of more than one parameter in a random manner can lead to multiple maxima for the focus measure.

For a planar object perpendicular to the optical axis, the blur circle radius R is a constant over the image of the object (this may not be obvious at first sight, but it can be proved easily). In this case the camera acts as a linear shift invariant system. Therefore g_i will be equal to the convolution of the focused image $f(x,y)$ with the corresponding point spread function. Convolution in the saptial domain corresponds to multiplication in the Fourier domain. Therfore, if G_i is Fourier transform of g_i, we can write

$$G_i(\omega,\nu) \;=\; H_a(\omega,\nu;\mathbf{e_i},u)\; F(\omega,\nu)\;. \tag{16}$$

Substituting for the right hand side of Eq. (16) from Eq. (9), we obtain

$$G_i(\omega,\nu) \;=\; 2\frac{J_1(R(\mathbf{e}_i;u)\;\rho(\omega,\nu))}{R(\mathbf{e}_i;u)\;\rho(\omega,\nu)}\; F(\omega,\nu)\;. \tag{17}$$

The left hand side of the above equation can be computed from the recorded images g_i. For a Gaussian PSF model, the expression corresponding to Eq. (17) is

$$G_i(\omega,\nu) \;=\; e^{-\frac{1}{2}\rho^2(\omega,\nu)\; r^2(\mathbf{e}_i;u)}\; F(\omega,\nu)\;. \tag{18}$$

Given the above defocusing model, the problem now is to define a focus measure which is a maximum for the best focused image among the sequence of images $g_i(x,y)$ and gradually decreases as the image blur increases. Except for the image sequence g_i which forms the input, no other information is assumed to be known such as the actual values of the camera parameters $\mathbf{e}_i$, or the optical transfer function.

Figure 6 shows cross sections of the circularly symmetric optical transfer functions H_a for various values of the blur circle radius R. These OTFs are sinc-like with a dominant main lobe followed by side lobes on either side. The OTF magnitude at the origin is 1 and decreases monotonically to zero at $\rho \approx 1.22\pi/R$ at the end of the main lobe. The first side lobe peak is approximately at $1.63\pi/R$ and has a value of about 0.132. The amplitude of the side lobes fall off rapidly (as $\rho^{-1.5}$ [3]) with ρ. If we consider the effect of only the dominant main lobe, we see that the OTFs have the general characteristics of a low pass filter. As the blur circle radius increases (i.e. blur increases), the higher frequencies are attenuated more. In the main lobe, the higher a frequency, the more the attenuation. Roughly speaking, the area under the main lobe increases with decreasing defocus. An obvious focus measure suggested by this observation is the volume integral

$$M_0 \;=\; \int\!\!\int |G_i(\omega,\nu)|\,d\omega\, d\nu \tag{19}$$

This focus measure is proved to be sound, monotonic, and unimodal, in Theorem 1 in the appendix assuming the OTF magnitude outside the main lobe to be zero. Theorem 2 proves the same result for the case of a Gaussian

OTF. Further, the maximum value of this focus measure is shown to be

$$\int\int |F(\omega,\nu)| d\omega d\nu \tag{20}$$

when the blur circle radius is zero (i.e. when the observed image is in perfect focus).

This focus measure involves the computation of the Fourier transform of the image which is computationally expensive ($O(N^2 logN)$ for an image of size $N \times N$). Fortunately, we can define other focus measures which can be computed more efficiently. For this purpose, we note that we can use the volume integral of any *function* of $|G_i(\omega,\nu)|$ which increases monotonically with increasing $|G_i(\omega,\nu)|$. This is proved in Theorems 3 and 4 in the appendix. Therefore, literally an infinite number of focus measures can be defined. However, based on computational requirements and noise sensitivity, we select only three measures. Before we proceed to the three measures, we raise the question as to whether a sound focus measure can be defined which is not a monotonic function of $|G_i(\omega,\nu)|$. We believe that any such function would exhibit local maxima for some $|F(\omega,\nu)|$.

5 Three Focus Measures

5.1 Image Energy

$$M_e \;=\; \int\int |G_i(\omega,\nu)|^2 d\omega d\nu \tag{21}$$

This does not require the computation of Fourier transform of the image because it can be computed efficiently using the Parseval's theorem:

$$M_e \;=\; \int\int |g_i(x,y)|^2 dx dy \tag{22}$$

An equivalent measure is the image grey level variance:

$$M_1 \;=\; \frac{1}{A}\int\int (g_i(x,y)-\mu_i)^2 dx dy \tag{23}$$

where μ_i is the mean of g_i and A is the area of image g_i. Note that

$$M_1 \;=\; M_e/A - \mu^2 \tag{24}$$

According to Theorems 1 and 2 in the appendix, M_e and therefore M_1 are monotonic, unimodal, and sound focus measures.

5.2 Energy of image gradient

$$M_2 = \iint |\vec{\bigtriangledown} g_i(x,y)|^2 \; dx dy \tag{25}$$

Note that

$$\mathcal{F}\left\{\frac{\partial g_i(x,y)}{\partial x}\right\} = j\omega G(\omega,\nu) \tag{26}$$

Therefore from Parseval's theorem,

$$\iint \left(\frac{\partial g_i(x,y)}{\partial x}\right)^2 \; dx dy = \iint \omega^2 |G_i(\omega,\nu)|^2 \; d\omega d\nu \tag{27}$$

Similarly,

$$\iint \left(\frac{\partial g_i(x,y)}{\partial y}\right)^2 \; dx dy = \iint \nu^2 |G_i(\omega,\nu)|^2 \; d\omega d\nu \tag{28}$$

Adding the above two relations we get

$$M_2 = \iint \left(\omega^2+\nu^2\right) |G_i(\omega,\nu)|^2 \; d\omega d\nu \tag{29}$$

Therefore we see that M_2 measures the image energy after $G_i(\omega,\nu)$ has been high pass filtered by

$$L_2(\omega,\nu) = \sqrt{\omega^2+\nu^2} = \rho \tag{30}$$

Therefore, according to Theorems 5 and 6 in the appendix, M_2 is a monotonic, unimodal and sound focus measure.

5.3 Energy of image Laplacian

$$M_3 = \iint \left(\bigtriangledown^2 g_i(x,y)\right)^2 \; dx dy \tag{31}$$

Note that

$$\mathcal{F}\left(\frac{\partial^2 g_i(x,y)}{\partial x^2}\right) = -\omega^2 G_i(\omega,\nu) \tag{32}$$

and

$$\mathcal{F}\left(\frac{\partial^2 g_i(x,y)}{\partial y^2}\right) = -\nu^2 G_i(\omega,\nu) \tag{33}$$

$$\Rightarrow \mathcal{F}\left(\bigtriangledown^2 g_i(x,y)\right) = -\left(\omega^2+\nu^2\right) G_i(\omega,\nu) \tag{34}$$

From Parseval's Theorem

$$\iint \left(\bigtriangledown^2 g_i(x,y)\right)^2 \; dx dy = \iint \left(\omega^2+\nu^2\right)^2 |G_i(\omega,\nu)|^2 d\omega d\nu \tag{35}$$

Therefore, M_3 measures the image energy after $G_i(\omega,\nu)$ has been high pass filtered by

$$L_3(\omega,\nu) = \left(\omega^2+\nu^2\right) \;=\; \rho^2 \tag{36}$$

See Figure 3 for a plot of $L_2(\omega,\nu)$ and $L_3(\omega,\nu)$. Therefore according to Theorems 5 and 6 in the appendix, M_3 is a sound, unimodal and monotonic focus measure.

6 Analysis of Side Lobe Effect

In the previous section and the appendix, we ignored the presence of side lobes in proving the monotonicity and the correctness of the focus measures. If the focused image of an object has high energy frequency content in the side lobes, then the focus measures discussed earlier may exhibit local maxima. This complicates the task of searching for the global maximum. In order to illustrate this, consider the following pathological case: an object whose Fourier spectrum is (see Figure 4)

$$|F(\rho)| = \delta\left(\rho - \frac{1.63\pi}{R}\right), \tag{37}$$

where δ is the *Dirac delta* function, i.e., the Fourier spectrum is zero everywhere except at the peak of the first side lobe. In this case, there is a local maximum for all the focus measures defined earlier. Both increasing and decreasing the blur circle radius causes the focus measures to decrease.

The above effect of the side lobe can be eliminated by an ideal low pass filter which removes the side lobes (see Figure 4). The cut off frequency ρ_c of the ideal LPF should be $1.22\pi/R_{max}$ where R_{max} is the maximum blur circle radius expected for the given operating range of the distance of the object. However, this solution for the side lobe effect is too drastic. The disadvantages of this solution are the computational cost of ideal lowpass filtering and the complete elimination of frequency content beyond the cut-off frequency ρ_c of the LPF. Such complete elimination of frequencies will make it impossible to focus on objects with non zero spectrum only for $\rho > \rho_c$. A better solution to the side lobe effect is to attenuate the side lobes relative to the main lobe rather than complete elimination of the side lobes. For this purpose we suggest a Gaussian lowpass filter. It can be implemented economically by convolving the images in the spatial domain with a Gaussian having a small spread (about $\sigma = 1.5$ to 2.5 pixels). Convolution can be implemented efficiently as two one-dimensional operations, first along rows, then along columns, because Gaussian is seperable. If k is the factor by which we wish to attenuate the peak of the first side lobe, the parameter r for the Gaussian low pass filter can be computed as follows.

$$
\begin{aligned}
e^{-\frac{1}{2}\rho^2 r^2} &= k && (38)\\
\Rightarrow r &= \frac{\sqrt{2}\,\sqrt{-log\ k}}{\rho} && (39)\\
&\approx \frac{\sqrt{2}}{1.63\pi}\,R_{max}\,\sqrt{-log\ k} && (40)\\
r &\approx 0.2762\sqrt{-log\ k}\ R_{max} && (41)\\
& && (42)
\end{aligned}
$$

If $k = 0.25$ and $R_{max} = 5$ pixels, then $r = 1.63$ pixels. The size of the one-dimensional Gaussian filter in this case is about 7.

Another advantage to low pass filtering the image is the reduction of noise and the attenuation of aliased frequencies near the high end of the spectrum. Aliasing will be a problemn if the focused image has spatial frequencies larger than the Nyquist frequency.

The noise magnitude spectrum usually remains the same in all images of the image sequence g_i . In this case , if the effect of the side lobes is neglected, then the noise has no effect on the performance of the *focus measures* . The focus measures remain monotonic and unimodal. Further, if the effect of side lobes are negligible, then even aliasing has no effect on the performance of the focus measures. However. if the frequency content in the side lobes is high due to noise, aliasing, or the focused image, then the focus measures may exhibit local maxima and the global maximum may be shifted. This makes the search for global maximum more difficult and introduces error in the final result. Therefore the attenuation of side lobes improves the behaviour of the focus measures.

According to Theorems 5 and 6 in the appendix, low pass filtering the image sequence does not affect the soundness, monotocity, or the unimodality of the focus measures.

7 Band Pass Filtering

Suppose that the image sequence $G_i(\omega,\nu)$ is filtered by the Gaussian low pass filter $H_b(\omega,\nu) = e^{-\frac{1}{2}(\omega^2+\nu^2)r^2}$. The Fourier spectrum of the resulting image sequence is $|H_b(\omega,\nu)|\,|G_i(\omega,\nu)|$. The focus measures M_2 and M_3 for this image sequence are :

$$
\begin{aligned}
M_2' &= \int\int |\,\vec{\nabla}(h_b(x,y) * g_i(x,y)\,|^2\,dx\,dy\\
&= \int\int (\omega^2+\nu^2)\,|H_b(\omega,\nu)|^2\,|G_i(\omega,\nu)|^2\,d\omega\,d\nu\\
M_3' &= \int\int \nabla^2\,(h_b(x,y) * g_i(x,y)))^2\,dx\,dy\\
&= \int\int (\omega^2+\nu^2)^2\,|H_b(\omega,\nu)|^2\,|G_i(\omega,\nu)|^2\,d\omega\,d\nu
\end{aligned}
$$

From the Fourier domain expressions for M_2' above, M_2' can be thought of as the result of first filtering $G_i(\omega,\nu)$ by a filter, $B_2(\omega,\nu)$, having the Fourier Magnitude Spectrum $|B_2(\omega,\nu)| = \sqrt{(\omega^2+\nu^2)}\,|H_b(\omega,\nu)|$ and then measuring the spectral energy of the resulting image. Similarly, M_3' can be thought of as the result of first filtering $G_i(\omega,\nu)$ by a filter , $B_3(\omega,\nu)$, having the Fourier Magnitude Spectrum $|B_3(\omega,\nu)| = (\omega^2+\nu^2)|H_b(\omega,\nu)|$ A plot of $|B_2(\omega,\nu)|$ and $|B_3(\omega,\nu)|$ are shown in Figure 7. We see that both these filters have band-pass characteristics. Experimental results presented later show that these band-pass filters make the focus measures to have sharp peaks while generally retaining monotonicity and unimodality. These band-pass filters have the desirable characteristic of attenuating low frequencies which contribute less to the focus measure and attenuating high frequencies affected by side lobes and noise, but emphasizing medium frequencies.

8 Discussion of other focus measures

A number of focus measures have been proposed in the literature (Ligthart and Groen, 1982, Jarvis, Krotkov 1987). Among these, Tenengrad (Tenenbaum 1970, Schlag et al, Krotkov 1987) has been found to be the best. More recently, a focus measure based on a modified Laplacian operator (SML operator) has been suggested to perform better than Tenengrad (Nayar '92).

Tenengrad (Schlag et al, Tenenbaum) is a measure of thresholded gradient magnitude. It is similar to M_2 except that only those points where the gradient magnitude is greater than a pre-specified threshold, are used in the calculation. The other points are not used in the calculation of the focus measure. Because of the thresholding operation, a Fourier domain filter analysis of this focus measure is not possible. More importantly, if the threshold is non-zero, this focus measure cannot be proved to be theoretically sound, i.e., it cannot be proved that the global maximum of the focus measure occurs for the best focused image. Moreover, this focus measure involves the selection of a threshold. For these reasons, and the fact that the focus measures discussed in this paper (M_1 to M_3 and M_1' to M'_3) performed very well in a large number of experiments, we do not recommend this method in actual applications.

Nayar (Nayar '92) has proposed a new focus measure based on a new operator named *sum-modified-Laplacian* (SML). It is defined as :

$$\nabla^2{}_M\ g_i(x,y) \quad = \quad \left|\frac{\partial^2 g_i}{\partial x^2}\right| + \left|\frac{\partial^2 g_i}{\partial y^2}\right|$$

The SML focus measure is defined as :

$$SMLF \quad = \quad \int\int\left(\left|\frac{\partial^2 g_i}{\partial x^2}\right| + \left|\frac{\partial^2 g_i}{\partial y^2}\right|\right)\,dx\,dy$$

SML differs from the usual Laplacian in that the magnitude of the second derivatives are summed instead of their actual values. Therefore it is a non-linear operator. The focus measure SMLF is a simple summation of the result of applying the SMLF operator; it does not involve squaring the integrand. For these reasons, it is not possible to provide a Fourier domain filter analysis of this focus measure. Further, as in the case of Tenengrad, we do not believe that this focus measure can be proved to be theoretically sound.

One of the main reason that Tenengrad and SMLF were said to be better focus measures in the past was that they gave sharper peaks. The sharper the peak produced by a focus measure the better it was thought because the location of the position of the maximum was thought to be more accurate. However it should be noted that a blunt peak of any focus measure can be sharpened by simply squaring the values of the focus measure. In fact, the peak can be sharpened to any desired degree by raising the values to some large positive power. (See Figure 5). Therefore, not only the sharpness of the peak, but also the smoothness (or monotonicity) is important. Sharpening the peak of a focus measure will magnify and reveal hidden local maxima. For bad focus measures, the amplitude of the local maxima will be large where as it will be small for good focus measures, thus reducing the uncertainty in locating the actual global maximum. In view of this observation, we believe that Tenengrad and SMLF have no particular advantage over some of the focus measures considered in this paper.

9 Discrete focus measures

Discrete versions of the focus measures were implemented on SPARCS. The expressions for the $N \times N$ image $g_i(x,y)$ are given here. Each image was first normalized with respect to mean. Magnification normalization was not done as the change in magnification was less than 2%.

1. *Variance*

Variance is computed as

$$M_1 = \frac{1}{N^2}\sum_x\sum_y (g_i(x,y)-\mu)^2 \tag{43}$$

$$= \frac{1}{N^2}\sum_x\sum_y g_i^2(x,y)-\mu^2 \tag{44}$$

$$\text{where } \mu = \frac{1}{N^2}\sum_x\sum_y g_i(x,y) \tag{45}$$

2. *Energy of image gradient*

This focus measure is computed as

$$M_2 = \sum_x\sum_y\left(g_x^2+g_y^2\right) \tag{46}$$

$$\text{where } g_x(x,y) = g_i(x+1,y)-g_i(x,y) \tag{47}$$

$$\text{and } g_y(x,y) = g_i(x,y+1)-g_i(x,y) \tag{48}$$

3. *Energy of Laplacian of the Image*

$$M_3 = \sum_x\sum_y (g_{xx}+g_{yy})^2 \tag{49}$$

where

$$g_{xx}+ \quad g_{yy} = \tag{50}$$

$$-g_i(x-1,y-1)-4g_i(x-1,y)-g_i(x-1,y+1) \tag{51}$$

$$-4g_i(x,y-1)+20g_i(x,y)-4g_i(x,y+1)$$

$$-g_i(x+1,y-1)-4g_i(x+1,y)-g_i(x+1,y+1)$$

4. *Variance of low-pass filtered image*

The image was first low-pass filtered by convolution with a two-dimensional Gaussian. Since the Gaussian is a seperable filter, it was implemented as two one-dimensional convolutions, first along rows and then along columns.

$$g_i'(x,y) = \sum_p\sum_q h_b(p,q)\ g_i(x-p,y-q) \tag{52}$$

$$= \sum_p\sum_q h_b'(p)h_b'(q)\ g_i(x-p,y-q) \tag{53}$$

$$= \sum_p h_b(p)\sum_q h_b(q)g_i(x-p,y-q) \tag{54}$$

The variance of this smoothed image is computed similar to M_1.

$$M_1' = \frac{1}{N^2}\sum_x\sum_y {g_i'}^2(x,y) - \frac{1}{N^2}\sum_x\sum_y g_i'(x,y) \tag{55}$$

5. *Energy of low-pass filtered image gradient*

The original image g_i was first low-pass filtered as in the previous case to obtain g_i'. Then M_2' was computed similar to M_2 except that g_i' was used in place of g_i.

6. *Energy of Laplacian of the low-pass filtered image*

M_3' was computed similar to M_3 except that the low-pass filtered image g_i' was used in place of the original image g_i.

10 SPARCS

The focus measures described here were implemented on a camera system named Stonybrook Passive Autofocusing and Ranging Camera System (SPARCS). SPARCS was built in the Computer Vision Laboratory at the Department of Electrical Engineering, State University of New York, Stony Brook. A block diagram of the system is shown in Figure 2. SPARCS consists of a SONY XC-711 CCD camera and an Olympus 35-70mm motorized lens. Images from the camera are captured by a frame grabber board (Quickcapture DT2953 of Data Translation). The frame grabber board resides in an IBM PS/2 (model 70) personal computer. The images taken by the frame grabber are processed in the PS/2 computer.

The focal length of the lens can be varied manually from about 35mm to 70mm. The F-number which is defined as the ratio of the focal length f to aperture diameter D can also be set manually to 4, 8, 22 etc.,. The lens system consists of multiple lenses and focusing is done by moving the front lens forward and backward. The lens can be moved either manually or under computer control. To facilitate computer control of the lens movement there is a stepper motor with 97 steps, numbered 0 to 96. Step number 0 corresponds to focusing an object at distance infinity and step number 96 corresponds to focusing a nearby object, at a distance of about 50cm from the lens as shown in Table 1. The motor is controlled by a microprocessor, which can communicate with the IBM PS/2 through a digital I/O board (Contec mPIO24/24). Pictures taken by the camera can be displayed in real time on a color monitor (SONY PVM-1342 Q). The images acquired and stored in the IBM PS/2 can be transferred to a SUN workstation.

The camera settings used in the experiment were

- Focal Length = 35mm.
- F- Number = 4.
- Camera Gain Control = +6dB.
- White Balance = Off.
- Gamma Compensation = Off.

11 Experiments

Experiments were conducted on a large number of test objects. The results on three of these objects will be presented. The three objects are

1) a picture of a cartoon,

2) face of a person, and

3) a small circular light source having large high spatial frequency content (the diameter of a pinhole: 2mm).

The third object was chosen to observe the effect of aliasing in the Camera OTF. The illumination of the first two experiments was about 500 lux and no illumination was given for the third experiment except its light source.

Each object was placed at a different distance(Cartoon: 65 steps, Human Face: 57steps, Light Source: 40 steps) in front of the camera and the program was run. The vertical lines in Figures 8 9 10 indicate the known distance of the object. SPARCS acquired one image of the object at each lens position. There are 97 lens positions corresponding to 97 steps of the lens stepper motor. From each image thus acquired, a 64×64 subimage of the object was extracted. Then all the 6 focus measures were computed and printed. This procedure was repeated for each of the three objects. The 6 focus measures were normalized to have the same peak values by dividing their values by their maximum values. The results are plotted in Figures 8 9 10. In the plots we see that all focus measures reach a peak at almost the same location. The only difference lies in the sharpness of peaks and the smoothness of the plotted curves.

In Figures 8 9 10 the vaiance curve shows a very slow slope, particularly almost same to that of signal power in the light source experiment. However, the curves of gradient and laplacian operators show very sharp slopes near the peak point while they are almost flat at the far position. Some misleadings were found in the experiments of the above two operators by one or two steps(2 or 4% error) at the focused position.

Furthermore, there is a large value of local maximum peak near the focused position in the light source experiment for the laplacian operator, which shows that aliasing is a serious problem in high pass filtering.

In the plots we see that the gaussian curve is flatter than the signal energy curve by the smoothing effect. In the experiment of the squared gradient-gaussian method, we got very sound results in the sharpness and the smoothness. The method improves the monotonic characteristic increasing the shapness of the focus measure curve. In the plots its curve shows the linearity quite well at all the object positions.

12 Conclusion

13 Appendix

In this appendix we use the following notation and relations.

$f(x,y)$ is the focused image of an object and $F(\omega,\nu)$ is its Fourier transform where ω, ν and $\rho = \sqrt{\omega^2+\nu^2}$ are spatial frequencies expressed in radians/unit distance. $g_1(x,y)$ and $g_2(x,y)$ are two normalized images of the object recorded by a camera with camera parameters $\mathbf{e}_1$ and $\mathbf{e}_2$. R_1 and R_2 are the normalized blur circle radii corresponding to g_1 and g_2 respectively. $H_1(\rho)$ and $H_2(\rho)$ are the camera OTFs corresponding to g_1 and g_2 respectively. $G_1(\omega,\nu)$ and $G_2(\omega,\nu)$ are the Fourier transforms of g_1 and g_2 respectively. Defocusing is a convolution operation and the following relations hold:

$G_1(\omega,\nu) = H_1(\rho)F(\omega,\nu)$

$G_2(\omega,\nu) = H_2(\rho)F(\omega,\nu)$

Definition of the focus measure $M_0(i)$:

$M_0(i) = \int\int \mid G_1(\omega,\nu) \mid d\omega d\nu$ for i=1,2.

$\alpha(x)$ is a monotonically increasing function such that if $0 \leq x_1 < x_2$ then $\alpha(x_1) < \alpha(x_2)$.

Definition of the focus measure $M_0^{'}(i)$:

$$M_0^{'}(i) = \int\int \alpha(\mid G_i(\omega,\nu) \mid) d\omega d\nu \quad \text{for } i=1,2. \tag{56}$$

$L(\omega,\nu)$ is the transfer function of a filter.
Definition of the focus measure $M_0^{''}(i)$:

$$M_0^{''}(i) = \int\int \mid L(\omega,\nu) \mid \alpha(\mid G_i(\omega,\nu) \mid) d\omega d\nu \qquad (57)$$

The first theorem shows that M_0 is a sound focus measure. It increases monotonically as the blur circle radius decreases (i.e. the image blur decreases) and reaches a maximum when the image is in best focus (i.e. the blur circle radius is zero). It also shows that M_0 has a unique maximum and therefore has no local maxima. In this theorem, the paraxial geometric optics model is used for the OTF, except that the magnitude of the OTF outside its main lobe is taken to be zero.

Theorem 1 *If*

$$H_1(\rho) = \begin{cases} \frac{2J_1(R_1\rho)}{R_1\rho} & \text{for } 0 \le R_1\rho \le 1.22\pi \text{ (inside main lobe)} \\ 0 & \text{elsewhere(outside main lobe)} \end{cases} \qquad (58)$$

$$H_2(\rho) = \begin{cases} \frac{2J_1(R_2\rho)}{R_2\rho} & \text{for } 0 \le R_2\rho \le 1.22\pi \\ 0 & \text{elsewhere} \end{cases} \qquad (59)$$

and $\mid R_2 \mid \ > \ \mid R_2 \mid \ > \ 0 \mid R_2 \mid \ > \ \mid R_1 \mid \ > 0 label eq:13 (59)$
then
(i) $M_0(1) > M_0(2)$ and
(ii)

$$\lim_{\mid R_1 \mid \to 0} M_0(1) = \int\int \mid F(\omega,\nu) \mid d\omega d\nu$$

.

(Proof)(i) $\mid R_1 \mid \ > \ \mid R_2 \mid$
$\Longrightarrow 1 \ > \ \mid H_1(\rho) \mid \ > \ \mid H_2(\rho) \mid \geq 0$ for $0 < \rho \leq \frac{1.22\pi}{R_1}$(inside main lobe)
$\Longrightarrow \mid H_1(\rho) \mid\mid F(\omega,\nu) \mid \ > \ \mid H_2(\rho) \mid\mid F(\omega,\nu) \mid$
$\Longrightarrow \mid G_1(\omega,\nu) \mid \ > \ \mid G_2(\omega,\nu) \mid$
$\Longrightarrow \int\int \mid G_1(\omega,\nu) \mid d\omega d\nu \ > \ \int\int \mid G_2(\omega,\nu) \mid d\omega d\nu$
$\Longrightarrow M_0(1) \ > \ M_0(2)$

(ii)

$$\lim_{\mid R_1 \mid \to 0} M_0(1) = \lim_{\mid R_1 \mid \to 0} \int\int \mid H_1(\rho) \mid\mid F(\omega,\nu) \mid d\omega d\nu$$

$$= \int\int \mid F(\omega,\nu) \mid d\omega d\nu$$

because

$$\lim_{\mid R_1 \mid \to 0} H_1(\rho) = 1$$

.

Theorem 2 is the same as Theorem 1, except that the OTF corresponds to a Gaussian. The proof of this theorem is similar to that of Theorem 1.

Theorem 2 *If*

$$H_1(\rho) = e^{-\frac{\rho^2 r_1^2}{2}} \qquad (60)$$

$$H_2(\rho) = e^{-\frac{\rho^2 r_2^2}{2}} \qquad (61)$$

$$r_1 = cR_1, \ \ r_2 = cR_2 \qquad c \text{ is a proportionality constant} \qquad (62)$$

and

$$\mid R_2 \mid \ > \ \mid R_1 \mid \ > \ 0 \qquad (63)$$

then
(i) $M_0(1) \ > \ M_0(2)$ *and*
(ii)

$$\lim_{\mid R_1 \mid \to 0} M_0(1) = \int\int \mid F(\omega,\nu) \mid d\omega d\nu$$

.

$M_0^{'}$ is a focus measure defined as the volume integral of a monotonically increasing function of $\mid G(\omega,\nu) \mid$. Typical examples of the monotonically increasing function $\alpha(x)$ are $\alpha(x) = x^2$ or $\alpha(x) = x^n$ for n>1. Theorem 3 shows that $M_0^{'}$ has properties similar to M_0. With minor exceptions, the proof of this theorem is similar to Theorem 1.

Theorem 3 *If the conditions 58 to* **??** *in Theorem 1 are satisfied,*
then

$$(i) M_0^{'}(1) \ > \ M_0^{'}(2) \qquad (64)$$

and

$$(ii) \lim_{\mid R_1 \mid \to 0} M_0^{'}(1) = \int\int \alpha(\mid F(\omega,\nu) \mid) d\omega d\nu. \qquad (65)$$

(Proof)
Similar to Theorem 1, the main step to be noted is that $\mid G_1(\omega,\nu) \mid \ > \ \mid G_2(\omega,\nu) \mid$
$\Longrightarrow \alpha(\mid G_1(\omega,\nu) \mid) \ > \ \alpha(\mid G_2(\omega,\nu) \mid)$.

Theorem 4 *If conditions 60 to 63 are satisfied, then relations 64 and 65 will be true.*

Theorem 5 shows that a focus measure which works correctly for an image sequence will also work correctly if the entire image sequence is filtered by the same filter $L(\omega,\nu)$. Therefore, band pass filtering image sequence will not affect the monotonicity or the location of the maximum of a focus measure.

Theorem 5 *If conditions 58 to* **??** *are satisfied, then*

$$(i) M_0^{''}(1) \ > \ M_0^{''}(2) \qquad (66)$$

and

$$(ii) \lim_{|R_1| \to 0} M_0^{''}(1) = \int \int \mid L(\omega,\nu) \mid\mid F(\omega,\nu) \mid d\omega d\nu. \tag{67}$$

(Proof)

Similar to the proof of Theorem 1; the main step to be noted is that

$\mid G_1(\omega,\nu) \mid \; > \; \mid G_2(\omega,\nu) \mid$

$\Longrightarrow \mid L(\omega,\nu) \mid\mid G_1(\omega,\nu) \mid \; > \; \mid L(\omega,\nu) \mid\mid G_2(\omega,\nu) \mid .$

Theorem 6 *If conditions 60 to 63 are satisfied, then relations 66 and 67 will hold.*

(Proof)

Similar to Theorem 5.

References

[1] J. D. Gaskill, *Linear Systems, Fourier Transforms, and Optics,* John Wiley & Sons, New York, 1978.

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

[3] B. K. P. Horn, *Robot Vision,* McGraw-Hill Book Company, 1986.

[4] R. A. Jarvis, "Focus optimization criteria for computer image processing", Microscope 24, pp.163, 1976.

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

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

[7] S. Nayar, "Shape from Focus System for Rough Surfaces", *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Champaign,Illinois, pp. ,June15-18,1992.

[8] A. P. Pentland, "A new sense for depth of field", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-9, No. 4, pp. 523–531, July 1987.

[9] J. F. Schlag, A. C. Sanderson, C. P. Neuman, and F. C. Wimberly, "Implementation of automatic focusing algorithms for a computer vision system with camera control", CMU-RI-TR-83-14, Robotics Institute, Carnegie-Mellon University, 1983.

[10] M. Subbarao, "Parallel depth recovery by changing camera parameters", *Second International Conference on Computer Vision,* Florida, USA, pp. 149-155, December 1988.

[11] M. Subbarao, "Efficient depth recovery through inverse optics", Editor: H. Freeman, *Machine Vision for Inspection and Measurement,* Academic press, Boston, pp. 101-126, 1989.

[12] M. Subbarao, "Computational methods and electronic camera apparatus for determining distance of objects, rapid autofocusing, and obtaining improved focus images", U.S. patent application serial number 07/373,996, June 1989 (pending).

[13] M. Subbarao, "Determining distance from defocused images of simple objects", Tech. Report No. 89.07.20, Computer Vision Laboratory, Dept. of Electrical Engineering, State University of New York, Stony Brook, NY 11794-2350.

[14] M. Subbarao, and A. Nikzad, "A model for image sensing and digitization in machine vision", *OE/BOSTON '90, SPIE conference*, Boston, November 1990.

[15] M. Subbarao, N. Agarwal, and G. Surya, "Application of Spatial-Domain Convolution/Deconvolution Transform for Determining Distance from Image Defocus", Tech. Report No. 92.01.18, Computer Vision Laboratory, Dept. of Electrical Engg., State University of New York, Stony Brook, NY 11794-2350, 1992.

[16] M. Subbarao, and T. Wei, "Depth from Defocus and Rapid Autofocusing: A Practical Approach", Tech. Report No. 92.01.17, CVL, Dept. of EE, SUNY, Stony Brook, NY 11794-2350, 1992. (An abridged version of this appears in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*,Champaign, Illinois, pp. , June 15-18,1992.

[17] M. Subbarao, and M. C. Lu, "Computer Modeling and Simulation of Camera Defocus", Tech. Report No. 92.01.16, CVL, Dept. of EE, SUNY, Stony Brook, NY 11794-2350, 1992.

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

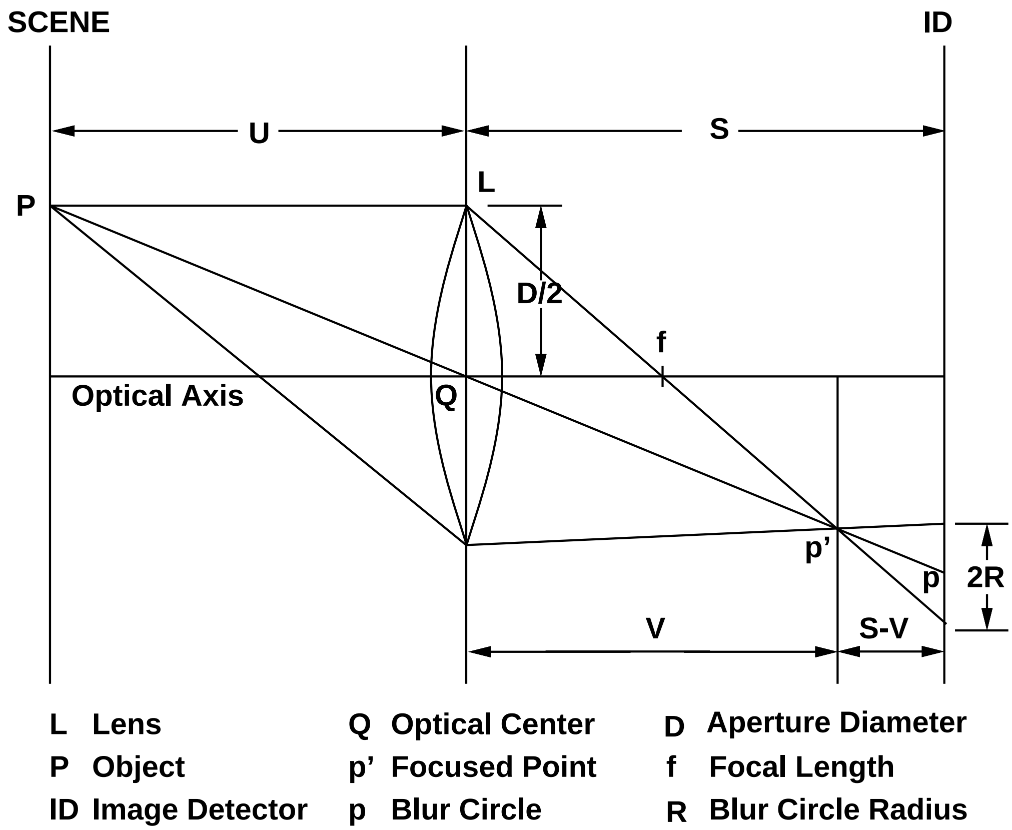


Figure 1: Image Formation in a Convex Lens.

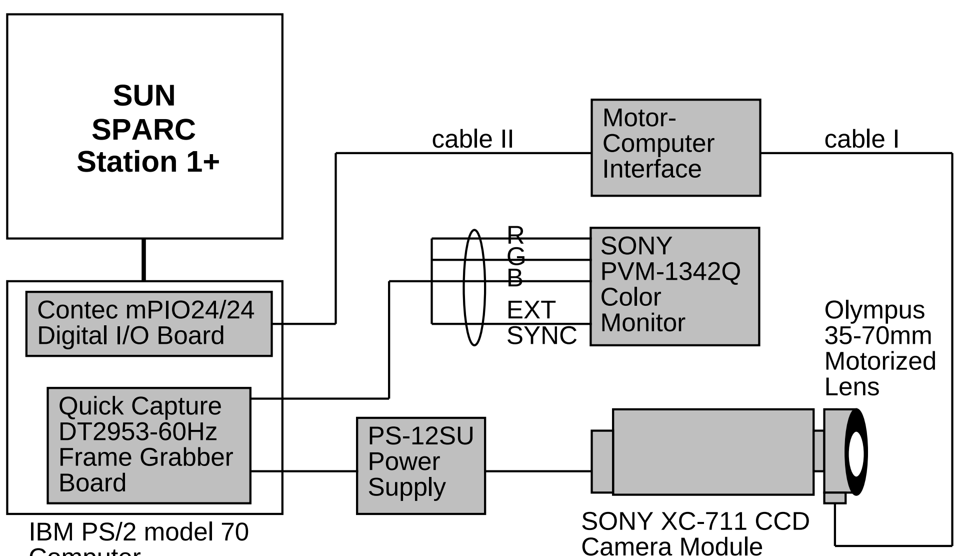


Figure 2: SPARCS.

Stonybrook Passive Autofocusing and Ranging Camera System-SPARCS - is a prototype camera system developed at the Computer Vision Labatory for experimental research in robotic vision, State University of New York at Stony Brook

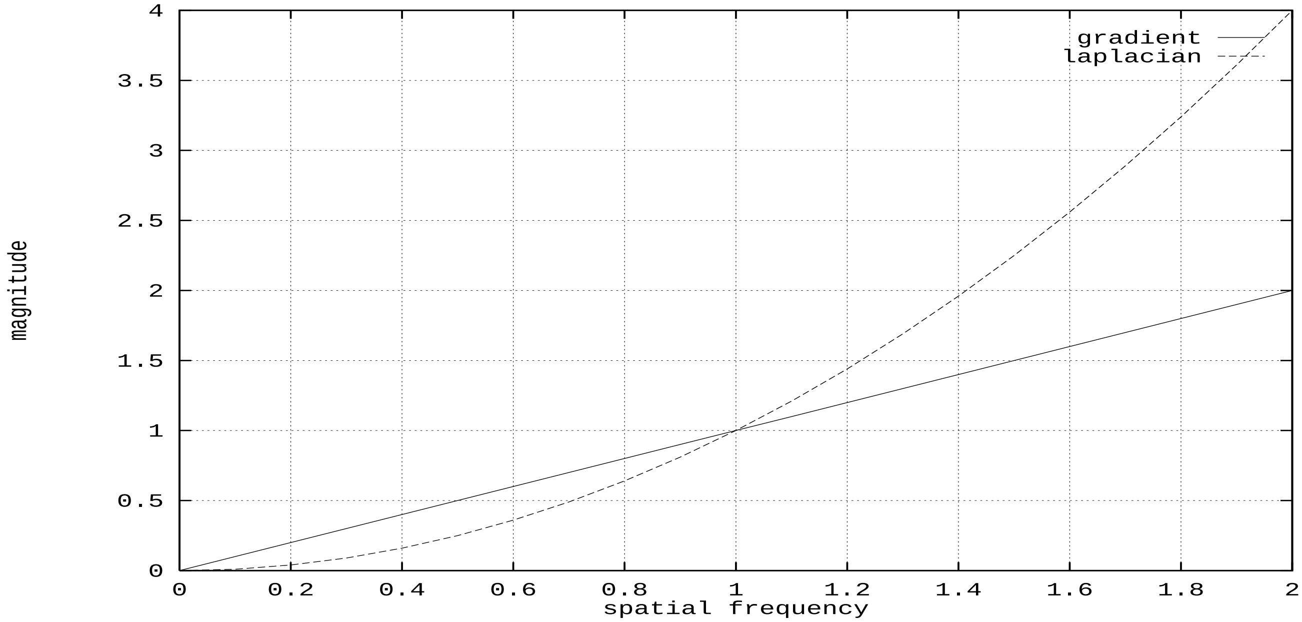


Figure 3: Radial profile of $L_2(\omega,\nu)$(solid line) and $L_3(\omega,\nu)$(dotted line) with respect to spatial frequency ρ where $\rho(\omega,\nu) = (\omega^2+\nu^2)^{\frac{1}{2}}$.

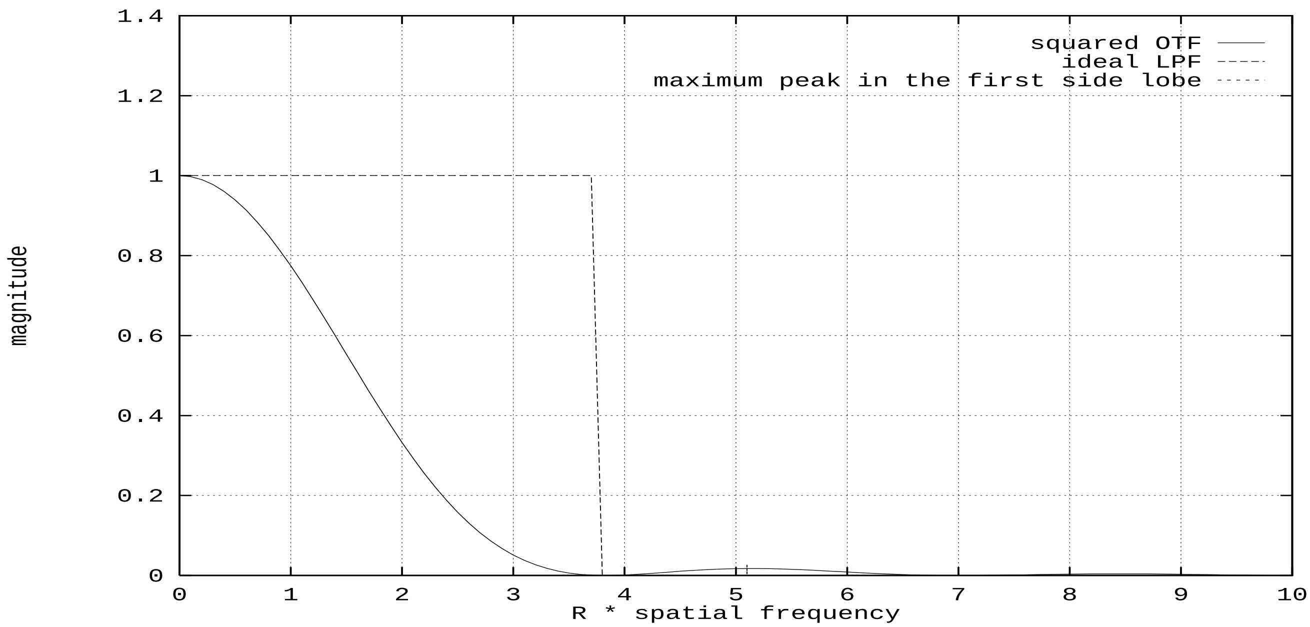


Figure 4: Radial profile of $\mid H(R\rho) \mid^2$ and the magnitude response of the ideal low pass filter.

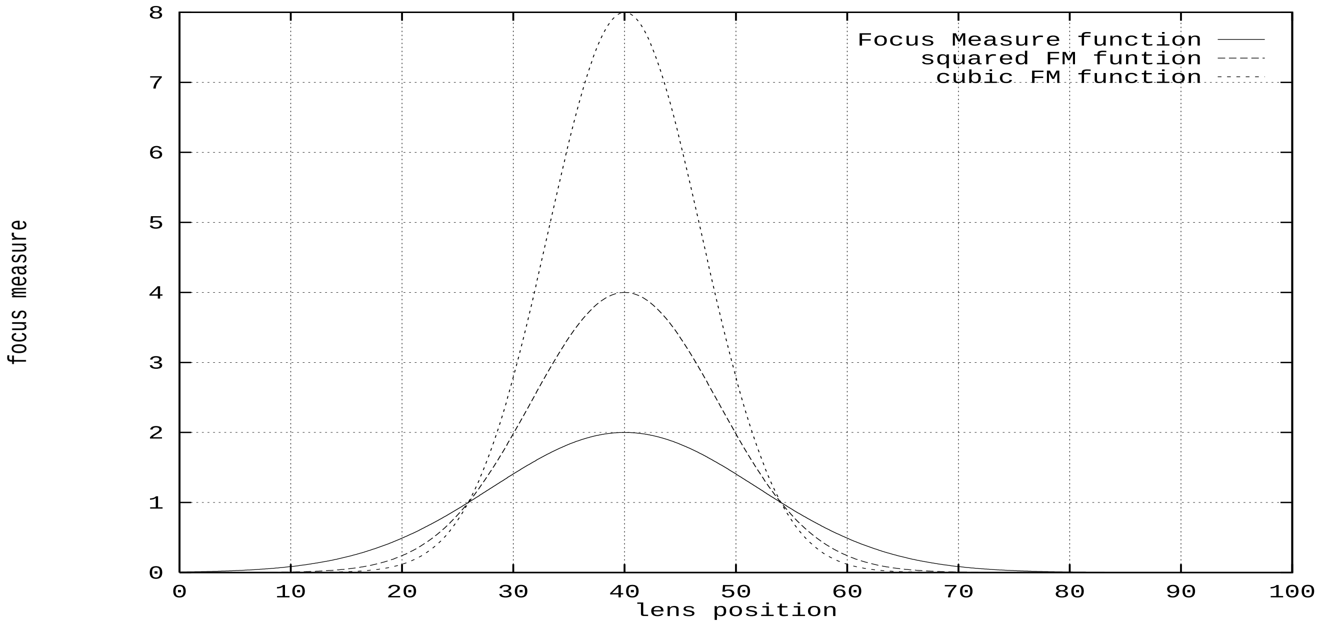


Figure 5: The effect of multisquared Focus Measure functions.

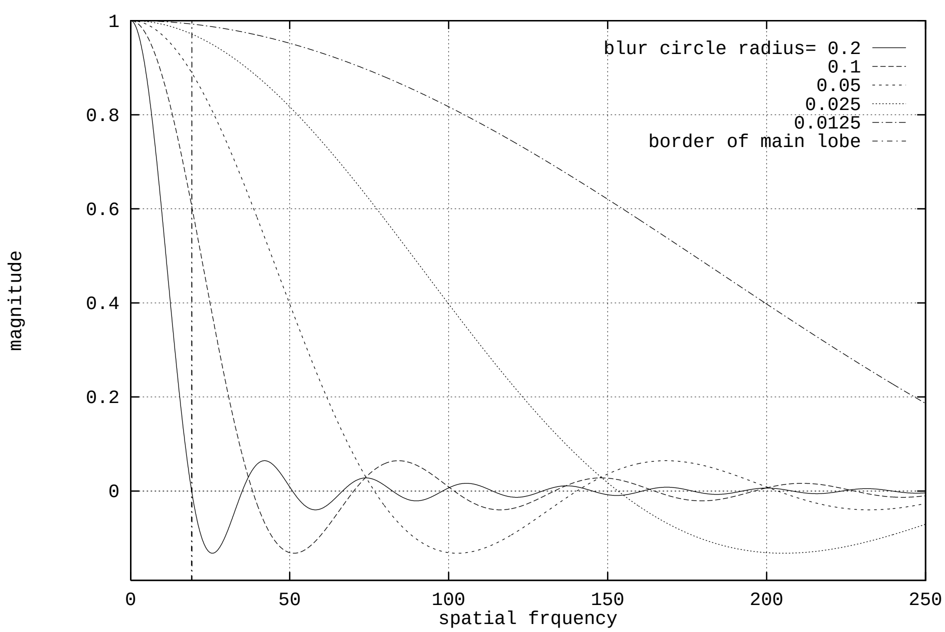


Figure 6: A sequence of the Optical Transfer Functions where the vertical line indicates the border of main lobe for a OTF curve with R_{max}= 0.2mm.

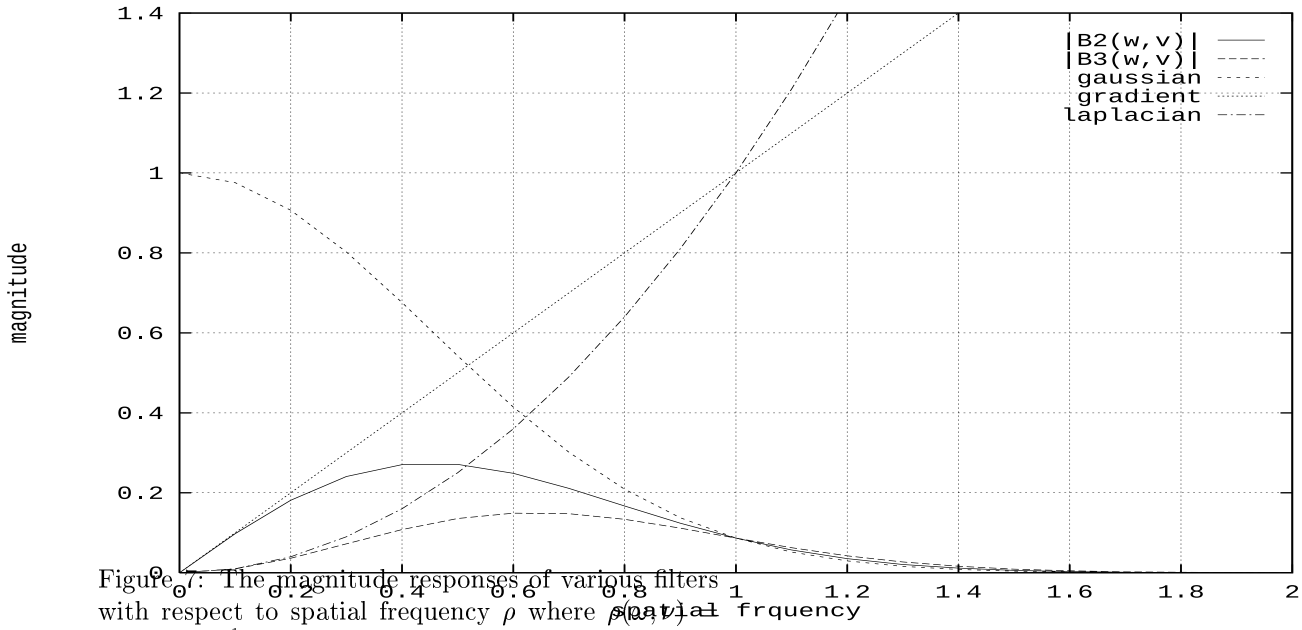


Figure 7: The magnitude responses of various filters with respect to spatial frequency ρ where $\rho(\omega,\nu) = (\omega^2 + \nu^2)^{\frac{1}{2}}$: $B_2(\omega,\nu), B_3(\omega,\nu), G_i(\omega,\nu), (\omega^2 + \nu^2)^{\frac{1}{2}}, and (\omega^2 + \nu^2)$.

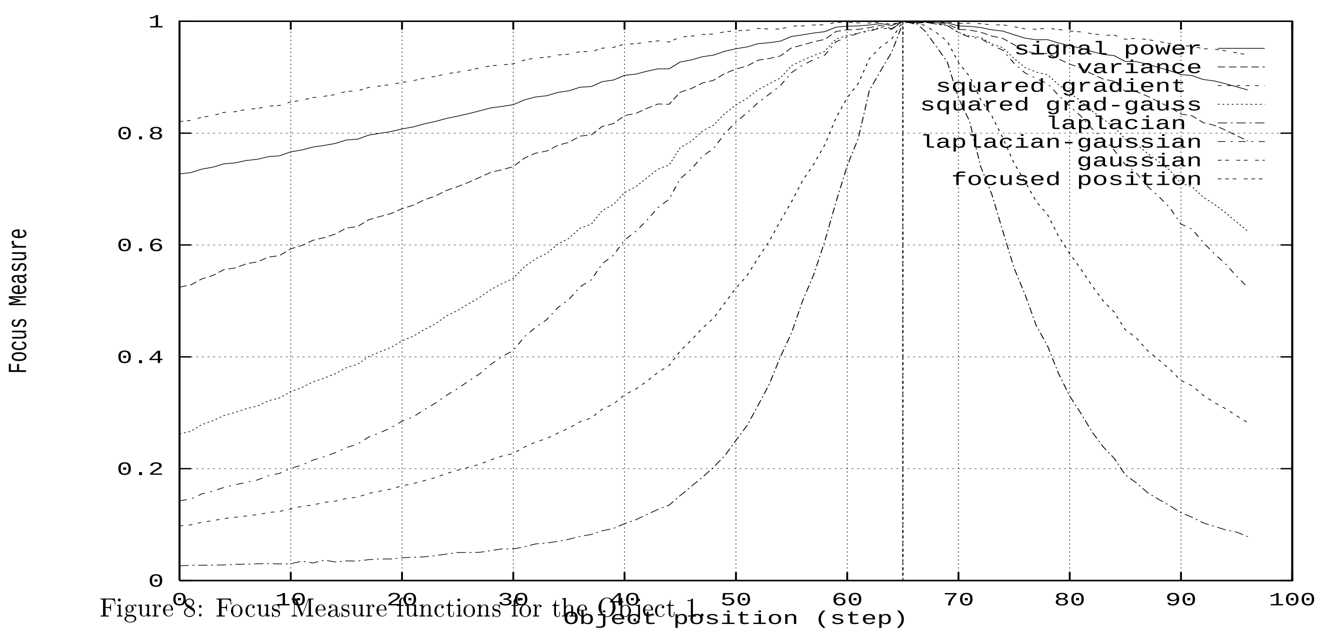


Figure 8: Focus Measure functions for the Object 1.

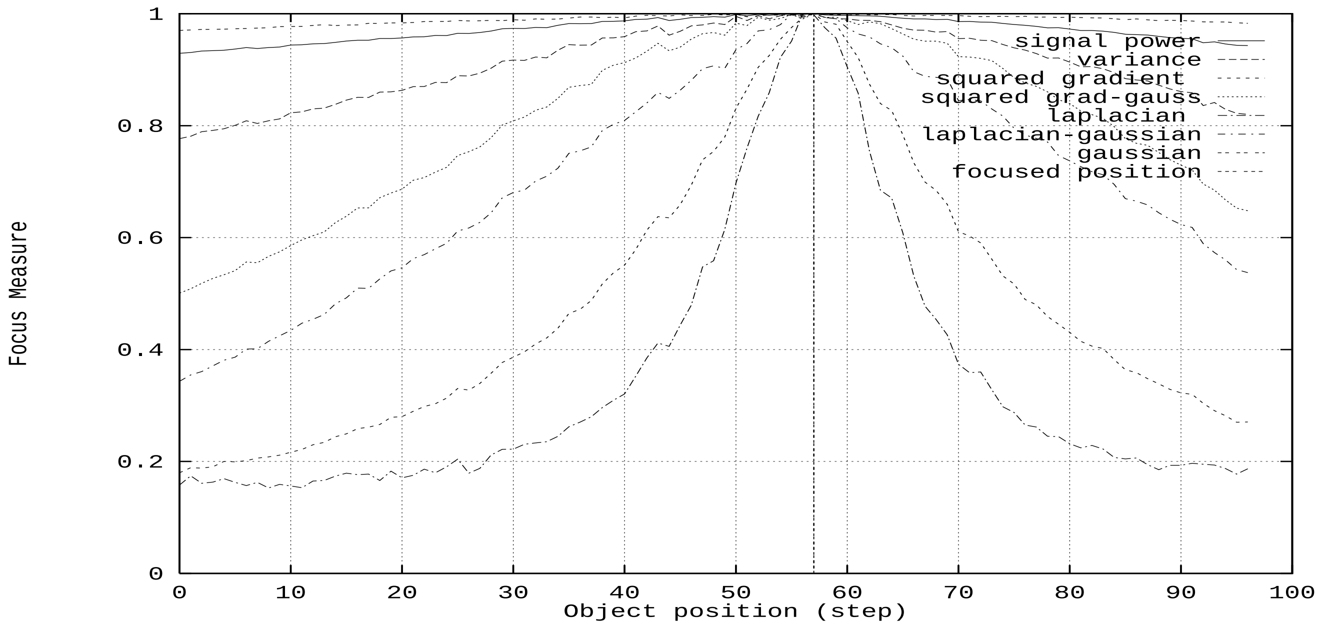


Figure 9: Focus Measure functions for the Object 2.

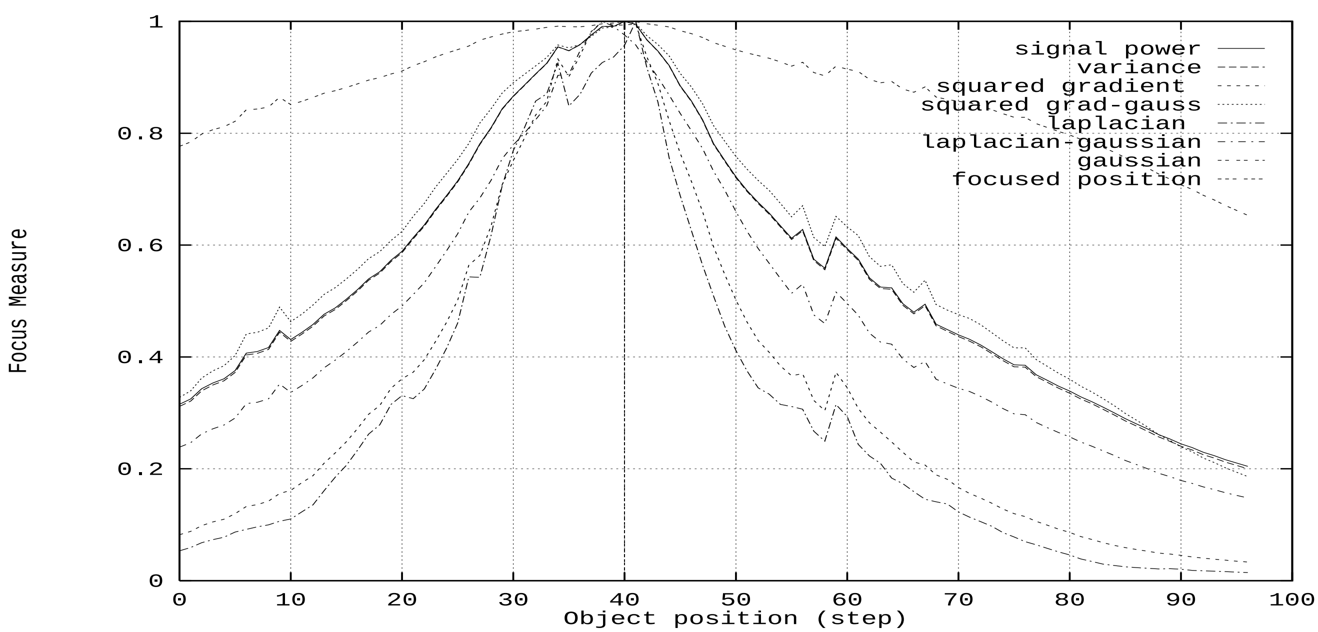
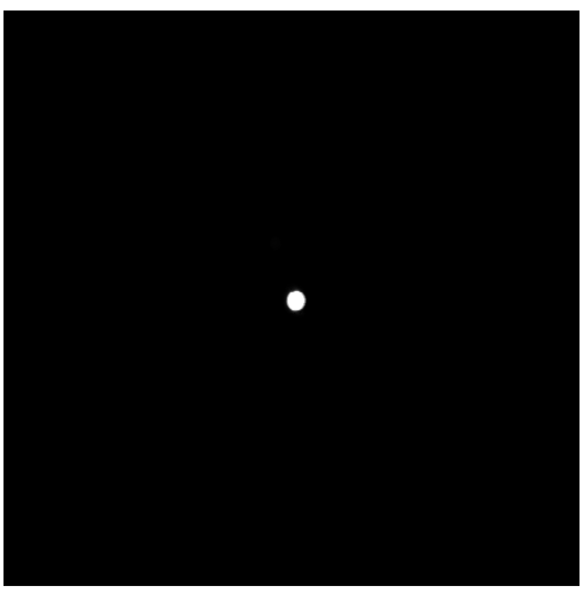


Figure 10: Focus Measure functions for the Object 3.

Figure 10: Object 1: Cartoon. Figure 11: Object 2: Human Face.



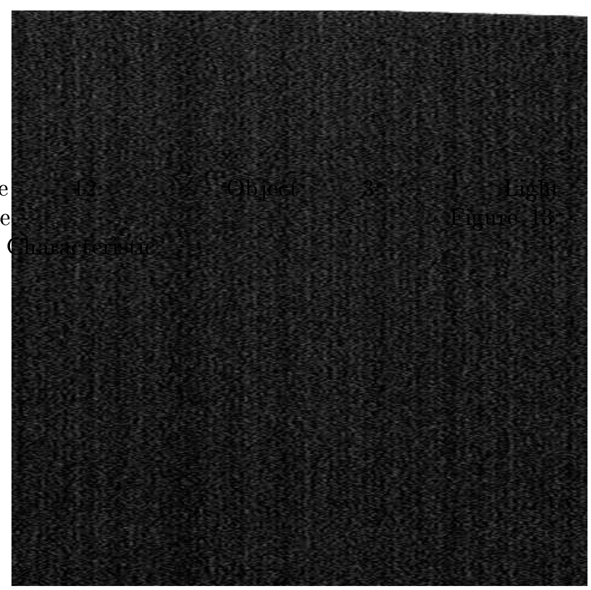


Figure 12: Object 3: Light Source. Figure 13: Noise Characteristic.