IMPLEMENTATION OF A PASSIVE AUTOMATIC FOCUSING ALGORITHM FOR DIGITAL STILL CAMERA

Je-Ho Lee, Kun-Sop Kim, Byung-Deok Nam,*
Jae-Chon Lee, Yong-Moo Kwon, and Hyoung-Gon Kim
Division of Electronics and Information Technology
Korea Institute of Science and Technology
Seoul 130-650, Korea
*Center of Precision Instruments
Samsung Aerospace Industries, Ltd.
Suwon 440-600, Korea

Abstract
A passive AF algorithm for the digital still camera is proposed. As a class of hybrid processing techniques, the proposed algorithm utilizes the two different focus measures by which the best focused image is extracted via two stages. With the paraxial geometric optics model of image formation, some well-known focus measures are investigated and a robust focusing algorithm is designed based on them. The proposed algorithm has been successfully tested on a prototype digital still camera.

I. Introduction
Digital technology is a rapidly evolving field with growing applications in science and engineering. Together with the growth in VLSI technology, this digital revolution has introduced the idea of digital multimedia system which is an assortment of home electronics, personal computers, and communication devices, etc. For instance, a low cost consumer electronic camera will be available not as a device by itself but as a part of an integrated camera-computer-television system which is capable of doing image acquisition, image processing, annotation, desktop publishing, and long distance transmission; a versatile portable multimedia system. This implies electronic imaging device will move from a stand alone system environment into an open system environment. To keep in step with this digital multimedia trend, multifunctional digital still cameras are being developed by many companies. These newly equipped digital still cameras will be quite different from the first version of them which were simple image capturing devices as an alternative to traditional optical cameras.

One of the most attractive advantages of the digital still camera over the conventional camera is probably the capability of the image processing using microprocessor. Color correction, image filtering, automatic exposure (AE), and even AF can be carried out through the image processing. Among this functions, AF is one of the hardest and most important features to be implemented on the digital still camera.

Many passive AF algorithms have been investigated and compared in the literature [1] [2] [3] where the object image has been modeled as a Gaussian filtered image. Under Gaussian approximation, the energy of image gradient is often defined as a focus measure whose value is supposed to be maximum in the best focused image. This approach sounds reasonable since high frequency components determine the sharpness of the object image dominantly. However, in the case of camera lens system having wide focusing and zooming ranges, the performances of such algorithms are sometimes not so good as it should be. This is mainly due to the failure of modeling the object image. Recently an interesting focusing aspect called the “sidelobe effect” has been discussed in [4] where the paraxial geometric optics model is used as the object image model instead of Gaussian. Image gradient-based focus measure is applied to the object image after attenuating sidelobe in the object image. This is done by a Gaussian low pass filter. However, it involves the selection of attenuating factors for the filter and this preprocessing is computationally very expensive.

In this paper, we propose a focusing technique based on two-stage focusing process where two different focus measures are used. In the proposed algorithm, well-focused images are roughly searched in the first stage, and the best focused image is extracted at the next
stage. To be implemented on real cameras, two focus measures for AF algorithm should have following properties.

- The focus measure for the first stage should be computationally simple and insensitive to the side-lobe effect.
- The second focus measure, which is used for extracting the best focused image, should be able to find the best focused (sharpest) image among many well-focused images.

The choice of focus measures in the proposed algorithm is confirmed by theoretically as well as experimentally. Also we show that the proposed algorithm works well either in Gaussian or the paraxial geometric optics model of the object image.

In Section II, we will introduce the focusing phenomenon with emphasis on the optical characteristics of the camera lens system and then investigate existing well-known AF algorithms. Section III gives analyses of the focus measures and describes the proposed algorithm. In Section IV, the focus measures are applied to real image taken by the prototype digital still camera and their performances are compared. The paper is concluded in Section V.

II. A focusing phenomenon and focus measures

It is observed that well focused images contain more information and detail than defocused images. Then, which information does give the degree of focus to the human brain? It is known that the sharpness of the image makes the human determine if the image is in focus. If one observes the frequency components, it will be noted that there will be comparatively few high frequency components when the image is not enough in contrast, i.e. blurred, and their amount will increase as the image is near in focus. This phenomenon can be understood theoretically by considering following simple image formation example shown in Fig. 1.

![Figure 1: Simple lens imaging.](image)

In this figure, \( f \) is the focal length, \( d_o \) is the distance between the object plane and the lens, and \( d_i \) is the distance between the lens and the image plane. Variations in any of above three parameters, \( f \), \( d_o \), and \( d_i \) will result in changing focus. Since the point \( p \) is focused in front of the image plane, the resultant image \( \tilde{p} \) will be blurred. According to geometric optics, \( \tilde{p} \) has the same shape as the lens aperture, and since we take the aperture to be circular, it is also a circle whose radius is \( R \). This is called a blur circle or a circle of confusion. The degree of focus can be expressed in terms of this \( R \); the smaller \( R \) is the better focus is. If it is possible to measure the size of \( R \) in the given image, the AF can be done by simply finding the image having the smallest value of \( R \). Unfortunately, however, it is impossible to measure \( R \) directly, and hence we have to look for other quantities alternative to \( R \).

If the lens is ideal, its ability to produce an ideal point source is limited only by a diffraction [5]. In this case, the image of \( p \) has the uniform brightness inside the circle and zero outside. Assume that light energy incident on the lens is one unit. The response of the camera lens system to an ideal point source is then called the point spread function, which is equivalent to the impulse response in linear system theory. The point spread function of ideal lens is defined as

\[
h_{\text{ideal}}(x, y) = h_{\text{ideal}}(r) = \begin{cases} \frac{1}{\pi R^2} & \text{if } r \leq R \\ 0 & \text{otherwise} \end{cases} \tag{1}
\]

where \( r^2 = x^2 + y^2 \). Defined as the Fourier transform of the point spread function, the optical transfer function provides information on how an electro-optical system element affects the spatial frequency amplitudes and phases. The optical transfer function corresponding to the preceding point spread function is

\[
H_{\text{ideal}}(u, v) = H_{\text{ideal}}(\rho) = 2J_1(R \cdot \rho) \tag{2}
\]

where the radial spatial frequency \( \rho^2 = u^2 + v^2 \) and \( J_1 \) is the first-order Bessel function of the first kind. Fig. 2 shows cross sections of the optical transfer function \( H_{\text{ideal}}(\rho) \) for various values of \( R \). Thus measuring the energy of the focused image might be one of the indirect methods for estimating the size of \( R \).

However, since most of camera lens systems consist of non-ideal multiple lens, above analysis of the lens may far from reality. In case of non-ideal multiple lens system, the overall point spread function is obtained by modeling the system as a cascade combination of linear subelements including lenses. For \( N \) independent, cascaded subelements in a system, the overall point spread function is

\[
h_{\text{total}}(x, y) = h_1(x, y) * h_2(x, y) * h_3(x, y) * \cdots * h_N(x, y) \tag{3}
\]
Characterization of the total system is then the problem of determining the point spread functions of individual subelements. However, it is not easy to accomplish the qualitative analysis of individual subelements of the actual camera system. Moreover even though the overall point spread function for a certain system is achieved by any ways, to build a system-independent AF algorithm, the generalization of the point spread function of the camera system should be established. A commonly used model for the generalized point spread function of the camera system is a Gaussian. The only parameter that needs to be determined for this simple model is the standard deviation in terms of $R$. The Gaussian is defined as

$$h_{Gauss}(r) = \frac{1}{\sqrt{2\pi \sigma}} \exp\left(-\frac{r^2}{2\sigma^2}\right) \quad (4)$$

where $\sigma$ is the standard deviation of the distribution of the point spread function. In practice it is found that $\sigma$ is proportional to $R$ [4].

$$\sigma = kR \quad (5)$$

where $k$ is a constant and it is approximated equal to $\sqrt{2}$ [4]. The Fourier transform of a Gaussian is another Gaussian. The resultant optical transfer function of (5) is

$$H_{Gauss}(p) = \exp\left(-\frac{R^2p^2}{4}\right) \quad (6)$$

Fig. 3 is the plot of cross section of $H_{Gauss}(\rho)$. As I mentioned earlier, both of two optical transfer functions, $H_{ideal}(p)$ and $H_{Gauss}(p)$ hold a concept that a lens can be understood as a low pass filter with variable cutoff frequency. The cutoff frequency is maximum when the lens is in the focused position and decreases as the lens goes to the defocused position. The main difference between these two optical transfer functions, $H_{ideal}(p)$ and $H_{Gauss}(p)$ is that sidelobes are not observed in case of the Gaussian. In this paper, we derive an AF algorithm taking into account characteristics of these two optical transfer functions. That is, no matter which optical transfer function is closer to real one, or even real optical transfer function is in between ideal and mathematically approximated one, the proposed algorithm works robustly.

In most of AF algorithms, a focus measure is defined that is a maximum for the image having the smallest size of $R$ (the best focused image) and it generally decreases as the size of $R$ increases. Constructing good AF algorithm depends on designing a good focus measure.

Since high frequency components determine the sharpness of the object image dominantly, the energy of image gradient is employed very often for the focus measure. Image gradients are based on the first partial derivatives and therefore they have high pass filter properties. Let $g(x, y)$ be the $N \times N$ grey-scale image which is taken by camera. The gradient of an image $g(x, y)$ at location $(x, y)$ is defined as the two-dimensional vector which is given by

$$\nabla g(x, y) = \begin{bmatrix} G_x \\ G_y \end{bmatrix} = \begin{bmatrix} \frac{\partial g}{\partial x} \\ \frac{\partial g}{\partial y} \end{bmatrix} \quad (7)$$

An focus measure which utilizes the above gradient function is defined as

$$FM_{grad} = \sum_x \sum_y \{\nabla g(x, y)\}^2 = \sum_x \sum_y \{G_x^2 + G_y^2\} \quad (8)$$

Among the many spatial gradient operations, the Roberts gradient is most common and historically earliest ones. This is given by

$$G_x = g(x, y) - g(x + 1, y + 1) \quad (9)$$

$$G_y = g(x + 1, y) - g(x, y + 1) \quad (10)$$

The advantage of using Roberts gradient is that it guarantees the fast computation. The numbers of addition and multiplication in equation (8) are $4(N - 1)^2 - 1$ and $2(N - 1)^2$ respectively, if Roberts gradient is used. Slight more complicated version of gradient operation is Sobel gradient which is given by

$$G_x = g(x - 1, y + 1) + g(x + 1, y + 1) - g(x - 1, y - 1) - g(x + 1, y - 1)$$

$$2g(x, y + 1) - 2g(x, y - 1) \quad (11)$$

$$G_y = g(x + 1, y - 1) + g(x + 1, y + 1) - g(x - 1, y - 1) - g(x - 1, y + 1)$$

$$2g(x + 1, y) - 2g(x - 1, y) \quad (12)$$

Using $3 \times 3$ area in the computation of the gradient has the advantage of increased smoothing over $2 \times 2$ Roberts gradient, tending to make the derivative operations less sensitive to noise. Moreover, the weight of Sobel gradient produce additional smoothing effects. It is possible to base gradient computations over larger area [6], but $3 \times 3$ area is by far the most popular because of advantages in computational speed and modest hardware requirements. The focus measure which utilizes Sobel gradient is commonly referred to as the Tenengrad function [7]. Total computation amounts are $11(N - 2)^2 - 1$ additions and $6(N - 2)^2$ multiplications. This fact indicates that Sobel gradient requires operations three times as many as Roberts gradient. The mathematical Laplacian is also often used for focus measure. The mathematical Laplacian of a two
dimensional function $g(x, y)$ is defined as

$$
\nabla^2 g(x, y) = \frac{\partial^2 g}{\partial x^2} + \frac{\partial^2 g}{\partial y^2} \quad (13)
$$

and focus measure based on Laplacian is represented by

$$
FM_{Lap} = \sum_x \sum_y \left( \nabla^2 g(x, y) \right)^2 = \sum_x \sum_y \left( \frac{\partial^2 g}{\partial x^2} + \frac{\partial^2 g}{\partial y^2} \right)^2 \quad (14)
$$

The Laplacian of image is calculated in the similar way to image gradient. One version of the Laplacian of image is given by

$$
\nabla^2 f = f(x, y) - f(x + 1, y) - f(x, y + 1) - f(x - 1, y) - g(x + 1, y) - g(x, y - 1)
$$

The Laplacian focus measure needs $5(N - 2)^2 - 1$ additions and $2(N - 2)^2$ multiplications.

On the other hand, unlike to above partial derivative-based focus measures, the image energy can be straightforwardly used for focus measure and is defined as

$$
FM_{energy} = \int_{-\infty}^{\infty} |G(u, v)|^2 \, du \, dv \quad (16)
$$

For the $N \times N$ grayscale image, (12) can be approximated as

$$
FM_{energy} = \sum_x \sum_y |g(x, y)|^2 \quad (17)
$$

using the Parseval's theorem. A closely related to the image energy, but slightly improved focus measure is the discrete variance which is given by

$$
FM_{var} = \frac{1}{N^2} \sum_x \sum_y |g(x, y) - \mu|^2 \quad (18)
$$

where

$$
\mu = \frac{1}{N^2} \sum_x \sum_y g(x, y) \quad (19)
$$

Because of the term $\mu$, $FM_{var}$ has an advantage of invariance to the change of mean image brightness.

III. Two-Stage AF algorithm

As seen in Section II, focus measures nothing but measure the energy of image after convolving image with various spatial filters. For instance, image gradient focus measures use the first partial derivative-based high pass filter. Thus focus measures reviewed in Section II can be expressed as

$$
FM_{grad} = \sum_x \sum_y \left| \nabla \{ h(r) \ast f(x, y) \} \right|^2 \quad (20)
$$

$$
FM_{Lap} = \sum_x \sum_y \left| \nabla^2 \{ h(r) \ast f(x, y) \} \right|^2 \quad (21)
$$

$$
FM_{var} = \sum_x \sum_y \left| h(r) \ast f(x, y) - \mu \right|^2 \quad (22)
$$

From the frequency domain expressions, $FM_{grad}$ can be thought of as the result of first filtering $F(u, v)$ by a filter having the Fourier magnitude spectrum $|H(p)|$ and then measuring the spectral energy of the resultant focused image. Plots of $|H_{ideal}(\rho)|$, $|\rho H_{ideal}(\rho)|$, and $\rho^2 |H_{ideal}(\rho)|$ are shown in Fig. 4. It is noted that because of the term $\rho$, low frequency components are attenuated while high frequency components are emphasized. It also be noted that $\rho$ emphasizes sidebands too, which can give rise to false focus measuring. This fact implies that $FM_{grad}$ will work fine if image is reasonably focused, that is, sidebands are negligible, but it can produce erroneous results when it deals with defocused images. $FM_{Lap}$ will work as the same way as $FM_{grad}$ does, but even due to square term of $\rho$, sidebands will be much more emphasized. On the other hand, $FM_{var}$ just measures the energy of $G(u, v)$ and therefore there is no special emphasis of high frequency components and sidebands. Thus it is predictable that $FM_{var}$ will provide very smooth results lacking sharp discrimination.

The above investigations of the focus measures implies that it is very hard to accomplish the AF function with one focus measure alone. In proposed AF algorithm, two different focus measures are thus employed, $FM_{var}$ and the energy of Sobel gradient, which is a good tradeoff between the smoothness of discrete variance and the accuracy obtained by Sobel gradient focus measure. In other words, the discrete variance is delegated to the secondary role of finding the vicinity of focus while the energy of Sobel gradient is applied only to the resultant some well-focused images. By doing this, we can keep the computational load low and preserve the accuracy of the focus measure. Moreover, since the image gradient focus measure will not applied to the defocused images, there is no more sidelobe amplifying phenomenon.

IV. Experiment

For experiment, the proposed algorithm is implemented on the prototype digital still camera. In the prototype system, the optical module consists of a 270,000 pixels...
A 3 CCD, a ×10 zoom lens whose typical moving range is between 10 (wide) and 250 (tele), and 0 to 7 mm motorized focusing lens, etc. The lens system consists of multiple lenses, and focusing is done by moving the rear lens forward and backward. The lens can be moved by manually or under DSP (Digital Signal Processor) control. To facilitate DSP control of the lens movement, there is a stepping motor with maximum 1200 steps. Each of CCD sensed images is digitized by 8-bit A/D converter, and resultant 24-bit RGB image undergoes color space conversion through a dedicated color converter chip into 16-bit YCbCr image (4:2:2 format), and is stored into a frame memory. Only 8-bit Y component is used for computing focus measures. The focused image is compressed according to JPEG (Joint Photographic Experts Group) standard, and is recorded in DOS formatted PCMCIA (Personal Computer Memory Card International Association) compatible memory card. This compressed still picture can be display on a TV monitor after decompression or printed out by a color video printer. Also it can be played back on a personal computer with proper software and a card driver.

We conducted experiments on a huge number of real volumetric objects and test planar objects with some well-known focus measures to validate our algorithm. Among them, we present the result for an object illustrated in Fig. 5. The image was taken with zoom lens position of 117 and the object was placed at 1.9 meters from the camera. Focus measures being tested were discrete variance, energy of image gradients (Robert and Sobel) and energy of Laplacian. Fig. 6 shows the performances of these focus measures. From experiments, we make following observations. The focus measure based on Laplacian seems to be worst since it gives rise to many local maxima. Also as we expected, this phenomenon has happened more severely in the highly defocused images. Moreover in some cases which are not plotted here, its global maximum is quite shifted from the actual best focused image. Energy of Robert gradient exhibits sharp peaks near the focused region but its peaks are somewhat noisy and rough. On the contrary, focus measure from the energy of Sobel gradient shows its accuracy especially in the vicinity of focus, but it also produces some local maxima in the highly defocused region. Discrete variance is very smooth in entire region, even in the defocused region. However, its peak points are too gradual and rough to estimate the best focused image.

V. Conclusion

The goal of this work was to develop and implement a practical AF algorithm for a prototype digital still camera. Since a low cost yet accurate AF technique is required for the digital still camera, employing a passive method based on the digital image processing is desirable. We tried to analyze existing AF methods in frequency domain relating them with the optical transfer function of camera system since most of passive AF is done by measuring and comparing high frequency components contained in object images. Under the assumption that the object image is formed by the paraxial geometric optics model, we found that focus measures based on the energy of image gradient have some problems. sidedge effects. Our solution is to build a two-stage AF algorithm where the image variance is used to find the vicinity of focus at the first stage, and the best focused image is detected at the second stage with the the focus measure based on the Sobel gradient. The proposed algorithm is also applicable to Gaussian approximated imagery and is easy to implement. The algorithm has been successfully tested on a prototype digital still camera.

References

Figure 2: The cross sections of $H_{\text{ideal}}(\rho)$.

Figure 3: The cross sections of $H_{\text{Gauss}}(\rho)$.

Figure 4: Plots of $|H_{\text{ideal}}(\rho)|$, $\rho|H_{\text{ideal}}(\rho)|$ and $\rho^2|H_{\text{ideal}}(\rho)|$.

Figure 5: Test object.

Figure 6: Results on the test object.