Performance Evaluation of Different Depth From Defocus (DFD) Techniques

Tao Xian, Murali Subbarao

Dept. of Electrical & Computer Engineering[*]
State Univ. of New York at Stony Brook, Stony Brook, NY, USA 11794-2350

ABSTRACT

In this paper, several binary mask based Depth From Defocus (DFD) algorithms are proposed to improve autofocusing performance and robustness. A binary mask is defined by thresholding image Laplacian to remove unreliable points with low Signal-to-Noise Ratio (SNR). Three different DFD schemes-- with/without spatial integration and with/without squaring-- are investigated and evaluated, both through simulation and actual experiments. The actual experiments use a large variety of objects including very low contrast Ogata test charts. Experimental results show that autofocusing RMS step error is less than 2.6 lens steps, which corresponds to 1.73%. Although our discussion in this paper is mainly focused on a spatial domain method STM1, this technique should be of general value for different approaches such as STM2 and other spatial domain based algorithms.

Keywords: Autofocusing, Depth From Defocus (DFD), binary mask, reliable point, spatial integration, squaring scheme

1. INTRODUCTION

Passive techniques of ranging or determining the distance of objects from a camera is an important problem in computer vision. Depth From Focus (DFF) [1, 2] is essentially a parameter searching procedure which requires acquiring and processing many images. The search involves many mechanical motions of camera parts, thus limiting the speed of autofocusing. Depth From Defocus (DFD) is an elegant passive autofocusing method. It needs only two or three images, and recovers the depth information by computing the degree of blur.

DFD methods for arbitrary objects have been proposed by some researchers. They can be classified as frequency domain approaches [3~5], spatial domain approaches [6~10] and statistical approaches [11, 12]. The frequency domain approaches generally need more computation and yield lower depth-map density than spatial domain methods. Statistical approaches normally involve optimization operation which requires more images and computing.

Spatial domain approaches have the inherent advantage of being local in nature. They use only a small image region and yield a denser depth-map than frequency domain methods. Therefore, they are better for some applications such as continuous focusing, object tracking focusing, etc. Subbarao [6] proposed a Spatial-domain Convolution/Deconvolution Transform (S Transform) for n-dimensional signals for the case of arbitrary order polynomials. Surya and Subbarao [7] presented STM to estimate the blur parameter in the spatial domain. There are two basic variations: STM1 changes the lens position or focus step and STM2 varies the aperture diameter. Ziou [10] fitted the images by a Hermite polynomial basis. They show that any coefficient of the Hermite polynomial computed using the more blurred image is a function of the partial derivatives of the other image and the blur difference.

In this paper, new binary mask based STM1 algorithms are proposed to improve autofocusing performance and robustness for arbitrary scenes. A binary mask is defined by thresholding image Laplacian to remove unreliable points with low Signal-to-Noise Ratios (SNR). Three different schemes -- with/without spatial integration and with/without squaring -- are investigated and evaluated, through both simulations and actual experiments. Experimental results show that the RMS step error for autofocusing is less than 2.6 lens steps, which is corresponds to 1.73%. Although our

---

[*] E-mail: {txian, murali}@ece.sunysb.edu; Tel: 1 631 632-9149; WWW: www.ece.sunysb.edu/~cvl

discussion in this paper is mainly deals with STM1, this technique should be useful in STM2 and other spatial domain methods.

2. STM OVERVIEW

The basic theory of STM is briefly reviewed here to introduce relevant formulas and to define the terms for future discussion.

2.1. S Transform

A Spatial-domain Convolution/Deconvolution Transform (S Transform) has been developed for images and n-dimensional signals for the case of arbitrary order polynomials [6].

If $f(x,y)$ is an image that is a two-dimensional cubic polynomial defined by:

$$f(x,y) = \sum_{m=0}^{3} \sum_{n=0}^{3-m} a_{mn} x^m y^n \tag{1}$$

where a_{mn} are the polynomial coefficients. The restriction on the order of f is made to be valid by applying a polynomial fitting least square smoothing filter to the image.

Let $h(x,y)$ be a rotationally symmetric point spread function (PSF). In a small region of the image detector plane, the camera system acts as a linear shift invariant system. The observed image $g(x,y)$ is the convolution of the corresponding focused image $f(x,y)$ and the PSF of the optical system $h(x,y)$:

$$g(x,y) = f(x,y) \otimes h(x,y) \tag{2}$$

where $\otimes$ denotes the convolution operation.

A spread parameter σ_h is used to characterize the different forms of PSF. It can be defined as the square root of the second central moment of the function h. For a rotationally symmetric function, it is given by:

$$\sigma_h^2 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x^2 + y^2) h(x,y) dx dy \tag{3}$$

The above deconvolution formula can be written as:

$$f(x,y) = g(x,y) - \frac{\sigma_h^2}{4} \nabla^2 g(x,y) \tag{4}$$

For simplicity, the focused image $f(x,y)$ and defocused images $g_i(x,y), i=1,2$ are denoted as f and g_i from now on.

2.2. STM Autofocusing

p: Object Point   OA: Optical Axis   ID: Image Detector
LF: Light Filter   P1: First Principal Plane s, f, D: Camera Parameters
AS: Aperture Stop Pn: Last Principal Plane v: Dist of Image Focus
L1: First Lens   Q1: First Principal Point p': Focused Image
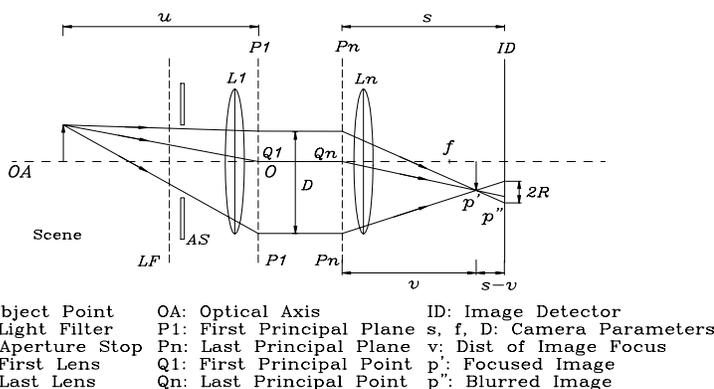Ln: Last Lens   Qn: Last Principal Point p": Blurred Image

Figure 1. Schematic diagram of camera system

A schematic diagram of a camera system is shown in Fig. 1. The Aperture Stop (AS) is the element of the imaging system which physically limits the angular size of the cone of light accepted by the system. The field stop is the element that physically restricts the size of the image. The entrance pupil is the image of the aperture stop (AS) as viewed from the object space, formed by all the optical elements preceding it. It is the effective limiting element for the angular size of the cone of light reaching the system. Similarly, the exit pupil is the image of aperture stop, formed by the optical elements following it. For a system of multiple groups of lenses, the focal length will be the effective focal length f_{eff};

the object distance u will be measured from the first principal point (Q_1), the image distance v and the detector distance s will be calculated from the last principal point (Q_n). Imaginary planes erected perpendicular to the optical axis at these points are known as the first principal plane (P_1) and the last principal plane (P_n) respectively.

If an object point p is not focused, then a blur circle p'' is detected on the image detector plane. The radius of the blur circle can be calculated:

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

where f is the effective focal length, D is the diameter of the system aperture, R is the radius of the blur circle, u, v, and s, are the object distance, image distance, and detector distance respectively. The sign of R here can be either positive or negative depending on whether $s \geq v$ or $s < v$.

After magnification normalization, the normalized radius of blur circle can be expressed as a function of camera parameter setting $\vec{e}$ and object distance u as

$$R'(\vec{e},u) = \frac{Rs_0}{s} = \frac{Ds_0}{2}\left[\frac{1}{f} - \frac{1}{u} - \frac{1}{s}\right] \tag{6}$$

Therefore from Eqn. (6) we have

$$\sigma = mu^{-1} + c \tag{7}$$

where

$$m = -\frac{Ds_0}{2k} \text{ and } c = -\frac{Ds_0}{2k}\left[\frac{1}{f} - \frac{1}{s}\right] \tag{8}$$

Let g_1 and g_2 be the two images of a scene for two different parameter settings $\vec{e}_1 = (s_1, f_1, D_1)$ and $\vec{e}_2 = (s_2, f_2, D_2)$.

$$\sigma_i = m_i u^{-1} + c_i, \quad i = 1,2 \tag{9}$$

Rewriting Eqn. (9) by eliminating u^{-1}:

$$\sigma_1 = \alpha\sigma_2 + \beta \tag{10}$$

where

$$\alpha = \frac{m_1}{m_2} \text{ and } \beta = c_1 - c_2\frac{m_1}{m_2} \tag{11}$$

From Eqn. (4), for each defocused image we can obtain:

$$f = g_i - \frac{\sigma_i^2}{4}\nabla^2 g_i, \quad i = 1,2 \tag{12}$$

Then equating the right side of Eqn. (12):

$$g_1 - \frac{\sigma_1^2}{4}\nabla^2 g_1 = g_2 - \frac{\sigma_2^2}{4}\nabla^2 g_2 \tag{13}$$

Under the third order polynomial assumption: $\nabla^2 g_1 = \nabla^2 g_2$. Therefore, in Eqn. (13), the $\nabla^2 g_1, \nabla^2 g_2$ can be replaced by: $\nabla^2 g = (\nabla^2 g_1 + \nabla^2 g_2)/2$ and rewritten as:

$$g_1 - g_2 = \frac{1}{4} G \nabla^2 g \tag{14}$$

where

$$G = \sigma_1^2 - \sigma_2^2 = \frac{4(g_1 - g_2)}{\nabla^2 g} \tag{15}$$

Now substituting for σ_1 in terms of σ_2 using Eqn. (10) into Eqn. (14), and using the definition of G in Eqn. (15), we have:

$$\sigma_2^2 (\alpha^2 - 1) + 2\alpha\beta\sigma_2 + \beta^2 = G \tag{16}$$

where the definition of α and β are same as in Eqn. (11).

There are two variations of STM, namely STM1 and STM2. In STM1, the lens position is changed during the acquisition of the two images g_1 and g_2. Therefore $\alpha = \frac{m_1}{m_2} = \frac{D_1}{D_2} = 1$, so we have:

$$\sigma_2 = \frac{G}{2\beta} - \frac{\beta}{2} \tag{17}$$

In STM2, only the diameter of camera aperture is changed in the image acquisition of two images g_1 and g_2. In this case we have $\beta = 0$ and $\alpha = \frac{D_1}{D_2}$. Therefore Eqn. (16) reduces to:

$$\sigma_2 = \pm\sqrt{\frac{G}{\alpha^2 - 1}} \tag{18}$$

3. BINARY MASK BASED STM1 ALGORITHMS

Because of camera noise, the original STM1 algorithm of Subbarao and Surya [7] uses the steps of squaring, spatial integration, and mode selection of histogram. This was done as the cameras used in the past were of poor quality compared to the modern digital still cameras. With the development of digital cameras, the traditional scheme should be revisited, and simpler schemes can be implemented for DFD autofocusing. These schemes improve robustness and performance of autofocusing. In this paper, we concentrate on STM1, but the same techniques can be applied to STM2.

3.1 Defining the binary mask

In previous mode selection, the histogram is built by computing σ_2 at each pixel in a 48*48 neighborhood, and then the mode of the histogram is regarded as the best estimation of σ_2. Low contrast image regions yield low image Laplacian values. Due to camera noise and quantization, Laplacian estimates have very low SNR leading to large error in the estimation of σ_2. Therefore, a new binary mask is introduced to improve the robustness of the STM algorithm. The binary mask is defined by thresholding the Laplacian values. This operation removes unreliable points with low Signal-to-Noise Ratio (SNR). The binary mask is defined as:

$$M_0(x,y) = \begin{cases} 1 & |\nabla^2 g| \geq T \\ 0 & o.w. \end{cases}, \quad (x,y) \in W \tag{19}$$

where T is a threshold on the Laplacian which can be determined experimentally. An averaging of σ_2 based on the binary mask is used instead of the mode of its histogram.

The Binary Mask based STM1 With Squaring and With Integration (BM_WSWI) can be expressed as:

$$G = S(g_1, g_2) \frac{4}{U} \sum \sum_{(x,y) \in W} M_0(x,y) \sqrt{\frac{\iint [g_1(x,y) - g_2(x,y)]^2 \, dxdy}{\iint [\nabla^2 g(x,y)]^2 \, dxdy}} \tag{20}$$

where $U = \sum \sum_{(x,y)\in W} M_0(x,y)$ is the weight of the binary mask. $S(g_1,g_2)$ is the sign function which is decided by the variance of two images $Var(g_1)$ and $Var(g_2)$.

$$S(g_1,g_2) = \begin{cases} -1 & , \quad Var(g_1) \geq Var(g_2) \\ +1 & , \quad Var(g_1) < Var(g_2) \end{cases} \tag{21}$$
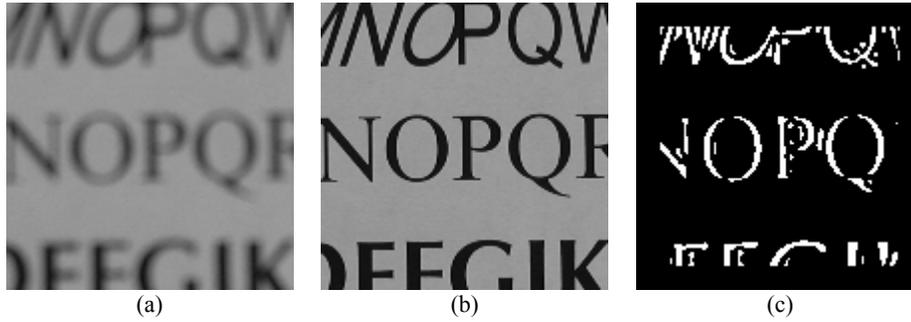

(a)   (b)   (c)

Figure 2. Binary Mask Formation
(a) Focusing window at lens step 35 (b) Focusing window at lens step 98
(c) Corresponding binary mask

3.2 Spatial integration
Spatial integration reduces random noise at the cost of sacrificing spatial resolution; moreover, without thresholding, it may take some unreliable points into account.

To understand the effect of spatial integration, a variation — Binary Mask based STM1 With Squaring and Without Integration (BM_WSOI), which does not integrate over a small region is calculated by:

$$G = S(g_1,g_2)\frac{4}{U}\sum\sum_{(x,y)\in W} M_0(x,y)\sqrt{\frac{[g_1(x,y)-g_2(x,y)]^2}{[\nabla^2 g(x,y)]^2}} \tag{22}$$

3.3 Squaring scheme
The squaring scheme was introduced in the original STM1 algorithm to reduce the effects of camera noise. Squaring permitted integration over an image region without canceling positive and negative values of image Laplacians during summation. However, squaring also loses the sign information. Therefore the sign function $S(g_1,g_2)$ in Eqn. (21) is used. However, integration over an image region reduces the spatial resolution of depth-map.

Another variation directly uses Eqn. (15). It is carried out without squaring and without integration. In the Binary Mask based STM1 of Without Squaring Without Integration (BM_OSOI), the average of G is calculated based on the binary mask:

$$G = \frac{4}{U}\sum\sum_{(x,y)\in W} M_0(x,y)\frac{[g_1(x,y)-g_2(x,y)]}{\nabla^2 g(x,y)} \tag{23}$$

The above algorithms are evaluated on both synthetic and real data.

4. SIMULATION

In practical experiments, there are many factors that are coupled together even in a single recorded image, such as lens aberration, vignetting, nonlinear sensor response, automatic gain control, and automatic white balance. In order to verify the theory itself and evaluate the different variations under the same conditions, a computer simulation system Image Defocus Simulator 1 (IDS1) is implemented for generating a series of test images. Due to advances in VLSI technology,

digital still cameras have improved in imaging capabilities compared to the video camera-image grabber architecture of the past. The original IDS of Lu [13, 14] has been simplified and updated to model a modern digital still camera.

A database of simulated images has been built for experiments. Images for synthetic database are displayed in Fig. 3. Fig. 4 shows 9 images of Boat arranged in 3 rows and 3 columns. The distance of the object increased (250 mm-950 mm) row-wise from top to bottom. The distance s between the lens and the image detector increases column-wise from left to right. The effective focal length and the F-Number are fixed at 19.5 mm and 2.8 respectively. In Fig. 4, the images are focused somewhere along the top-left to bottom-right diagonal direction. The image focus decreases on either side of the diagonal direction. This is consistent with the fact that image blur should increase when either the object is moved farther or closer from its focused position, or when the image detector is moved farther or closer from its focused position.

To compare the performance of BM_WSWI, BM_WSOI and BM_OSOI, The focal length of the camera is 19.5 mm, and the F number is set to 2.8. For each algorithm, a focusing window whose size is 144x144 is placed at the center of the test image. The size of the Gaussian filter and LoG filter are both 15x15 pixels. The sigma table corresponding BM_WSWI, BM_WSOI and BM_OSOI are shown in Fig. 5(a), (b) and (c) respectively.

The rms step errors of the three variations are compared in Fig 5(d). Comparing Fig. 5 (a) and (b), the scheme using spatial integration (BM_WSWI) has higher RMS step error than the scheme without spatial integration (BM_WSOI) in the range 270 mm to 2200 mm. The RMS step error of BM_WSOI increases dramatically at the far field and near field. From Fig. 5 (b) and (c), variations BM_WSOI and BM_OSOI behave similar in RMS step error although some large step errors happen in some situations for BM_WSOI.

Another observation from Fig. 5 is that the RMS step error can be limited to be low at long range by suitably selecting the step interval for image capture or by using a third image.

5. EXPERMINTAL RESULTS

Binary Mask STM algorithms described above are implemented on the Olympus C3030 camera. The camera is controlled by a host computer (Pentium 4 2.4GHz) from a USB port. The lens' focus motor of C3030 ranges in position from step 0 to step 150; Step 0 corresponds to focusing a nearby object at a distance of about 250 mm from the lens and Step 150 corresponds to focusing an object at a distance of infinity.

The lens designs in current digital cameras have several focusing modes such as Macro mode and Standard mode to improve the autofocusing performance for different distance range. The relative position of lens elements changes when the focusing mode switches from one to another. The relationship between the lens step number and the reciprocal of the best-focused distance is no longer linear, and the practical object distance vs. focus step curve needs to be measured using Depth From Focus (DFF) technique. A double three step DFF algorithm is used to avoid local maxima in the best focus measure searching procedure [2]. The results are shown in Fig. 6. There exist two roughly linear segments, the first for Macro mode and the second for Standard mode. The transition area is around 787 mm (31").

To generate the sigma-step lookup tables for different variations, the defocused images of a calibration object are acquired at 20 different distances. At each distance, two defocused images are obtained with the focus step numbers 35, and 98. Then σ_2 is estimated by the algorithms — BM_WSWI, BM_WSOI and BM_OSOI. The results are plotted in Fig. 7. There are common "flat" areas in the sigma-step curves that approximately lie in the range from Step 60 to Step 110 for Olympus C3030. In this area, we find that a small variation in sigma may cause a large fluctuation in the focusing step. However the corresponding change in focusing measure or image sharpness is not significant. This means that if we use focus step number as error metric, large errors are expected in the step 60 to 110 interval. This is misleading since the error in image sharpness will be small. Therefore, focus step is not the best error metric as it overstates the error, but focus measure difference is a better measure of performance as it corresponds to image sharpness in autofocusing. However, to be conservative, we use the metric of lens step number with a hopeful note that the actual defocus error will be lower.
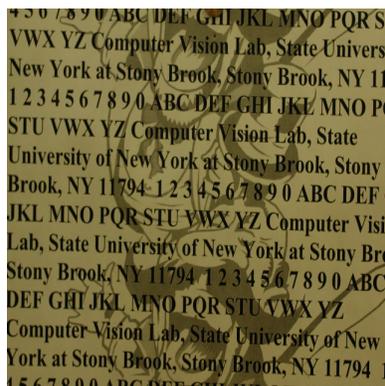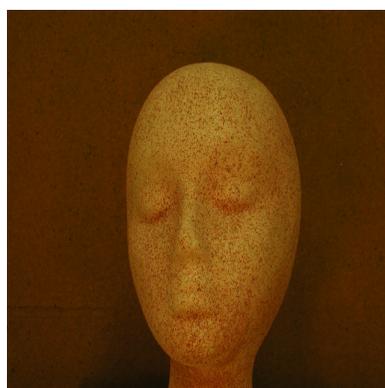
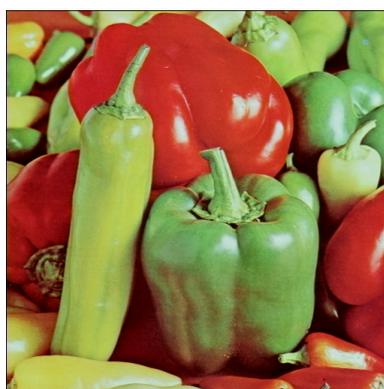**Figure 3.** Image database for IDS1
(a) Letter (b) DrawLetter (c) CD Rom (d) Vacuum
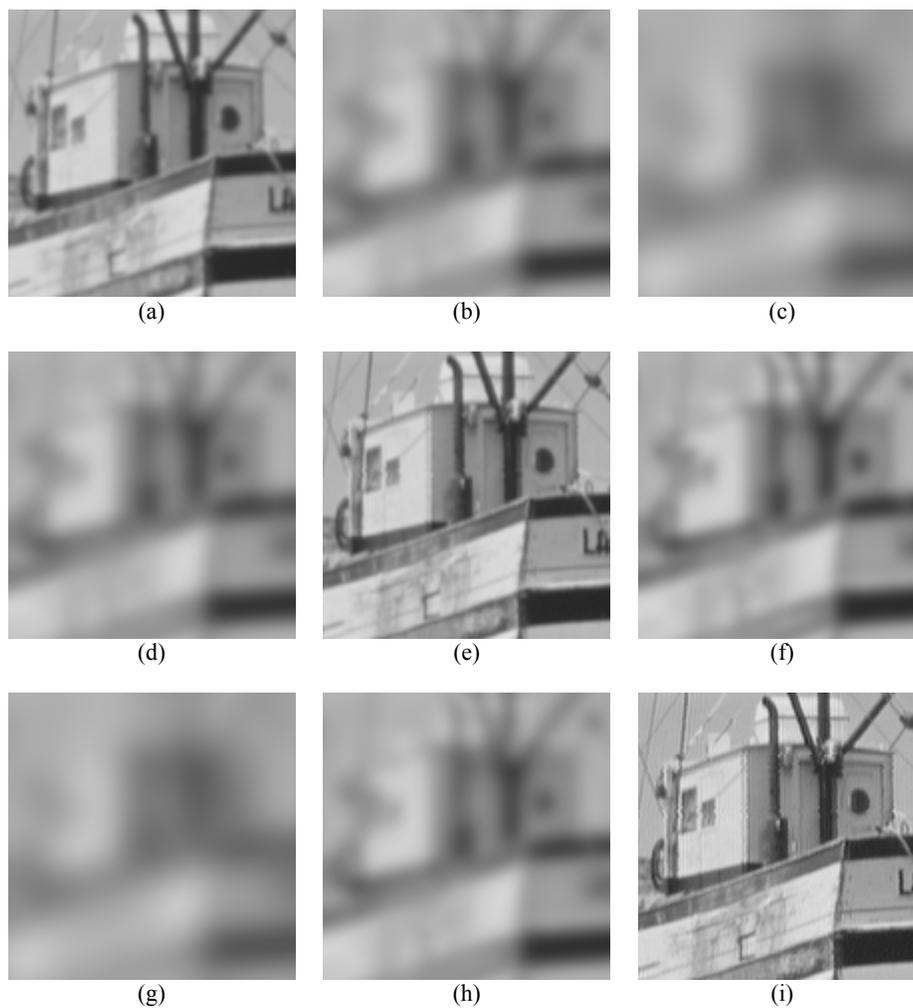(e) Head (f) Monarch (g) Peppers (h) Boat (i) Lena

Figure 4. Sample output of IDS1 simulation system.
Object distance u=250 mm, (a) step= 40, (b) step=80, (c) step=120;
When u=450 mm, (d) step= 40, (e) step=80, (f) step=120;
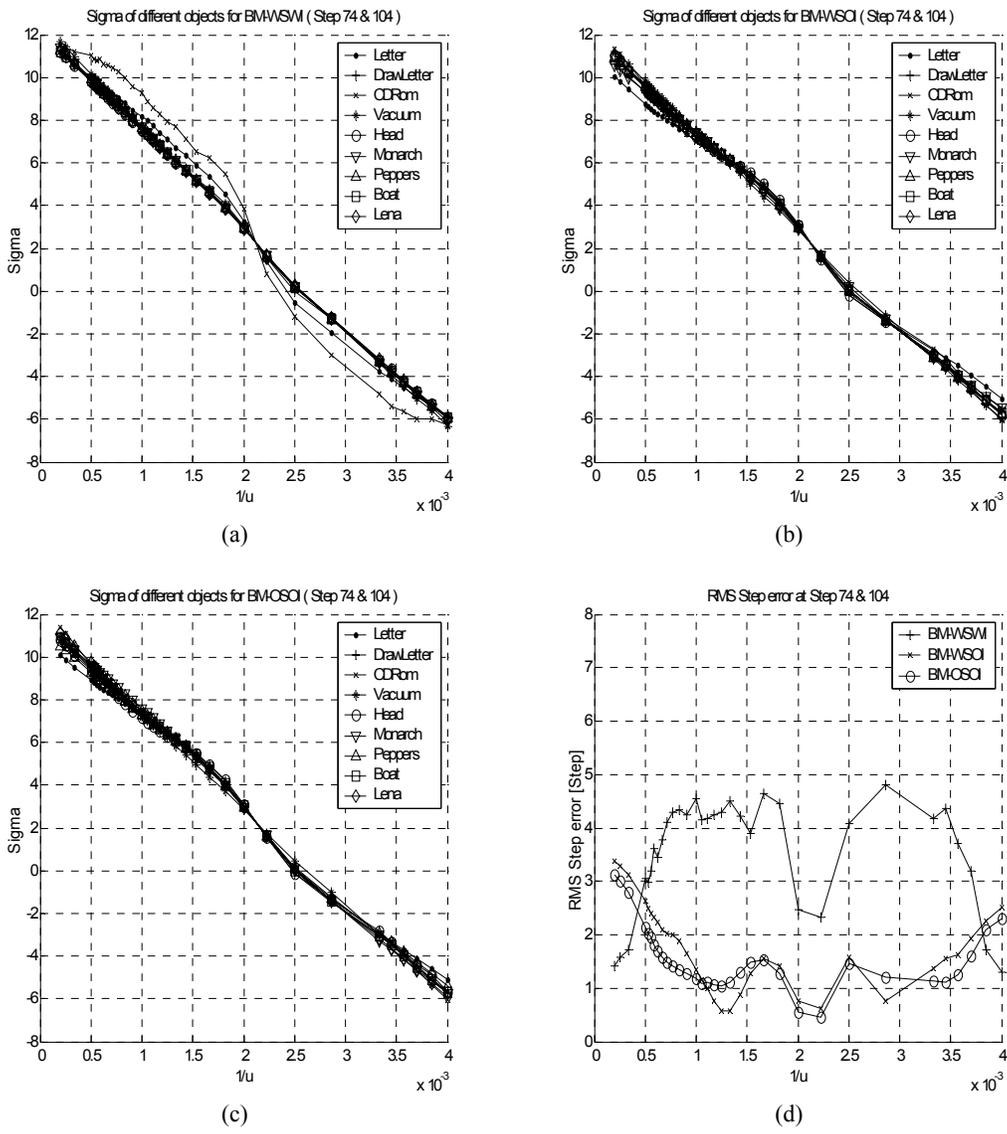When u=950 mm, (g) step= 40, (h) step=80, (i) step=120.

Figure 5. Sigma Table and RMS step error of different algorithms (simulation)
(a) BM_WSWI (b) BM_WSOI (c) BM_OSOI (d)RMS step error for (a), (b), (c)
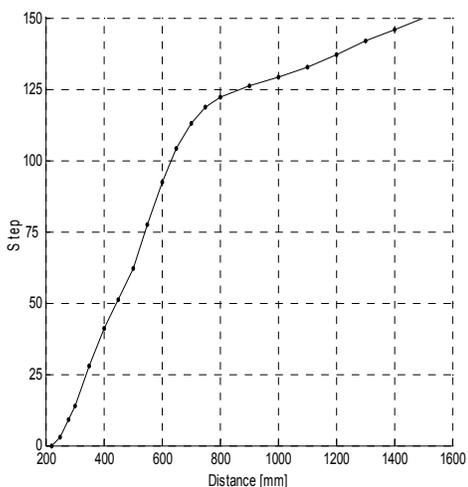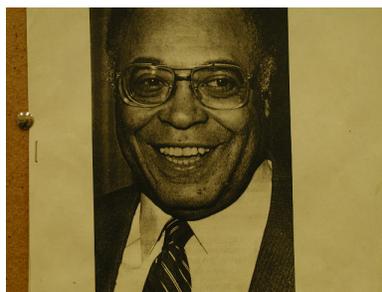
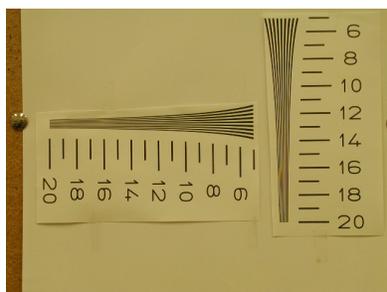**Figure 6.** Step number vs. object distance
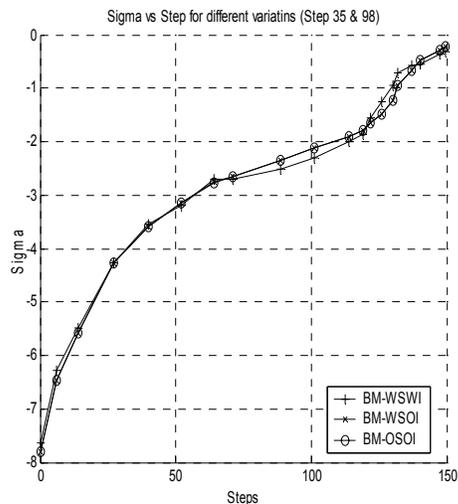
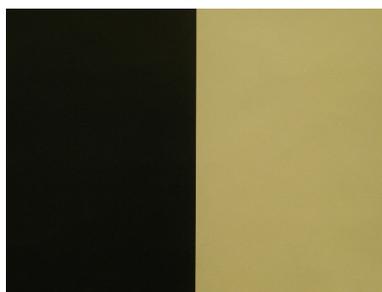**Figure 7.** Sigma vs. step number
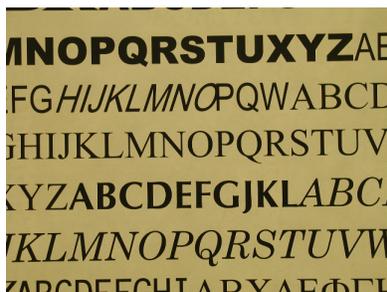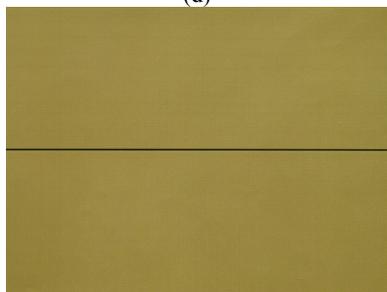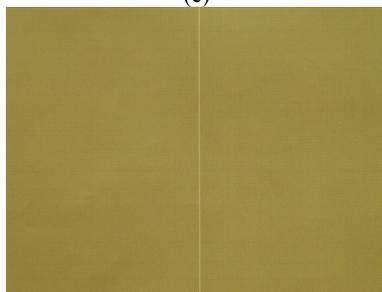

(a)

(b)

(c)

(d)

(e)

(f)

(g)

(h)

Figure 8. Test Objects
(a) Letter (b) Head (c) DVT (d) Chart (e) Ogata Chart 1
(f) Ogata Chart 2 (g) Ogata Chart 3 (h) Ogata Chart 4

To bring out the capability of DFD algorithms, experiments are performed on eight objects that are relatively difficult to measure, as shown in Fig. 8. Eight positions are randomly selected. The distance and the corresponding steps are listed in Table 1. One of the test objects at those positions are shown in Fig. 9. The F-number is set to 2.8, and the focal length is set to 19.5 mm. A focusing window is located at the center of the scenes. Its window size is 96x96. The Gaussian smoothing filter and LoG filter are 9x9 pixels. The sensor nonlinear response compensation [15] is utilized.

Measurement results and rms error for (a) BM_WSWI (b) BM_WSOI, and (c) BM_OSOI are plotted in Fig. 10 and Fig. 11 respectively. Comparing the schemes, With Spatial Integration (BM_WSWI) perform better than the schemes WithOut Spatial Integration (BM_WSOI and BM_OSOI) at far field positions (position 8), but sometimes they may give large errors due to unreliable points. The schemes without squaring perform better than schemes with squaring at some positions (position 2). This can also be observed from simulation results in the previous sections.

Errors at position 8 could be large as the first image processed in DFD will be highly blurred (first image is captured at a lens position of 35 whereas the focused step position is around 145). In this case, a third image closer to 145 should be recorded and processed for better accuracy. Taking all factors into account such as accuracy, computational requirements, simplicity of algorithm, resolution of depth-map, etc., we suggest BM_OSOI for use in practical applications. Even when very low contrast objects such as those in Fig. 8g and 8h are present, an RMS error of about 3 steps can be expected which gives very sharp focused images in autofocusing applications.

6. CONCLUSION

A new binary mask is defined based on thresholding of image Laplacian to remove unreliable points with low Signal-to-Noise Ratio (SNR) in DFD applications. This mask is exploited in different DFD schemes such as with/without spatial integration and with/without squaring, and their performances are investigated and evaluated both with simulation and actual experiments. Experimental results show that the autofocusing RMS step error is roughly similar for the different schemes. However, taking several factors such as accuracy and computational resources into account suggests that the DFD scheme of without squaring and without spatial integration (BM_OSOI) is best suited for practical applications. While this paper deals with STM1, the conclusions here should be applicable to other spatial domain DFD methods such as STM2.
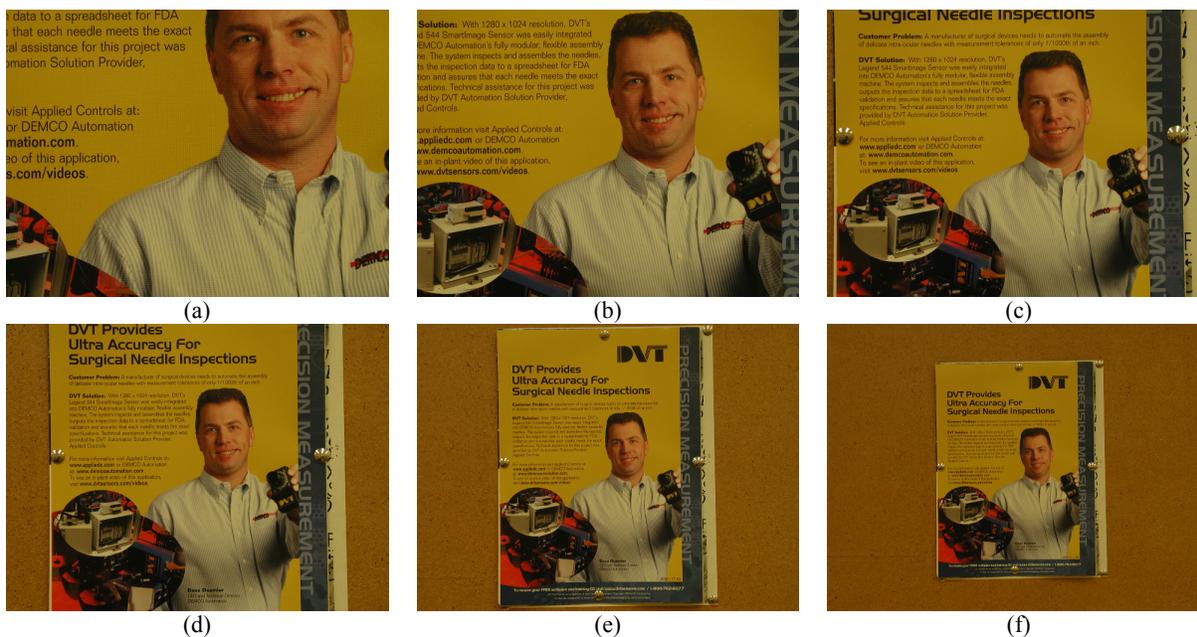

Figure 9. Test object at different positions
(a)Position 1 (b) Position 2 (c) Position 3
(d) Position 4 (e) Position 6 (f) Position 8

	Pos. 1	Pos. 2	Pos. 3	Pos. 4	Pos. 5	Pos. 6	Pos. 7	Pos. 8
Distance [mm]	325.2	473.5	626.1	782.8	913.1	1055.3	1232.6	1350.7
Step	19.00	55.00	96.50	120.50	126.00	131.25	139.00	144.75

Table 1. Object Positions in DFD Experiment
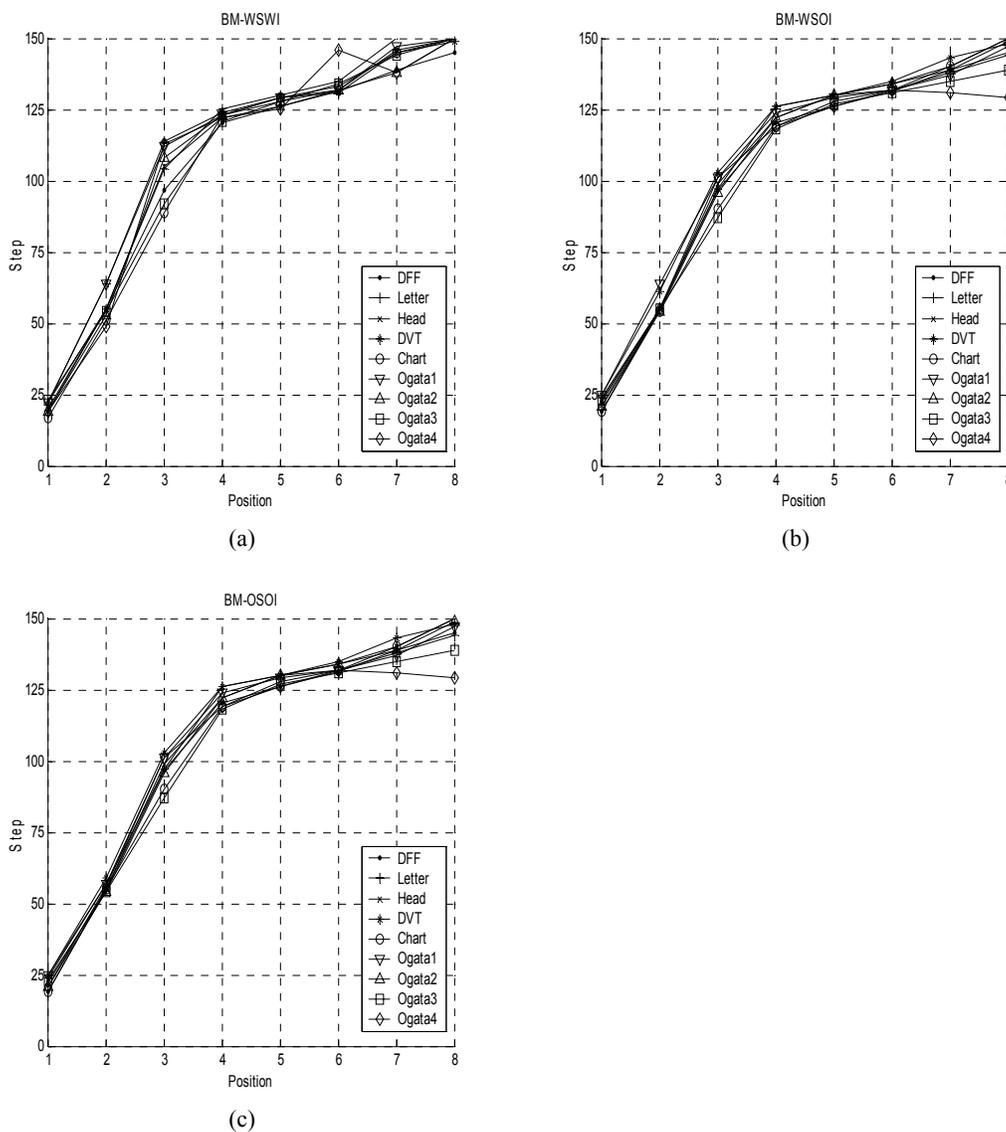

(a)

(b)

(c)

Figure 10. Measure result for different algorithms
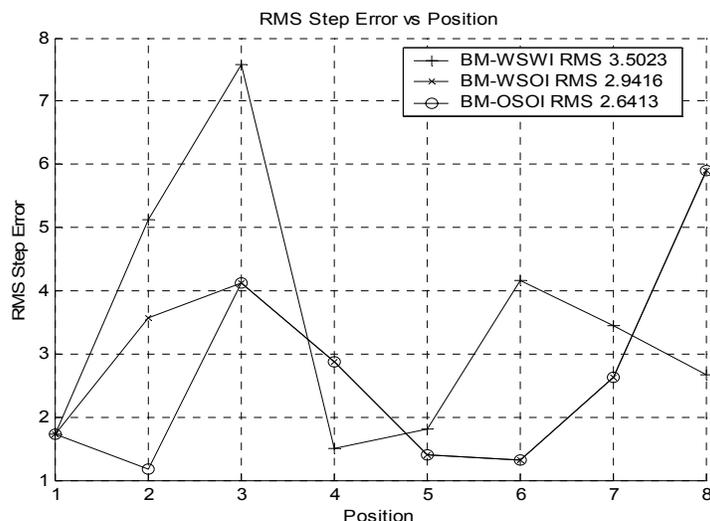(a) BM_WSWI (b) BM_WSOI (c) BM_OSOI

Figure 11. RMS step error for different algorithms (real data)
(a) BM_WSWI (b) BM_WSOI (c) BM_OSOI

REFERENCES

1. Eric P. Krotov, "Focusing", *International Journal of Computer Vision*, p223-237, 1987
2. M. Subbarao, T. S. Choi, and A. Nikzad, "Focusing Techniques," *Optical Engineering*, p2824-2836, 1993.
3. Yalin Xiong, Steve Shafer, "Depth from Focusing and Defocusing", *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, p68-73, 1993
4. P. Penland, "A new sense for depth of field", *IEEE Trans. on Pattern Analysis and Machine Intelligence,* Vol. 9(4), p.523-531, 1987
5. M. Watanabe and S. K. Nayar, "Rational Filters for Passive Depth from Defocus", *International Journal of Computer Vision*, Vol.27, No.3, p. 203-225, 1998
6. M. Subbarao, "Spatial-Domain Convolution/Deconvolution Transform", Tech. Report 91.07.03, Computer Vision Laboratory, Dept. of Electrical Engineering, SUNY at Stony Brook.
7. M. Subbarao and G. Surya, "Depth from Defocus: A Spatial Domain Approach", *International Journal of Computer Vision*, Vol. 13, No. 3, p. 271-294,1994
8. M. Subbarao, and G. Surya, "Depth from Defocus by Changing Camera Aperture: A Spatial Domain Approach", *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, p. 61-67, New York, 1993
9. M. Subbarao, T. Wei, and G. Surya, "Focused Image Recovery from Two Defocused Images Recorded with Different Camera Settings", *IEEE Trans. on Image Processing*, Vol. 4, No. 12, 1995
10. Djemel Ziou, Francois Deschênes, "Depth from Defocus Estimation in Spatial Domain", *Computer Vision and Image Understanding*, Vol.81 No. 2, p.143-165, 2001
11. A.N. Rajagooalan, S. Chaudhuri, "Recursive computation of maximum likelihood function for blur identification from multiple observations", *IEEE Trans. Image Processing*, Vol.7 No. 7, pp 1075-1079, 1998
12. A.N. Rajagooalan, S. Chaudhuri, "An MRF Model-Based Approach to Simultaneous Recovery of Depth and Restoration from Defocused Images", *IEEE Trans. on Pattern Analysis and Machine Intelligence,* Vol. 21(7), p.577-589, 1999
13. M. Subbarao and A. Nikzad, "A Model for Image Sensing and Digitization in Machine Vision," *Proc. of SPIE*, Vol.1385, p70-84, 1990
14. M. Subbarao, and M. C. Lu, "Computer Modeling and Simulation of Camera Defocus," *Machine Vision and Applications*, Vol.7, p277-289, 1994
15. T. Xian and M. Subbarao, "Camera Calibration and Performance Evaluation of Depth From Defocus (DFD)", *Proc. of SPIE*, Boston, Oct., 2005