

Depth-from-Defocus: Blur Equalization Technique

Tao Xian and Murali Subbarao

Dept. of Electrical and Computer Engineering
State Univ. of New York at Stony Brook, NY 11794-2350

ABSTRACT

A new spatial-domain Blur Equalization Technique (BET) is presented. BET is based on Depth-from-Defocus (DFD) technique. It relies on equalizing the *blur* or *defocus* of two different images recorded with different camera parameters. Also, BET facilitates modeling of images locally by higher order polynomials with lower series truncation errors. The accuracy of BET is further enhanced by discarding pixels with low Signal-to-Noise ratio by thresholding image Laplacians, and relying more on sharper of the two blurred images in estimating the blur parameters. BET is found to be superior to some of the best comparable DFD techniques in a large number of both simulation and actual experiments. Actual experiments used a large variety of objects including very low contrast digital camera test charts located at many different distances. In autofocusing experiments, BET gave an RMS error of 1.2% in lens position.

Keywords: Depth-from-Defocusing, spatial-domain, Blur Equalization Technique

1. INTRODUCTION

Passive techniques of ranging from a camera is an important problem in computer vision. 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 investigated by many researchers. There are two general types of DFD algorithms: statistical and deterministic. Maximum likelihood and Markov Random field methods have been used in the statistical category require more images and computing. Deterministic algorithms can be classified as frequency domain approaches,¹⁻⁴ and spatial domain approaches.⁵⁻⁸ The frequency domain approaches are generally computation expensive and yield lower depth-map density.

The 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 the frequency domain counterparts. They can be used for the analysis of shift-variant defocused images using a piecewise shift-invariance approximation. Therefore, they are better suited in many applications in 3D vision, and in digital imaging (autofocusing). Spatial domain methods require less computing than Fourier domain methods. Therefore they are preferred in real-time applications such as tracking and autofocusing of moving objects.

A Spatial-domain Convolution/Deconvolution Transform (S Transform) for n-dimensional signals for the case of arbitrary order polynomials was proposed.⁹ Surya and Subbarao⁵ utilized S Transform to estimate the blur parameter in the spatial domain. There are two basic variations of STM: STM1 involves the change of lens position and STM2 involves the change of aperture diameter. Ziou and Deschenes⁷ fitted the images to a Hermite polynomial basis. They show that any coefficient of the Hermite polynomial computed using the blurred image is a function of the partial derivatives of the other image and the blur difference. Favaro⁸ use a singular value decomposition technique to estimate the operator that spans a left null space from a number of deblurred images.

In this paper, we present a new spatial-domain depth-from-defocus method named Blur Equalization Technique (BET). BET relies on equalizing the *blur* or *defocus* of two different images recorded with different camera

Further author information:

T. X.: E-mail: txian@metrologic.com

M. S.: E-mail: murali@ece.sunysb.edu

parameters. It uses the S transform in the convolution mode. The theory underlying BET and experimental results are presented. The accuracy of BET is further enhanced by discarding pixels with low Signal-to-Noise ratio through thresholding image Laplacians, and relying more on sharper of the two blurred images in estimating the blur parameters. These steps, combined with careful calibration for sensor response, vignetting, and magnification correction, makes BET a very accurate and robust technique useful in practical applications such as depth recovery and autofocusing. Its performance is compared with three highly effective and significantly improved versions of STM1.⁶ BET is found to have more accurate and robust.

Since BET is related to STM, the basic principles of STM are briefly reviewed in Section 2. The theory and computational algorithm for BET is presented in Section 3. In Section 4, experimental results are presented.

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

S Transform has been developed for images and n-dimensional signals for the case of arbitrary order polynomials.⁹

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

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

where a_{mn} are the polynomial coefficients. The restriction on the order of f can be satisfied by applying a polynomial fitting least square smoothing filter to the image. In practice, it is found to be better to apply a Gaussian smoothing filter.

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 $h(x, y)$:

$$g(x, y) = f(x, y) \otimes h(x, y) \quad (2)$$

A spread parameter σ_h is used to characterize the different forms of PSF. For a rotationally symmetric function, it is given by:

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

The deconvolution formula based on S Transform⁹ can be written as:

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

For simplicity, the focused image $f(x, y)$ and defocused images $g(x, y)$ are denoted by f and g respectively.

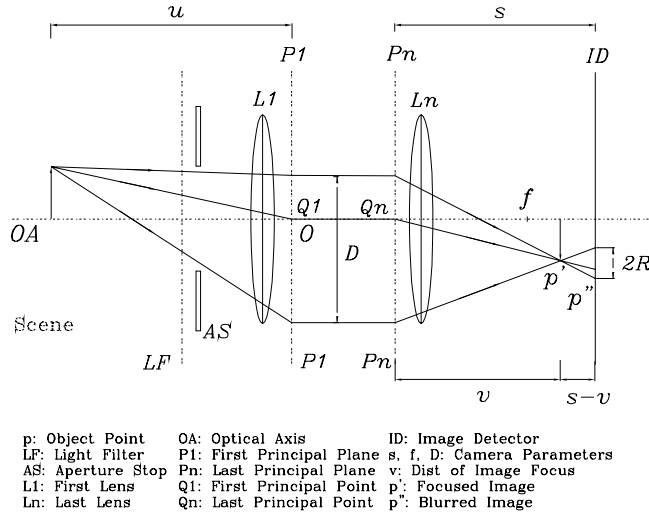


Figure 1. Schematic diagram of camera system

2.2. Depth-from-Defocus: STM

A schematic diagram of a camera system is shown in Fig. 1. 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 to be:

$$R = \frac{Ds}{2} \left(\frac{1}{f} - \frac{1}{u} - \frac{1}{s} \right) \quad (5)$$

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 and object distance u as

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

Therefore from Eqn.(6), and $\sigma = kR'$ for some constant k ($k \approx \sqrt{2}$) we obtain:

$$\sigma = mu^{-1} + c \quad (7)$$

where

$$m = -\frac{kDs_0}{2} \quad \text{and} \quad c = -\frac{kDs_0}{2} \left(\frac{1}{f} - \frac{1}{s} \right) \quad (8)$$

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

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

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

$$\sigma_1 = \alpha \sigma_2 + \beta \quad (10)$$

where

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

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

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

Under the third order polynomial assumption in Eqn. (1), for the same object, we have:

$$g_1 - g_2 = \frac{1}{4} G \nabla^2 g \quad (13)$$

where $\nabla^2 g = \nabla^2 g_1 = \nabla^2 g_2$, and

$$G = \sigma_1^2 - \sigma_2^2 = \frac{4(g_1 - g_2)}{\nabla^2 g} \quad (14)$$

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

$$\sigma_2^2(\alpha^2 - 1) + 2\alpha\beta\sigma_2 + \beta^2 = G \quad (15)$$

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

In STM1, lens position is changed during the acquisition of the two images g_1 and g_2 (STM1), but the aperture diameter remains unchanged. Therefore $\alpha = \frac{m_1}{m_2} = \frac{D_1}{D_2} = 1$, and we obtain:

$$\sigma_2 = \frac{G}{2\beta} - \frac{\beta}{2} \quad (16)$$

3. BLUR EQUALIZATION TECHNIQUE

3.1. Blur Equalization

In STM, the spatial-domain deconvolution equation Eqn. (12) is used to eliminate the focused image term to obtain Eqn. (15) for the blur parameter. This is equivalent to deblurring the two defocused images g_1 and g_2 and equating the resulting focused image f . This operation may be termed *deblur* or *focus* equalization. Since this deconvolution amplifies high-frequency content, it may reduce the Signal-to-Noise Ratio (SNR). This operation, when combined with the cubic-model restriction for f as in Eqn. (1) requires that ($\nabla^2 g_1 = \nabla^2 g_2$). However, in practical applications, when the Laplacians are computed for the two images, they are almost never equal, and sometimes significantly different from each other. Using higher order terms with higher image derivatives may reduce SNR. Therefore, unnecessary errors are introduced in depth recovery, particularly when the object contrast is too low or too high.

BET is an alternative to STM. In BET, in contrast with STM, we use *blur* or *defocus* equalization by convolving the two defocused images with suitable PSFs that correspond to *attenuating* high-frequency content which improves SNR. Remarkably, this removes the requirement of ($\nabla^2 g_1 = \nabla^2 g_2$). Following the previous notation, two defocused images $g_i(x, y), i = 1, 2$ can be expressed as:

$$g_i(x, y) = f(x, y) \otimes h_i(x, y) \quad i = 1, 2 \quad (17)$$

where $h_i(x, y)$ is the PSF of corresponding defocused image at position i . We have

$$g_1(x, y) \otimes h_2(x, y) = [f(x, y) \otimes h_1(x, y)] \otimes h_2(x, y) \quad (18)$$

$$g_2(x, y) \otimes h_1(x, y) = [f(x, y) \otimes h_2(x, y)] \otimes h_1(x, y) \quad (19)$$

From the commutative property of convolution, the right side of Eqn. (18) equals the right side Eqn. (19), that is:

$$g_1(x, y) \otimes h_2(x, y) = g_2(x, y) \otimes h_1(x, y) \quad (20)$$

Using Forward S Transform for convolution:

$$g_1(x, y) \otimes h_2(x, y) = g_1(x, y) + \frac{\sigma_2^2}{4} \nabla^2 g_1(x, y) + \frac{\sigma_2^4}{24} (\nabla^2)^2 g_1(x, y) + R(O^6) \quad (21)$$

$$g_2(x, y) \otimes h_1(x, y) = g_2(x, y) + \frac{\sigma_1^2}{4} \nabla^2 g_2(x, y) + \frac{\sigma_1^4}{24} (\nabla^2)^2 g_2(x, y) + R(O^6) \quad (22)$$

Combining Eqn. (20), (21) and (22), and ignoring the higher order terms $R(O^4, O^6)$, we have:

$$g_1(x, y) + \frac{\sigma_2^2}{4} \nabla^2 g_1(x, y) = g_2(x, y) + \frac{\sigma_1^2}{4} \nabla^2 g_2(x, y) \quad (23)$$

In deriving the above equation, the approximation may be much less than it appears to be, because, we are only neglecting the *difference* of fourth and higher order terms between the left and the right sides of the equation. We are not neglecting the terms on only one side, or separately on the two sides. Experimental results support this observation.

Using Eqn. (9) we get:

$$a_1 \sigma_1^2 + b_1 \sigma_1 + c_1 = 0 \quad (24)$$

where the coefficients are defined as:

$$a_1 = \frac{\nabla^2 g_2}{\nabla^2 g_1} - 1, \quad b_1 = 2\beta, \quad c_1 = - \left[\frac{4(g_1 - g_2)}{\nabla^2 g_1} + \beta^2 \right] \quad (25)$$

3.2. SNR Based Thresholding

Signal-to-Noise Ratio (SNR) of an image is affected by several factors such as object contrast, sensor noise, sampling rate, and quantization error. We found that image Laplacians have very low SNR leading to large errors in the estimation of σ_2 . Therefore, thresholding on the Laplacian values is introduced to improve the robustness. This operation removes unreliable points with low SNR. Laplacian Mask $M_0(x, y)$ is formed by:

$$M_0(x, y) = \begin{cases} 1 & \nabla^2 g_2 \geq T \\ 0 & \text{otherwise} \end{cases} \quad (26)$$

A Delta Mask $M_1(x, y)$ is also used here to guarantees the real property of the solution.

$$M_1(x, y) = \begin{cases} 1 & \Delta_1 \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (27)$$

where $\Delta_1 = b_1^2 - 4a_1c_1$.

The final binary mask $M_{f1}(x, y)$ is obtained from the BIT-AND operation:

$$M_{f1}(x, y) = M_0(x, y) \& M_1(x, y) \quad (28)$$

The computation of σ_1 is guided by $M_{f1}(x, y)$, as displayed in Fig. 2(d). The best estimation of σ_1 is considered as the average based on $M_{f1}(x, y)$.

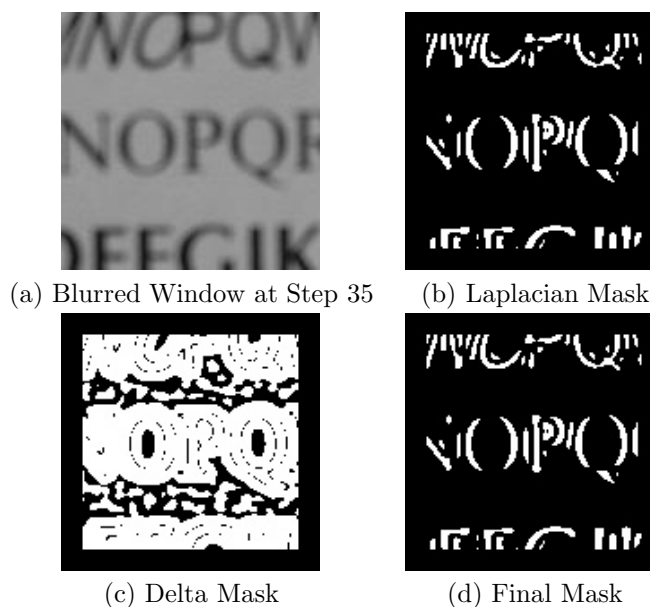


Figure 2. Mask Formation

3.3. Image Switching Based on Sharpness Measure

Another quadratic equation in σ_2 can be derived as:

$$a_2\sigma_2^2 + b_2\sigma_2 + c_2 = 0 \quad (29)$$

and the coefficients are:

$$a_2 = 1 - \frac{\nabla^2 g_1}{\nabla^2 g_2}, \quad b_2 = 2\beta, \quad c_2 = -\left[\frac{4(g_1 - g_2)}{\nabla^2 g_2} + \beta^2\right] \quad (30)$$

Similarly the binary mask $M_{f2}(x, y)$ can be formed.

Mathematically, Eqn. (24)-(25) and Eqn. (29)-(30) are identical. However, computationally we find that the two equations have different noise sensitivity, perhaps due to noisy image Laplacians appearing in the denominator in the equations. It seems that solving for the blur parameter of the sharper or more focused image gives more accurate results. In this case, the absolute value of the image Laplacians appearing in the denominators will be larger than otherwise. A simple focus or sharpness measure is computed for the two images. We define the Sum of Laplacian in the Focusing Window (SLFW) $L_i = \sum_x \sum_y |\nabla^2 g_i(x, y)|$, $i = 1, 2$ as the sharpness measure. The sharper images will have a smaller blur parameter σ . Now the equation to be solved is determined as follows:

$$\begin{cases} a_1\sigma_1^2 + b_1\sigma_1 + c_1 = 0, & L_1 \geq L_2 \\ \sigma_2 = \sigma_1 + \beta \\ a_2\sigma_2^2 + b_2\sigma_2 + c_2 = 0 & L_1 < L_2 \end{cases} \quad (31)$$

A computer simulation is conducted using a high contrast object “Letter”. The reliability measure SLFW and the corresponding switching scheme are plotted in Fig. 3. In Fig. 3 (b), “+” curve is calculated from Eqn. (24)-(25), and “*” curve is obtained by Eqn. (29)-(30).

4. EXPERIMENTAL RESULTS

BET algorithm described above was 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 position ranges from Step 0 to

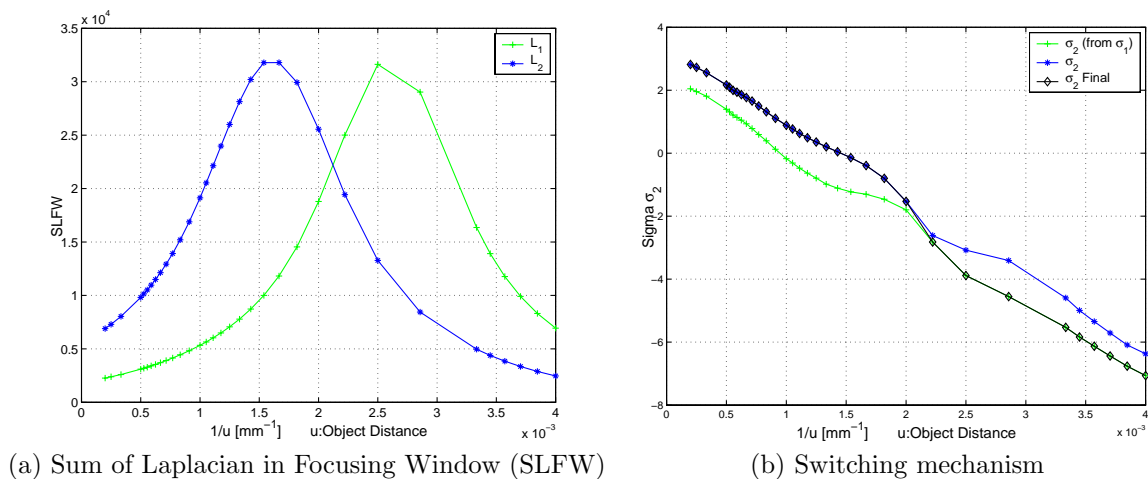


Figure 3. Switching Mechanism

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. Each lens step number corresponds to a unique object distance that will be in best focus for that lens position. In the experiments, the relation between the step number and object distance was obtained using a Depth-from-Focus (DFF) technique.

To generate the sigma-step lookup tables (LUT) for different variations of DFD, the defocused images of a calibration object are acquired at 20 different distances. For each object distance, two defocused images are recorded with lens positions at step numbers 35 and 98. Then σ_2 is estimated by the BET DFD algorithm. The results are plotted in Fig. 4.

The performance of DFD algorithms are evaluated with real experiments using eight different objects shown in Fig. 5. Some of them are standard camera test charts used in evaluating the autofocus performance of new digital camera models by industry. The last two objects are of very low contrast that are very difficult to autofocus. Each object was placed at eight different distances in the range of 325 mm to 1350 mm at roughly 150 mm intervals. The distance and the corresponding steps are listed in Table 1. The steps were obtained using the DFF algorithm. The F-number was set to 2.8, and the focal length was set to 19.5 mm. The focusing window size was 96×96 located at the center of the scene. At the beginning, a Gaussian smoothing filter with a sigma parameter of 1.8 pixels and size 9×9 was used to smooth images. The image Laplacians were also computed using the corresponding LOG filter of size 9×9 . The image Laplacians were thresholded to weed out low contrast pixels with low SNR. The non-linear sensor response was calibrated and corrected as described in.¹⁰ One value of depth or blur parameter sigma was estimated in each pixel. The estimates were averaged in the 96×96 window.

Position	1	2	3	4	5	6	7	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 Experiments

Measurement results and RMS error for BET are plotted in Fig. 6 respectively. In order to compare BET's performance with some of the best existing DFD techniques, three substantially improved variations of STM (see⁶ for details) were tested in the same set of experiments. These variations use "deblur" equalization as opposed to "blur" equalization as explained earlier. The results for BET and the three variations of STM are plotted together in Fig. 6. The RMS step error for BET is 1.8 lens steps, which corresponding 1.2% of step range. This error of 1.8 steps may be close to the best achievable as the step quantization error itself is 0.5 steps,

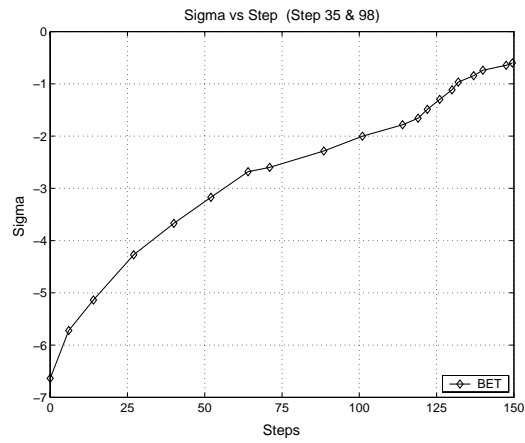
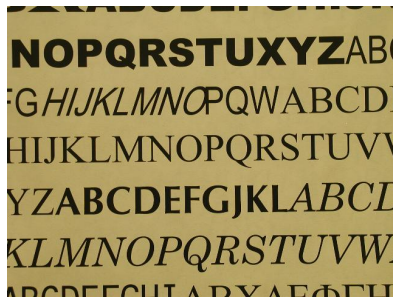
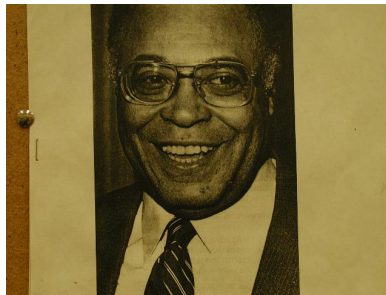


Figure 4. Sigma vs. step number



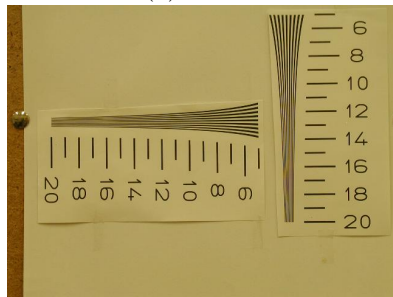
(a) Letter



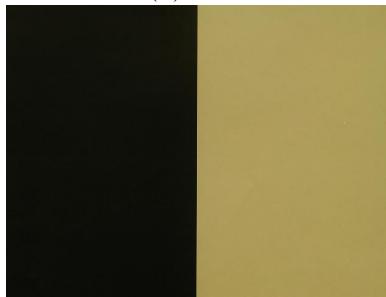
(b) Head



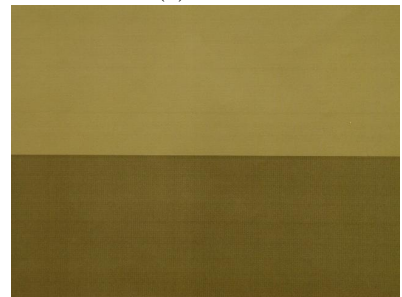
(c) DVT



(d) Chart 1



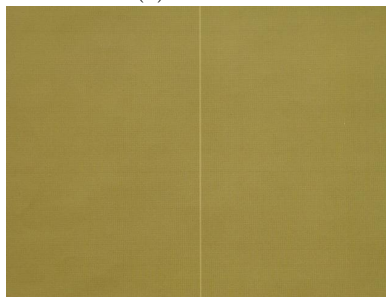
(e) Chart 2



(f) Chart 3



(g) Chart 4



(h) Chart 5

Figure 5. Test Objects

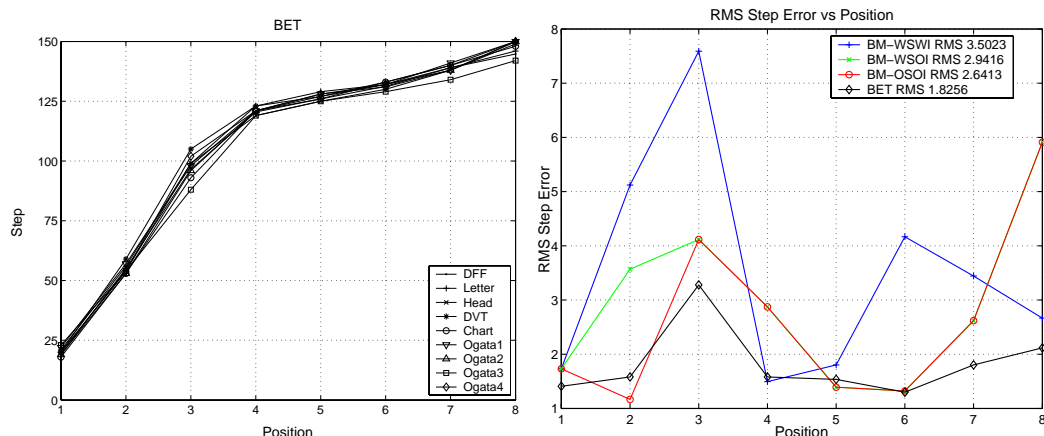


Figure 6. Measurement result for test objects

and additional error is expected due to sampling, grey level quantization, and image noise. BET is better than the other STM methods which have RMS errors of 2.6, 2.9, and 3.5 step errors. BET has a particularly better performance at far field such as positions 6, 7 and 8 when one of the image will be highly blurred. Therefore, since the computational requirement of BET and other STM methods are the same, BET is recommended for practical applications.

The maximum RMS step error of BET occurs at position 3. As Fig. 4 shows, position 3 is located in “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 focus or image sharpness measure is not significant. In this sense, BET’s performance is somewhat better than what is indicated by the RMS error in lens step position. BET was also tested and compared in extensive camera simulation experiments. The results and conclusions were similar to actual experiments reported here.

5. CONCLUSION

A new Depth-from-Defocus technique named Blur Equalization Technique is presented. It is a spatial-domain technique that can provide dense depth estimates using modest local computations in small image windows (*e.g.* 9×9). The theoretical basis of BET is different from other Depth-from-Defocus techniques. The computational algorithm is relatively simple. In both simulation and actual experiments, BET has been demonstrated to perform better than some of the best existing techniques, on many objects at different distances, including some standard test objects and very low contrast objects. The accuracy and robustness of BET seems to be close to the best possible. Therefore, it is recommended for use in practical applications. While the BET presented here uses the Spatial domain convolution/deconvolution transform, the basic idea of “blur equalization” can be extended to Fourier domain and other DFD techniques found in the current literature.

REFERENCES

1. M. Subbarao and T. Wei, “Depth from defocus and rapid autofocusing: A practical approach,” in *Proc. of IEEE Conf. on Comp. Vision and Pat. Recogn. (CVPR '92)*, pp. 773–776, June 1992.
2. J. Ens and P. Lawrence, “An investigation of methods for determining depth from focus,” *IEEE Trans. Pattern Anal. Mach. Intell.* **15**(2), pp. 97–108, 1993.
3. Y. Xiong and S. Shafer, “Depth from focusing and defocusing,” in *Proc. of IEEE Conf. on Comp. Vision and Pat. Recogn. (CVPR '93)*, pp. 68–73, June 1993.

4. M. Watanabe and S. K. Nayar, "Rational filters for passive depth from defocus," *Int. J. of Comp. Vision* **27**(3), pp. 203–225, 1998.
5. M. Subbarao and G. Surya, "Depth from defocus: A spatial domain approach," *Int. J. of Comp. Vision* **13**(3), pp. 271–294, 1994.
6. T. Xian and M. Subbarao, "Performance evaluation of different depth from defocus (DFD) techniques," in *Proc. of SPIE: 2- & 3-D Methods for Inspect. and Metro. III (Optics East 2005)*, October 2005.
7. D. Ziou and F. Deschenes, "Depth from defocus estimation in spatial domain," *Computer Vision and Image Understanding* **81**(2), pp. 143–165, 2001.
8. P. Favaro, A. Mennucci, and S. Soatto, "Observing shape from defocused images," *Int. J. of Comp. Vision* **52**(1), pp. 25–43, 2003.
9. M. Subbarao, "Spatial-domain convolution/deconvolution transform," Tech. Rep. 91.07.03, Computer Vision Laboratory, Dept. of Electrical Engineering, SUNY at Stony Brook, 1991.
10. T. Xian and M. Subbarao, "Camera calibration and performance evaluation of depth from defocus (DFD)," in *Proc. of SPIE: 2- & 3-D Methods for Inspect. and Metro. III (Optics East 2005)*, October 2005.