Robust Depth-from-Defocus for Autofocusing in the Presence of Image Shifts

Younsik Kang [a], Xue Tu [a], Satyaki Dutta[b], Murali Subbarao [a]

[a] {yskang, tuxue, murali}@ece.sunysb.edu, [b] sunny@math.sunysb.edu
[a] Dept. of Electrical and Computer Eng., State University of New York at Stony Brook, NY 11794;
[b]Dept. of Mathematics, SUNY at Stony Brook, NY 11794

ABSTRACT

A new passive ranging technique named Robust Depth-from-Defocus (RDFD) is presented for autofocusing in digital cameras. It is adapted to work in the presence of image shift and scale change caused by camera/hand/object motion. RDFD is similar to spatial-domain Depth-from-Defocus (DFD) techniques in terms of computational efficiency, but it does not require pixel correspondence between two images captured with different defocus levels. It requires approximate correspondence between image regions in different image frames as in the case of Depth-from-Focus (DFF) techniques. Theory and computational algorithm are presented for two different variations of RDFD. Experimental results are presented to show that RDFD is robust against image shifts and useful in practical applications. RDFD also provides insight into the close relation between DFF and DFD techniques.

Keywords: Depth-from-Defocus, Depth-from-Focus, Autofocus, Passive Ranging, Camera/Hand shake.

1. INTRODUCTION

Passive ranging and autofocusing techniques based on image focus and defocus analysis have been investigated by many researchers [1-11]. Depth-from-Focus (DFF) [1,3] is a technique based on image focus analysis. It requires capturing and computing focus measures for many (about 10 in a typical application) images. It involves searching for a camera parameter setting that maximizes the focus measure in a small image block. It is slow and provides a low resolution depth (one estimate in each image block), but it does not need camera calibration or pixel correspondence. It needs only image region correspondence. DFF is relatively robust and accurate. Depth-from-Defocus (DFD) [4-11] is a technique based on image defocus analysis. It uses only two or three images captured with different camera parameter settings. It compares the change in blur to estimate depth. It is fast and efficient but requires camera calibration. Spatial domain DFD techniques provide a denser depth-map (upto one depth estimate at each pixel) than DFF techniques, but they need pixel correspondence between different image frames. DFD is also less accurate and robust than DFF unless more images (about 5) are used. Fourier domain DFD [10] techniques provide lower resolution depth-maps and require more computation than spatial domain DFD techniques [5,11].

In this paper a new passive ranging technique named Robust Depth-from-Defocus (RDFD) is presented for autofocusing in digital still/video cameras. RDFD is similar to spatial-domain DFD techniques [5,11] in terms of computational efficiency and the use of only two images, but it does not require pixel correspondence between two images captured with different defocus levels. It requires approximate correspondence between small image regions in different image frames as in the case of DFF techniques. This is an important advantage because, in autofocusing applications, small image shifts between different captured image frames are caused by many unavoidable factors such as hand shake, camera vibration, magnification change (due to lens motion), and object motion. Both RDFD and spatial-domain DFD use a local deconvolution formula involving image derivatives for estimating an image defocus parameter. However, unlike DFD techniques, RDFD relies on computations at corresponding *small image regions* instead of corresponding *pixels*. RDFD is suitable for real-time autofocusing in digital still and video cameras where the number of image frames captured and computational resources are limited, and accurate pixel correspondence is not available.

It should be noted that Fourier domain DFD techniques [10] do not need pixel correspondence but only region correspondence (as with DFF) and therefore less sensitive to image shift caused by camera vibration, etc. However, Fourier domain techniques have other problems. First, they require more computation as Fourier coefficients will have to be computed. Next, unlike local spatial domain DFD techniques such as STM [5,11], Fourier domain techniques are

global and less accurate as they ignore the image overlap problem. Fourier domain techniques use windowing to extract image blocks and compute Fourier coefficients of the image blocks as an estimate of local Fourier coefficients. As blurring just outside the border of the image blocks changes the images inside the image blocks, errors will be introduced in these techniques. RDFD uses local spatial domain computations similar to DFD-STM [5,11] at each pixel, and is therefore less affected by the image overlap problem.

This paper presents the theory and a computational algorithm for two different variations of RDFD. One of the variation uses Greene's identity to avoid the use of second-order image derivatives which are highly noise sensitive. Instead, it relies more on first-order image derivatives which are a lot less sensitive to noise. Greene's identity provides a relation between the area integral involving second-order image derivatives to a boundary integral involving only first-order image derivatives. Experimental results are presented to compare the performance of RDFD with DFD in [5,11] (STM). These results show that RDFD is more resistant to small image shifts than DFD-STM, and it is useful in practical applications.

2. ROBUST DEPTH FROM DEFOCUS (RDFD)

RDFD theory is an extension of the DFD theory based on the S Transform Method (STM) [5,11]. Let $g_1(x,y)$ and $g_2(x,y)$ be two defocused images captured with two different camera parameter settings such as two different lens positions. Let $f(x,y)$ be the focused image and $h_1(x,y)$, and $h_2(x,y)$, be the two point spread functions (PSFs) so that

$$g_1(x,y) = h_1(x,y) * f(x,y) \qquad (1), \text{ and } \qquad g_2(x,y) = h_2(x,y) * f(x,y) \qquad (2)$$

where * denotes the convolution operation. Let σ_1 and σ_2 be the blur parameters given by the square root of the second-central moment of the PSFs $h_1(x,y)$, $h_2(x,y)$, respectively (see Eq. (A11) in Appendix). The spatial domain DFD technique STM [5,11] uses the following expression to compute a change in the blur parameters (assuming a local cubic polynomial model of the images):

$$\sigma_1^2 - \sigma_2^2 = \frac{4[g_1(x,y) - g_2(x,y)]}{(\nabla^2 g_1(x,y) + \nabla^2 g_2(x,y))/2} \qquad (3)$$

Spatial domain DFD methods such as STM [5,11] require pixel correspondence between two images captured with different camera settings. Pixel correspondence is needed because the expression used in solving for depth include mixed-image terms such as $g_1(x,y) - g_2(x,y)$ in the above equation. These mixed terms need the value of two different images g_1 and g_2 at the same corresponding pixel position specified by (x,y). In RDFD, such terms involving both images at the same corresponding pixel are avoided. Instead focus measure terms computed over corresponding image regions are used as in the case of DFF methods. The expressions for the focus measures are derived from the local spatial-domain deconvolution (inverse S Transform) expressions [5,11] (see Appendix):

$$f(x,y) = g_1(x,y) - \frac{\sigma_1^2}{4}\nabla^2 g_1(x,y) + \text{ Higher Order Terms (H.O.T.s)} \qquad (4)$$

$$f(x,y) = g_2(x,y) - \frac{\sigma_2^2}{4}\nabla^2 g_2(x,y) + \text{ Higher Order Terms (H.O.T.s)} \qquad (5)$$

The right hand sides in the above two equations express the same focused image, and equating them is called the Focus Equalization Technique. Equating the two right hand sides and squaring both sides to compute focus measures, and then neglecting the higher order terms, we obtain:

$$\iint_R \left(g_1(x,y) - \frac{\sigma_1^2}{4}\nabla^2 g_1(x,y) \right)^2 dx\,dy \quad = \quad \iint_R \left(g_2(x,y) - \frac{\sigma_2^2}{4}\nabla^2 g_2(x,y) \right)^2 dx\,dy \tag{6}$$

We use squaring to avoid adding positive and negative Laplacian quantities to get a small sum that is sensitive to noise. We integrate over an image region *R* to avoid needing pixel correspondence to solve for depth. We need only image region correspondence for the region *R*. Larger the size of image region *R*, smaller is the sensitivity of the solution to errors in establishing accurate region correspondence. Increasing the size of *R* also reduces spatial-resolution of depth-map in 3D shape recovery applications.

Next consider an equally blurred image $g_3(x,y)$ obtained from $g_1(x,y)$ and $g_2(x,y)$:

$$g_3(x,y) = g_1(x,y) * h_2(x,y) = h_1(x,y) * h_2(x,y) * f(x,y) \tag{7}$$

$$g_3(x,y) = g_2(x,y) * h_1(x,y) = h_1(x,y) * h_2(x,y) * f(x,y) \tag{8}$$

$$g_1(x,y) * h_2(x,y) = g_2(x,y) * h_1(x,y) \tag{9}$$

$$g_3(x,y) = g_1(x,y) + \frac{\sigma_2^2}{4}\nabla^2 g_1(x,y) + \text{ H.O.T.s} \tag{10}$$

$$g_3(x,y) = g_2(x,y) + \frac{\sigma_1^2}{4}\nabla^2 g_2(x,y) + \text{ H.O.T.s} \tag{11}$$

The right hand sides in Eq. (10) and Eq. (11) express the same defocused image $g_3(x,y)$, and equating them is called the Blur or Defocus Equalization Technique (BET/DET) [7]. Equating the two right hand sides, squaring both sides, integrating to compute focus measures, and then neglecting higher order terms, we obtain:

$$\iint_R \left(g_1(x,y) + \frac{\sigma_2^2}{4}\nabla^2 g_1(x,y) \right)^2 dx\,dy = \iint_R \left(g_2(x,y) + \frac{\sigma_1^2}{4}\nabla^2 g_2(x,y) \right)^2 dx\,dy \tag{12}$$

Subtracting Eq. (12) from Eq. (6) and simplifying, we obtain

$$G = \sigma_1^2 - \sigma_2^2 = 8 \times \frac{\iint_r g_1(x,y)\nabla^2 g_1(x,y)\,dxdy - \iint_r g_2(x,y)\nabla^2 g_2(x,y)\,dxdy}{\iint_r (\nabla^2 g_1(x,y))^2\,dxdy + \iint_r (\nabla^2 g_2(x,y))^2\,dxdy} \tag{13}$$

Note the similarity between Eq. (3) and Eq. (13). Equation (13) forms the basis of RDFD for computing the quantity $\sigma_1^2 - \sigma_2^2$. It is the same quantity used in DFD-STM [5,11], but computed using a different approximation over a small image region, and hence more robust. DFD based on STM uses a local cubic approximation for images which results in the Laplacians of the two images to be equal, i.e. $\nabla^2 g_1 = \nabla^2 g_2$. But, RDFD uses a better approximation allowing the two Laplacians to be different, i.e. $\nabla^2 g_1 \neq \nabla^2 g_2$. If we substitute $\nabla^2 g_1 = \nabla^2 g_2$ in Eq. (13) of RDFD, then Eq. (13) becomes similar to Eq. (3) except for integration. Equation (13) can be solved by substituting for σ_1 in terms of σ_2 using the known camera constants (obtained by calibration) α and β as:

$$\sigma_1 = \alpha\sigma_2 + \beta \tag{14}$$

We get a quadratic equation in σ_2 which can be solved in closed-form. If the camera aperture diameter is not changed, then $\alpha = 1$ and the equation becomes linear with a unique solution. In all our experiments, aperture diameter remains the same and therefore $\alpha = 1$, and a unique solution is obtained for σ_2. From a solution for σ_2, we obtain the distance *u* of the object from the relation

$$\sigma_2 = m_2\, u^{-1} + c_2 \tag{15}$$

and the known camera constants m_2 and c_2 (obtained by calibration). In the experiments, we use a lookup table obtained through calibration to solve for σ_2 and u after computing the quantity *G* from images using Eq. (13). Experimental results comparing the DFD with RDFD implemented directly using Eq. (13) is presented later. In the next section we present another interesting variation of RDFD.

2.1 RDFD based on Greene's Identity

Image noise is dramatically amplified in the computed derivatives of images, and it becomes progressively hopeless as the order of image derivatives increases. Equation (13) for implementing RDFD involves the computation of second-order image derivatives (Laplacians) which yield very noisy estimates. In order to reduce the effects of noise, a new technique was used to compute the right hand side of Eq. (13). It uses the first-order image derivatives which are much less noise sensitive in comparison with the second-order derivatives (Laplacian). The well-known Greene's identity in Calculus is used to reduce the reliance on computing second order image derivatives by replacing some terms with equivalent terms involving only the first-order derivatives. Greene's identity is

$$\iint_R g\, \nabla^2 g = \int_B g\, (\vec{\nabla} g \bullet \vec{n}) - \iint_R (\vec{\nabla} g \bullet \vec{\nabla} g) \tag{16}$$

where R is an image region and B is the boundary of R, $\vec{n}$ is the unit outward normal vector to the boundary, and $\vec{\nabla} g$ denotes the gradient vector of the blurred image $g(x, y)$. This identity is widely used in the theory of partial differential equations. In the case of one dimension this result can be derived as follows. Assume $g : [a, b] \to \Re$ be a twice differentiable function from an interval $[a, b]$ to the real numbers. An expression of the derivative of the product term $(g.g')'$ can be rearranged to obtain

$$g.g'' = (g.g')' - (g')^2 \tag{17}$$

Integrating,

$$\int_a^b g.g''dx = \int_a^b (g.g')'dx - \int_a^b (g')^2 dx \;= g(b)g'(b) - g(a)g'(a) - \int_a^b (g')^2 dx \tag{18}$$

Applying the two-dimensional equivalent of this formula to Eq. (13), we obtain

$$G = \sigma_1^2 - \sigma_2^2 = 8\, \frac{\int_B g_1(\vec{\nabla} g_1 \bullet \vec{n}) - \iint_R \vec{\nabla} g_1 \bullet \vec{\nabla} g_1 - \int_R g_2(\vec{\nabla} g_2 \bullet \vec{n}) + \iint_R \vec{\nabla} g_2 \bullet \vec{\nabla} g_2}{\iint_M (\nabla^2 g_1)^2 + \iint_M (\nabla^2 g_2)^2} \tag{19}$$

In the numerator of the above equation, terms in the original equation (Eq. (13)) with area integrals involving image Laplacians (second-order derivatives) have been replaced by terms with boundary and area integrals of only the first-order image derivatives. Computationally, it may be more robust, but this conclusion is not certain because boundary integral of first-order derivatives may be as noise sensitive as area integral of second-order derivatives, especially when there is image shift due to camera shake or motion. Experimental results did not show significant improvement of accuracy, but the method is included here as future improvements may be possible.

2.2 Two Variations of DFD: FET and BET

Two other variations of DFD were investigated. Both are closely related to DFD STM [5,7,11] but vary in some detail. The first one, Focus Equalization Technique (FET), is based on solving the following equation obtained by equating the right hand sides of equations (4) and (5):

FET:
$$4(g_1(x,y) - g_2(x,y)) = \sigma_1^2 \Delta g_1(x,y) - \sigma_2^2 \Delta g_2(x,y) \tag{20}$$

The second one, Blur Equalization Technique (BET) [7], is based on solving the following equation obtained by equating the right hand sides of equations (10) and (11):

BET:
$$4(g_1(x,y) - g_2(x,y)) = \sigma_1^2 \Delta g_2(x,y) - \sigma_2^2 \Delta g_1(x,y) \tag{21}$$

The above equations are solved by substituting σ_2 with $\alpha\sigma_1 + \beta$ (with $\alpha = 1$) and solving for σ_1. Effect of noise is reduced by computing σ_1 at only those pixels where the Laplacian magnitude is more than a preset threshold. σ_1 is averaged over a small image window to obtain a better estimate. These two methods give results that are comparable or sometimes better than the DFD STM, and they are both sensitive to image shifts as is the case with DFD STM. But these two methods are of interest as alternatives to DFD in the absence of image shifts.

2.3 Relation Between DFF and the Blur Parameter σ of DFD

DFF techniques [3] are based on computing a focus measure of blurred images. An example of a focus measure is the energy (i.e. integral of square) of image intensity or energy (integral of square) of derivatives of image intensity. The local convolution expression derived in the Appendix section can be used to derive a relation between these focus measures and the blur parameter σ on which the the DFD techniques are based. Using Eq. (A12) in the Appendix we obtain the following expression for the focus measure based on image energy:

$$\iint_R g^2(x,y)\, dx\, dy = \iint_R \left[f(x,y) + \frac{\sigma^2}{4} \nabla^2 f(x,y) + H.O.T.s \right]^2 dx\, dy \tag{22}$$

Similarly, for the focus measure based on the energy of image gradient magnitude, we obtain

$$\iint_R (g_x^2 + g_y^2)\, dx\, dy = \iint_R \left[f_x^2 + f_y^2 + \frac{\sigma^2}{4} (\nabla^2 f_x + \nabla^2 f_y) + H.O.T.s \right]^2 dx\, dy \tag{23}$$

And for the focus measure based on the energy of image Laplacian we obtain

$$\iint_R (\nabla^2 g)^2\, dx\, dy = \iint_R \left[\nabla^2 f + \frac{\sigma^2}{4} \nabla^2 (\nabla^2 f) + H.O.T.s \right]^2 dx\, dy \tag{24}$$

In all the above three cases, the focus measure of the blurred image can be expressed in terms of the focused image and the blur parameter sigma as a polynomial of σ^2

$$FM(\sigma^2) = F_0 + F_2\sigma^2 + F_4\sigma^4 + H.O.T.s \tag{25}$$

where F_0, F_2, F_4, are constants for a given focused image. For example, we obtain

$$\iint_R g^2(x,y)\, dx\, dy = \iint_R f^2(x,y)\, dx\, dy + \frac{\sigma^4}{16} \iint_R [\nabla^2 f(x,y)]^2\, dx\, dy + \frac{\sigma^2}{2} \iint_R f(x,y)\, (\nabla^2 f(x,y))\, dx\, dy + \text{H.O.T.s} \tag{26}$$

Eq. (25) is a good approximation for small blur (i.e. small σ) which implies that the camera parameter setting is close to the focus setting. This equation establishes an explicit relation between DFF and DFD techniques. It also provides a polynomial model for fitting the focus measures to a function of camera parameters. Such polynomial fitting is often done close to the focus setting (therefore σ will be small) to determine the value of the camera parameter setting that maximizes the focus measure. A good approximation to the relation between sigma and the camera parameters (denoted by -- *s*: lens to image detector distance, *D*: aperture diameter, *f:* focal length, and *u*: object distance) is given by

$$\sigma = \frac{Ds}{2\sqrt{2}}\left[\frac{1}{f} - \frac{1}{s} - \frac{1}{u}\right] = mu^{-1} + c \tag{27}$$

3. EXPERIMENTS

Many experiments were carried-out on both simulation and real image data to evaluate the performance of RDFD in relation to DFD. Some results on real image data are presented here. Two planar test objects— Checkerboard and BigLetter-- shown in Fig. 1 were used. These objects were placed at 9 different distances in the range 300 mm to 600 mm from a digital camera. The digital camera had a computer controlled stepper motor for changing the lens position to focus at different distances. The lens step positions ranged from Step 0 to Step 1500. For each object and distance, two images of the test object were captured, one with the lens at step position 800, and another at step position 1100. Aperture diameter was not changed. The image captured at Step 1100 was shifted or scaled by different amounts to simulate the effect of camera/object motion or camera/hand shake. Note that image scaling simulates the effect of relative motion between the camera and the object along the direction of view (optical axis) as this motion results in changes in image magnification. The shifting and scaling amounts were limited to be small as in the case of actual applications. The image captured at Step 800 was not changed. The resulting images (size 600x480) were used for estimating the blur parameter σ and object distance for autofocusing. This estimate was done using both the RDFD and the DFD (STM) methods. The results are shown in Fig. 2 and Fig. 3.

Fig. 1. Test objects. Left: "Checkerboard". Right: "BigLetter".

In Fig. 2(a), Fig. 2(b), horizontal axis represents the reciprocal of object distance in meters, and the corresponding blur parameter σ is plotted along the vertical axis in pixel units. For an ideal thin-lens camera, this plot is expected to be roughly linear, but for practical cameras some deviation from linear behavior is usual. In Fig. 2, the plots labeled "DFD by 0" and "RDFD by 0" represent the results for σ when there is no image shift. This is the reference plot and other plots should be compared with this plot. For example, the plot labeled "DFD by 10" and "RDFD by 10" correspond to shifting the image at Step 1100 by 10 pixels. We see that the plot labeled "DFD by 10" is quite far from the plot labeled "DFD by 0", thus indicating a large error in the presence of 10 pixel image shift. This error for DFD generally increases as the shift increases. Therefore the DFD technique is very sensitive to image shifts. On the other hand, the plots labeled "RDFD by 0" and "RDFD by 10" almost coincide at different distances thus indicating that RDFD is robust with respect to 10 pixel image shifts. This is true even up to 20 pixel image shifts for the objects considered in our experiments.

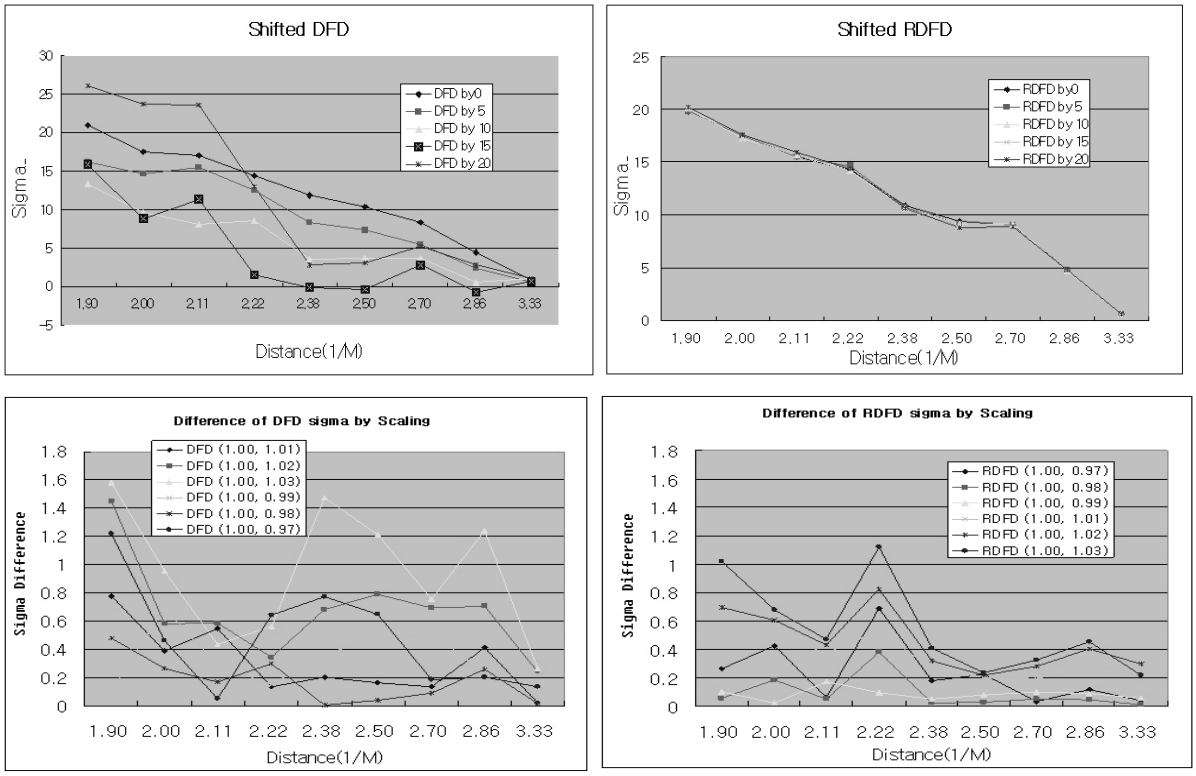

Fig. 2. Results of DFD and RDFD with "Checkerboard". (a) Top left: DFD results with image shift. (b) Top right: RDFD results with image shift. (c) Bottom left: DFD result with scaling. (d) Bottom right: RDFD results with scaling.

Fig 2(c) and Fig. 2(d) show the results in the case of scale/magnification changes for the image captured at lens Step 1100. The magnification was changed by up to 3 percent in the range [0.97, 1.03] where a factor of 1.0 represents no change. This simulates change in magnification due to different amounts of relative motion between the camera and the object. This motion is expected to occur in the elapsed time interval between the capture of the first image at lens step 800 and the second image at step 1100. For DFD, Fig. 2(c) shows the deviation of computed σ with respect to the case when there is no magnification change (as shown in Fig. 2(a) with label "DFD by 0"). Similarly, Fig. 2(d) shows the results for RDFD. We see that the errors for DFD are usually much more than for RDFD leading to the conclusion that RDFD is robust against small changes in scaling.

Fig. (3) shows the results for the other object BigLetter. The interpretation of the plots is similar to that for Fig. 2. Again, the conclusion is the same, that RDFD is far more robust with respect to image shifts and scale changes in comparison with DFD. As expected, the actual errors are not the same as in Fig. 2. This indicates that the actual errors are dependent on the object pattern and contrast.

A limited set of experiemnts were carried out for the second version of RDFD that uses Eq. (19). The new version did not improve the results in comparison with the first version. In some cases it was worse than the first version. So these results are not presented.

4. CONCLUSION

A new spatial domain depth-from-defocus technique-- RDFD – is presented for autofocusing in digital cameras. It is adapted to be robust against image shifts due to camera/object motion and camera/hand shake. Its performance has been compared with a spatial domain DFD technique and its advantages in the presence of image shift and scale change have been demonstrated. RDFD shares some characteristics of both DFF and spatial domain DFD techniques. Investigation of RDFD leads to the derivation of an explicit relation between DFF and DFD techniques.

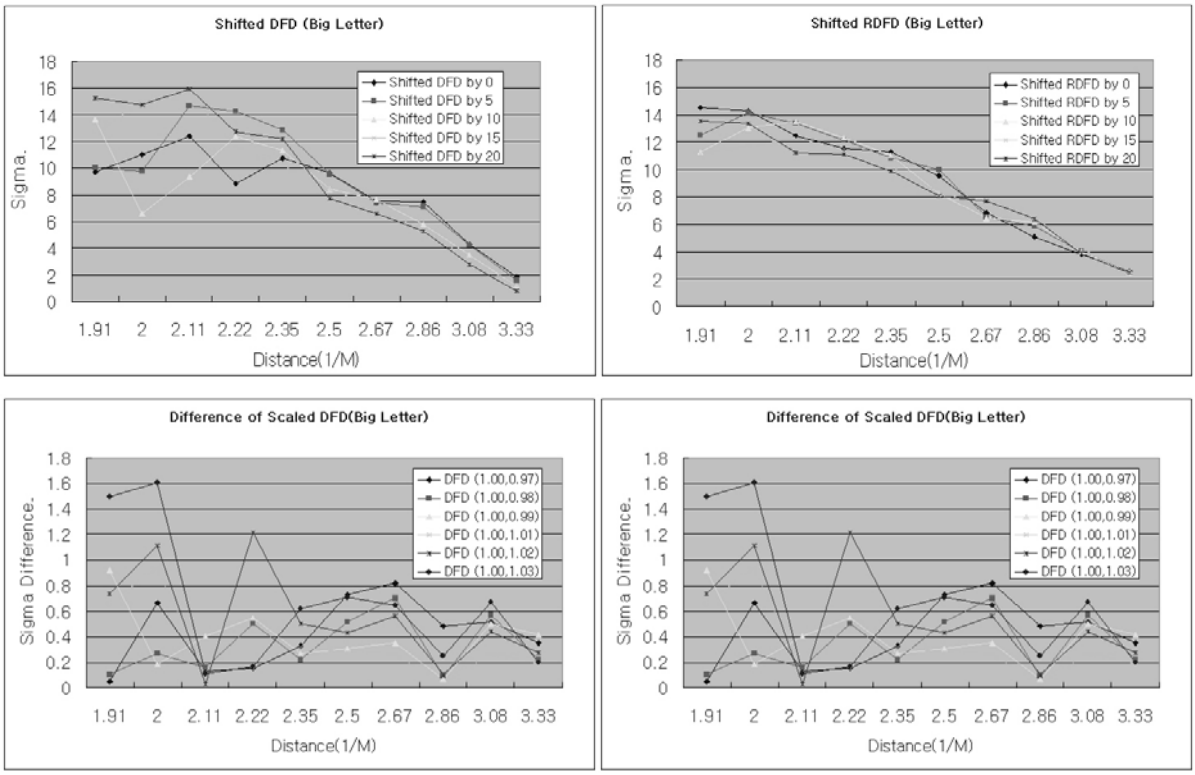

Fig. 3. Results of DFD and RDFD with "BigLetter". (a) Top left: DFD results with image shift. (b) Top right: RDFD results with image shift. (c) Bottom left: DFD result with scaling. (d) Bottom right: RDFD results with scaling..

5. APPENDIX

A brief summary of the Spatial-domain Convolution/Deconvolution Transform or S-Transform is presented here. S-Transform [11] has been developed for images and *n*-dimensional signals for the case of arbitrary order of polynomials. It provides a completely local expression for the convolution and deconvolution of an image with respect to a filter or Point Spread Function (PSF) in the spatial domain. Using a local polynomial approximation of an image, spatial domain formulas are derived for convolution (corresponds to blurring or defocusing) and deconvolution (corresponds to deblurring or image restoration) with respect to a PSF.

Let $h(x, y)$ be a rotationally symmetric PSF. This assumption of rotational symmetry is a useful one, but it can be removed easily for theoretical analysis if needed. If we assume the camera to be a lossless system (i.e., no light energy is absorbed by the camera system) then

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x, y)\, dx\, dy = 1. \tag{A1}$$

The moments of the PSF are defined by

$$h_{m,n} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^m\, y^n\, h(x, y)\, dx\, dy. \tag{A2}$$

Since the PSF is rotationally symmetric, many odd moments are zero and it can be shown that

$$h_{0,1} = h_{1,0} = h_{1,1} = h_{0,3} = h_{0,3} = h_{3,0} = h_{1,2} = h_{2,1} = 0,\ h_{2,0} = h_{0,2}, \text{ and } h_{0,0} = 1 \tag{A3}$$

A blurred image $g(x, y)$ captured by a camera system can be modeled by the convolution of the focused image $f(x, y)$ with the corresponding PSF $h(x, y)$ as:

$$g(x,y) = f(x,y) * h(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x-\varsigma, y-\eta)\; h(\varsigma,\eta)\; d\varsigma\, d\eta \tag{A4}$$

If we approximate the focused image to be a cubic polynomial locally in a small image region corresponding to the size of the blur circle region at a pixel, then $f(x-\varsigma, y-\eta)$ can be expanded up to order 3 in a Taylor series as:

$$f(x-\varsigma, y-\eta) = \sum_{0\le m+n\le 3} \frac{(-\varsigma)^m}{m!}\frac{(-\eta)^n}{n!} f^{m,n}(x,y) + \text{ Higher Order Terms (H.O.T.s)}. \tag{A5}$$

where $f^{m,n}(x,y) = \dfrac{\partial^m}{\partial x^m}\dfrac{\partial^n}{\partial y^n} f(x,y)$. Substituting Eq. (A5) into Eq. (A4) and simplifying, we obtain

$$\begin{aligned} g(x,y) &= \sum_{0\le m+n\le 3} \frac{(-1)^{m+n}}{m!n!} f^{m,n}(x,y)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \varsigma^m \eta^n\; h(\varsigma,\eta)\, d\varsigma\, d\eta + \text{H.O.T.s} \\ &= \sum_{0\le m+n\le 3} \frac{(-1)^{m+n}}{m!n!} f^{m,n}(x,y)\, h_{m,n} + \text{H.O.T.s} \end{aligned} \tag{A6}$$

This corresponds to the Forward S-Transform. In order to derive the inverse S-Transform, Eq. (A6) is rewritten as

$$f(x,y) = g(x,y) - \sum_{1\le m+n\le 3} \frac{(-1)^{m+n}}{m!n!} f^{m,n}(x,y)\, h_{m,n} + \text{H.O.T.s}. \tag{A7}$$

Therefore the derivatives of the focused image are given by

$$f^{p,q}(x,y) = g^{p,q}(x,y) - \sum_{1\le m+n+p+q\le 3} \frac{(-1)^{m+n}}{m!n!} f^{m+p,n+q}(x,y)\, h_{m,n} + \text{H.O.T.s}. \tag{A8}$$

Recursively substituting for derivatives of $f(x,y)$ using Eq. (A8) in Eq. (A7), we obtain the inverse S transform which gives an expression for the focused image $f(x,y)$ in terms of the derivatives of the blurred image $g(x,y)$ and the moments $h_{m,n}$ of the PSF. In particular we obtain

$$f(x,y) = g(x,y) - \frac{h_{2.0}}{2}\left(g^{2,0}(x,y) + g^{0,2}(x,y)\right) + \text{H.O.T.s} = g(x,y) - \frac{h_{2.0}}{2}\nabla^2 g(x,y) + \text{H.O.T.s} \tag{A9}$$

where ∇^2 is the Laplacian operator. If we define blur parameter as

$$\sigma^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x^2+y^2)\, h(x,y)\, dx\, dy \tag{A10}$$

then $h_{2,0} = h_{0,2} = \sigma^2/2$. Therefore Eq. (A9) can be written as

$$f(x,y) = g(x,y) - \frac{\sigma^2}{4}\nabla^2 g(x,y) + \text{H.O.T.s} \tag{A11}$$

Using these results, we can write an expression for the blurred image $g(x,y)$ in terms of the focused image $f(x,y)$ using Eq. (A6) as

$$g(x,y) = f(x,y) + \frac{\sigma^2}{4}\nabla^2 f(x,y) + \text{H.O.T.s} \tag{A12}$$

REFERENCES

[1] E. Krotkov, “Focusing”, *international Journal of Computer Vision*, 1, 223-237 (1987).

[2] S. K. Nayar, “Shape from Real-time Range Camera”, *Proceeding of IEEE computer Society Conference on Computer Vision and Pattern Recognition*, San Diego, California, June 1992.

[3] M. Subbarao, T. S. Choi, and A. Nikzad, "Focusing Techniques," *Journal of Optical Engineering* , pp. 2824-2836 Nov. 1993.

[4] M. Subbarao and T. Wei, “Depth form Defocus and Rapid Autofocusing: A practical Approach”, *Proceeding of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Champaign, Illinois, June 1992.

[5] M. Subbarao and G. Surya, "Depth from Defocus: A Spatial Domain Approach," *International Journal of Computer Vision*, Vol. 13, No. 3, pp. 271-294, 1994.

[6] 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, 1987.

[7] T. Xian, Three-Dimensional Modeling and Autofocusing Technology for New Generation Digital Cameras, Ph. D. thesis, Dept. of Electrical and Computer Engg., May 2006, State University of New York, Stony Brook, NY 11794-2350, USA..

[8] P. Favaro, and S. Soatto, “A geometric approach to shape from defocus”, IEEE Trans. on PAMI, Vol. 27, Issue 3, pp. 406-417, March 2005.

[9] S. Chaudhuri and A.N. Rajagopalan, ``Depth from Defocus: A Real Aperture Imaging Approach”, Springer-Verlag, New York, 1999.

[10] Tse-Chung Wei, "Three-Dimensional Machine Vision using Image Defocus," Ph.D. Thesis, Dept. of Electrical Engineering, SUNY at Stony Brook, Dec.1994.

[11] Gopal Surya, "Three-Dimensional Scene Recovery from Image Defocus," Ph.D. Thesis, Dept. of Electrical Engineering, SUNY at Stony Brook, Dec. 1994.