Improving Shape from Focus Using Defocus Information

K.S. Pradeep A.N. Rajagopalan
Image Processing and Computer Vision Laboratory
Department of Electrical Engineering
Indian Institute of Technology Madras, India
ee04s018@ee.iitm.ac.in, raju@ee.iitm.ac.in

Abstract

Shape from focus (SFF) method determines the degree of focus in a sequence of observations to estimate the shape of a 3-D object. Existing SFF algorithms use an ad hoc *interpolation strategy to account for the error due to the finite step-size by which the translational table is moved while capturing the images. We propose an improved SFF method that uses relative defocus blur derived from actual image data to arrive at the final estimates of the shape of the object. A space-variant image restoration scheme is also proposed to obtain a focused image of the 3-D object. The shape estimates as well as the quality of the restored image using the proposed method are superior to that of traditional SFF.*

1 Introduction

The shape from focus (SFF) technique requires a sequence of images of an object in order to compute its shape. This is done by moving the translational table in finite steps and searching the camera positions at which different object elements appear in focus ([1, 2, 3]). A focus measure profile is obtained for each object pixel in the image sequence and the depth estimate corresponding to a pixel is obtained by interpolating a Gaussian curve [1] over a few values near the peak value of the focus measure curve. In [2], the focus measure profile is modeled by a low-order polynomial. The location of the local maximum of the fitted polynomial curve is taken to be the final estimate of depth. However, the accuracy of the traditional SFF technique is limited by whether or not the focus measure curve gets adequately sampled in the vicinity of its peak, the only region where interpolation is known to hold [1].

We propose a new strategy to improve the estimates of SFF by using the actual relative defocus blur among the observations. Since our approach is data-driven it leads to a better estimate of the shape of an object. One of the yardsticks to evaluate the accuracy of computed depth profile is the quality of the focused image that can be estimated using it. Owing to the fact that the defocus blur distribution is space-variant for all observations, estimating the focused image of the object is a challenging problem. In fact, no real-aperture camera can bring all points on a 3-D object into perfect focus! We propose a method for space-variant restoration of the focused image of the object. Experimental results are given on synthetic as well as real data to demonstrate the improvement in shape estimation using the proposed method.

2 Traditional Shape from Focus

Shape from Focus (SFF) ([1, 2]) uses several images of an object with different regions on the surface of the object coming into focus in different images. The degree of focus is the only cue used for shape computation. To capture the images, the object of interest is placed on a table over which a camera, with its parameters fixed, is translated in the z-direction in steps of Δd (also called the step-size) till it sweeps across the entire 3-D object. An image is captured at each step to obtain a sequence of defocused images.

According to the lens law, a point source at a distance D from a lens of focal length f and radius R will form an image at distance p from the lens. In a camera, if the sensor plane is at a distance v ($v \neq p$) from the lens, then the image formed on the sensor will not be a point image but will be a circular blob of radius r_D given by $r_D = Rv\left[\frac{1}{f} - \frac{1}{v} - \frac{1}{D}\right]$, or equivalently, $r_D = Rv\left[\frac{1}{w_d} - \frac{1}{D}\right]$ where w_d is called the working distance ([4]). By finding the position of the camera which brings a particular part of the object into focus on the sensor, the corresponding shape estimate can be obtained.

To find the degree of focus over small regions in each of the captured images, a focus measure operator is used. Many focus operators have been proposed in the literature ([1], [2], [3]). Following [1], we use the sum-modified-

Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06)
0-7695-2521-0/06 $20.00 © 2006 IEEE

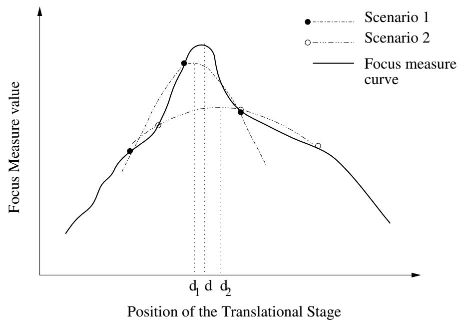


Figure 1. Gaussian interpolation for estimating depth in traditional SFF.

Laplacian (SML) operator. The resulting focus measure values $F(m,n)$ yield a focus measure profile for every pixel location (m,n) in the image sequence. A few of the strongest focus measure values in this profile are interpolated using a Gaussian fit in the vicinity of the peak to compute the final depth estimate.

If the focus measure values that are used for interpolation do not all lie in the vicinity of the peak, then the interpolation strategy will yield erroneous estimates of shape. This is better explained with Fig. 1 in which two sampling scenarios are shown that correspond to inspecting a sample with two different step-sizes and initial positions of the translational stage. It is worth noting that in neither of these scenarios has the continuous focus profile been sampled adequately in the vicinity of its actual peak, the only region where the Gaussian fit is valid [1]. Whereas the true depth is d, the two sampling scenarios yield d_1 and d_2, respectively. As a result, the shape estimates are in error.

3 The Proposed Method

Since interpolation is an *ad hoc* way of filling in missing data, the derived depth estimates can be erroneous as discussed above. The estimate will be more reliable if it is made to depend on actual image data. In the strategy proposed in this paper, for every point on the object, a pair of images is chosen from the sequence with respect to which its image is the sharpest. These correspond to the two positions of the translational table at which the point in question was closest to the focused plane of the camera. The actual relative blur between these images is computed and using this defocus information, the final depth of the point is calculated.

In the literature, defocus blur is typically modeled as a Gaussian given by $h(m,n) = \frac{1}{2\pi\sigma^2}\exp\left\{-\frac{m^2+n^2}{2\sigma^2}\right\}$. We also make this approximation in our work and further validate it in the experimental section. We show that the results using the proposed method are superior to interpolation. The blur parameter σ can be related to the blur radius r_D as $\sigma = \rho r_D = \rho R v\left[\frac{1}{w_d} - \frac{1}{D}\right]$ where ρ is the camera constant and must be calibrated for a given camera [4].

Assume that there are M frames $I_1, I_2, \ldots, I_M$ in the SFF image sequence. Let $g_k(m,n)$ and $g_{k+1}(m,n)$ denote the two sub-images in I_k and I_{k+1} from the captured sequence which result in the two strongest focus measure values for a point P. As illustrated in Fig. 2, I_k and I_{k+1} correspond to the two images which were captured when the translational table was in the position at which P was closest to the focused plane of the camera. Without loss of generality, let us assume that $g_{k+1}(m,n)$ is more blurred than $g_k(m,n)$. The formation of these sub-images can then be described as the convolution of the unknown focused sub-image $i_f(m,n)$ of point P with the corresponding PSFs $h_{k+1}(m,n)$ and $h_k(m,n)$ whose blur parameters are σ_{k+1} and σ_k, respectively. Even though the images in the sequence actually suffer from space-variant defocus blur, over small regions the blur can be assumed to be locally space-invariant so that convolution holds [4]. The relative blur between the two images can then be expressed in the Fourier domain as

$$\frac{G_{k+1}(u,v)}{G_k(u,v)} = e^{-(u^2+v^2)(\sigma_{k+1}^2-\sigma_k^2)/2} \quad (1)$$

where $G_{k+1}(u,v)$ and $G_k(u,v)$ are the N-point 2-D Fourier transforms of $g_{k+1}(m,n)$ and $g_k(m,n)$, respectively, while u and v are the spatial frequency components. From equation (1), the relative blur is again a 2-D Gaussian function with blur parameter σ_r where $\sigma_r = \sqrt{\sigma_{k+1}^2 - \sigma_k^2}$. One can estimate σ_r using equation (1) as

$$\sigma_r = \frac{1}{N^2}\sum_{u,v=0}^{N-1}\sqrt{\frac{-2\ln\frac{G_{k+1}(u,v)}{G_k(u,v)}}{(u^2+v^2)}} \quad (2)$$

This estimator works very well in a noise-free synthetic case. However, for a real camera, the PSF of the relative blur will not be a true Gaussian and using equation (2) may not yield a good estimate of σ_r. We propose to estimate σ_r by fitting a 2-D Gaussian curve to the underlying data $\frac{G_{k+1}(u,v)}{G_k(u,v)}$ in the least-squares sense. The BFGS (Broyden-Fletcher-Goldfarb-Shanno) quasi-Newton method [5] with a mixed quadratic and cubic line search procedure is used for this purpose. The standard deviation returned by the curve fitting method is taken to be σ_r.

The blur parameter σ_r can be translated into depth as follows. From Fig. 2, we have

$$\sigma_{k+1} = \rho R v\left[\frac{1}{w_d} - \frac{1}{w_d - x}\right]$$
$$\text{and}\ \ \sigma_k = \rho R v\left[\frac{1}{w_d} - \frac{1}{w_d + \Delta d - x}\right] \quad (3)$$

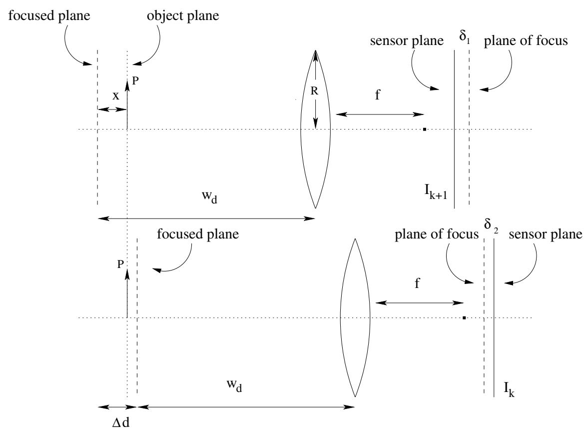


Figure 2. An object point P that does not appear in focus in any of the captured images.

where x denotes the distance between the position of the translational table when I_{k+1} was captured and where it actually should have been in order to bring the point P in perfect focus (Fig. 2). Thus, the value of σ_r derived from image data along with the knowledge of lens parameters is then used in equation (3) for computation of x which yields a refined estimate of the depth of the point P. Since we use actual image data in the form of the relative defocus blur, our method is data-driven and leads to a more accurate shape recovery. The improvement is particularly significant as Δd increases whence interpolation is less and less dependable.

3.1 Point-wise Focused Image Recovery

The motivation for carrying out image restoration is three-fold. (i) No real-aperture camera can yield a uniformly focused image of a 3-D object. (ii) The quality of the restored image can be used to evaluate the performance of shape estimation techniques since an accurate shape profile must lead to a good quality restored image. (iii) The restored image serves the purpose of verifying the assumptions and approximations made in the algorithm.

For a given object point P (Fig. 2), we pick that image I_k in the sequence $I_1, I_2, \ldots, I_M$ which yields the strongest focus measure value. The formation of the defocused sub-image $g_k(m,n)$ in image I_k can be described as

$$g_k(m,n) = i_f(m,n) * h_k(m,n) + \eta(m,n) \qquad (4)$$

where '$*$' represents 2-D convolution and $\eta(m,n)$ is additive noise. From the shape estimate D_P of the object point P, the blur parameter corresponding to $h_k(m,n)$ can be computed as

$$\sigma_k = \rho R v \left[\frac{1}{w_d} - \frac{1}{D_P}\right]. \qquad (5)$$

If $H_r(u,v)$ is the Fourier transform of the restoration filter, then the process of restoring $i_f(m,n)$ can be described as $i_f(m,n) = FT^{-1}\{G_k(u,v)H_r(u,v)\}$ where FT^{-1} denotes the 2-D inverse Fourier transform. In order to accommodate the space-variant nature of the blur distribution, we propose to retain only the central pixel of $i_f(m,n)$ thus obtained and not the whole sub-image $i_f(m,n)$. This is due to the fact that during the formation of $g_k(m,n)$, the central pixel of the focused sub-image $i_f(m,n)$ contributes the most to the pixel value at that location. By repeating the above process at every pixel of the observed image, the entire focused image is reconstructed. Note that for every object point, there is a corresponding unique image in the SFF sequence for which the focus measure value is the strongest. The sub-image $g_k(m,n)$ chosen for a point can come from any of the images in the sequence depending on its depth. Hence, the entire sequence of images is utilized in the reconstruction of the focused image and this leads to a good quality focused image. Since the Wiener filter is optimal in the statistical sense, we employ it for restoration. The restoration filter is given by

$$H_r(u,v) = \frac{e^{-(u^2+v^2)\sigma_k^2/2}}{e^{-(u^2+v^2)\sigma_k^2} + \frac{S_\eta(u,v)}{S_f(u,v)}} \qquad (6)$$

where σ_k is given in terms of D_P by equation (5), while $S_\eta(u,v)$ and $S_f(u,v)$ are the power spectral densities of the noise and the focused image, respectively.

4 Experimental Results

In this section, we demonstrate the performance of the proposed method on synthetic as well as real images. The results are also compared with traditional SFF.

In the first experiment, a sequence of images corresponding to a object with 'ramp' depth of total height of 500 μm was synthetically generated. Since this experiment is a simulated one, the blur kernel was strictly Gaussian. The following camera parameters were assumed: $R = 10$ mm, $f = 30$ mm, $w_d = 75$ mm, and pixel dimension on the sensor array $= 2.5$ μm sq. The depth map was estimated using the Gaussian interpolation method as well as the proposed scheme. Table 1 shows a comparison of the rms errors incurred using the two algorithms. Clearly, the proposed method performs better than traditional SFF for all values of Δd. It scores significantly over the traditional SFF technique as the step-size becomes larger and larger.

In the next experiment, we considered the case of real images. Images of a solder ball (about 550 μm in height)

Table 1. RMS errors in depth estimation for traditional SFF and the proposed scheme.

Δd in μm	Traditional SFF	Proposed Scheme
75	5.1018	4.2841
100	9.5935	4.9093
125	12.3878	5.2374

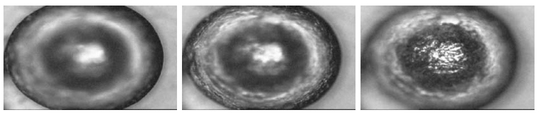
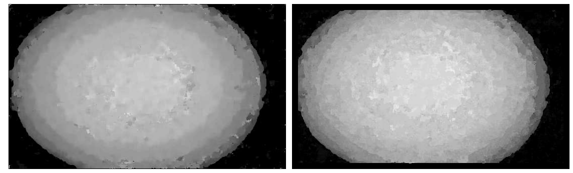


Figure 3. Sample images of a solder ball.

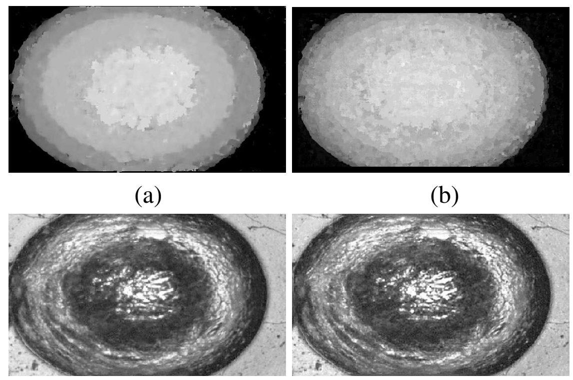
on a Ball Grid Array (BGA) were captured in the lab using a real-aperture camera attached to a Nikon LV-150 metallurgical microscope with a z-translational table. They were captured for different step-sizes. Some of the sample images are shown in Fig. 3.

The estimated shape profile of the solder ball using the interpolation strategy and the proposed method are shown in Fig. 4 for $\Delta d = 75\ \mu m$. Even though the shape of the solder ball is smooth, the depth profile obtained using traditional SFF exhibits concentric rings. These become even more pronounced when Δd is increased to 100 μm (Fig. 5). In contrast, the proposed method yields smoother and accurate estimates than traditional SFF. The improvement becomes strikingly evident as Δd increases. While the traditional SFF technique deteriorates rapidly with increase in Δd, the proposed method degrades gracefully. A quantitative error analysis is not possible for this experiment since the ground truth is not known. We next recovered the focused image of the 3-D object using these two methods. The restored focused focus images are shown in Fig. 5 (c,d) for the case $\Delta d = 100\ \mu m$. From the figure, we note that the quality of the restored image using the proposed method is visibly sharper than that of traditional SFF. The fact that the quality of the restored image is good also serves to justify modeling of the camera PSF by a 2-D Gaussian.



(a) (b)

Figure 4. Depth profile of the BGA solder ball shown as gray scale images for $\Delta d = 75\ \mu m$. (a) traditional SFF and (b) proposed method.



(c) (d)

Figure 5. Depth profile for $\Delta d = 100\ \mu m$ using (a) traditional SFF and (b) proposed method. (c,d) Restored images for traditional SFF and proposed method, respectively.

5 Conclusions

In this paper, we proposed a new method for improving the accuracy of the shape from focus scheme. It is data-driven and uses the actual relative blur between suitably chosen images for improving the accuracy of the estimated shape of a 3-D object. A space-variant image restoration method was also proposed that enables evaluation of the accuracy of the shape estimates based on the quality of the recovered focused image. Results on synthetic as well as real images clearly demonstrate the superiority of the proposed method over traditional SFF.

References

[1] S. K. Nayar and Y. Nakagawa, "Shape from focus," *IEEE Trans. PAMI,* vol. 16, pp. 824-831, 1994.

[2] M. Subbarao and T. Choi, "Accurate recovery of three-dimensional shape from image focus," *IEEE Trans. PAMI,* vol. 17, pp. 266-274, 1995.

[3] J. Kautsky, J. Flusser, B. Zitova and S. Simberova, "A new wavelet-based measure of image focus," *Pattern Recognition Letters 23,* pp. 1785-1794, 2002.

[4] S. Chaudhuri and A. N. Rajagopalan, *Depth from defocus: A real aperture imaging approach,* Springer Verlag, 1999.

[5] D. F. Shanno, "Conditioning of quasi-Newton methods for function minimization," *Mathematics of Computing,* vol. 24, pp. 647-656, 1970.