

An occlusion insensitive adaptive focus measurement method

Tarkan Aydin^{1,2,4} and Yusuf Sinan Akgul^{1,3}

¹ *GIT Vision Lab, <http://vision.gyte.edu.tr>, Department Of Computer Engineering, Gebze Institute Of Technology, Gebze, Kocaeli 41400 Turkey*

² *Department of Mathematics and Computer Science, Bahcesehir University, Besiktas, Istanbul 34353 Turkey*

³ *akgul@bilmuh.gyte.edu.tr*

⁴ *tarkan.aydin@bahcesehir.edu.tr*

Abstract: This paper proposes a new focus measurement method for Depth From Focus to recover depth of scenes. The method employs an all-focused image of the scene to address the focus measure ambiguity problem of the existing focus measures in the presence of occlusions. Depth discontinuities are handled effectively by using adaptively shaped and weighted support windows. The size of the support window can be increased conveniently for more robust depth estimation without introducing any window size related Depth From Focus problems. The experiments on the real and synthetically refocused images show that the introduced focus measurement method works effectively and efficiently in real world applications.

© 2010 Optical Society of America

OCIS codes: (110.6880) Three-dimensional image acquisition; (100.2000) Digital image processing; (150.5670) Range finding; (150.6910) Three-dimensional sensing.

References and links

1. M. Subbarao and G. Surya, "Depth from defocus: A spatial domain approach," *Int. J. Comput. Vis.* **13**, 271–294 (1994).
2. A. Pentland, S. Scherrock, T. Darrell, and B. Girod, "Simple range cameras based on focal error," *J. Opt. Soc. Am. A* **11**, 2925–2934 (1994).
3. S. K. Nayar, M. Watanabe, and M. Noguchi, "Real-time focus range sensor," *IEEE Trans. Pattern Anal. Mach. Intell.* **18**, 1186–1198 (1996).
4. A. N. Rajagopalan and S. Chaudhuri, "A variational approach to recovering depth from defocused images," *IEEE Trans. Pattern Anal. Mach. Intell.* **19**, 1158–1164 (1997).
5. V. Aslantas and D. T. Pham, "Depth from automatic defocusing," *Opt. Express* **15**, 1011–1023 (2007).
6. P. Favaro, S. Soatto, M. Burger, and S. J. Osher, "Shape from defocus via diffusion," *IEEE Trans. Pattern Anal. Mach. Intell.* **30**, 518–531 (2008).
7. E. Krotkov, "Focusing," *Int. J. Comput. Vision* **1**, 223–237 (1987).
8. J. V. Michael Bove, "Entropy-based depth from focus," *J. Opt. Soc. Am. A* **10**, 561–566 (1993).
9. S. Nayar and Y. Nakagawa, "Shape from Focus," *IEEE Trans. Pattern Anal. Mach. Intell.* **16**, 824–831 (1994).
10. M. Subbarao and T. Choi, "Accurate recovery of three-dimensional shape from image focus," *IEEE Trans. Pattern Anal. Mach. Intell.* **17**, 266–274 (1995).
11. Y. Schechner and N. Kiryati, "Depth from defocus vs. stereo: How different really are they?" *Int. J. Comput. Vis.* **39**, 141–162 (2000).
12. J. A. Marshall, C. A. Burbeck, D. Ariely, J. P. Rolland, and K. E. Martin, "Occlusion edge blur: a cue to relative visual depth," *J. Opt. Soc. Am. A* **13**, 681–688 (1996).
13. N. Asada, H. Fujiwara, and T. Matsuyama, "Seeing behind the scene: Analysis of photometric properties of occluding edges by the reversed projection blurring model," *IEEE Trans. Pattern Anal. Mach. Intell.* **20**, 155–167 (1998).
14. S. S. Bhasin and S. Chaudhuri, "Depth from defocus in presence of partial self occlusion," *Computer Vision, IEEE International Conference on* **1**, 488 (2001).

15. P. Favaro and S. Soatto, "Seeing beyond occlusions (and other marvels of a finite lens aperture)," *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on* **2**, 579 (2003).
16. T. Aydin and Y. Akgul, "A new adaptive focus measure for shape from focus," in "BMVC08," (2008).
17. H. Nair and C. Stewart, "Robust focus ranging," in "Computer Vision and Pattern Recognition, 1992. Proceedings CVPR '92, 1992 IEEE Computer Society Conference on," (1992), pp. 309–314.
18. J. M. Tenenbaum, "Accommodation in computer vision," Ph.D. thesis, Stanford, CA, USA (1971).
19. T. M. Subbarao and A. Nikzad, "Focusing technique," *Image Sig. Process. Anal.* **32**, 2824–2836 (1993).
20. S. Jutamulia, T. Asakura, R. D. Bahuguna, and P. C. DeGuzman, "Autofocusing based on power-spectra analysis," *Appl. Opt.* **33**, 6210–6212 (1994).
21. Y. Xiong and S. Shafer, "Moment and hypergeometric filters for high precision computation of focus, stereo and optical flow," *Int. J. Comput. Vis.* **22**, 25–59 (1997).
22. J. Kautsky, J. Flusser, B. Zitov, and S. Simberov, "A new wavelet-based measure of image focus," *Pattern Recognit. Lett.* **23**, 1785–1794 (2002).
23. M. Kristan, J. Pers, M. Perse, and S. Kovacic, "A bayes-spectral-entropy-based measure of camera focus using a discrete cosine transform," *Pattern Recognit. Lett.* **27**, 1431–1439 (2006).
24. J. Meneses, M. A. Suarez, J. Braga, and T. Gharbi, "Extended depth of field using shapelet-based image analysis," *Appl. Opt.* **47**, 169–178 (2008).
25. W. Huang and Z. Jing, "Evaluation of focus measures in multi-focus image fusion," *Pattern Recognit. Lett.* **28**, 493–500 (2007).
26. M. Subbarao and J.-K. Tyan, "Selecting the optimal focus measure for autofocusing and depth-from-focus," *IEEE Trans. Pattern Anal. Mach. Intell.* **20**, 864–870 (1998).
27. Y. Tian, K. Shieh, and C. F. Wildsoet, "Performance of focus measures in the presence of nondefocus aberrations," *J. Opt. Soc. Am. A* **24**, B165–B173 (2007).
28. M. Subbarao, T.-C. Wei, and G. Surya, "Focused image recovery from two defocused images recorded with different camera settings," *IEEE Trans. Image Process.* **4**, 1613–1628 (1995).
29. P. Favaro and S. Soatto, *3-D Shape Estimation and Image Restoration: Exploiting Defocus and Motion-Blur* (Springer London, 2007).
30. D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int. J. Comput. Vis.* **47**, 7–42 (2002).
31. R. Sakurai, "Irisfilter," <http://www.reiji.net/> (2004).
32. J. Chen, S. Paris, and F. Durand, "Real-time edge-aware image processing with the bilateral grid," in "SIGGRAPH 07," (ACM, New York, NY, USA, 2007), p. 103.
33. N. Joshi, R. Szeliski, and D. Kriegman, "PSF estimation using sharp edge prediction," *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on* pp. 1–8 (2008).
34. N. Joshi, C. Zitnick, R. Szeliski, and D. Kriegman, "Image deblurring and denoising using color priors," *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on* pp. 1550–1557 (2009).

1. Introduction

Images taken with finite aperture lens cameras contain depth cues about the real world due to the limited depth of field properties of the finite aperture lenses. Limited depth of field induces depth related blur in the images, except for the focused scene points. This information is effectively utilized by many methods in order to estimate depth of the scenes from its images. Such methods include depth from defocus (DFD) [1–6] and depth from focus (DFF) [7–10].

DFF uses a number of images (DFF set) of the same scene taken by focusing at different depths. It then determines the best focused image for each scene point. Once the best focused images are determined, the depth map of the scene can easily be computed by utilizing the general rules of optical geometry. As a result, comparable measurement of focus quality of image sections becomes the key task of DFF.

Focus quality of images is evaluated by a focus measure which traditionally analyzes the intensity variations in the images either in spatial or frequency domain assuming that the best focused image sections of the scene points have the highest intensity variations. In order to increase robustness of measurement against the problems such as noise, edge bleeding, and magnification, computed focus measures are aggregated in a local window. This approach implicitly assumes that the surfaces of the objects in the scene can be approximated by surface patches parallel to the image plane. Although this assumption makes the focus quality measurement practical, it is not always valid in the real world because object surfaces can have very

complex structure including depth discontinuities. As a result, the conventional DFF methods fail to yield accurate results around depth discontinuities.

Another phenomenon induced by the finite aperture lenses is the occlusion effect which makes occluded object visible through the occluding object (Fig. 1). The amount of the visibility of the occluded object depends on the focal settings and the relative distances of the occluded and occluding objects. The conditions in which the occlusion occurs are presented by Schechner and Kiryati [11]. In such cases, scene points near the occlusion regions may appear focused even focus is set to occluding or occluded objects. Consequently, determination of best focused image sections and respective depth estimation for these regions will be ambiguous [11]. The occlusion effects have been studied by several approaches [12–15], however, it is mostly unaddressed in DFF methods.

In this paper, we introduce a novel focus measure for the detection of best focused image for a given scene point. In order to address focus measure ambiguity that occurs in the presence of the occlusions, we employ all-focused image of the scene, which has the radiance information of only the visible parts of the scenes. Quality of focus is computed by measuring the similarity between the images in the DFF set and the all-focused image assuming that best focused image points have similar intensity variations with the corresponding all-focused image points.

We also employ adaptively weighted support windows [16] to aggregate the computed focus measure. The adaptive support windows would address the depth discontinuity problem by assigning different weights for pixels with different depth values. Since the pixel depths are not available, we use image intensities to determine whether given two neighboring pixels have similar depth values. Our adaptive measurement method allows larger window sizes for a more robust depth estimation without introducing any window size related DFF problems. The large support window sizes help with the edge bleeding problem as well [17]. Our method also has the advantage of recovering the depth values of the fine structure with considerably large support windows. The experiments performed on synthetically refocused images with ground truth depth values show that our method produced much more accurate depth values compared with traditional DFF methods. Other experiments with real images show the effectiveness of our method with sharp depth discontinuities and fine image structure.

The rest of the paper is organized as follows. Section 2 gives background of the DFF method and explains the occlusion problem in DFF systems. Section 3 describes the details of our novel focus measure operator. We describe the experiments performed in Section 4. Finally, we provide concluding remarks in Section 5.

2. Previous Work

2.1. Depth From Focus

DFF is a method for recovering depth from a set of N images (DFF set) taken with gradually changing focus settings. In other words, for each scene point viewed, there are N pixels taken with different focus setting. For each pixel of the same scene point, the quality of the focus is measured locally and the best focused image is determined. The focus setting of the selected image is then used for depth estimation of the corresponding point. The relation between the scene point depth and focal length of the lens is found by using

$$\frac{1}{f} = \frac{1}{u} + \frac{1}{v}, \quad (1)$$

where f is the focal length of the lens, u is distance between lens and image plane, and v is the depth of the scene point. DFF can be implemented by changing either f , u , v , or any combination of them.

The first objective of DFF methods is therefore to find the best focused image points in the DFF set. In order to measure the focus quality of an image region around a pixel, a focus measure operator is used. The operator should produce high response to high frequency variations of the image intensity, since better focused image has higher frequency components. It should produce maximal response to the perfectly focused image regions.

After the focus measure computation, the focus setting f for each pixel (x,y) of the same scene point is simply selected as

$$f_i(x,y) = \arg \max_i (FM_i(x,y)), \quad (2)$$

where $i = 1..N$ and N is the number of images in the set. $FM_i(x,y)$ is the focus measure computed for the image region around the pixel position (x,y) of the i^{th} image in the set.

A wide range of focus measures has been proposed in the literature. Simple and frequently used focus measure operators include Tenengrad [18], Variance [19], and Laplacian [7]. Since the x and y components of the Laplacian operator may cancel out each other, the operator might produce incorrect responses, even in the presence of high frequency variations in image intensities. In order to prevent cancelation of components, Nayar and Nakagawa [9] introduced a new focus operator called Modified Laplacian (ML) which sums the absolute values of the components. There are also several other focus measures based on the entropy [8], power spectra analysis [20], moment filters [21], wavelets [22], discrete cosine transforms [23], and shapelet decompositions [24]. Comparison of different focus measures under certain circumstances can be found in [25–27].

2.2. Focus Measure Ambiguity due to Occlusion

The focus measures of the literature analyze the intensity variations around the local image regions to detect the degree of focus. The main assumption of these methods is that the real aperture lenses would produce blurry observations for object points unless they are in perfect focus. The blurriness of defocused image regions is due to the lack of high frequency image components. In other words, an object point is in its best focus on an image that produces the highest frequency components. The above assumption is realistically acceptable for smooth and continuous surfaces. However, it is problematic around the depth discontinuities where the analysis of intensity variations would be affected from object points at different depths.

Another problem with the frequency analysis based focus measures is the occlusion problem. In the presence of occlusion, real aperture lenses exhibit a complex optical phenomenon that the pinhole model cannot explain. They make the occluded scene points visible through the occluding objects. As a result, the scene points on the occluding region may appear focused when focus is set to occluding or occluded objects. In such circumstances, existing focus measures would encounter difficulty in detecting which focus setting would produce the best focused images around these regions. Therefore depth estimation for these regions will be ambiguous [11].

A sample image formation process explaining the occlusion effect in real aperture lenses is illustrated in Fig. 1. Two images of the same scene are taken one with focus is set to occluding and the other with focus is set to occluded. When the focus is set to occluded object (background object), the image points near the occluding boundary receive a mixture of light both from focused occluded and blurred occluding regions [Fig. 1(a) and 1(c)]. As a result, the occluding object may appear to be in focus even though the focus is set to the occluded object. Interestingly, the same occluding object would also be in focus if the focus is set on it [Fig. 1(b) and 1(d)]. Therefore, it is not sufficient to estimate the focus quality of occluding objects by simply performing an intensity variation analysis. More scene radiance based information needs to be incorporated into the focus measurement process.

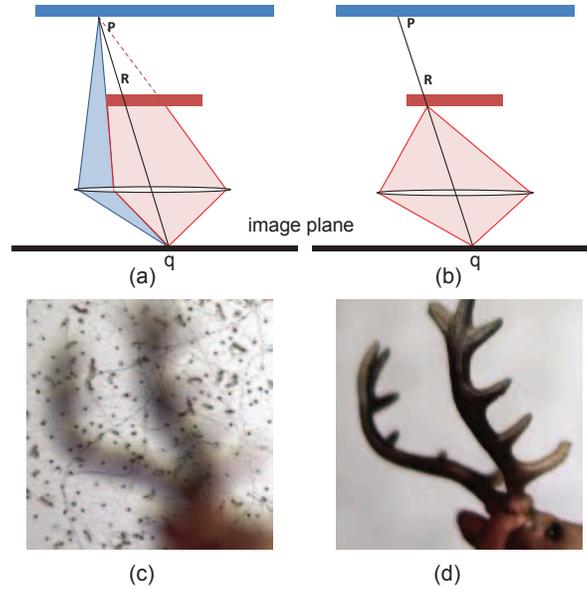


Fig. 1. Image formation process of a scene with two fronto parallel layers where focus is set to (a) background (occluded) and (b) foreground (occluding). The image point q may appear focused when focus is set to foreground or background. Real images of a sample scene taken with focus setting set to (c) occluded and (d) occluding. Although the antlers are defocused in (c), its corresponding image regions contain higher frequency variation than in (d). The images in (c) and (d) are cropped from the images of the scene in Fig. 8

Image formation process is approximated by convolving the scene irradiance with a point spread function. Point spread function would be a Dirac δ function for focused cases and a gaussian blur kernel for defocused cases. Consequently, the image formation process for image point q in Fig. 1(b) can be written as

$$I_q = L_{occluding} * \delta(R), \quad (3)$$

where $L_{occluding}$ is the radiance of the occluding object, and $*$ is the convolution operator. Similarly, the image formation process for image point q in Fig. 1(a) can be written as

$$I_q = \alpha L_{occluded} * \delta(P) + L_{occluding} * H_R, \quad (4)$$

where $L_{occluded}$ is the radiance of the occluded object, $0 < \alpha < 1$ is the attenuation in the light rays emanating from the point R due to the blocking of occluding object, and H_R is the gaussian blur kernel at the point R . It can be easily inferred from Eq. (4) and Eq. (3) that image point q seems to be focused for both cases. Therefore, depending on the frequency variations in the scene radiance at point P and R , the image point q can be assigned to either depth of P or R .

3. Robust Focus Measure

In order to address the ambiguity problem present in existing focus measures, radiance of the scene can be examined so that contribution of occluded object to the corresponding image intensities can be detectable. Radiance of the scene can be obtained by using multi-focus image fusion on the DFF image set [28]. However, multi-focus image fusion methods are also based on the image intensity variations and eventually they suffer from the occlusion problem as well.

As a practical solution, we infer the radiance information of the visible part of the scene from its all-focused image. The all-focused image can be obtained optically by setting the diameter of the lens aperture to a small value. In this case, the camera acts like a pin-hole camera and the captured image closely represents the radiance of the visible scene. All-focused image is also employed by some DFF and DFD methods in order to measure image focus quality around the image edge regions [17] or measure the degree of blur in the images [8].

We measure the image focus quality without suffering from the occlusion problem by exploiting the scene radiance information. Because the intensities of the focused image regions should resemble the intensities of the all-focused image, it is used as a reference in our focus measure. Therefore, we define our focus measure as the similarity between the all-focused image and the DFF image set. This focus measurement is robust against the occlusion problem because it does not depend on the degree of local intensity variations. However, measuring the image similarity is not trivial because different aperture and focus settings would result in brightness changes between images of the same scene. In order to address this situation, we use normalized cross-correlation (NCC) as the similarity metric because of its insensitivity to the brightness differences.

$$NCC_{\Omega}(A, B) = \frac{\sum(A - \bar{A})(B - \bar{B})}{\sqrt{\sum(A - \bar{A})^2 \sum(B - \bar{B})^2}}, \quad (5)$$

where Ω is the local evaluation window in which the similarity is computed. By using Eq. (5), the focus measure of image point (x, y) is defined as

$$FM_i(x, y) = NCC_{\Omega}(I_f(x, y), I_i(x, y)), \quad (6)$$

where I_f is the all-focused image of the scene and I_i is the i^{th} image in the DFF set.

Another optical phenomenon which occurs with the finite aperture lens is image magnification due to the different aperture and focus settings used by each image. The magnification causes misregistration between all-focused image and images of the DFF set. This problem can be avoided by estimating the magnification factor [29]. Images in the DFF set are registered with respect to the all-focused image.

The focus measure for an image pixel should be aggregated in a finite support window so that it is robust against the problems due to the noise, edge bleeding, etc. The size of the window must be large enough in regard to blur radius to be free from edge bleeding [17] and to handle noisy and homogeneous image parts. This approach assumes that the depth of the scene does not change inside the window. However, larger support windows located on depth discontinuities contain pixels from different depths which violate the depth constancy assumption. Therefore, depth estimation around these regions would be inaccurate.

In order to recover scene depth not only around depth-homogeneous regions but also around depth discontinuities, a proper support window should be selected for each pixel adaptively by considering the shapes of the objects in the scene. However the structure of the scene is not known a priori, hence it should be inferred from a reasonable source. Assuming that the scene radiance contains cues about the scene structure, it is possible to produce an adaptive support window for each pixel [16]. Adaptivity of windows is achieved by assigning weights to each pixel in the rectangular support window. A different support window is computed for each pixel in the all-focused image. The weights are assigned to the support window pixels according to similarity and proximity scores between the pixels of the window and the pixel for which the window is computed. Similar and closer pixels get larger weights in the assumption that they probably lie on the same surface

Weights of the support window ω_{x_0, y_0} centered around the pixel x_0, y_0 are computed using all-focused image I_f according to formula,

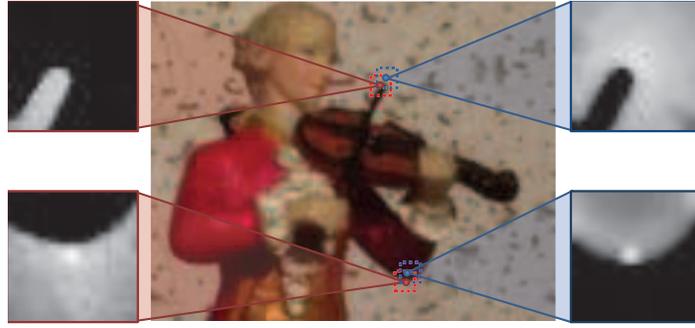


Fig. 2. Computed adaptive support windows for four different pixels in an image. The brighter pixel means larger weight. The weights for darker regions are close to zero.

$$\omega_{x_0,y_0}(x,y) = \exp(-(\Delta d/\gamma_1 + \Delta I_f/\gamma_2)), \quad (7)$$

where

$$\Delta d = [(x - x_0)^2 + (y - y_0)^2]^{1/2}, \quad (x,y) \in \Omega_{x_0,y_0} \quad (8)$$

and

$$\Delta I_f = \|I_f(x,y) - I_f(x_0,y_0)\| \quad (9)$$

where γ_1 and γ_2 are constant parameters to supervise relative weights. Δd can be considered as the euclidian distance in spatial domain and ΔI_f can be considered as the euclidian distance in color space. Figure 2 shows sample adaptive support windows computed for four different pixel regions of a sample image.

Using the introduced focus measure operator (FM) focus measures are computed for all images of the DFF set. Then the aggregated Adaptive Focus Measures (AFM) for pixel x_0, y_0 is computed using support window produced by Eq. (7).

$$AFM_i(x_0,y_0) = \sum \omega_{x_0,y_0}(x,y) FM_i(x,y) \quad i = 1..N \quad (10)$$

where N is the number of images in the DFF set.

Once the new focus measures are computed for each image in the set, the best-focused image region and its corresponding focus setting f are obtained for each pixel. Depth of the pixels are computed by substituting the focus settings in Eq. (1) as in the traditional depth from focus methods.

Note that, the all-focused image of the scene has to be taken with a small aperture diameter to produce a sharp image at all depth values. However, producing an equally sharp image of a scene is not trivial for real aperture lenses because even for small aperture values the lenses cannot behave like a perfect pinhole camera. It means that, for some sections of the images, the best focused images in the DFF set become sharper than the all-focus image. In this case, our method cannot detect the best focused image section. It selects the image near the best focused image, which creates problems for Eq. (6) [see Fig. 4(a) and 4(b)]. We conveniently address this problem by finding the best focus setting from Eq. (10) and then searching for the nearest peak of the Modified Laplacian Operator in a neighborhood of 3 focus settings. Although the ML operator usually produces two peaks in the presence of occlusions, our approach considers only the peak near the result estimated by our focus measure. This solution makes our final results very robust against the practical problems of obtaining real all-focused images.

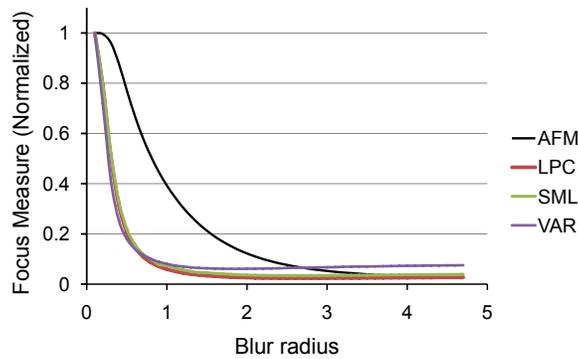


Fig. 3. Comparison of different focus measures for synthetically generated images. The focus measures are the proposed adaptive focus measure (AFM), sum of laplace (LPC), sum of modified laplace (SML), and variance (VAR)

4. Experiments

In order to test and verify the accuracy and effectiveness of our method, we performed several experiments using the DFF sets with synthetic images, synthetically refocused real images with ground truth depth values, and real images of a scene with occlusion. We took γ_1 as 5 and γ_2 as half of the window size for weight computation parameters in Eq. (7).

As the first experiment, we produced a highly textured synthetic image and its blurred images in order to test the response of our focus measure with respect to the degree of blurring. The blurring of images are performed by convolving the synthetic image with gaussian blur kernels whose sizes are gradually increased. Figure 3 shows the graph of computed focus measures with respect to the degree of blurring (blur radius). We compared our adaptive focus measure (AFM) with sum of laplace (LPC), sum of modified laplace (SML), and variance (VAR) focus measures. As seen in Fig. 3 our focus measure shows monotonic behavior like existing focus measures (i.e. focus measure values decrease as the degree of blurring increases).

Our second experiment uses a real DFF set to show the robustness of our focus measure under occlusion positions. Figure 4(a) shows two sample regions that cause problems for the existing focus measures, which produce either ambiguous or incorrect peak positions. Our focus measure, on the other hand, produces unambiguous peaks at the correct depth values as shown in Fig. 5

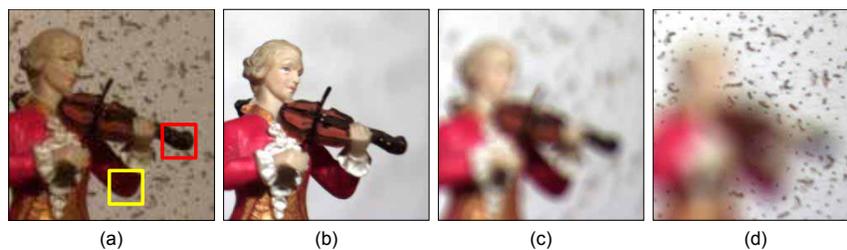


Fig. 4. All-focused image of a sample real scene (a) and its images with different focus settings (b-d). Sample occlusion regions causing focus measure ambiguity are marked with red and yellow squares in (a).

It is difficult to validate a DFF method numerically because there is no publicly available

ground truth data for DFF methods. For our third experiment, we partially addressed this problem by using the Middlebury stereo data set [30] with ground truth depth values and produced synthetically refocused images using the depth map by Iris Filter [31]. Iris Filter is a software plug-in meant to reproduce the depth of field and bokeh (defocused image regions) texture. Iris Filter needs the all-focused image, the depth map of the image, the focus depth setting, and the aperture size. We produced around 30 images with different focus settings for two data sets (Venus and Cones).

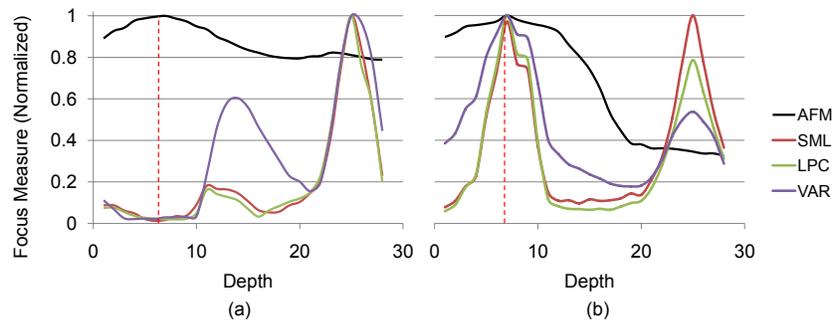


Fig. 5. Performance of different focus measures around two occlusion regions marked in Fig. 4(a). Three different existing focus measures and proposed focus measure (AFM) are used. (a) is the graph for the red window and (b) is the graph for the yellow window. The dashed lines show the manually detected true depth values. Depending on the texture in the occluded and occluding region, existing focus measures produce ambiguous or incorrect responses.

We chose images of scenes with significant depth discontinuities in order to show the performance of our method and DFF methods with different focus measures around these regions. Note that the surface continuity assumption is severely violated in these images, therefore the classical DFF methods produce erroneous results as expected. Figure 6 shows the all-focused image, true depth image, the result of the proposed method, and the result of the DFF methods with different focus measures for the Venus data set. Figure 7 shows the same images for the Cones data set. As seen in the resulting depth images in Fig. 6 and Fig. 7, at depth discontinuities, our method performs much better than other DFF methods.

We calculated the RMS errors of the estimated depth values with the ground truth. Table 4 and Table 4 show these results with various support window sizes for the DFF sets in Fig. 6 and Fig. 7 respectively. We observe that our results are as good as the best DFF methods for small window sizes. However, our error rates decrease much faster than other methods as the window sizes increase. Smaller window sizes result in inaccuracies for all methods due to the edge bleeding and insufficient texture. However, other DFF methods cannot take the advantage of larger window sizes because larger window sizes magnify the effect of depth discontinuities for other methods and hence error rates increase with the increasing window size after some point. Particularly, the performance of our method near depth discontinuities is much better than other DFF methods because our method can preserve arbitrarily shaped depth discontinuities very well due to its adaptive nature.

Our final experiment is on two DFF sets of a challenging real scene that includes sharp depth discontinuities and occlusions. Test images were captured in the lab by using a real-aperture camera. 30 images of the same scene were obtained at different focus settings. All-focused image of the scenes were captured with smallest aperture size of the lens. The true depth maps

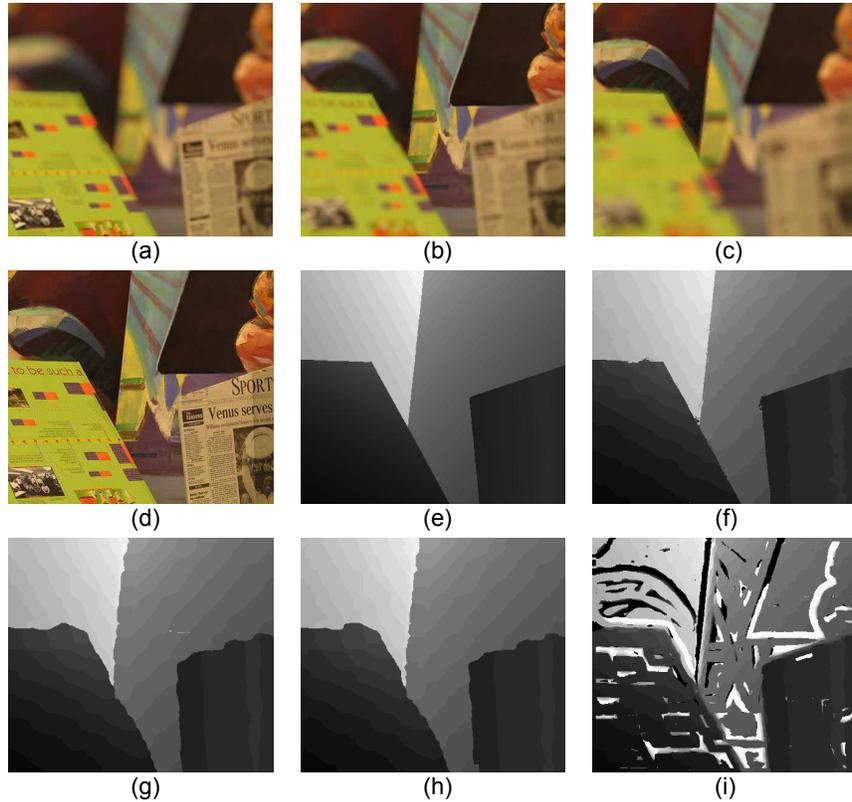


Fig. 6. Synthetically refocused images (a-c) generated with Iris filter from the all-focused image (d) with true depth map (e). Estimated depth map of the scene using our method (f) and traditional DFF with SML (g), LPC (h), and VAR (i) focus measures. The closer objects are represented by darker regions in the depth images.

Table 1. RMS errors of depth values with various window sizes on synthetically refocused real images of Fig. 6.

Method Name	Window Sizes				
	25x25	21x21	15x15	11x11	9x9
SML	14	12	12	13	14
LPC	14	13	13	13	15
VAR	59	56	55	55	56
AFM	5	6	7	13	15

for real images are unknown. Figure 8 and Fig. 9 shows the all-focused image of the scene and depth maps obtained by other DFF methods and our method. Visual inspection of the estimated depth maps shows that our method handles the sharp depth discontinuities and the fine structures very well. The traditional DFF methods smear up the depth map around the fine structures. Although DFF method which also uses adaptive windows with ML [16] performs well around these regions, it produces incorrect results around occlusion regions as traditional DFF methods do.

Results of our method and DFF with traditional focus measures using various window sizes

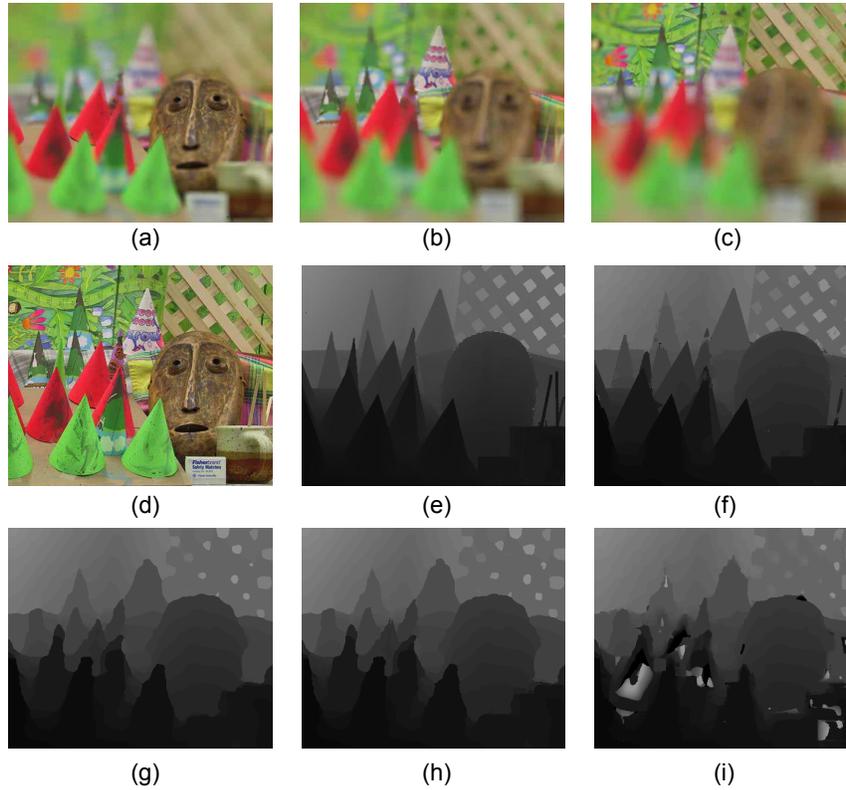


Fig. 7. Synthetically refocused images (a-c), all-focused image (d), and true depth image (e). Estimated depth image of the scene of our method (f) and traditional DFF with SML (g), LPC (h), and VAR (i) focus measures.

Table 2. RMS errors of depth values with various window sizes on synthetically refocused real images of Fig. 7.

Method Name	Window Sizes				
	25x25	21x21	15x15	11x11	9x9
SML	12	11	10	9	9
LPC	12	11	10	9	9
VAR	21	20	19	19	22
AFM	6	6	6	7	8

show that our method successfully recovers depth discontinuities. Larger window sizes produce better result with our method whereas the DFF methods produce erroneous results with the larger window sizes. In addition, our method is able to recover depth of scene points around occlusion regions robustly, which is not possible with intensity variations based focus measures.

5. Conclusions

We introduced a new adaptive focus measurement method for the depth estimation of a scene for depth from focus. Unlike the classical focus measures, our method employs all-focused image of the scene to address the focus measure ambiguity problem which occurs in the presence

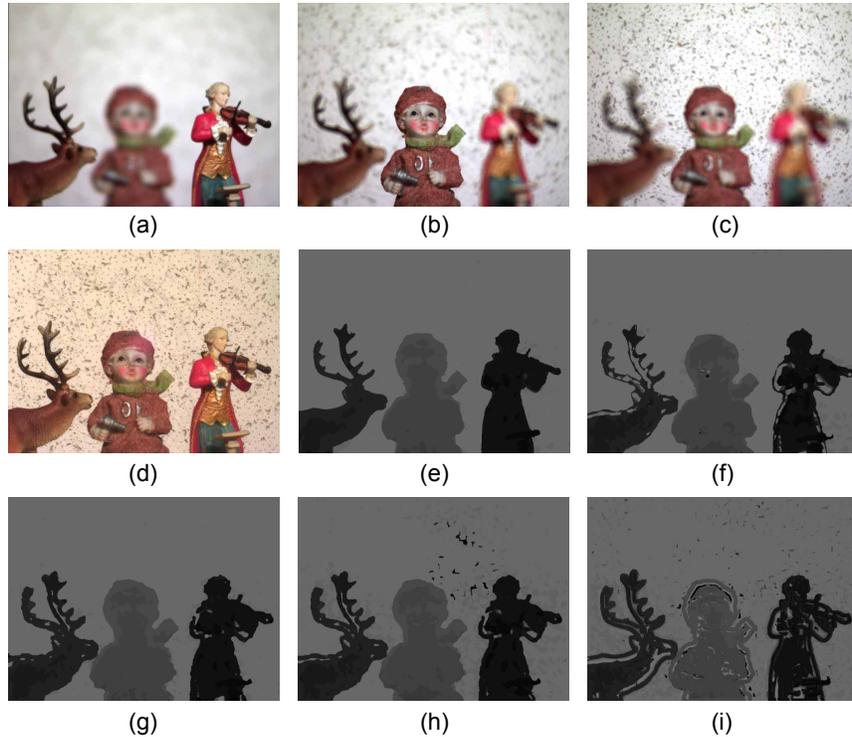


Fig. 8. All-focused image of real scene (a), sample images in DFF set (b-d) and its depth image of our method (e). Estimated depth image of the scene by DFF using adaptive windows with ML focus measure (f) and traditional DFF with SML (g), LPC (h), and VAR (i) focus measures.

of occlusion. Computed focus measures are aggregated in adaptively shaped and weighted support windows to address the problems of the depth discontinuities which is difficult to handle for traditional depth from focus techniques. Adaptivity of a window is achieved by assigning different weights to the pixels of the window. The weights of the support windows are adjusted according to similarity and proximity criteria by utilizing the all-focused image of the scene.

Results of the experiments show that the performance of the method around the depth discontinuities and occlusion regions is much better than the existing DFF methods because the proposed method can preserve arbitrarily shaped depth discontinuities without any problems due to adaptive windows. In addition, larger support window sizes make the depth estimation process even more robust without introducing any side effects as in traditional DFF methods.

Dependence of the method on the intensity values of the all-focused image does not pose any problems even for high texture surfaces because the focus measure operators perform robustly for these kinds of regions without any large window support. Consequently, for these regions our method performs at least as good as the traditional DFF methods. If the all-focused image region includes depth discontinuities without any intensity change, then our method will use an incorrect support window. However, since the same window will be used by non-adaptive methods, our method will never perform worse than traditional non-adaptive methods.

The proposed method can be interpreted as the depth map estimation of an all-focused image. Depth values for the same image can be estimated using some other technique, such as stereo or depth from shading. The all-focused image may function as an implicit registration medium

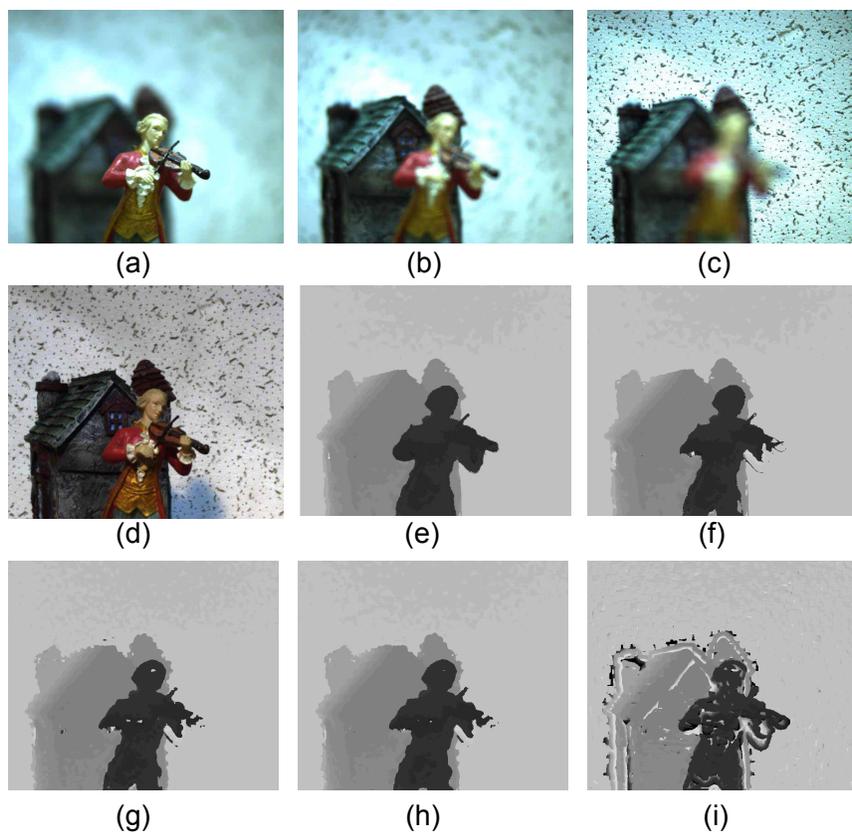


Fig. 9. All-focused image of another real scene (a), sample images in DFF set (b-d). Estimated depth image of the scene by our method (e), DFF using adaptive windows with ML focus measure and traditional DFF method with SML (g), LPC (h), and VAR (i) focus measures.

between the depth estimation methods. As a result, a robust fusion of depth values from focus and other depth estimation or shape from X methods becomes feasible.

One disadvantage of our method is the relatively higher computational requirements due to weight calculations. However this problem can be addressed by employing data structures like the bilateral grid [32]. In addition, the all-focused image requirement of our method increases the overall image acquisition time. However, increase in the time is not significant, because most of the time is used for changing focus settings and capturing images for DFF set. Another issue with the proposed method is the practical problems with all-focused image, which cannot be obtained equally sharp for the whole scene. Although we successfully address this problem by making a SML based small local search, other methods that would increase the sharpness of the all-focused image can be employed [33, 34].

Acknowledgements

This work was conducted at the Computer Vision Laboratory at Gebze Institute of Technology. It was supported by TUBITAK Career Project 105E097.