A New Adaptive Focus Measure for Shape From Focus

Tarkan Aydin Yusuf Sinan Akgul
GIT Vision Lab
http://vision.gyte.edu.tr
Department Of Computer Engineering
Gebze Institute Of Technology
Cayirova, Gebze, Kocaeli 41400 Turkey
{taydin,akgul}@bilmuh.gyte.edu.tr

Abstract

This paper proposes a new focus measure operator for Shape From Focus to recover a dense depth map of a scene. The method can handle depth discontinuities effectively by using adaptively shaped and weighted support windows. The support window shapes and weights are determined from the image characteristics of the all-focused image of the scene. Similar and closer pixels in the support window get higher weights, which inhibits the problems caused by the depth discontinuities. The size of the support window can be increased conveniently for a more robust depth estimation without introducing any window size related Shape From Focus problems. The large support window sizes also addresses the edge bleeding problem. The experiments on the real and synthetically refocused images show that the introduced ideas work effectively and efficiently in real world applications.

1 Introduction

Recovering the 3D structure of a scene from multiple images is one of the main research areas of Computer Vision. Shape from defocus (SFD) and shape from focus (SFF) are two of these traditional approaches which use focus as a cue for depth extraction of a scene. SFD[14][3] is the problem of reconstructing the depth map of a scene from a set of images by estimating the relative blurriness of images. In contrast, the SFF estimates the depth by searching for the best focused scene points from image set taken with different focus settings. For each image in the sequence, the quality of focus is computed and the best focused image in the sequence is determined for each pixel. The focus setting of the selected image is used for depth estimation of the corresponding pixel. The relation between the object depth and focal length of the lens is found by using

$$\frac{1}{f} = \frac{1}{u} + \frac{1}{v}, \tag{1}$$

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

Inferring shape form focus can be posed as a problem of measuring the focus quality of image pixels. Although there are different SFF approaches, such as [5][8][13], almost all of them employ local windows to measure the focus quality because it is very difficult to measure the focus quality using a single pixel. The local window approach assumes that the surfaces of the objects in the scene can be approximated by surface patches parallel to the image plane (fronto planar or equifocal surface assumption). 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 and non-equifocal surfaces. As a result, the conventional SFF method fails to yield accurate results around depth discontinuities and slanted surfaces.

Another major problem with the traditional SFF methods is called edge bleeding which is caused by the occlusions due to the sharp depth discontinuities when imaged with finite aperture lenses[6] (Fig.1 (a) and (b)). Edge bleeding affects the image regions even at perfect focus and it cannot be easily addressed by traditional SFF methods because it is visible in every image in the whole image set for image areas near sharp depth discontinuities. Edge bleeding also causes focus measure ambiguities around the layered object positions [11](such as the ambiguity between scene points P and R in Figure 1 (b) and (d)). Errors due to edge bleeding can be reduced by increasing focus measure window sizes at least as large as the size of the support of the widest blur kernel expected in the setup[7, 12]. However, using larger window sizes in focus measure produces erroneous results around depth discontinuities. Nair and Stewart [7] use an all-focused image of the scene to address the edge bleeding problem. They extract the surface intensity edge positions from the all-focused image. The extracted edges are used in an SFF depth estimation method and the depth values are propagated to the other image sections. This method, however, has problems with image sections with no dominant edge features, e.g., smooth and low textured regions. On the other hand, accurate depth recovery around discontinuities can be possible using coaxial projector-camera system by utilizing projector defocus model [17].

In this paper, we propose to employ adaptively shaped support windows to measure the focus quality of a given image region from a set of images taken with different known focus settings. The adaptive support windows would address the depth discontinuity problem if they assign 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. In order to address the edge bleeding problem, we obtained the image intensity values from the all-focused image of the scene, which can be taken by a finite aperture lens with a small aperture size setting. Our adaptive measurement method allows larger window sizes for a more robust depth estimation without introducing any window size related SFF problems. The large support window sizes help with the edge bleeding problem as well. 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 SFF methods. Other experiments with real images shows the effectiveness of our method with sharp depth discontinuities and fine image structure.

The rest of the paper organized as follows. Section 2 gives background of the SFF method. Section 3 describes the details of our adaptive focus measure operator. We describe the experiments performed in Section 4. Finally, we provide concluding remarks

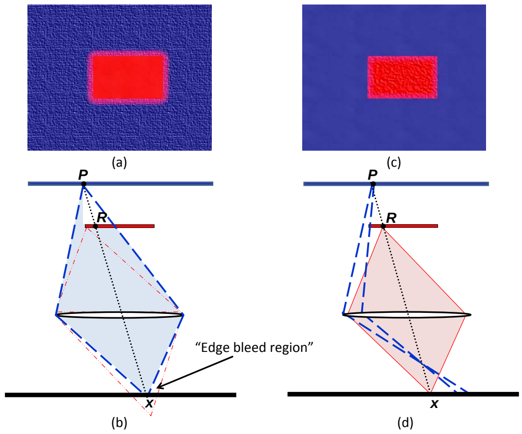

Figure 1: *(a)Picture of a scene with two equifocal layers where focus is set to background. The boundary between the background and foreground blurred because the image points near the boundary receive a mixture of light from both focused background and blurred foreground. (b)The image formation process of (a). (c) The image of the same scene when focus is set to foreground. (d)The image formation process of (c). The point x may appear focused when focus is set to both foreground or background (focus measure ambiguity between the scene points P and R).*

in Section 5.

2 Previous Work

2.1 Focus Measure Operators

A focus measure operator is used to measure the focus quality of an image region. 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. A simple and frequently used operator is Laplacian (∇^2)

$$\nabla^2 I(x,y) = \frac{\partial^2 I(x,y)}{\partial x^2} + \frac{\partial^2 I(x,y)}{\partial y^2}.$$

If the image is highly textured at pixel level, using Laplacian for depth estimation would produce satisfactory results. However, this situation is not realistic and rarely happens in real problems. To alleviate this problem, focus measures are summed up in a finite support window. For the Laplacian operator case, focus measure of the pixel x_0, y_0 is formulated for the as

$$FM(x_0,y_0) = \sum_{(x,y)\in\Omega_{x_0,y_0}} \nabla^2 I(x,y),$$

where Ω_{x_0,y_0} is the support window. Since the x and y components of the Laplacian operator may cancel out each other ($I_{xx} = -I_{yy}$), the operator might produce incorrect responses, even in the presence of high frequency variations in image intensity. In order to prevent cancelation of components, Nayar and Nakagawa [8] introduce a new focus operator called Sum of Modified Laplacian (SML) which is the sum of Modified Laplacian operator (ML) values in a neighborhood of local support window Ω. Discrete implementation of the ML and the SML operators can be easily applied in the image domain such that

$$\begin{aligned} ML(I(x,y)) &= |-I(x+s,y)+2I(x,y)-I(x-s,y)| \quad + \\ & \quad |-I(x,y+s)+2I(x,y)-I(x,y-s)|, \end{aligned}$$

where $s = 1,2,..,N$.

$$SML(x_0,y_0) = \sum_{(x,y)\in\Omega_{x_0,y_0}} ML(I(x,y)) \quad for\, ML(I(x,y)) \geq T,$$

where T is a threshold value and Ω_{x_0,y_0} is the support window centered around the pixel x_0, y_0.

2.2 Shape From Focus

SFF is a method for recovering depth from an image sequence which are taken with different focus settings. After the focus measure computation, the focus setting f for each pixel (x,y) is simply selected as

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

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

Depth values for each pixel is computed from the estimated focus settings using the lens formula (Equation 1). Although SFF works effectively for some applications, e.g, microscopic objects[8], since the planar surfaces assumption is not realistic in most cases, the traditional SFF method does not yield accurate results especially with depth discontinuities and slanted surfaces.

Subbarao and Choi[13] proposed a new concept called Focused Image Surface (FIS) which is a piecewise planar approximation of the shape of object in 3-D image space. Starting from the initial estimate of FIS, focus measure is recomputed for all possible planar surface patches in the small cubic volume formed by the image sequence. Planar surface patches with maximum focus measure are searched to extract FIS. The search is based on brute force approach therefore it is computationally expensive. Ahmad and Choi[1] use dynamic programming optimization to reduce the computational complexity of FIS. Initial estimate of FIS is obtained by using traditional SFF methods. Therefore, FIS results is dependent on the SFF results.

The piecewise planar approximation does not give accurate results for some objects with complex geometry, therefore Yun and Choi [16] propose Curved Window Focus Measure which uses piecewise curved search windows to find accurate shape of FIS. To interpolate the piecewise curved window, they introduced Nine Control Points (NCP) model and constructed Lagrange polynomial for the surface equations.

Both planar and curved surface approximation methods ignore the surface discontinuities, therefore produce erroneous results around these regions. The next section explains how we address the surface discontinuity problem effectively by introducing adaptively shaped windows for focus measure computation.

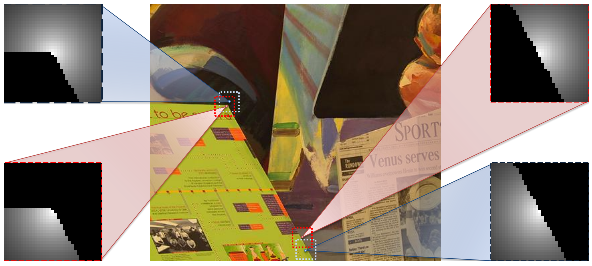

Figure 2: *Computed adaptive support windows for four different pixels in an image. The brighter pixel means larger weight and the weights for darker region are close to zero.*

3 Adaptive Focus Measuring

In traditional SFF, the focus measure for an image pixel is computed over a finite support window [4][7] with the assumption that depth of the scene should not change inside the window. The size of the window must be large enough in regard to blur radius to be free from edge bleeding [7][12] and to handle noisy and homogeneous image parts. However, larger support windows located on depth discontinuities contain pixels from different depths which violates the equifocal surfaces assumption. Therefore, depth estimation around these regions would be in accurate.

In order to recover depth of surfaces not only around homogeneous and noisy regions but also around depth discontinuities, a proper support window should be selected for each pixel adaptively. Assuming that image characteristics contains cues about scene structure, it is possible to produce an adaptive support window for each pixel based on the local image characteristics. Therefore, we need to examine intensity variations of the whole image. Unfortunately, conventional cameras with large aperture lenses tend to blur image areas around depth discontinuities even for perfectly focused surfaces, which is visualized in Figure 1. As a result, it is impractical to build adaptive support window from partially focused images in the sequence. Instead of using partially focused images, we employed the all-focused image to generate image context dependent adaptive support windows. Note that employing the all-focused image does not add any computational complexity to the system. It is also very convenient to obtain this image using the same setup that produces the image sequence for the SFF by stopping down the aperture of the lens.

Adaptivity of windows is achieved by assigning weights to each pixel in the support window. A different support windows are 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 enclosed by the window and the pixel for which 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 the following formula.

$$\omega_{x_0y_0}(x,y) = e^{-(\triangle d/\gamma_1 + \Delta I_f/\gamma_2)}, \tag{2}$$

where $\Delta d = \sqrt{(x-x_0)^2+(y-y_0)^2}$, $(x,y) \in \Omega_{x_0,y_0}$, and

$$\Delta I_f = \sqrt{(I_f^r(x,y)-I_f^r(x_0,y_0))^2+(I_f^g(x,y)-I_f^g(x_0,y_0))^2+(I_f^b(x,y)-I_f^b(x_0,y_0))^2}.$$

Δd is euclidian distance in spatial domain and ΔI is euclidian distance in color space where I_f^r, I_f^g, and I_f^b are the intensity values of image for red, green, and blue channels. γ_1 and γ_2 are constant parameters to supervise relative weights. Note that this function is the combination of two Gaussian on space distance(Δd) and range distance(ΔI). Figure 2 shows sample adaptive support windows computed for four different pixels for Venus image of the Middlebury Image Base[10].

Using one of the standard focus measure operators (FM) initial focus measures are computed for all partially focused images. Then the new focus measure AFM for pixel x_0, y_0 is computed using support window produced using Equation 2.

$$AFM(x_0,y_0) = \sum_{(x,y)\in\Omega_{x_0,y_0}} \omega_{x_0y_0}(x,y)\,FM(x,y). \tag{3}$$

Once the new focus measures are computed for each image in the sequence, the best-focused frame and its corresponding focus settings f are obtained for each pixel. Depth of the pixels are computed by substituting the focus settings in Equation 1 as in the traditional shape from focus method.

Our adaptive weighting method resembles the Bilateral Filtering [15] which is designed to perform edge preserving smoothing of images. It uses the concept of Gaussian smoothing by weighting the filter coefficients using corresponding image context. Weights are assigned according to similarity and proximity criteria which adaptively reshapes the filter. Bilateral filter was also successfully applied to stereo, inpainting, computer graphics, etc.

4 Experiments

In order to test and verify accuracy and effectiveness of our method, we performed experiments using partially focused real images and synthetically refocused real images with ground truth depth values. All experiments for the SFF method and our method were carried out using the Modified Laplacian (ML) as initial focus measure operator. We used a step size of 3 and threshold value of 5 as the ML parameters. We took γ_1 as 5 and γ_2 as half of the window size for weight computation parameters.

In order to show the accuracy of our method numerically, we performed experiments on synthetically refocused real images. The refocused images were generated with iris filter[9] using true depth map of the scene. 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. A realistic image of the scene is produced using the inputs. We produced around 30 images with different focus settings. We obtained the image and its true depth map from Middlebury Image Base[10].

In order to test the performance of our method, we chose an image of a scene with significant depth discontinuities. Note that the equifocal surface assumption is severely violated in this image, therefore the classical SFF method produces erroneous result as

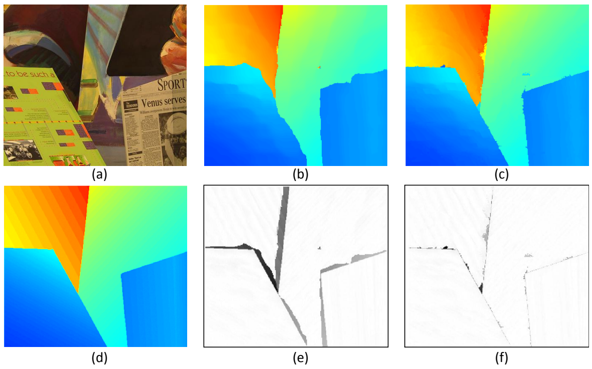

Figure 3: *Synthetic refocused images were generated with iris filter from the all-focused image(a) with true depth map(d). Estimated depth map of the scene using (b)SFF and (c) our method with window size 25.(e) Difference image of (b) and (d). (f) Difference image of (c) and (d)*

	RMS errors				
Window Size	25x25	21x21	15x15	11x11	9x9
SFF	14,21	12,97	12,24	13,29	14,45
Our Method	5,32	5,44	7,5	12,31	14,46

Table 1: RMS errors of produced depth map with various window sizes.

expected. Figure 3 shows the all-focused image, the result of the SFF method, and the result of the proposed method. We also show the difference images between the true depth map and the estimated depth maps for both method. As seen in the difference images in Figure 3, the SFF method and our method performs comparably around continuous depth regions. However, at depth discontinuities our method performs much better than SFF.

We calculated the RMS errors between our result and the ground truth. The same errors are calculated for the SFF method. Table 1 shows these results with various support window sizes. For our method error rates decrease as the window sizes increase, because smaller window sizes results in inaccuracies due to the edge bleeding and insufficient texture. In contrast, SFF suffers from the depth discontinuities and hence error rates increases with the increasing window size. Particularly, the performance of our method near depth discontinuities is much better than the SFF because our method can preserve arbitrarily shaped depth discontinuities very well. We tested our method also on challenging real scenes including sharp depth discontinuities. Four sets of test images were captured in the lab using a real-aperture camera. For each set, 30 images were obtained with different focus settings. All-focused image of the scenes were captured with smallest aperture size of the lens. The true depth maps for real images are unknown. Figure 4 shows the all-focused image of each set and their depth maps obtained by SFF and our method. Visual inspection of the estimated depth maps shows that our method handles the sharp depth discontinuities very well. We also observe that the fine structure in rows 1 and 3 (tree branches) is handled properly with our method. The traditional SFF method smears up

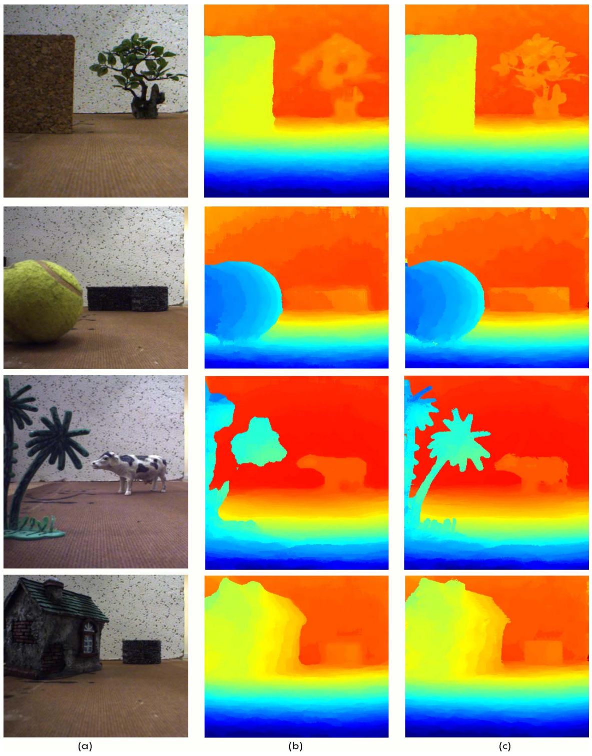

Figure 4: *Real test images used in experiments Column (a) all-focused image (b) results of traditional SFF with window size 21x21 (c) results of proposed method with the same window size.*

the depth map around the fine structures. Our last experiment is also on a real scene taken with another setup with slightly higher depth of field. Some of the images in the set are shown in the top row of the Figure 5. Results of our method and SFF with various window size shows that our method successfully recovers depth discontinuities. Larger window sizes produce better result with out method whereas the SFF method produces erroneous results with the larger window sizes.

5 Conclusions

We introduced a new focus measure method for the depth estimation of a scene for shape from focus. The method employs novel adaptively shaped and weighted support windows to addresses the problems due to the depth discontinuities and edge bleeding, which are difficult to handle for traditional shape from focus techniques. An all-focused image of the scene is used to adjust the weights of the support windows according to similarity and proximity criteria.

Results of the experiments show that the performance of the method around the depth discontinuities is much better than the SFF method because the proposed method can pre-

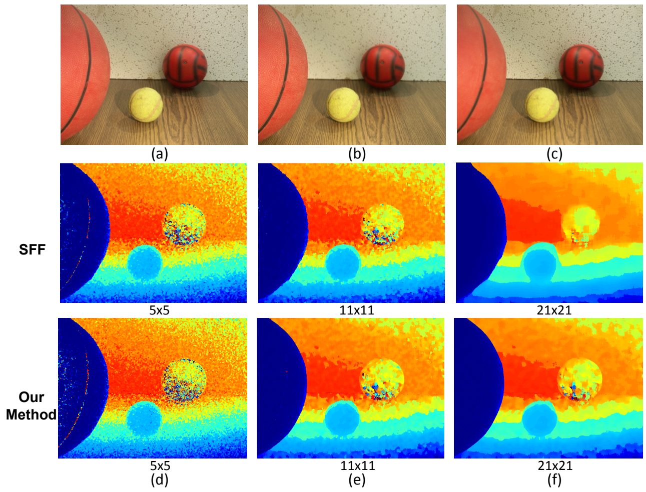

Figure 5: *(a) All-focused image of the scene. Partially focused images with focus is set to near object(b) and far object(c). Second row shows the results of SFF method and third row shows results of proposed method using support region of (d) 5x5,(e)11x11,(f)21x21 pixels.*

serve arbitrarily shaped depth discontinuities without any problems. In addition, larger support window sizes makes the depth estimation process even more robust without introducing any side effects as in traditional SFF methods.

The methods dependence 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 SFF 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 methods 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 shape from shading. The all-focused image may function as an implicit registration medium between the depth estimation methods. As a result, a robust fusion of depth values from focus and other depth estimation methods becomes feasible.

Compared with traditional the SFF methods, our method seems to have higher computational complexity due to computation of the weights for the support windows. Fortunately, using proper data structures, like the bilateral grid [2], very fast implementation of the algorithm is possible.

6 Acknowledgements

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

References

[1] Muhammad Bilal Ahmad and Tae-Sun Choi. A heuristic approach for finding best focused shape. *IEEE Trans. Circuits Syst. Video Techn.*, 15(4):566–574, 2005.

[2] Jiawen Chen, Sylvain Paris, and Frédo Durand. Real-time edge-aware image processing with the bilateral grid. In *SIGGRAPH 07*, page 103, New York, NY, USA, 2007. ACM.

[3] Paolo Favaro, Stefano Soatto, Martin Burger, and Stanley J. Osher. Shape from defocus via diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(3):518–531, 2008.

[4] R. Jarvis. A perspective on range-finding techniques for computer vision. *Pattern Analysis and Machine Intelligence*, 3:122139, 1983.

[5] E.P. Krotkov. Focusing. *IJCV*, 1(3):223–237, October 1987.

[6] J.A. Marshall, C.A. Burbeck, D. Ariely, J.P. Rolland, and K.E. Martin. Occlusion edge blur: A cue to relative visual depth, 1996.

[7] H.N. Nair and C.V. Stewart. Robust focus ranging. In *CVPR92*, pages 309–314, 1992.

[8] S.K. Nayar and Y. Nakagawa. Shape from focus. *PAMI*, 16(8):824–831, August 1994.

[9] R. Sakurai. Irisfilter. *http://www.reiji.net/*, 2004.

[10] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1-3):7–42, April 2002.

[11] Y.Y. Schechner and N. Kiryati. Depth from defocus vs. stereo: How different really are they? *International Journal of Computer Vision*, 39(2):141–162, 2000.

[12] C.V. Stewart and H. Nair. New results in automatic focusing and a new method for combining focus and stereo. In *In Proc. SPIE Sensor Fusion II: Human and Machine Strategies*, volume 1198, page 102113, 1989.

[13] M. Subbarao and Tao Choi. Accurate recovery of three-dimensional shape from image focus. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 17(3):266–274, Mar 1995.

[14] M. Subbarao and G. Surya. Depth from defocus: A spatial domain approach. *IJCV*, 13(3):271–294, December 1994.

[15] Carlo Tomasi and Roberto Manduchi. Bilateral filtering for gray and color images. In *ICCV*, pages 839–846, 1998.

[16] Joungil Yun and T.S. Choi. Accurate 3-d shape recovery using curved window focus measure. *ICIP 99*, 3:910–914 vol.3, 1999.

[17] L. Zhang and S. K. Nayar. Projection Defocus Analysis for Scene Capture and Image Display. *ACM Trans. on Graphics (also Proc. of ACM SIGGRAPH)*, Jul 2006.