Chapter 1
Depth From Moving Apertures

Mahmut Salih Sayar[1], Yusuf Sinan Akgül[2].

Abstract Two new focus measure operators for Shape From Focus to estimate the three-dimensional shape of a surface are proposed in this paper. The images are formed by a camera using moving apertures. The well-focused image pixels are identified by frequency analysis or matching with the all-focused image of the scene. Frequency analysis involves the usage of classical focus measure operators and summing up for each aperture to find the focus quality to be maximized. The lesser method uses the match-points and error ratios of any matching algorithm used within an all-focused image region and all same-focus-different-aperture images to find the total displacement. The inverse of total displacement can be used as a focus quality measure. The experiments on real images show that the introduced ideas work effectively and efficiently.

1.1 Introduction

In recent years, the research area of retrieving the 3D structure of a scene has gained much attention in Computer Vision. Shape From Focus [1] (SFF) and Shape From Defocus [2] (SFD) are two of the popular approaches in this area. SFF is a method to find the best-focused image in the image-set for each pixel. SFD is a method that can estimate a blurriness value for each image pixel, defining a focus measure operator. SFF uses any focus measure operator for each image and selects the image that gives the maximum response. Depth of the object can be estimated by using

$$\frac{1}{f} = \frac{1}{u} + \frac{1}{v}, \tag{1.1}$$

[1] M. S. Sayar (✉)
Gebze Institute of Technology, Computer Vision Lab., Gebze, 69121 Kocaeli, Turkey
e-mail: salelltd@yahoo.com

[2] Y. S. Akgül (✉)
Gebze Institute of Technology, Computer Vision Lab., Gebze, 69121 Kocaeli, Turkey
e-mail: akgul@bilmuh.gyte.edu.tr

where f is the focal length of the lens, u is the distance between the camera lens and the image plane, v is the depth of the object.

Despite the number of algorithms in the literature that use SFF, the problem of finding a focus measure that can find a ground-truth depth map for all scene conditions is still a major challenge in computer vision. A classical method called Sum of Modified Laplacians (SML) that finds depth map by using pixel frequencies was proposed by Nayar and Nakagawa [1]. Aydin and Akgul[4] proposed a new method called Adaptive Focus Measure (AdaFocus) that uses all-focused image information around the pixel as a cue to improve the results for traditional SFF methods.

Most of the methods in literature use full aperture [1-7], and/or rely on special scene conditions [5]. In this paper, we propose a new method that finds the depth map of the scene by the use of moving apertures. Unlike [4], our method uses all-focused image information to measure the displacements of each point in all images received using different apertures. However, our method can also make use of frequency analysis, since focus blurring effects occur on every bad-focused image part regardless of apertures. Thus, it is also possible to use [1] or [4] in conjunction with our method.

The rest of the paper is organized as follows. Section 2 gives background of the focus measure operators in literature. Section 3 describes the methodology that we have defined, and the focus measure operators that we have developed. We describe the experiments performed in Section 4. Finally, we provide concluding remarks in Section 5.

1.2 Previous Work

In classical image formation process that uses full aperture, the light rays pass through the lens of camera before they reach the sensors and form the image. A sharp image region is formed around well-focused objects, and blurry image region is formed around bad-focused objects; but the size or the center of the object's image does not change.

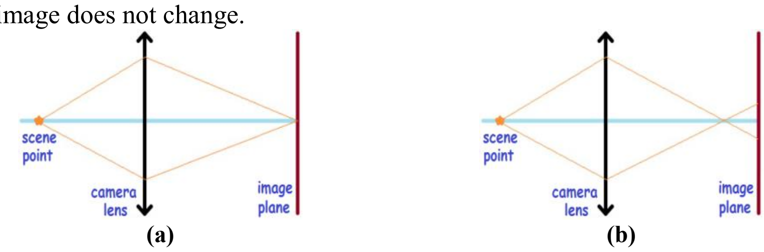


Fig. 1.1. Camera model for classic image formation. (a) Well-focused on the object. (b) Bad-focused on the object.

The objective of a focus measure operator is to evaluate the blurriness of all pixels on images retrieved under different focus settings. SML[1] is a classical focus measure operator that can measure the blurriness of an image region by means of frequency analysis. The focus quality value of a pixel at (x_0,y_0) is found using

$$\begin{aligned} ML\big(I(x,y)\big) \;=\; & \mid -I(x+s,y)+2I(x,y)-I(x-s,y)\mid + \\ & \mid -I(x,y+s)+2I(x,y)-I(x,y-s)\mid \quad , \end{aligned} \tag{1.2}$$

$$SML(x_0,y_0)=\sum_{(x,y)\in\Omega_{x_0,y_0}} ML\big(I(x,y)\big)\ for\ ML\big(I(x,y)\big)\geqslant T, \tag{1.3}$$

where s is the step-size, T is a threshold value, and Ω(x_0,y_0) is a window around the pixel at (x_0,y_0).

Since well-focused regions usually have higher frequencies, and bad-focused regions usually have lower frequencies; it is possible to estimate the depth map of a scene using SML[1]. The best focus setting for each pixel is selected as the focus setting that gives the maximum(or minimum) focus measure response, and the depth of the pixel can be found using Eq.1.1.

However, SML[1] is proven to be weak around depth discontinuities (due to edge-bleeding[6]), and textureless regions. Nair and Stewart [7] tried to eliminate the edge-bleeding problem using an all-focused image of the scene. Aydin and Akgul[4] developed a method called Adaptive Focus Measure (AFM) that uses all-focused image information to address the edge-bleeding problem. The all-focused image of a scene is obtained by using a lens with an aperture as small as possible. According to [4], the all-focused image information can be mixed with any focus measure operator using

$$AFM\big(x_{0,}y_0\big)=\sum_{(x,y)\in\Omega_{x_0,y_0}}\omega_{x_0,y_0}(x,y)FM(x,y), \tag{1.4}$$

$$\omega_{x_0,y_0}(x,y)=e^{-\left(\frac{\Delta d}{\gamma_1}+\frac{\Delta I_f}{\gamma_2}\right)}, \tag{1.5}$$

where FM(x,y) is the focus measure value for the pixel at (x,y), $\omega_{x0,y0}$(x,y) is the weight of the pixel (x,y) inside the region $\Omega_{x0,y0}$, Δd is the distance between (x,y) and (x_0,y_0), and ΔIf is the effective color difference between the all-focus pixels at (x,y) and (x_0,y_0).

The key difference of the proposed methods in this paper and most of the state-of-the-art approaches in literature is the effective usage of moving-apertures. While the classic approaches rely on only one image of the scene for each focus setting, our robust methods rely on many different-apertured images of the same scene. Moreover; like [4], our flexible methods can make use of all-focus image information and other focus measure operators. Unlike [1], [4] and [5]; our method is also capable of making use of matching methods.

1.3 Depth from Moving Apertures

Moving apertures can be implemented by using an apparatus in front of camera having differently-positioned holes. As it can be observed in Fig.1.2, the concept of moving apertures causes the light to form different-apertured images in the same focus setting which may have different center points. It is possible to model a focus measure operator based on the displacements of the center points. Moreover, the frequency of image pixels can still be used as another clue since focal blurring is not avoided. In this way, we avoid all kinds of restrictions on scene conditions; the flexibility of our approach is also proven by the ability of our method to operate with SFF, SML[1], AdaFocus[4] and matching algorithms.

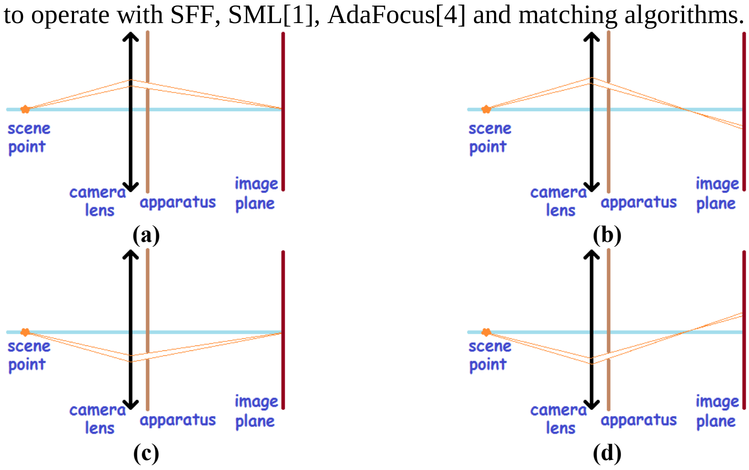


Fig. 1.2. Camera model for moving-aperture image formation. (a, c) Well-focused on the object. All light-rays meet at a certain point regardless of aperture position. (b, d) Bad-focused on the object. Aperture position determines the image position.

1.3.1 Aperture Integrals Focus Measure

The position or blurriness of a scene point in an image changes with the aperture position or size. Any classical method such as SML[1] or AdaFocus[4] can be used to measure the blurriness of each point in each image that were taken using different apertures on the same focus setting. The Aperture Integrals Focus Measure (AIFM) operator is:

$$AIFM(x,y) = \sum_{i=1}^{N_d} FM(x,y)[i], \tag{1.6}$$

where FM(x,y)[i] is the result of a classical focus measure operator for image pixel at (x,y) for aperture i and N_d is the number of apertures used.

However, frequency of the pixels is also inversely proportional to the distance of the pixel to the center of each aperture. (Fig.1.3.) Fortunately, mean value of the frequencies for all apertures can be used to address this problem. Since the number of apertures is equal for all images, summing up is enough to determine a correct focus measure value.

1.3.2 Matching Focus Measure

Template Matching can be used to find how much a pixel has moved from where it was in the all-focused image. For each focus setting; a small all-focused image window around the pixel is matched within a larger window in the images for all different aperture settings. The sum of displacement values for valid matches is used as a focus measure. Note that the template matching method is required to match correctly under different lighting conditions and/or different aperture sizes and/or focus blur effect. In this paper, we used Normalized Cross Correlation (NCC) [3] as our template matching algorithm.

Some image regions that have very low intensity values cannot be matched properly using the NCC algorithm, thus we define zero-threshold to reject these regions. Zero-threshold is used to avoid errors on very dark regions, so it is usually a very small intensity value. If the scene does not contain very dark regions, zero-threshold is not necessary.

Template matching algorithms may find bad matches due to noise, blur, bad lighting, scaling, rotation, etc. The failure of matching can be determined by using a matching error-threshold. Matching error threshold avoids wrong matches and thus, it avoids wrong additions to the focus measure value. However, a very strict error threshold also causes wrong results due to the lack of displacement information. Therefore, it is hard to estimate a good matching error threshold value for all experiments. On the other hand, it is also possible to try different error-thresholds for each image.

Since the matching method may not be able to match the all-focused image region with many different-apertured images, it becomes necessary to define aperture-threshold that rejects the current focus setting if the number of accepted matches is lower than the threshold value. Otherwise, the small number of apertures also makes a small focus measure value. Taking average solves the issue a bit, but very small number of apertures may not make a good average either. Moreover, the real reason for the inability to match is that the regions are not similar. (In our case, probably too much focal-blurring, scaling, or movement.) In fact, it is this behavior which should remind us that this focus should not be the "best" focus for the current pixel. Aperture-threshold is usually selected as a small integer such as 3, 4 or 5.

The Matching Focus Measure (MFM) is

$$MFM(x,y) = \begin{cases} \infty & , if\ c(x,y) < T_a, \\ TFM(x,y), & otherwise, \end{cases} \tag{1.7}$$

where

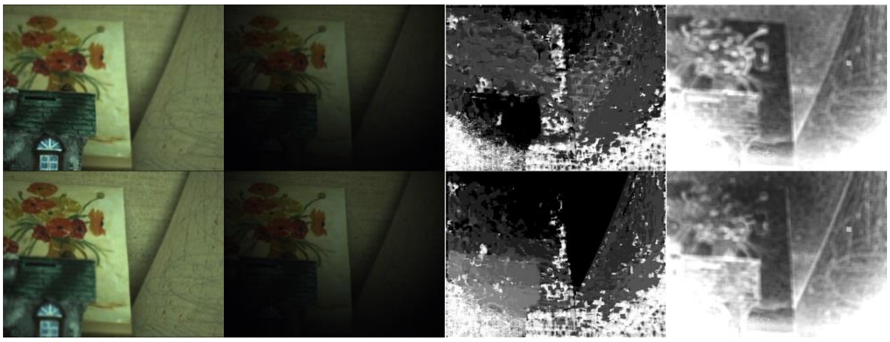
$$N_c(n,x,y) = \begin{cases} 1, if\ E_n(x,y) < T_e \wedge I_f(x',y') > T_z, \\ 0, \qquad otherwise, \end{cases} \tag{1.8}$$

$$c(x,y) = \sum_{i=1}^{N_d} N_c(i,x,y), \tag{1.9}$$

$$TFM(x,y) = \sum_{i=1}^{N_d} N_c(i,x,y)\, d\big(I_i(x,y)\big), \tag{1.10}$$

where $d(I_i(x,y))$ is the distance from the image pixel $I_i(x,y)$ to its matched pixel in all-focused image $I_f(x',y')$, $E_i(x,y)$ is the resulting absolute error value for matching between all-focused image(I_f) and the image retrieved using aperture $i(I_i)$, Nd is the number of apertures used, T_e is the matching error-threshold, T_z is the zero-threshold, and T_a is the aperture-threshold.

(a)
(b)



(1) (2) (3) (4)

Fig. 1.3. Template matching results for: (a) Focus setting 6 using aperture 6. (b) Focus setting 29 using aperture 6. (1) w/full aperture. (2) w/aperture 6. The region around the center of aperture responds better in terms of frequency i.e. it has less blur effect. (3) Displacements of pixels. (Normalized) (4) Error results of NCC. (Normalized) All-focus image can be seen in Fig.1.5(a1).

1.4 Experiments

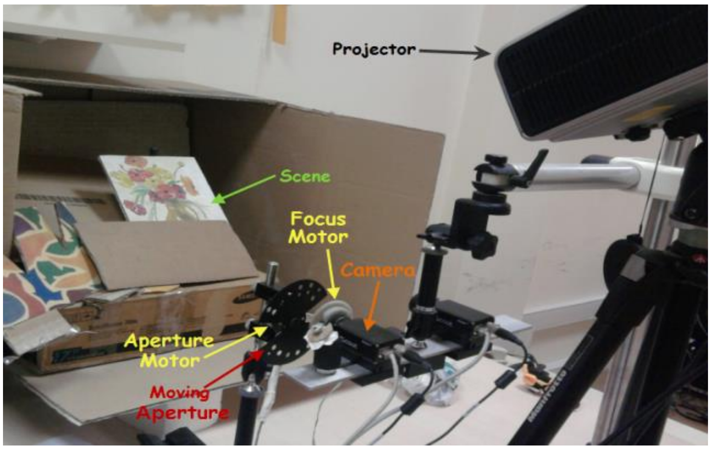


Fig. 1.4. Experimental setup.

A mechanical setup involving two step motors and a control board was used, as well as a Basler camera, and a projector to get the ground-truth information using Scharstein-Szeliski[5] method. The apparatus had 12 holes, and we had a total of 78 different positions by moving it. Some positions were ignored due to the lack of scene's visibility. We used an intensity threshold of 10 to determine the visibility of a certain pixel. 54 aperture positions for 1st experiment, 55 for 2nd experiment were selected by using a threshold of 10% to visibility score. In order to reduce noise and get more reliable images, temporal average of 3 images was used. The image resolution of our camera was 640x480, the number of focus steps used in our experiments was 30. 8 bits binary coding, exposure times of 50ms, 100ms and 150ms were used for [5] with bitwise threshold of 5 and codewise threshold of 2. For SML[1] and AdaFocus[4], a step size of 1,2 and 3 were used seperately, with a constant window size of 25 and focus measure threshold of 5. For Ada-Focus[4]; 5 was used as γ_1, and 12 was used as γ_2. For template matching, an all-focus window size of 15, and aperture window size of 29 were used, with a zero-threshold of 3, aperture-threshold of 3 and an error threshold of %0.5. The mean value of all results for all 3 step-sizes were used to map the ground-truth result of [5] into our system and modified by hand. Resulting depth maps for two experiments are shown in Fig.1.4 and their error ratios in Table 1.1.

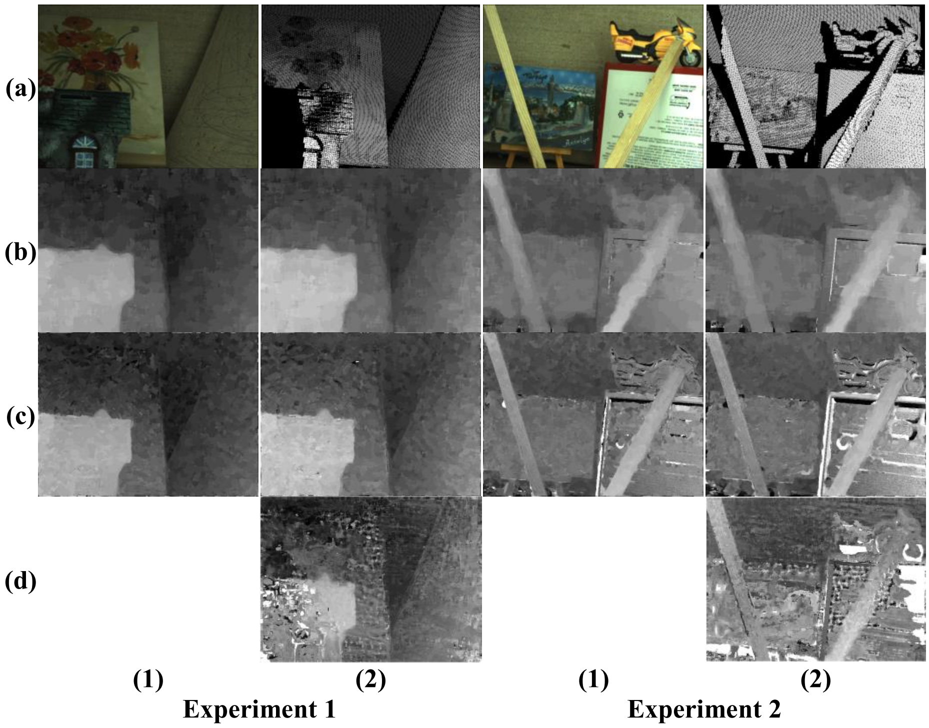


Fig. 1.5. a1 : All-focus image. a2 : Ground truth depth map. b : SML. c : AdaFocus. d : Template Matching. 1 : w/full aperture. 2 : w/moving apertures

Table 1.1. Average errors for experiments.

Method	Avg. Error Ratio (Exp1)	Avg. Error(Abs) (Exp1)	Avg. Error Ratio (Exp2)	Avg. Error(Abs) (Exp2)
SML with apertures(step=1)	3.51%	1,05	2.84%	0,85
SML with apertures(step=2)	3.93%	1,18	3.24%	0,97
SML with apertures(step=3)	4.28%	1,28	3.42%	1,03
AdaFocus with apertures(step=1)	3.64%	1,09	4.24%	1,27
AdaFocus with apertures(step=2)	4.08%	1,22	5.00%	1,5
AdaFocus with apertures(step=3)	4.48%	1,34	5.34%	1,6
SML without apertures(step=1)	3.78%	1,13	2.60%	0,78
SML without apertures(step=2)	3.59%	1,08	2.59%	0,78
SML without apertures(step=3)	3.49%	1,05	2.75%	0,82
AdaFocus without apertures(step=1)	3.68%	1,1	3.90%	1,17
AdaFocus without apertures(step=2)	3.54%	1,06	3.49%	1,05
AdaFocus without apertures(step=3)	3.47%	1,04	3.59%	1,08
Template Matching	6.73%	2,02	8.48%	2,54

1.5 Conclusions

In this paper, we have defined a new concept called moving apertures to acquire precise depth maps. The resulting depth maps and error rates prove that the concept of moving apertures can improve the results of any focus measure operator depending on the number of apertures used. Our method can work with methods that use frequency analysis or matching without any restrictions for the scene conditions.

However; improvements and changes, such as the selection and parameters of matching algorithm, are necessary for MFM. It is possible to improve the results of both AIFM and MFM by applying different frequency analysis methods, focus measure operators, or matching methods.

For future work, we plan to use our system for producing ground truth depth maps for SFF maps.

Acknowledgments **This work was supported by TUBITAK Career Project 105E097.**

1.6 References

1. S.K. Nayar, Y. Nakagawa: Shape from focus. Pattern Analysis and Machine Intelligence, 16(8):824–831, August 1994.
2. M. Subbarao, G. Surya: Depth from defocus: A spatial domain approach. International Journal of Computer Vision, 13(3):271–294, December 1994.
3. W.K. Pratt: Digital Image Processing. Wiley, New York, 1978.
4. T. Aydin, Y.S. Akgul: A New Adaptive Focus Measure for Shape From Focus. British Machine Vision Conference, Leeds, September 2008.
5. D. Scharstein, R. Szeliski: High-accuracy stereo depth maps using structured light. Computer Vision and Pattern Recognition, p.195-202, Wisconsin, June 2003.
6. Y.Y. Schechner, N. Kiryati: Depth from defocus vs. stereo: How different really are they? International Journal of Computer Vision, 39(2):141–162, 2000.
7. H.N. Nair, C.V. Stewart: Robust focus ranging. Computer Vision and Pattern Recognition, p.309–314, 1992.