Depth of General Scenes from Defocused Images Using Multilayer Feedforward Networks

Veysel Aslantas and Mehmet Tunckanat

Erciyes University, Faculty of Engineering, Department of Computer Eng., 38039
Melikgazi, Kayseri, Turkey
{Aslantas, Tunckanat}@erciyes.edu.tr

Abstract. One of the important tasks in computer vision is the computation of object depth from acquired images. This paper explains the use of neural networks to calculate the depth of general objects using only two images, one of them being a focused image and the other one a blurred image. Having computed the power spectra of each image, they are divided to obtain a result which is independent from the image content. The result is then used for training Multi-Layer Perceptron (MLP) neural network (NN) trained by the backpropagation algorithm to determine the distance of the object from the camera lens. Experimental results are presented to validate the proposed approach

1 Introduction

The depth of the visible surface of a scene is the distance between the surface and the sensor. Obtaining this information of a scene from its two-dimensional images can assist numerous applications such as object recognition, navigation and inspection.
Based on focus blur information, three different depth computation techniques have been developed: Depth from Automatic Focusing (DFAF) [1-3], Depth from Defocusing (DFD) [4-16] and Depth from Automatic Defocusing (DFAD) [17].

Previous work on depth from defocusing usually required knowledge of the relation among lens parameters, depth and blur circle radius (e.g., in [3, 5]). The main limitation of this is that those parameters are based on the ideal thin lens model and can never be measured precisely for any camera since real imaging systems have several lenses.

Most of the DFD algorithms found in the literature are also based on assuming that the PSF of the camera is either a Gaussian or a circularly symmetric function to obtain a relation between depth and the spread parameter of the PSF which is potentially the most erroneous part of many of the earlier methods.

NNs have found many applications in vision-related areas. Recently, they have also been employed to compute depth using focus blur [15, 16, 18, 19]. In the proposed method, a neural network [20] is employed to compute the depth of a scene. The technique requires two images of the same scene one of which is sharp the other is defocused. Power spectrum of each images are computed. The spectrums are then divided to acquire a result which is independent from the image content. The result is employed for training NN trained by the BP algorithm to determine the distance of

F.A. Savacı (Ed.): TAINN 2005, LNAI 3949, pp. 41 – 48, 2006.

the object from the camera lens. Therefore, the proposed method is independent of the PSF of the camera and does not require a relation between blur and depth.

The technique does not require special scene illumination and needs only a single camera. Therefore, there are no correspondence and occlusion problems as found in stereo vision and motion parallax techniques or intrusive emissions as with active depth computation techniques.

The remainder of the paper comprises five sections. Section 2 presents the theory underlying the proposed technique. Section 3 explains the experimental procedure for acquiring data to train and test MLPs in depth computation. Section 4 discusses the results obtained. Section 5 concludes the paper.

2 Problem Formulation

The transformation effected by an optical system can be modeled as a convolution operation [21]. The image of a blurred object may then be written as:

$$I(x,y) = \int\int H(x-\xi, y-\eta, d(\xi,\eta))S(\xi,\eta)\, d\xi d\eta df \qquad (1)$$

where x and y are image coordinates, ξ and η are two spatial variables, $S(x,y)$ and $I(x,y)$ are the sharp and defocused images of the source object respectively, $d(x,y)$ is the distance from the object to the plane of best focus (PBF) and $H(x,y,d)$ is the PSF. If the distance from the object to the PBF is constant, then the PSF $H(x,y,d)$ can be written as $H(x,y)$ and the defocusing process is defined as a convolution integral:

$$I(x,y) = \int\int H(x-\xi, y-\eta)S(\xi,\eta)\, d\xi d\eta \qquad (2)$$

The convolution operation is usually denoted by the symbol $\otimes$. Therefore, Equation (2) can be abbreviated as:

$$I(x,y) = H(x,y) \otimes S(x,y) \qquad (3)$$

In the Fourier domain, Equation (3) can be expressed as:

$$I(u,v) = H(u,v)S(u,v) \qquad (4)$$

where: $\{I(x,y), I(u,v)\}$, $\{H(x,y), H(u,v)\}$ and $\{S(x,y), S(u,v)\}$ are Fourier pairs. The function $H(u,v)$ is often referred to as the optical transfer function (OTF).Most of the focus based techniques assume that the distance function $d(x,y)$ is slowly varying, so that it is almost constant over local regions. The defocus is then modeled by the convolution integral over these regions.

2.1 Form of Point Spread Function

Figure 1 shows the basic geometry of image formation. Each point in a scene is projected onto a single point on the focal plane, causing a focused image to be formed on it. However, if the sensor plane does not coincide with the focal plane, the image formed on the sensor plane will be a circular disk known as a *"circle of confusion"* or *"blur circle"* with diameter 2R, provided that the aperture of the lens is also circular.

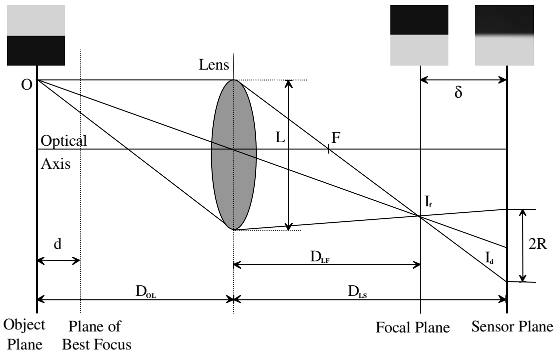

Fig. 1. Basic Image Formation Geometry

According to geometrical optics, the intensity distribution within the blur circle is assumed to be approximately uniform i.e., the PSF is a circular *"pillbox"*. In reality, however, diffraction effects and characteristics of the system play a major role in forming the intensity distribution within the blur circle. After examining the net distribution of several wavelengths and considering the effects of lens aberrations the net PSF is best described by a 2D Gaussian function [5, 6]:

$$H(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}} \tag{5}$$

where σ is the spread parameter which is proportional to the radius R of the blur circle.

$$R = k\sigma \tag{6}$$

The proportionality constant k depends on the system and can be determined through calibration. The optical transfer function $H(u,v)$ for geometrical optics can be obtained by taking the Fourier transform (FT) of $H(x,y)$:

$$H(u,v) = e^{-\frac{(u^2+v^2)\sigma^2}{2}} \tag{7}$$

By substituting Equation (7) into Equation (4) and solving for σ, an equation can be written as:

$$\sigma^2 = \frac{-1}{(u^2+v^2)} \ln\left(\frac{I(u,v)}{S(u,v)}\right) \tag{8}$$

where $I(u,v)$ and $S(u,v)$ are the powers of $I(x,y)$ and $S(x,y)$ respectively.

In principle, it is sufficient to calculate σ at a single point (u,v) by employing Equation (7). However, a more accurate value can be obtained by averaging σ over some domain in the frequency space:

$$\sigma^2 = \frac{-1}{A}\iint_P \frac{1}{(u^2+v^2)} \ln\left(\frac{I(u,v)}{S(u,v)}\right) dudv \tag{9}$$

where P is a region in the (u,v) space containing points where $\big(I(u,v)/S(u,v)\big) > 0$ and A is the area of P [3].

2.2 Relating Depth to Camera Parameters and Defocus

The object may be either in front of or behind the plane of best focus on which points are sharply focused on the focal plane. From Figure 1, by using similar triangles, a formula for a camera with a thin convex lens of focal length F can be derived to establish the relationship between the radius R of the blur circle and the distance D_{OL} from a point in a scene to the lens [5]:

$$D_{OL} = \frac{FD_{LS}}{D_{LS} - F - 2fk\sigma} \tag{10}$$

where D_{LS} is the distance between the lens and the sensor plane, f is the f-number of a given lens. When the object is in front of the PBF, Equation (10) becomes:

$$D_{OL} = \frac{FD_{LS}}{D_{LS} - F + 2fk\sigma} \tag{11}$$

Equations (10) and (11) relate the object distance D_{OL} to σ.

2.3 Window Effect

The convolution operation uses pixels that are out of the local region and this computation causes leaks. In order to reduce the effect of these image overlap problem [23], the image region must be multiplied by a center weighted function such as Gaussian mask:

$$w(x,y) = \frac{1}{2\pi s^2} e^{-\frac{x^2+y^2}{2s^2}} \tag{12}$$

where s is the spread parameter of the Gaussian and depends on the window size. s can have a value of the half of the window size [23]. The mask must be centered at the region of interest. The resulting image is then used for depth computation. However, erroneous depth result can be obtained if the relations given above are used.

3 Collection and Preparation of MLP Training and Test Data

An object can have any form and intensity distribution. Therefore, it is very difficult for a neural network to learn intensity data since a large amount of data corresponding to different object will be needed for each distance. If an image content independent data set is obtained for a particular distance, it is possible to teach a NN using this data. Hence, the following process was made for computing this kind of data set.

For the simulations in this paper, a 100x100 binary random-dot-pattern (RDP) image (Figure 2(a)) was blurred by a 2-D Gaussian function with $\sigma = 1$. The result of this process was a grey level image which was used as the original focused image (Figure 2(b)).

Fig. 2. Images used for the simulations (a) Binary Random Dot Pattern Image (b) Blurred image of (a) (=1)

In the simulations, the camera was set such that the best focused plane was at infinity (thus the object would always be between the best focused plane and the lens). The focal length and *f*-number of the lens used in the experiments were set to 40mm and 2.8 respectively. The distance of the object (Figure 2(b)) from the camera varied for different images but the camera parameters were the same for all images.

For each distance, the defocused images were generated by convolving the original image with a 2-D Gaussian function given in Equation (5). The spread parameter of the Gaussian function is computed from Equation (11). The camera constant (*k*) had a value of 0.71. The value is a good approximation in most practical cases [9]. To reduce image overlapping effects, the images are multiplied with a Gaussian mask using Equation (12).

Image regions were selected randomly from the same location on sharp and defocused images. The size of the regions was 16x16 and 32x32 pixels for different experiments. The power spectrum of the each image region was calculated. The spectrums are then divided to acquire a result which is independent from the image content. Therefore, a set of data was obtained for a certain distance. Because the result is symmetric along *u* and *v* axes, only first quarter of the data was used in the experiments. The data size was reduced further by taking the square of each coefficient and summing them column wise. Hence, the size of data for the NN was reduced from 16x16 and 32x32 to 1x8 and 1x16 respectively. The calculation was made for each distance ranging from 600mm to 1400mm at intervals of 10mm. The data set computed for each depth was used as input to train the NN. In the experiments, all neural networks used had one output (for depth). The test data set was obtained in the same way from the images given in Figure 3.

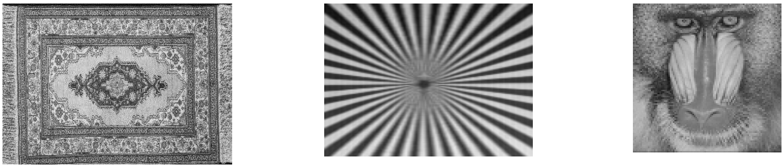

Fig. 3. Test Images (a) Rug (b) Test Pattern (c) Baboon

4 Experiments

For each distance, ten image regions were selected randomly from the sharp and blurred RDP image shown in Figure 2(b). 10 dB zero-mean Gaussian noise was added to the five of them. The computations described in the previous section were applied on each image region. The data were entered in rows in a text file to prepare the training set. The distance values corresponding to each row were added to the file to act as desired outputs. Several MLP structures were used to search for the best result. In the investigations reported here, the activation functions of the neural networks were tangent (T), linear (P) and logarithmic (L) functions. All the NNs used in this work are trained by the Levenberg-Marquardt (LM) algorithm [22]. After the training process had completed, each neural network was tested with the training and test data sets. The NN structures used for each experiment and their results for the noise free images can be seen in Table 1.

Table 1. Calculation and NN Parameters for each experiment

Exp. No.	Input Neurons	Activation Functions	Hidden Neurons	%Error			
				Rug	Test Pattern	RDP	Baboon
1	16	T-T-P	5–13	0.4702	0.2915	0.2760	0.3054
2	16	T-T-P	20–16	0.1164	0.2266	0.1394	0.0574
3	16	L-T-T-P	16–12–14	0.3187	0.2394	0.1657	0.2056
4	8	T-T-P	16–12	0.4285	0.9963	0.1008	0.6718
5	8	T-L-T-P	8–4–2	0.5215	0.2964	0.4245	1.9330
6	8	L-T-R-P	13–15–17	0.5727	0.4076	0.1141	0.4597
7	16	T-T-T-P	5–13–4	0.0459	0.2008	0.0103	0.0477
8	8	L-L-L-P	18–12–10	0.1957	0.3159	0.1177	0.2501
9	16	T-T-T-L-P	5–13–4–6	0.3589	0.0937	0.2410	0.2005
10	8	L-L-T-P	19–11–13	0.5180	0.2239	0.1368	0.0752

The results of 10 dB noisy images for the seventh and ninth experiments are given in Table 2. It can be noted that the neural networks produced low percentage errors corresponding to high depth computation accuracies. The results for noisy test data sets produced from the Baboon image are plotted in Figure 4. It can be noted that the error in depth computation was about 0.6%.

Table 2. The results of 10 dB noisy images

Exp No.	Rug	Test Pattern	RDP	Baboon
7	0.4678	0.5407	0.3594	0.2310
9	0.4616	0.6480	0.4402	0.3127

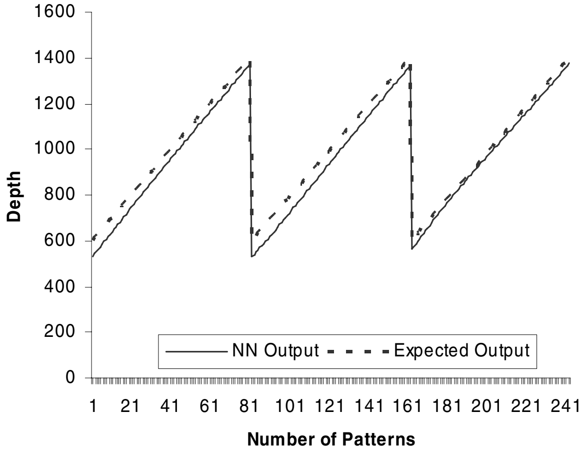

Fig. 4. The result of the noisy Baboon image for experiment 7

5 Discussion

In this paper, a simple method for computing the depth information of general objects from a sharp and a blurred image of them using a neural network. The experiments results given show that it is possible to compute depth of objects with reasonable accuracy. Previous work on DFD usually required the relation among lens parameters, the depth and blur circle radius. The main limitation of this is that those parameters are based on the ideal thin lens model and can never be measured precisely for any camera since real imaging systems have several lenses. Most of the DFD algorithms found in the literature are also based on assuming that the PSF of the camera is either a Gaussian or a circularly symmetric function to obtain a relation between depth and the spread parameter of the PSF which is potentially the most erroneous part of many of the earlier methods. The proposed method is independent of the PSF of the camera and does not require a relation between blur and depth. The method was implemented to compute the depth of general scenes using their sharp and defocused images. Experimental results have shown that the error was approximately 0.6%.

References

1. Krotkov E.P.: Focusing, International Journal of Computer Vision, 1(3),(1987),.223-237.
2. Nair H.N., and Stewart C.V.: Robust Focus Ranging, IEEE Conference on Computer Vision and Pattern Recognition, Champaign, Illinois, (1992),309-314
3. Subbarao M., and Choi T.: Accurate Recovery of Three Dimensional Shape from Focus, IEEE Transaction on Pattern Analysis and Machine Intelligence" 17(3), (1995), 266-274
4. Grossmann P.: Depth from Focus, Pattern Recognition Letters, 5(1), (1987), 63-69
5. Pentland A.P.: A New Sense for Depth of Field, IEEE Transaction on Pattern Analysis and Machine Intelligence 9(4), (1987), 523-531

6. Subbarao M., and Gurumoorthy N.: Depth Recovery from Blurred Edges, IEEE Conference on Computer Vision and Pattern Recognition, Ann Arbor, Michigan, (1988),498-503
7. Holeva L.F.: Range Estimation from Camera Blur by Regularised Adaptive Identification, International Journal on Pattern Recognition and Artificial Intelligence 8(6), (1994), 1273-1300
8. Pentland A.P., Scherock S., Darrell T. and Girod B.: Simple Range Cameras based on Focal Error, Journal of Optical Society of America, 11(11),(1994),2925-2934
9. Subbarao M., and Surya G.: Depth from Defocus: A Spatial Domain Approach, International Journal on Computer Vision, 13(3), (1994), 271-294
10. Xu S., Capson D.W. and Caelli T.M.: Range Measurement from Defocus Gradient, Machine Vision Applications, 8, (1995), 179-186
11. Asada N., Fujiwara H. and Matsuyama T.: Edge and depth from focus, International Journal on Computer Vision, 26(2), (1998), 153-163
12. Watanabe M. and Nayar S.K.: Rational filters for passive depth from defocus, International Journal on Computer Vision, 27(3), (1998), 203-225
13. Asada N., Fujiwara H. and Matsuyama T.: Particle depth measurement based on depth-from-defocus, Optics & Laser Technology, 31(1),(1999),95-102
14. Chaudhuri S. and Rajagopalan A.N.: Depth from defocus: A real aperture imaging approach, Springer-Verlag New York, Inc. (1999)
15. Pham D.T. and Aslantas V.: Depth from defocusing using a neural network, Pattern Recognition, 32(5), (1999), 715-727
16. Asif M. and Choi T.S.: Shape from focus using multilayer feedforward neural networks, IEEE Transaction on Image Processing, 10(11), (2001), pp.1670-1675
17. Aslantas V.: Depth from automatic defocusing, sent for publication
18. Aslantas V.,Tunckanat M.: Depth From Image Sharpness Using a Neural Network, ICSP, Canakkale, Turkey, (2003) 260-265
19. Aslantas, V.: Estimation of Depth From Defocusing Using A Neural Network. ICIS International Conference on Signal Processing. Çanakkale, Turkey. (2003) 305-309
20. Rumelhart D.E. and McClelland J. L.: Parallel Distributed Processing: Explorations in the Microstructure of Cognition, MIT Press, Cambridge, MA., (1986)
21. Horn, B.K.P.: Robot Vision. McGraw-Hill, New York (1986)
22. Gill P. E., Murray W., and Wright H.: Practical Optimization, New York: Academic Pres., 1981
23. Wei C., Three Dimensional Machine Vision Using Image Defocus, Ph.D. Dissertation, Electrical Eng., State University of New York at Stony Brook, 1994