Using Radial Basis Function Networks to Approach the Depth from Defocus

Shyh-Ming Jong and Jeng-Sheng Huang

Department of Mechanical Engineering, Chung Yuan Christian University, Chung Li, Taiwan, R.O.C.

In range finding, the depth from defocus (DFD) is a simple and effective method. The DFD yields the absolute depth, and does not have the image-to-image matching and occlusion problems. Therefore, we use the DFD method to analyze the defocused images to obtain depth information using Gaussian blurred function. In order to find the range of objects, a sigma value of the Gaussian function due to edges out of focus is necessary. Because the sigma value of the Gaussian function depicts on the intensity of images grabbed by imaging devices, we employ an approximate method, the radial basis function networks (RBFN), to approach the sigma value directly in the spatial domain. The RBFN regularizes the center position and the sigma value of the Gaussian function to fit the profile of the defocused image by three layers of neural networks based on the radial basis function. It has accurate ranging results with less than 8% of the root mean square error in sigma value approaching and 5% of the relative error in ranging, imaging system ranges from 220 mm to 355 mm and focuses at 400 mm.

Journal of Imaging Science and Technology 45: 400–406 (2001)

Introduction

In biological systems, the images that fall on the retina are quite badly focused everywhere except within the central fovea.[1] There is a gradient of focus, ranging from nearly perfect focal at the point of regard to almost complete blurred at points on distance objects. It is called the depth of field.[2] Different lenses have variant magnification of depth of field in optics. We regard it in focus when an image is located within the depth of field as well as defocus when image is out of the depth of field. The magnitude of defocus is formed by an object's position and the optical parameters (e.g., focal length, aperture, etc.). It can be measured by imaging devices, e.g., CCD camera. When images are taken by imaging system, the profile of the defocus is visible in gray level and approached by mathematic analysis. The distance of the objects can be ranged if the camera parameters are known and defocus model is set in a certain distance. The depth from defocus (DFD)[3–5] method in range finding field is well known. The range finding methods can be divided into two categories: active and passive. There are several passive imaging methods, such as stereopsis,[6] structure from motion,[7] shape from shading,[8,9] depth from focus (DFF)[10] and depth from defocus (DFD).[11–14] Among these passive methods, the DFD yields the absolute depth that can be recovered with only two images. Besides, DFD has not the image-to-image matching and occlusion problems. Therefore, we use DFD method to analyze the defocused images to obtain depth information using the Gaussian spread function, which is a passive range-finding method. In order to find the range of objects, a sigma value of the Gaussian function due to edges out of focus is necessary. The sigma value definitely affects the accuracy of the distance between object and image, which is estimated in spatial or frequency domain. Most approach methods in spatial domain are focused on the polynomial estimation[4] that has the drawbacks, such as polynomial terms must be compatible with the defocus algorithm governed by camera model, existence of the transfer and accuracy of the estimation. Because the sigma value of the Gaussian function depicts on the intensities of image grabbed by imaging system, it needs an approximate algorithm to find the sigma value. Multi-layer feed-forward neural networks are regarded as the universal approximators[15] and have the capability to approach nonlinear input–output relationships of a continuous and multivariate function. We employ a direct approximate algorithm, the radial basis function networks (RBFN),[16–18] to evaluate the sigma value in the spatial domain of the defocused image. The radial basis function networks comprise three layers, input, output and hidden layer respectively. When the networks fit the target, the pixel numbers grows toward the intensities of the image respectively, the weights vector of the hidden layer show the center position of the Gaussian function as well as the weights vectors of the output layer show the sigma value of the Gaussian function. The RBFN is a direct method to measure the depth of the defocus images by approaching the sigma value from the Gaussian blurred edges.

Analysis Method

In order to find the depth of the objects and the camera, we have to know the camera parameters as well as the blurred parameters. The camera parameters are based on the geometric camera model and blurred parameters based on point-spread function (PSF). Then we can get back the depth between objects and camera by radial basis function networks (RBFN) and the camera model.

Original manuscript received June 5, 2000

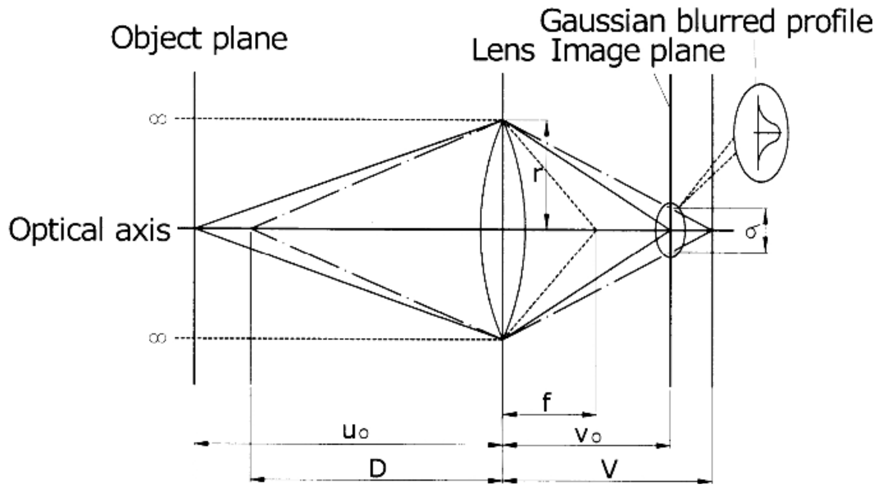


Figure 1. Geometry of imaging and Gaussian blurred profile.

Geometric Camera Model

The amount of defocus or blurring depends solely on the distance to the surface of exact focus and the characteristics of the lens system. As the distance between the imaged point and the surface of the exact focus increases, the image objects become progressively more defocused. If we could measure the amount of blurring at a given point in the image, therefore it seems possible that we could use the parameters of the lens system to compute the distance to the corresponding point in the scene.

Based on the similar triangles, the distance D (see Fig. 1) to an imaged point is related to the parameters of the lens system and the amount of defocus by the following equations[4] .

$$D = \frac{fv_0}{v_0 - f - \sigma F} \quad \text{if } D > u_0 \tag{1.a}$$

and

$$D = \frac{fv_0}{v_0 - f + \sigma F} \quad \text{if } D < u_0 \tag{1.b}$$

where D is the distance from the lens to the object.

v is the distance from the lens to the defocused image.

f is the focal length.

F is called F-number and is equal to f / 2r, r is the radius of the aperture.

u_0 is the distance between the lens and the locus of perfect focus.

v_0 is the distance between the image plane and the lens in perfect focus.

The lens law can be expressed as

$$\frac{1}{u_0} + \frac{1}{v_0} = \frac{1}{f} \tag{2}$$

when u_0 is infinite, then v_0 is equal to f.
σ is the blur parameter in length units (mm) can be written as

$$\sigma = \sigma_p \times k \tag{3}$$

where σ_p is a blur parameter in pixel units, and k is the camera parameter.

The camera parameter k is positive and is a constant of proportional characteristic of a given camera. When σ is very small, in which case diffraction effects dominate, let $k = 1/\sqrt{2}$ is a good approximation in most practical cases.[19] The camera parameter k has to be computed by camera calibration first.[20]

From the above description, the distance of object D depends on parameters: f, v, σ, and r. When an object is located at a fixed position, the blur parameter σ is varying with f, v and r. Most of the researchers have focused on comparing two images by adjusting the lens position[12] v_0 or aperture[14] r to measure the depth of object. When the robot moves and the camera parameters are free of adjustment, the blur parameter is only varying by moving in the distance D. It means that the depth recovery is based on the blurred parameter of the point spread functions, such as uniform, line and Gaussian, Uniform spread function exists when no light energy is absorbed by the camera system.[14] Line spread function depends on a perfect lens for avoiding the diffraction effects. We are interested in the most common case and then PSF can be approximated by the Gaussian function will be very close to the real lens.

Depth Recovery

There are two categories for depth recovery by depth from defocus method: spatial domain and spatial frequency domain. The popular frequency domain for depth recovery is the Fourier transform[27] and the well-known S-transform.[14] The polynomial approach[14,25] and inverse filter[4,13,20] are famous depth recovery in spatial domain. It is difficult for the frequency domain to deconvolute the defocus operator from the scene and to model it. The inverse filter method is inaccurate and constrained in both windowing and border effects. The polynomial approach in the spatial domain by the maximum likelihood method[4] occurs errors when the Laplacian of the convolution of raw image intensities and Gaussian point spread funtion is zero. Taking natural log of zero will make the program halted. We will present the radial basis function netwroks to find the sigma σ directly in the spatial domain in the following section.

Radial Basis Function Networks (RBFN)

We propose a neural network for finding the sigma value of the Gaussian point spread function in robot vision. Neural networks are divided into supervised and unsupervised learning network. The former with target values, the latter without any target. In this article, the image intensities of each pixels are the target of networks, so we need a supervised learning networks. The back-propagation networks (BPN) is the popular model in supervised learning networks, but it is slow in learning rate and is not the best approximation for continuous function.[25] In contrary, the RBFN provides the optimal approximate ability in supervised networks. The network architecture and radial basis functions is described as follow:

The RBFN's Architecture

A radial basis function networks may be depicted as shown in Fig. 2. The architecture consists of three layers, an input layer, a hidden layer and an output layer.

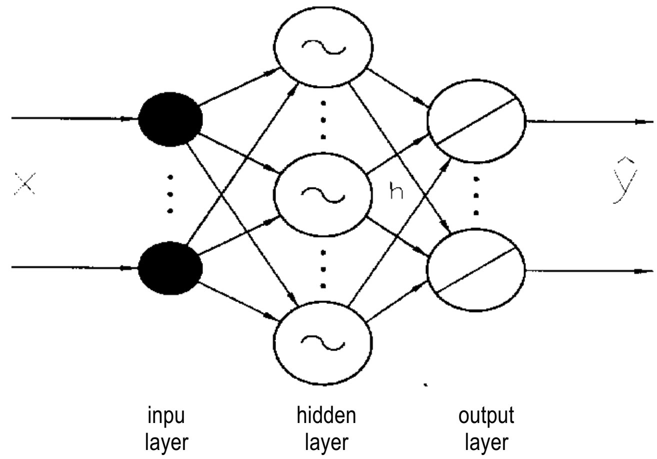


Figure 2. A multi-input/multi-output of the radial basis function networks.

(i) The Input Layer: The layer consists of I elements and forwards the input signals vector $x = [x_1, \ldots, x_I]$ to each neuron of the hidden layer. In this article, the input vector x are the pixel numbers of the blurred edge image (see Figs. 6a,c,e,g).

(ii) The Hidden Layer: The layer is composed by J elements with radial basis function and calculates the ouput h_j , $j \in \{1, \ldots, J\}$ from the input signals x. Feeding the input vector x onto each neuron as center c_j, and its distance between x and c is called the Euclidean norm, $a_j = ||x - c_j||$. The output of the hidden layer is:

$$h_j(x; c_j) = \phi(||x - c_j||)$$

where ϕ is a continuous differential function, the Euclidean norm can be modified by shape factor s (see Fig. 3).

$$h_j(x; c_j) = \phi(v_j = \frac{||x - c_j||}{s_j})$$

The unit of transferring function based on argument v is called the radial basis function unit. Because each neuron in the hidden layer owns different center c and shape factor s, the output h variant even with the same input x.

(iii) The Output Layer: This layer consists of K elements, each output is comprised of weights w_{kj} and the output h_j of the hidden layer in linear superposition as follow:

$$\hat{y}_k(x; P) = \sum_{j=1}^{J} w_{kj} h_j(x; c_j, s_j) \tag{4}$$

where P is all the training parameters, such as centers, shape factors and weights.

The output of the RBFN can be formed in matrix:

$$\begin{aligned}
\hat{y}(x;P) &= \begin{bmatrix} \hat{y}_i(x;P) \\ \cdot \\ \cdot \\ \hat{y}_k(x;P) \end{bmatrix} \\
&= \begin{bmatrix} \sum_{j=1}^{J} w_{1j} h_j(x; c_j, s_j) \\ \cdot \\ \cdot \\ \sum_{j=1}^{J} w_{kj} h_j(x; c_j, s_j) \end{bmatrix} \\
&= \begin{bmatrix} w_{11} & \cdots & w_{1J} \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ w_{K1} & \cdots & w_{KJ} \end{bmatrix} \begin{bmatrix} h_1(x; c_1, s_1) \\ \cdot \\ \cdot \\ h_J(x; c_J, s_J) \end{bmatrix} \\
&\equiv Wh(x; C, S)
\end{aligned} \tag{5}$$

where

$$C = [c_1, \ldots, c_J]'$$
$$S = [s_1, \ldots, s_J]'$$
$$W = \begin{bmatrix} w_{11} \cdot & \cdot w_{1J} \\ \cdot \quad \cdot & \cdot \quad \cdot \\ \cdot \quad \cdot & \cdot \quad \cdot \\ w_{K1} \cdot & \cdot w_{KJ} \end{bmatrix}$$

and P ≡ [C S W'].

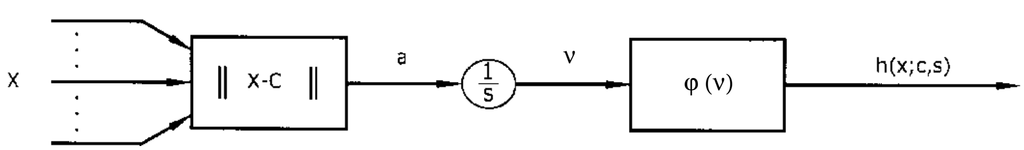


Figure 3. The radial basis function unit.

Hence, a RBFN can be written as the following mathematic forms strictly:

If $\{f = (f_1,\ldots, f_k);\ R^I \rightarrow R^K;\ f_k(x) = \sum_{j=1}^{J} w_{kj}\phi(\nu(x;c_j,s_j))$,

$1 \le k \le K;\ x,\ cR^I,\ s_j,\ w_{kj}R\}$, then f is called the RBFN and w_{kj} is the weight.

The Radial Basis Functions

The hidden layer consists of a set of radial basis functions. The hidden layer neuron calculates the Euclidean distance between the center c_j and the network input vector and then passes the result to a radial basis function. All the radial basis functions in the hidden layer neurons are usually of the same type. Typical choices of the radial basis functions are:

(i) The thin-plate-spline function (TSF):

$$\phi(\nu) = \nu^2 \times \log(\nu),$$

(ii) The Gaussian potential function (GPF):

$$\phi(\nu) = e^{-\nu^2},$$

(iii) The multi-quadric function (MQF):

$$\phi(\nu) = (\nu^2 + 1)^{1/2},$$

(iv) The inverse multi-quadric function:

$$\phi(\nu) = (\nu^2 + 1)^{-1/2},$$

where ν is a non-negative number and is the distance from the input vector x to the radial basis function center c. In radial basis function networks, the TSF[29,30] and MQF [16,17] are divergent functions when $\nu \rightarrow \infty$ as well as GPF[16,17] and inverse MQF are convergent functions when $\nu \rightarrow \infty$. The GPF, convergent function, is close to the Gaussian point spread function. Therefore, we use the Gaussian radial basis function network to approach the Gaussian blurred parameter as sigma value in defocus image by adjusting the centers and weights. When the output fits the targets which are the intensities of the blurred edge's profile, the weight of the vectors w_{kj} in output layer is proportional to sigma σ. Having estimated σ, Eqs. 1a and 1b can now be used to calculate the distance D to the image point. There are two possible solutions, one corresponding to a point in front of the locus of exact focus, the other corresponding to a point behind it. This ambiguity is generally unimportant because we focus at a fixed distance and the camera moves closely to the objects.

Results and Discussion

Using the natural light source to get the images that could be processed as ranging information is the one of the characteristics of the passive depth ranging methods.

Simulation with Standard Function

In order to check the exactness and accuracy of the algorithms described in the previous section, we used a step function to simulate the intensity of sharp edges and convoluted it with Gaussian function from σ = 1 to 20. At the beginning of the training procedure, we create a radial basis network with given inputs as number of pixels, and targets as gray level of pixels. Initially the hidden layer of the RBFN has no neurons. We set the error goal before network training, and the error is the difference between the output and the target in each epoch of training step. The following steps are repeated until the network's mean square error falls below goal.

(i) The network is simulated.
(ii) The input vector with the greatest error is found.
(iii) A radial basis function neuron is added with weights equal to that vector.
(iv) The output layer weights are redesigned to minimize error.

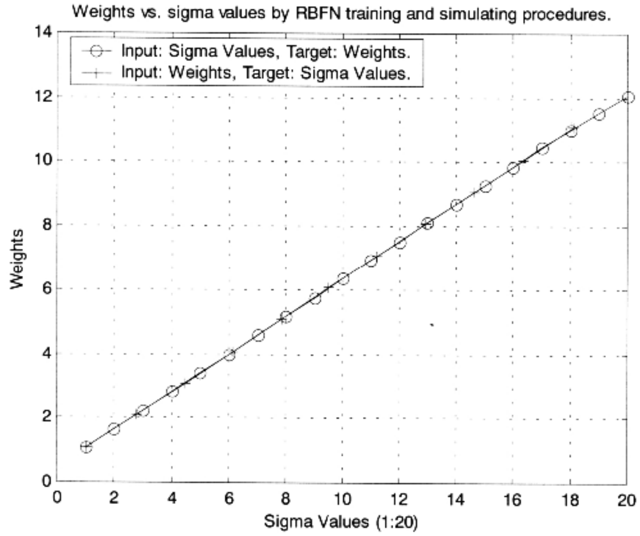


Figure 4. The relationship between weights and sigma values in radial basis function networks.

After the training procedure, we obtain the weights. Then, we simulate the RBFN with the weights by reversing the inputs and targets. We make the forwarding and reversing processes with different scale for avoiding data covered. From Fig. 4, it is obvious that the sigma value is proportional to the output layer weights. It shows that our algorithms are exact for realizing ideal images (noiseless and high quality images).

Depth Ranging Architecture

The implemental equipments used for this research are described as follow:

- Image grabber card – Imagenation Co., CX100 frame grabber card with 512 × 486 resolution at 8 bits.
- Lens – Canon Co, 1/2 inch sensor, aperture F1.6–16, focal length f = 16 mm.
- Camera – Ikegami Co., ICD–47, 1/2 inch B/W CCD with 768(H) × 494(V), S/N ratio 48db, 0.02lux/F1.6 (AGC on).
- IBM PC—Intel Pentium 200Mhz PC for data acquisition control and processing.

We set the aperture diameter to maximum (10mm) for grabbing useful images under natural lighting. From the sensor size of the CCD camera and the resolution of the grabber card, we get the camera parameter k equals to 0.0125. When we want to simulate the moves of the

TABLE I. The Calculated Distances of Objects and Their Relative Errors

Distance D (mm)	355	310	265	220
Sigma σ(pixels)	1.3709	5.5856	9.4653	14.8327
Calculated D (mm)	369.5937	299.5953	255.1150	211.6445
Root Mean Square error (%)	7.3500	2.0036	1.6300	1.1806
Relative error (%)	4.1137	3.3563	3.7302	3.7979

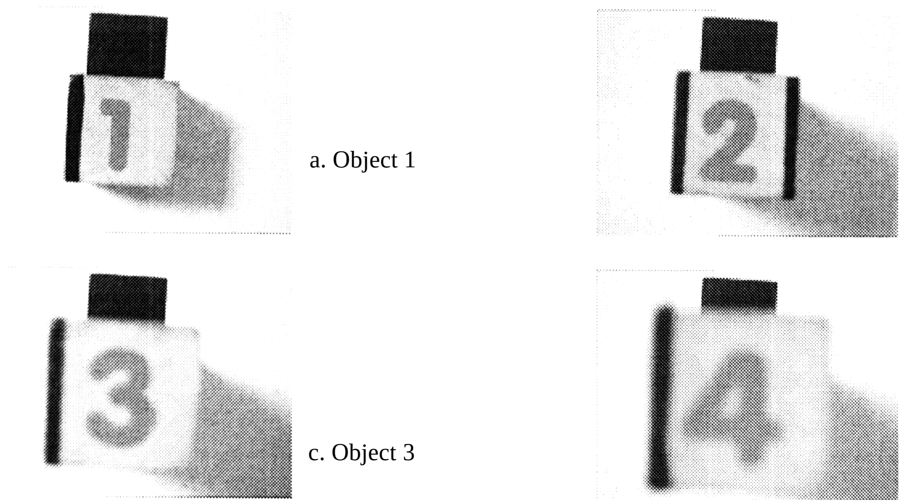

a. Object 1
b. Object 2
c. Object 3
d. Object 4

Figure 5. Pictures of objects grabbed by ranging implementation.

robot towards close objects, we place the cubic blocks, which each side is 45 mm, in front of the CCD camera. The lens is focused at 400 mm and stacks the cubic blocks one by one for simulating the robot moves from 335 mm to 220 mm with 4 steps close to objects (see Figs. 5a–d). Figures 6a, 6c, 6e, and 6g show the 3D profile on the blurred edges of the objects in front of camera at different positions: 355 mm, 310 mm, 265 mm and 220 mm respectively. Figures 6b, 6d, 6f, and 6h present the generalized mean profile of the blurred edges approached by RBFN. The experimental results are listed in Table I. From Table I, the RBFN has accurate ranging results with less than 8% of the root mean square error in sigma value approaching and 5% of the relative error in ranging, imaging system ranges from 220 mm to 355 mm and focuses at 400 mm.

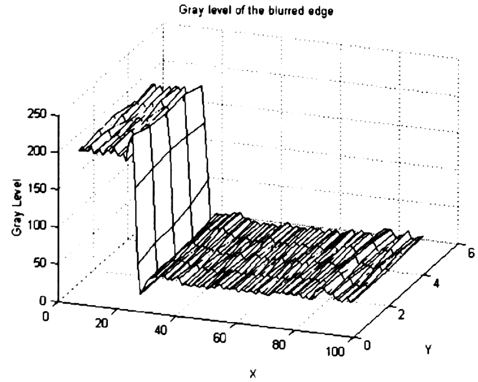
Discussion

Assuming that the point spread function of the defocused images is a Gaussian function, we have shown that the depth of the scene can be measured only from one image obtained from natural lighting and used the RBFN to approach the sigma value. Our implement consists of three steps, the network training procedure, the simulating step to approach the sigma value and the depth range step. In the first step, the larger shape factor is the smoother the function approximation will be. The larger shape factor requires a lot of neurons to fit a fast changing function. On the contrary, a smaller shape factor requires many neurons to fit a smooth function, and the network may not generalize well. We find the best shape factor in this implement is equal to one, as the Gaussian function is smooth except the object 1. For gaining the best results in approaching the sigma value, we set an error goal equal to zero during the training procedure of the ideal Gaussian point spread function. We use the Gaussian function to approach the Gaussian blurred edge of objects. Therefore, no transfer is needed like some DFD methods to deal the spatial domain and polynomial approximation. In training procedure, the RBFN spends more time for computing, but our case takes only one calculating period in simulation. The RBFN has ranging results with less than 8% of the root mean square error in sigma value approaching and 5% of the relative error in depth ranging. Therefore the RBFN is so accurate in ranging and fast in computing.

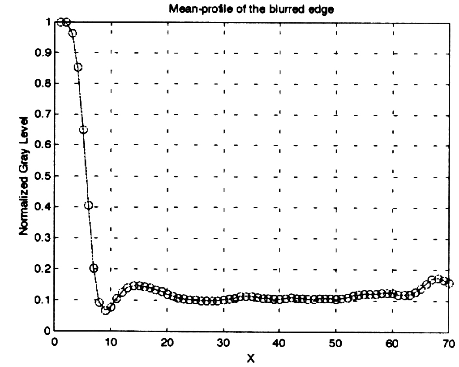
Conclusion

We employ radial basis function networks to evaluate the sigma σ; it has relevant ranging accuracy as well as fast computing. Because we do not need a lot of images like the depth from focus method for ranging, the algorithm evaluated in this article is an accurate and a fast method for depth ranging. Focusing is the most important part of the imaging system in computer vision, because it affects the shape and size of the objects in the image. The algorithm is also useful as the pre-processing of an imaging system for a pattern recognition and classification. With this accurate depth ranging system, we can develop an auto-focusing system. This is now required for any imaging system to allow us to get sharp focused images quickly and efficiently.

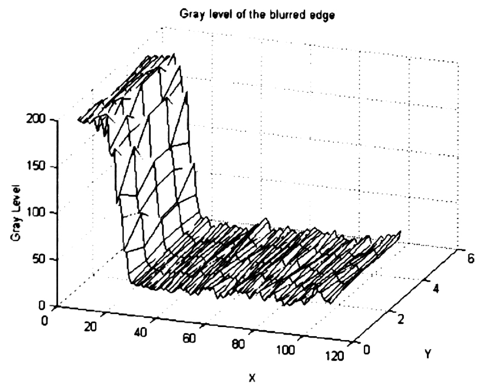
Acknowledgment. Support of this work by the National Science Council of the Republic of China contract NSC-89-2212-E-033-022 is gratefully acknowledged.



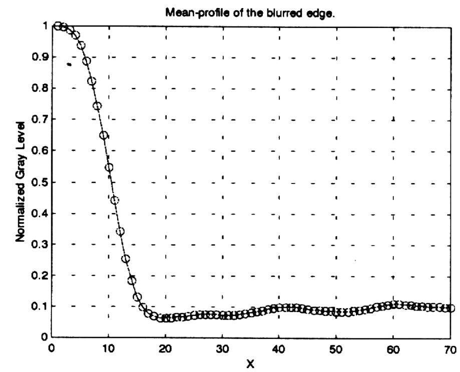
a. Object1 located at D=355mm.



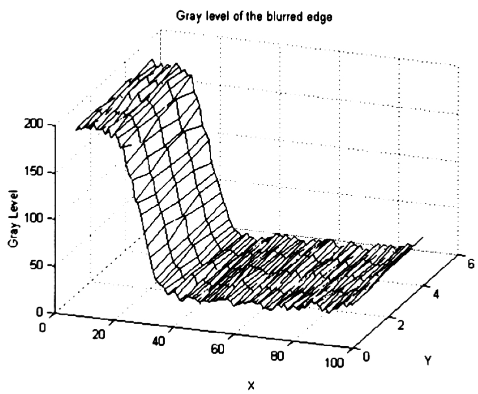
b. Object1's profile approached by RBFN.



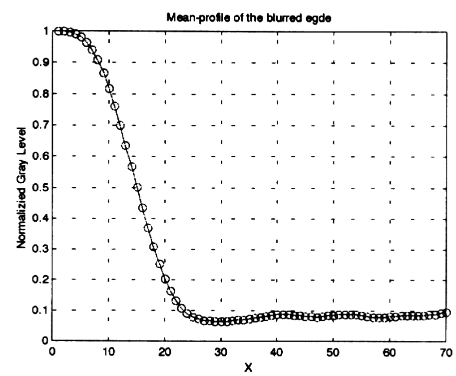
c. Object2 located at D=310mm.



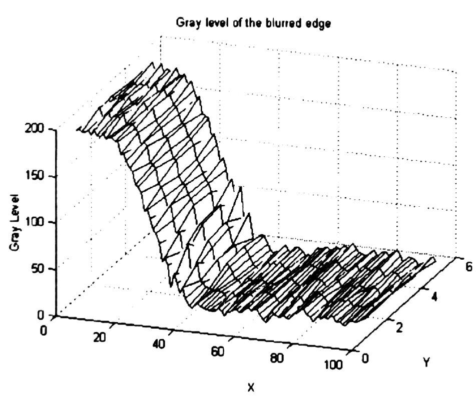
d. Object2's profile approached by RBFN.



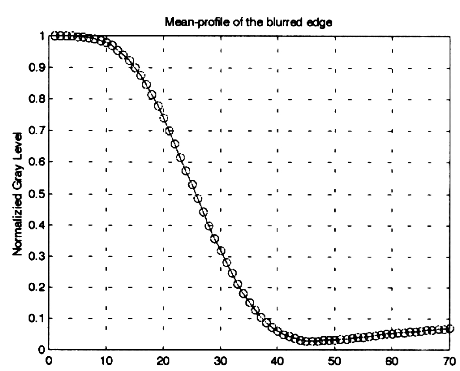
e. Object3 located at D=265mm.



f. Object3's profile approached by RBFN



g. Object4 located at D=220mm.



h. Object4's profile approached by RBFN.

Figure 6. The 3D profile of objects' blurred edge and the Gaussian function approached by RBFN.

References

1. M. Born and E. Wolf, *Principles of Optics,* Pergamon Press*,* London (1965).
2. A. P. Pentland, Depth of Scene from Depth of Field, unpublished paper presented at Image Understanding Workshop, Palo Alto, CA, 1982.
3. P. Grossman, Depth from focus, *Pattern Recognition Letters,* **5,** 63–69 *(*1987).
4. A. P. Pentland, A New Sense for Depth of Field, *IEEE Transactions on Pattern Analysis and Machine Intelligence. (PAMI)*, **9,** 523–531 *(*1987).
5. M. Subbarao, Parallel Depth Recovery by Changing Camera Parameters, in *Proc. IEEE, ICCV*, 149–155 (1988).
6. B. K. P. Horn, *Robot vision*, MIT Press, Cambridge, Massachusetts, 1986.
7. R. M. Haralick, Digital Step Edges from Zero Crossing of Second Directional Derivatives, *IEEE. Trans. PAMI*, 58–68 (1984).
8. B. K. P. Horn, *Obtaining Shape from Shading Information*, McGraw-Hill, New York, 1975*.*
9. M. Bichsel and A. P. Pentland, A simple algorithm for shape from shading, *Proc. IEEE CVPR*, 459 –465 (1992).
10. E. Krotkov, Focusing, *International J. of Computer Vision*, **1**, 223–237 (1987).
11. A. P Pentland, T. Darrel, M. Turk and W. Huang, A Simple, Real Time Range Camera, *Proc. IEEE, CVPR*, 256–261. (1989).
12. T. l. Hwang, J. J. Clark and A. L. Yuille, A Depth Recovery Algorithm Using Defocus Information, *Proc. IEEE, CVPR*, 476–482 (1989).
13. J. Ens and P. Lawrence, An Investigation of Methods for Determining Depth from Focus, *IEEE Transactions on Pattern Analysis and Machine Intelligence,* **15**, 97–108 (1993).
14. G. Surya and M. Subbarao, Depth from defocus by changing camera aperture: a spatial domain approach, in *Proc. IEEE, CVPR*, 61–67 (1993).
15. D. E. Rumelhart, G. E. Hinton and R. J. Williams, *Parallel Distributed Processing*, MIT Press, Cambridge, Massachusetts, 1986.
16. F. Girosi and T. Poggio, Networks and the Best Approximation Property, *Biological Cybernetics*, **63,** 169–176 (1990).
17. D. S. Broomhead and D. Lowe, Multivariable Functional Interpolation and Adaptive Networks, *Complex Systems*, **21**, 321–355 (1988).
18. T. Poggio and F. Girosi, Network for Approximation and Learning, *Proc. IEEE*, **78**, 1481–1498 (1990).
19. M. Subbarao and N. Gurumoorthy, Depth recovery from blurred edges, *Proc. IEEE, CVPR,* 498–503 (1998).
20. Y. Xiong and S. A. Shafer, Depth from Focusing and Defocusing, *Proc. IEEE, CVPR*, 68–73 (1993).
21. A. Rosenfeld and A. C. Kak, *Digital Picture Processing*, 2nd ed,Academic Press, London, 1982.
22. R. A. Jarvis, A Perspective on Range Finding Techniques for Computer Vision, *IEEE. Trans. PAMI*, 122–139 (1983).
23. T. Y. Young and K. S. Fu, Eds. *Handbook of Pattern Recognition and Image Processing,* Academic Press, London, 1986.
24. R. C. Gonzales and P. Wintz, *Digital Image Processing,* 2nd ed,Addison-Wesley, New York, 1987.
25. H. C. Andrews and B. R. Hunt, *Digital Image Restoration*, Prentice-Hall: Englewood Cliffs, NJ, 1987.
26. J. L. C. Sanz, *Advances in Machine Vision,* Springer-Verlag, New York, 1989.
27. M. Subbarao and T. C. Wei, Depth from Defocus and Rapid Autofocusing: a Practical Approach, *Proc. IEEE, CVPR*, 773–776 (1992).
28. S. Lai and C. Fu, A Generalized Depth Estimation Algorithm with a Single Image, *IEEE. Trans. PAMI*, **14,** 405–411 (1992).
29. S. Chen, S. A. Billings, C. F. N. Cowan and P. W. Grant, Practical Identification of Narmax Models Using Radial Basis Functions, *Int. J. Control*, **55**, 1327–1350 (1990).
30. S. Chen, S. A. Billings and P. W. Grant, Recursive Hybrid Algorithm for Non-linear System Identification Using Radial Basis Function Network, *Int. J. Control*, **55**, 1051–1070 (1992).