Shape From Focus Using Multilayer Feedforward Neural Networks

Muhammad Asif and Tae-Sun Choi, *Senior Member, IEEE*

***Abstract*—The conventional shape-from-focus (SFF) methods have inaccuracies because of piecewise constant approximation of the focused image surface (FIS). We propose a scheme for SFF based on representation of three-dimensional (3-D) FIS in terms of neural network weights. The neural networks are trained to learn the shape of the FIS that maximizes the focus measure.**

***Index Terms*—Focused image surface (FIS), neural networks, optimization, shape from focus (SFF), 3-D shape.**

I. Introduction

The conventional shape-from-focus (SFF) methods use a sequence of images taken by a single camera at different focus levels to compute depth of the objects. Some measure of quality of image focus, e.g., Laplacian or gray-level variance, is used to find the best focused image frame for a particular point in the image space. The camera parameter settings for that image frame are used to compute the distance of corresponding object point. Nayar and Nakagawa suggested using Gaussian interpolation to compute more accurate depth estimate [1].

The conventional SFF methods are based on the assumption that the object shape is piecewise constant in a small window around each pixel position. Based on this assumption, the focus measure is defined and computed over pixels from individual planar image frames. In most practical situations, this approximation is not valid due to complex geometry of objects in the scene. The conventional methods thus fail to provide accurate shape estimation for objects with complex geometry.

Subbarao and Choi [2] proposed an accurate method for SFF based on a new concept named focused image surface (FIS). They defined FIS of an object as a surface formed by the set of points at which the object points are focused by a camera lens. They also proved that there exists a one-to-one correspondence between the object shape and corresponding FIS. So SFF problem can be regarded as the one of finding accurate shape of FIS in the image space. The method called "SFF.FIS" used piecewise planar approximation of FIS in small three-dimensional (3-D) windows around each pixel [2]. This method provides more accurate results than the traditional SFF methods.

In another work, Choi *et al.* [4] proposed the use of Lagrange polynomials to approximate the shape of 3-D FIS. This method provides even better 3-D shape estimation, however, the computational cost is increased tremendously. Another disadvantage of this method is that an increase in the order of Lagrange polynomial results in undesirable oscillations in the approximated surface.

Manuscript received July 20, 1999; revised July 12, 2001. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Bangalore S. Manjunath.

The authors are with Signal and Image Processing Laboratory, Department of Mechatronics, Kwangju Institute of Science and Technology, Kwangju 500-712, Korea (e-mail: asif@sipl.kjist.ac.kr; tschoi@kjist.ac.kr).

Publisher Item Identifier S 1057-7149(01)09363-0.

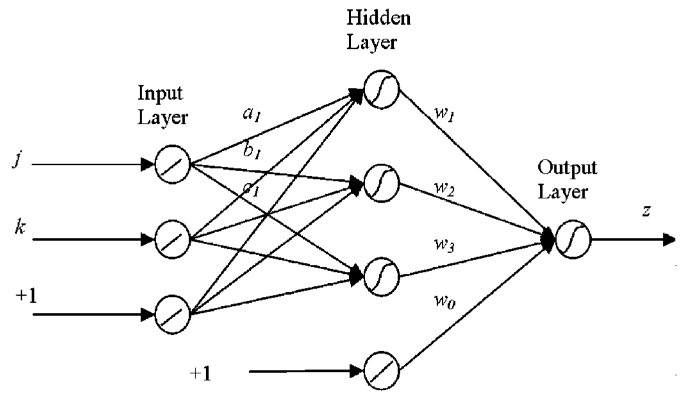


Fig. 1. Multilayer neural network structure for representation of FIS.

We propose the use of neural networks to learn shape of FIS by optimizing the focus measure over small 3-D windows. Due to their nonlinear characteristics, neural networks can be used to approximate any arbitrary function [3], [6], [7].

II. Shape From Focus Using Neural Networks

In SFF methods based on FIS, the objective is to find the most general form of function to approximate the FIS. A full search for a general function in 3-D space is computationally expensive because we need to compute the focus measure for all possible shapes of the function in the 3-D window and then search for the one that gives the best focus measure. The focus measure operators usually involve large amount of computations so the overall cost of computation is unacceptably high.

In this paper, we suggest a method to approximate the shape of FIS using neural networks. The neural network weights can be updated such that the focus measure is optimized when computed over the 3-D FIS learned by the network. Determination of SFF can thus be converted to an optimization problem in which a cost function (which is focus measure in this case) is to be optimized with respect to the network weights.

A. Neural Network Model for Shape From Focus

We use a three-layer neural network model like the one shown in Fig. 1 to approximate 3-D FIS in a small window. The first

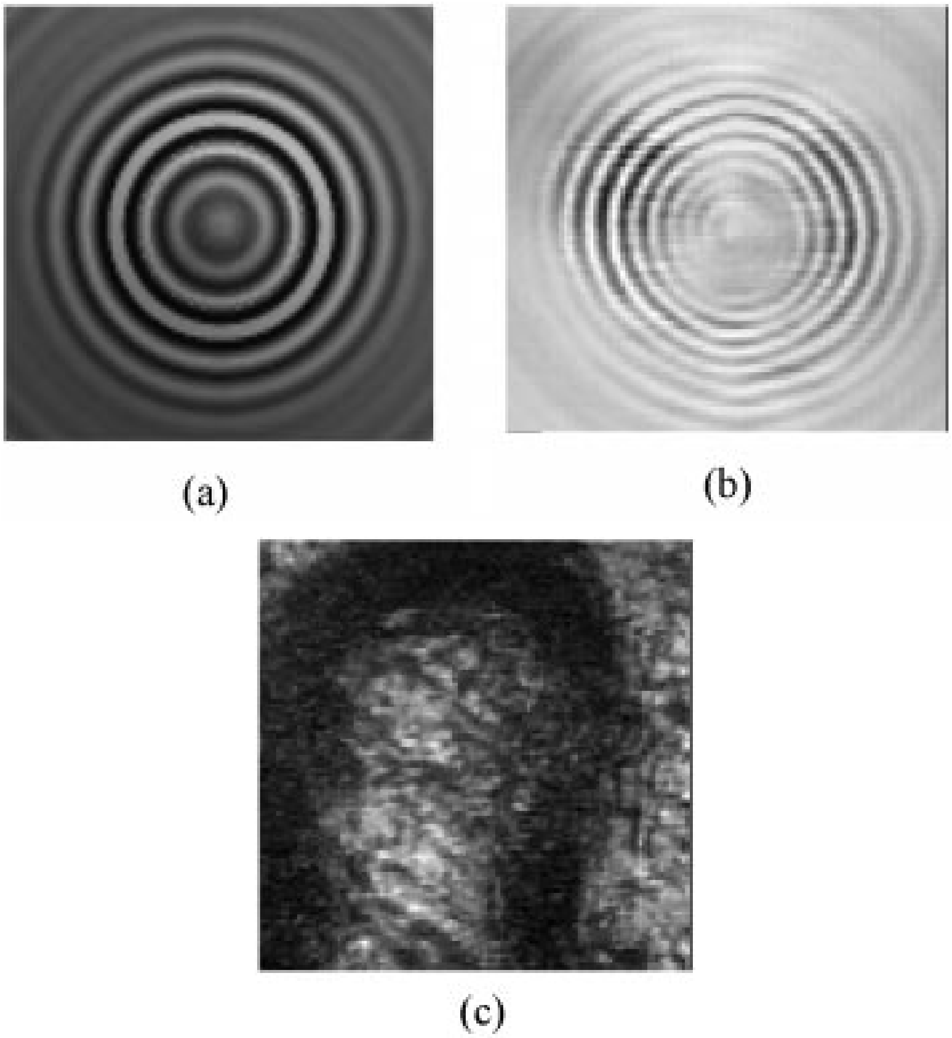


Fig. 2. Images of objects for experiments: (a) simulated cone, (b) real cone, and (c) microscopic object.

layer called *input layer* consists of linear processing elements or *neurons*. The next layer consisting of sigmoidal neurons is called *hidden layer*. The last layer called *output layer* consists of only one sigmoidal neuron. The elements with a constant input of 1 in the input layer and hidden layer are called as the *bias units*. Utilizing this three layer feedforward network structure we can express the shape of FIS in terms of the image frame number $z = z(j, k)$ as

$$z = \phi\left(\sum_{n=1}^{N} w_n \phi(a_n j + b_n k + c_n) + w_o\right) \tag{1}$$

where N is the number of elements in the hidden layer, $\{w_n, a_n, b_n, c_n\}$ are the network weights, j and k, are the indexes of the image point in the window under consideration, and $\phi(\bullet)$ is the semilinear (sigmoid) function defined as

$$\phi(x) = 1/(1 + e^{-x}) \tag{2}$$

where x is a dummy variable. The shape of the FIS has thus been represented as a function of weights of the neural network as given in (1).

B. Focus Measure

We define the focus measure F (gray level variance) in 3-D image window, centered around the pixel position (x, y), as

$$F = \sum_{j=-m}^{m} \sum_{k=-m}^{m} [I(x + j, y + k, z(j, k)) - \mu]^2 \tag{3}$$

where j and k are the local indexes of the pixel positions in the window around the pixel (x, y), $I(x + j, y + k, z(j, k))$ is the image gray level at the image position $(x + j, y + k)$ in frame number given by $z(j, k)$, and μ is the average gray level of the pixels on the window surface defined by $z(j, k)$. The map $z(j, k)$ is the estimate of the FIS as given by (1). This is actually the surface generated by the neural network model shown in Fig. 1 [8].

C. Network Training

We update the weights of the neural network according to the gradient ascent rule given as

$$\Delta W = \beta(\partial F / \partial W) \tag{4}$$

where ΔW is the net change in the weights vector W (which contains all the weights w_n, a_n, b_n and c_n as its elements) and

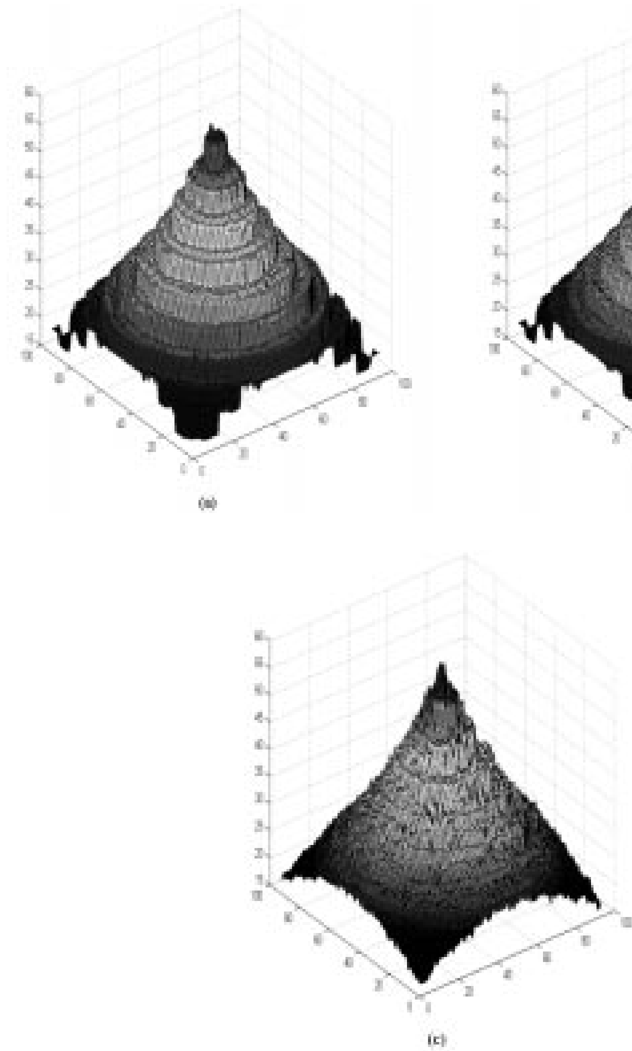


Fig. 3. Reconstructed 3-D depth map for simulated cone object by: (a) traditional SFF method, (b) SFF using Gaussian interpolation, and (c) SFF using neural networks.

β is the learning rate. By the chain rule of differentiation, (4) can be written as

$$\Delta W = \beta \left(\frac{\partial F}{\partial z}\right)\left(\frac{\partial z}{\partial W}\right). \quad (5)$$

By using (3), we can write (5) as

$$\Delta W = 2\beta \sum_{j=-m}^{m} \sum_{k=-m}^{m} \left([I(x+j, y+k, z(j, k)) - \mu] \cdot \frac{\partial}{\partial z}[I(x+j, y+k, z(j, k)) - \mu] \right) \left(\frac{\partial z}{\partial W}\right). \quad (6)$$

Since the mean of the pixels on the window surface is constant, so we can write this equation as

$$\Delta W = 2\beta \sum_{j=-m}^{m} \sum_{k=-m}^{m} \left([I(x+j, y+k, z(j, k)) - \mu] \cdot \frac{\partial}{\partial z}[I(x+j, y+k, z(j, k))] \right) \left(\frac{\partial z}{\partial W}\right). \quad (7)$$

The factor $(\partial/\partial z)I(x + j, y + k, z(j, k))$ represents the gradient of the image intensity with respect to frame number and the factor $(\partial z/\partial W)$ in (7) can be derived using (1) in a similar (but not the same) fashion as that for backpropagation algorithm.

In order to accelerate the learning process, we introduce a *momentum* term [3], [5] in (7). The weight update equation can then be written as

$$W(n+1) = W(n) + \beta\left(\frac{\partial F}{\partial W(n)}\right) + \alpha \Delta W(n-1) \quad (8)$$

where α is called the momentum constant and $\Delta W(n-1)$ is the previous value of the correction in the weights. The values of α and β varied throughout the experiments, typical values being 0.2 and 0.1, respectively. The final weights for a window are saved in memory and the same procedure is repeated for the next windows in the image space. This scheme can also be implemented in parallel.

III. Experimental Results

In this section, we will discuss the experiments and their results for 3-D shape recovery from image focus using neural networks. For the method of shape from focus using neural net-

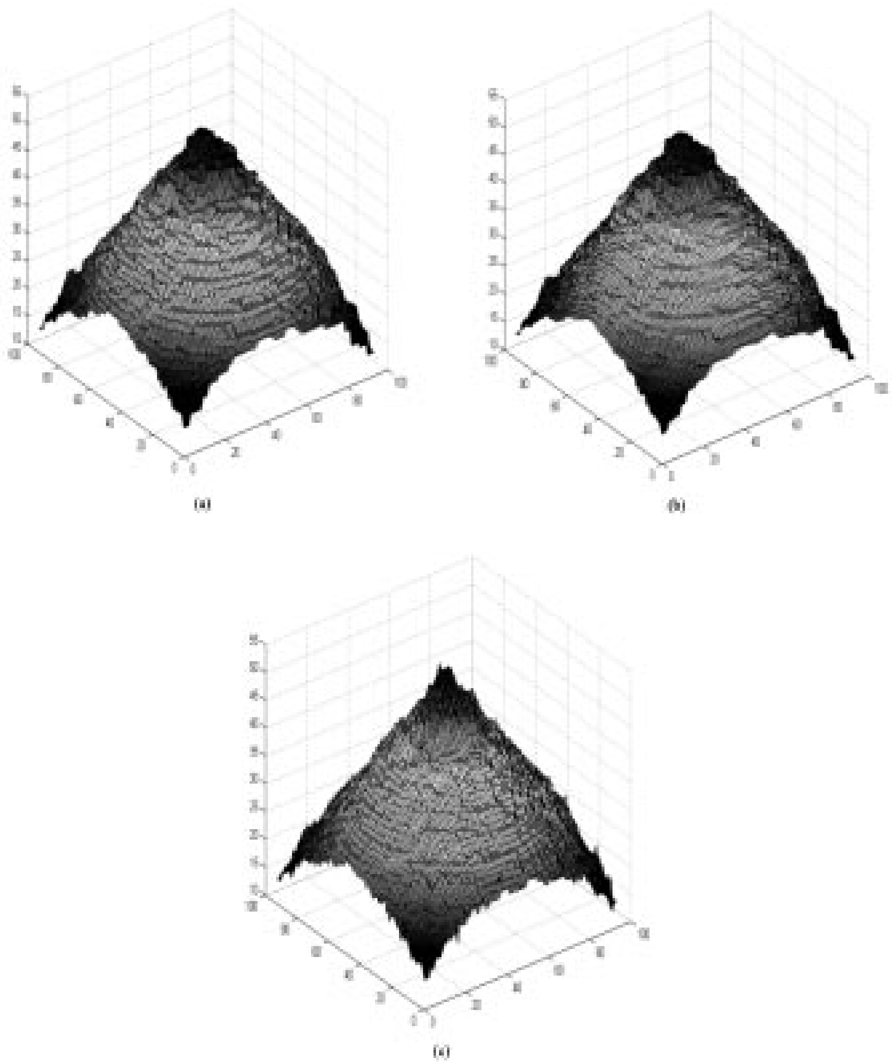


Fig. 4. Reconstructed 3-D depth map for real cone object by (a) traditional SFF method, (b) SFF using Gaussian interpolation, and (c) SFF using neural networks.

works, a rough estimate of FIS was obtained using the traditional shape from focus method. Gray level variance was used to compute the focus measure for all the experiments. The window size throughout the experiments was 7×7 pixels. This rough depth map was then used to compute the weight update ΔW. After updating the weights, a depth map was generated again and the derivatives were computed to calculate the next weight update. This process continued until certain performance goal had been achieved. The performance goal to stop the training was a certain number of epochs, e.g., 100. After the training was complete for one window, the same process was repeated for the next one and so on.

Three different types of objects were selected for experiments to show the performance of the proposed scheme. The first object is a simulated cone whose images were generated using a camera simulation software. A detailed description of the images for simulated cone object can be found in [2] and [4], since we have used the same images described there. The second object is a real cone whose images were taken using a CCD camera system [4]. The real cone object of length 90 cm and base diameter 14 cm was made of hardboard with black and white stripes drawn on the surface so that a dense texture of ring patterns is viewed in images. The third object is a microscopic object. Microscopic images of the head part of Lincoln statue on one cent coin taken at different lens steps were used for experiments. Image sequences consisting of 60 images each were used for the experiments. The size of each image was 100×100 pixels with 256 gray levels. Fig. 2 shows one image each of the three objects at lens step number 37. In addition to the proposed method, we also implemented the traditional SFF method and SFF using Gaussian estimation proposed by Nayar *et al.* [1] for comparison.

The results of 3-D shape reconstruction for simulated cone object, real cone object and the microscopic object are shown in Figs. 3–5, respectively. In each of the three figures, (a) shows results by traditional SFF, (b) shows results by SFF using Gaussian estimation, and (c) shows results by SFF using neural networks.

We can see a general trend in the results of the shape estimation by the three methods. For the depth estimate by traditional SFF method shown in (a) of Figs. 3–5, the estimated shape is not smooth and depth seems to change in large jumps instead of varying gradually. In case of simulated cone and real cone ob-

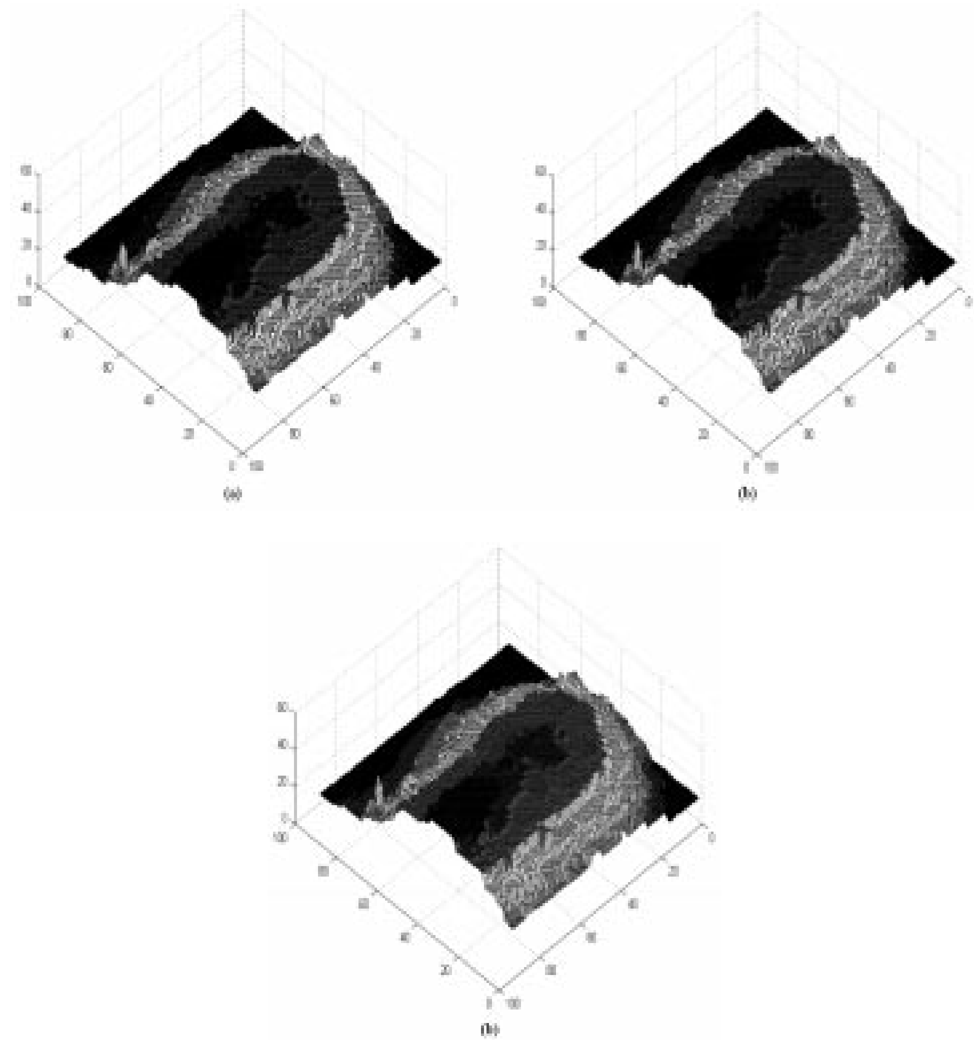


Fig. 5. Reconstructed 3-D depth map for microscopic object by: (a) traditional SFF method, (b) SFF using Gaussian interpolation, and (c) SFF using neural networks.

jects [Figs. 3(a) and 4(a)], the tip of the cone is not very sharp, apparently due to inaccurate approximation of FIS. SFF using Gaussian interpolation improves the depth estimation by making the depth step smaller as shown in (b) of Figs. 3–5, however, the improvement is not that significant in case of the simulated cone object [see Fig. 3(b)]. The results of shape estimation using neural networks are presented in (c) of Figs. 3–5. Here, we observe that the surface is smoother and the depth varies in even smaller steps. Also the tips of cone objects are sharper in case of SFF using neural networks [Figs. 3(c) and 4(c)] as compared to those in case of SFF by Gaussian interpolation [Fig. 3(b) and 4(b)].

IV. Conclusions

We have discussed a new method for SFF using neural networks. The traditional SFF methods do not yield accurate depth estimates because they compute the focus measure over the pixels in only single image frame. The proposed algorithm is based on representation of the FIS of the object in terms of neural network weights and then optimization of the focus measure over the network weights. Gradient ascent method is used to train the neural network to learn the shape of FIS. Two conventional SFF methods were also implemented to compare the shape estimates with those from the proposed scheme. Experimental results demonstrate that the proposed method gives better estimate of 3-D shape as compared to traditional SFF methods.

References

[1] S. K. Nayar and Y. Nakagawa, "Shape from focus," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 16, pp. 824–831, Aug. 1994.

[2] M. Subbarao and T. S. Choi, "Accurate recovery of three-dimensional shape from image focus," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 17, pp. 266–274, Mar. 1995.

[3] G.-Q. Wei and G. Hirzinger, "Learning shape from shading by a multilayer network," *IEEE Trans. Neural Networks*, vol. 7, pp. 985–995, July 1996.

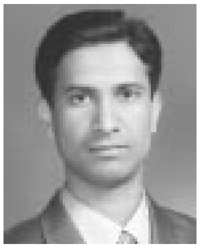
[4] T. Choi and J. Yun, "Three-dimensional shape recovery from focused image surface," *Opt. Eng.*, vol. 39, May 2000.

[5] A. Cichocki and R. Unbehauen, *Neural Networks for Optimization and Signal Processing*. New York: Wiley, 1992.

[6] F. Hornik, "Multilayer feedforward networks are universal approximators," *Neural Networks*, vol. 2, pp. 359–363, 1989.

[7] D. T. Pham and V. Aslantas, "Depth from focus using a neural network," *Pattern Recognit.*, vol. 32, no. 5, pp. 715–727, 1999.

[8] A. N. Rajagopalan and S. Chaudhuri, "A variational approach to recovering depth from defocused images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, pp. 1158–1164, Oct. 1997.



Muhammad Asif received the M.Sc. degree in electronics from Quaid-i-Azam University, Islamabad, Pakistan, in 1992, and the M.S. degree in mechatronics from Kwangju Institute of Science and Technology, Kwangju, Korea, in 1999. He is currently pursuing the Ph.D. degree in the Signal and Image Processing Laboratory, Department of Mechatronics, Kwangju Institute of Science of Technology.

His research interests are image processing, neural networks, pattern recognition, and machine vision.

Tae-Sun Choi (S'88–M'93–SM'99) received the B.S. degree in electrical engineering from Seoul National University, Seoul, Korea, in 1976, the M.S. degree in electrical engineering from the Korea Advanced Institute of Science and Technology in 1979, and the Ph.D. degree in electrical engineering from the State University of New York at Stony Brook in 1993.

He is currently an Associate Professor with the Department of Mechatronics, Kwangju Institute of Science of Technology, Kwangju, Korea. His research interests include image processing, machine/robot vision, and visual communications.