Superresolution for Ultrasonic Imaging in Air Using Neural Networks

Jack H. Winters

AT&T Bell Laboratories
Holmdel, New Jersey 07733

ABSTRACT

In this paper we study ultrasonic imaging in air using an array of transducers. We describe a superresolution technique that uses the fact that most surfaces act as perfect reflectors to ultrasonic pulses in air to generate accurate surface maps for object identification. The technique involves the minimization of a quadratic objective function subject to a nonlinear equality constraint. We show that this minimization can be accomplished by a two step penalty function method, which, although not practical on a general purpose computer, can operate in real time on a pair of neural networks. Results demonstrate that the technique generates accurate surface maps even with low receive signal-to-noise ratios.

I. INTRODUCTION

Three-dimensional machine vision using ultrasonic pulses may, in some cases, offer an attractive alternative to optical techniques because of the low cost of the transducers and the fact that depth information is directly available from the received signal. Furthermore, time gating of the received pulses can be used to examine only those objects in the range of interest, eliminating the background clutter that is often present in optical vision systems.

Since most objects act as perfect reflectors to ultrasonic pulses in air, ultrasonic vision systems generate maps of the surfaces of objects (e.g., range maps). Distorted or noisy data must, therefore, be processed to generate an image that contains a nonzero value at no more than one range for each look angle. In fact, this surface constraint can be used during image processing to greatly enhance image quality through the use of superresolution techniques [1]. Superresolution techniques are particularly useful in ultrasonic imaging since only a small portion of objects can be seen because of the specular reflections, and since it is usually desirable from a cost and physical size viewpoint to keep the number and size of the transducers to a minimum. An earlier analysis [2] showed that with superresolution, ultrasonic vision systems had tremendous potential in machine vision. However, superresolution techniques require an extensive amount of signal processing, especially with a three-dimensional image. Therefore, in previous papers on machine vision using ultrasonic pulses in air [3-8], which are primarily concerned with real time image processing on general purpose digital computers, only image processing algorithms that generate surface maps with much less accuracy than superresolution techniques have been analyzed.

Neural networks [9-12], however, which use an interconnected structure of simple elements for parallel computation, can be used to implement signal processing algorithms that operate orders of magnitude faster than on general purpose digital computers. The trick is to cast the algorithm into a form that can be solved using neural networks. In our case, superresolution can be achieved by finding the resolved image that minimizes the mean square error with the received image subject to the surface constraint [13]. Specifically, the problem can be expressed as minimizing a quadratic objective function subject to a nonlinear constraint. The minimization of a quadratic objective function can be implemented by neural networks as shown in [12]. Here we show how this neural network can be modified to include the nonlinear constraint.

II. SYSTEM MODEL

The vision system that we wish to consider is the same as that in [2] and shown in Figure 1, which, for simplicity, is drawn in two dimensions only (azimuth *or* elevation and

range). A surface (which may consist of one or more objects) is illuminated from the right by a spherical wave produced by an ultrasonic point source. The characteristic dimensions of the objects are assumed to be large compared with the ultrasonic wavelength, and the acoustic impedance of the objects is assumed to be much larger than that of air, so that the surface of the objects appears like a perfect mirror to the ultrasonic illumination. The scalar field at the points where the wave is reflected by the surface presents an input to the vision system which is then transformed, in accordance with the principles of Fourier optics [14], into an image along the image sphere. In practice, the thin lens may be replaced by an array of small receptors; for now, we assume the existence of many small receivers located along the image sphere which present a continuum of the image to the vision processor. We assume that all surfaces are located in the far field of the lens and that the range of interest is small enough so that the image is focused for all points on the surface that are illuminated.

Let $F(\alpha_a,\alpha_e,r)$ be the return from the surface in polar coordinates, where α_a and α_e are the azimuth and elevation angles, respectively, of the surface, and r is the range. Note that under the assumption that all surfaces are perfect reflectors, on the surface, for every α_a and α_e there is only one r, i.e., $r=h(\alpha_a,\alpha_e)$, or

$$F(\alpha_a,\alpha_e,r) = \begin{cases} E(\alpha_a,\alpha_e)\delta(r-h(\alpha_a,\alpha_e)) & \alpha_a,\alpha_e \in \Omega \\ 0 \qquad \text{elsewhere} \end{cases} , \tag{1}$$

where Ω is the region where the surface exists, and $E(\alpha_a,\alpha_e)$ is the return from the surface at look angles α_a and α_e. Thus, following the analysis of [2], which uses the principles of Fourier optics [14], we can show that, when the surface is illuminated by a pulse $\bar{p}(t)$, the received image is given by

$$\bar{S}(\theta_a,\theta_e,t) = F(\alpha_a,\alpha_e,r) \ast A(\alpha_a,\alpha_e) \ast \bar{p}(t) \tag{2}$$

where $\ast$ denotes convolution, θ_a and θ_e are the azimuth and elevation angles, respectively, of the received image, $A(\alpha_a,\alpha_e)$ is the azimuth and elevation aperture function of the lens, and the overbar denotes a complex valued function.

Next, consider the matrix form for the discretized version of Eq. (2). Since the received image has been bandlimited by the pulse bandwidth and lens aperture size, the resolution of surface map (without any *a priori* constraints) is approximately the inverse of the Nyquist rate, i.e., the Rayleigh or diffraction limit. Thus, a discretized version of the source spaced at approximately the inverse of the Nyquist rate will generate a received image similar to that obtained from the continuous source[1]. Note that since the discrete source points approximate the continuous source function under illumination by a pulse, with a nonzero pulse carrier frequency, the source points can take on complex values. Thus, the return matrix $\overline{\mathbf{F}}$ is a one, two, or three dimensional complex matrix, depending on the dimensionality of the problem. Therefore, to express Eq. (2) in matrix form we use the direct product method [1], i.e., we convert the 3-dimensional image to a 1-dimensional image vector by stacking the columns of $\overline{\mathbf{F}}$ on top of each other in the two and three dimensional cases. Thus, the received image is given by

$$\overline{\mathbf{S}} = \overline{\mathbf{H}}\overline{\mathbf{F}} \quad , \tag{3}$$

where $\overline{\mathbf{H}}$ is the impulse response matrix, calculated from $A(\alpha_a,\alpha_e)$ and $\bar{p}(t)$. With additive noise at the receiver, the observed complex image vector is then given by

$$\overline{\mathbf{G}} = \overline{\mathbf{H}}\overline{\mathbf{F}} + \overline{\mathbf{n}} \tag{4}$$

where $\overline{\mathbf{n}}$ is the received noise vector. Finally, since the techniques analyzed in the next section

1 Since the image resolution can be increased by superresolution techniques using *a priori* constraints, more closely spaced source points are used with these techniques.

require real rather than complex vectors and matrices, let

$$\mathbf{H} = \begin{bmatrix} \mathrm{Re}\{\overline{\mathbf{H}}\} & -\mathrm{Im}\{\overline{\mathbf{H}}\} \\ \mathrm{Im}\{\overline{\mathbf{H}}\} & \mathrm{Re}\{\overline{\mathbf{H}}\} \end{bmatrix} \tag{5}$$

$$\mathbf{G} = \begin{bmatrix} \mathrm{Re}\{\overline{\mathbf{G}}\} \\ \mathrm{Im}\{\overline{\mathbf{G}}\} \end{bmatrix} \tag{6}$$

and similarly for $\overline{\mathbf{F}}$ and $\overline{\mathbf{n}}$, so that the image equation is given by

$$\mathbf{G} = \mathbf{HF} + \mathbf{n} \quad . \tag{7}$$

Note that with d dimensions and M points per dimension, $\mathbf{H}$ contains $4M^{2d}$ elements, and, therefore, $\mathbf{H}$ is very large even with just a few points in each dimension.

III. SUPERRESOLUTION

We now consider methods to solve the imaging problem: given $\mathbf{G}$ $(=\mathbf{HF}+\mathbf{n})$, estimate $\mathbf{F}$. In this paper we study the calculation of an estimate of $\mathbf{F}$, $\hat{\mathbf{F}}$, that minimizes the mean squared error between the observed and degraded image, i.e., minimizes $||\mathbf{G}-\mathbf{H}\hat{\mathbf{F}}||^2$, subject to the surface constraint that the resolved image $\hat{\mathbf{F}}$ has, for each range, at most one nonzero element. We assume that the elements of the noise vector are independent Gaussian random variables with zero mean and variance σ^2.

Mathematically, we want to find the $\hat{\mathbf{F}}$ that minimizes the objective function $O(\hat{\mathbf{F}}) = ||\mathbf{G}-\mathbf{H}\hat{\mathbf{F}}||^2$, with the surface constraint. Since,

$$\begin{aligned} O(\hat{\mathbf{F}}) = ||\mathbf{G}-\mathbf{H}\hat{\mathbf{F}}||^2 &= (\mathbf{G}-\mathbf{H}\hat{\mathbf{F}})^T(\mathbf{G}-\mathbf{H}\hat{\mathbf{F}}) \\ &= \mathbf{G}^T\mathbf{G}-2\mathbf{G}^T\mathbf{H}\hat{\mathbf{F}}+\hat{\mathbf{F}}^T\mathbf{H}^T\mathbf{H}\hat{\mathbf{F}} \quad , \end{aligned} \tag{8}$$

where the superscript T denotes transpose, the objective function to be minimized can also be expressed as

$$O(\hat{\mathbf{F}}) = \hat{\mathbf{F}}^T\mathbf{H}^T\mathbf{H}\hat{\mathbf{F}}-2\mathbf{G}^T\mathbf{H}\hat{\mathbf{F}} \tag{9}$$

(since $\mathbf{G}^T\mathbf{G}$ is independent of $\hat{\mathbf{F}}$). The surface constraint can be expressed as

$$|\mathbf{F}|^T\mathbf{D}_1|\mathbf{F}| = 0 \quad , \tag{10}$$

where $|\mathbf{F}|$ denotes the absolute value of the elements of $\mathbf{F}$ (i.e., the absolute values of $\mathrm{Re}\{\overline{\mathbf{F}}\}$ and $\mathrm{Im}\{\overline{\mathbf{F}}\}$) and

$$\mathbf{D}_1 = \begin{bmatrix} \mathbf{D} & \mathbf{D} \\ \mathbf{D} & \mathbf{D} \end{bmatrix} \quad , \tag{11}$$

where $\mathbf{D}$ is an $M^d \times M^d$ matrix given by

$$\mathbf{D} = \begin{bmatrix} \mathbf{D}' & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{D}' \end{bmatrix} \tag{12}$$

where $\mathbf{D}'$ is a $M^{d-1} \times M^{d-1}$ matrix of all 1's except for 0's along the diagonal.

The minimization of an objective function subject to a nonlinear equality constraint can be accomplished by the penalty function method [13,15]. In this method, the objective function is modified so that a large positive value is added to it when the equality constraint is not satisfied. In particular, to reduce the problem of numerous local minima, a multiple-step penalty function method can be used whereby minimization of the objective function is repeated with, at each step, increasing values added when the equality is not satisfied. We found that a two step penalty function was satisfactory. First, we minimize the objective function

$$O(\hat{\mathbf{F}}) = \hat{\mathbf{F}}^T\mathbf{H}^T\mathbf{H}\hat{\mathbf{F}}-2\mathbf{G}^T\mathbf{H}\hat{\mathbf{F}} \quad . \tag{13}$$

Then, for the second step, we minimize the objective function

$$O(\hat{\mathbf{F}}) = \hat{\mathbf{F}}^T\mathbf{H}^T\mathbf{H}\hat{\mathbf{F}} - 2\hat{\mathbf{F}}^T\mathbf{H}^T\mathbf{G} + \beta_1/2\,|\hat{\mathbf{F}}|^T\mathbf{D}_1|\hat{\mathbf{F}}| \quad . \tag{14}$$

In our test cases, we found that $\beta_1 = 10$ ensured that the constraint was met.

IV. NEURAL NETWORK IMPLEMENTATION

Now, consider the neural network implementation of the two step penalty function method. We will consider the neural network implementation of the second step, Eq. (14), noting that the neural network implementation of the first step, Eq. (13), is the same network, but with $\beta_1 = 0$. For a neural network implementation, we need to find a network that has the objective function as a Liapunov function. That is, the time behavior of the network is such that the objective function is minimized. This can be done by using a method similar to that used by Hopfield [12]. Specifically, consider the objective (energy) function given by

$$E = -\frac{1}{2}\mathbf{v}^T\mathbf{T}\mathbf{v} - \mathbf{v}^T\mathbf{I} + \frac{1}{2}\beta_1|\mathbf{v}|^T\mathbf{D}_1|\mathbf{v}| + \frac{1}{R}\sum_{i=1}^{N}\int_0^{v_i} g^{-1}(v_i)\,dv_i \quad , \tag{15}$$

where

$$v_i = g(u_i) = \begin{cases} \lambda u_i & |u_i| \le 2/\lambda \\ 2 & u_i > 2/\lambda \\ -2 & u_i < -2/\lambda \end{cases} \quad , \tag{16}$$

where we have arbitrarily limited the maximum magnitude of the return to 2 (in practice, other maximum magnitudes can be used). Note that if we let $\mathbf{T} = -2\mathbf{H}^T\mathbf{H}$, $\mathbf{I} = 2\mathbf{H}^T\mathbf{G}$, and $\mathbf{v} = \hat{\mathbf{F}}$, the energy function is the same as the objective function, except for the last term. This term can be made negligible by choosing λ large enough. Converting to scalar notation, we can show that

$$\frac{dE}{dt} = -\sum_{i=1}^{N}\frac{dv_i}{dt}\left(\sum_{j=1}^{N} T_{ij}v_j - u_i/R + I_i - \mathrm{sgn}(v_i)\beta_1\sum_{j=1}^{N} D_{1_{ij}}|v_j|\right) \tag{17}$$

where

$$\mathrm{sgn}(v_i) = \begin{cases} 1 & \text{if } v_i \ge 0 \\ -1 & \text{if } v_i < 0 \end{cases} \quad . \tag{18}$$

If we design a network such that

$$C\frac{du_i}{dt} = \sum_{j=1}^{N} T_{ij}v_j - u_i/R + I_i - \mathrm{sgn}(v_i)\beta_1\sum_{j=1}^{N} D_{1_{ij}}|v_j| \quad , \tag{19}$$

then

$$\frac{dE}{dt} = -C\sum_{i=1}^{N}\left(\frac{dv_i}{dt}\right)^2\left(\frac{du_i}{dv_i}\right). \tag{20}$$

Thus, since u_i is a monotonically increasing function of v_i, and C and $\left(\frac{dv_i}{dt}\right)^2$ are positive,

$\frac{dE}{dt} \leq 0$, i.e., the network dynamics are such that the objective function is minimized, and the stable points in the network are the local minima of E (E is a Liapunov function for the system).

Figure 2 shows the network that implements the two step penalty function method. Note that since resistors can only implement positive values, both the inverted and noninverted output of the amplifiers are used, with the primed resistors (e.g., R'_{1N}) used for negative values and the unprimed resistors used for positive values. Also, in the network, the output of the f amplifiers is equal to the input and the response time of the output of the f amplifiers to changes in the input is assumed to be negligible compared to the response time of the g amplifiers with RC circuits on the inputs. In operation, after the top network settles to the solution of the first step, the bottom switch is thrown to the right and the bottom network settles to the solution of the second step (the final solution).

To study the performance of the neural network, consider an example case of a flat plate with $M=5$ and the points spaced at twice the inverse of the Nyquist rate. The performance of the network was determined using computer simulation. For solution on a general purpose minicomputer, even for this simple case the CPU time is about 30, 600, and 15,000 seconds for 1, 2, and 3 dimensional smearing, respectively. On the other hand, the neural network implementation of the algorithm requires a pair of networks with 10, 50, and 250 neurons (and 100, 2500, and 6.25×10^4 interconnecting resistors) for 1, 2, and 3 dimensional smearing, respectively, and the solution is obtained in less than 10 RC time constants (which could be on the order of microseconds) in all three cases.

Figure 3 shows the probability of error, P_e, in the resolved image ($P_e = .2$ corresponds to one point on the resolved image of the flat plate having the wrong range) versus the signal-to-noise ratio, S/N, in the received signal. Only one and two dimensional results are shown because a prohibitive amount of computer time is required to generate results for the three dimensional case. The results show that for $S/N \geq 15$ and 10 dB with one and two dimensional smearing, respectively, the P_e is very small. Note that at such low signal-to-noise ratios, it is nearly impossible to identify the flat plate in the unresolved image (at $S/N = 10$ dB, the standard deviation of the noise is 32% of the maximum signal level). Thus, our results show that the neural network can generate high accuracy images from noisy, smeared images in just a few microseconds.

V. SUMMARY AND CONCLUSIONS

In this paper we have studied imaging using ultrasonic pulses in air with an array of transducers. We described an algorithm that generates a surface map from the 3-dimensional observed image by minimizing the mean squared error between the observed and degraded image, i.e., it fits a surface to the observed image. We then showed that the algorithm can be implemented by a pair of neural networks which can resolve the image in microseconds. Results showed that the algorithm was robust against noise corruption of the received image. Thus, with these techniques, an array of ultrasonic transducers can be used to generate very accurate surface maps even with low received signal-to-noise ratios.

REFERENCES

[1] W. K. Pratt, *Digital Image Processing*, John Wiley and Sons, New York, 1978.

[2] A. S. Acampora and J. H. Winters, "Three Dimensional Ultrasonic Vision for Robotic Applications," submitted to *IEEE Transactions on Pattern Analysis and Machine Intelligence.*

[3] J. M. Richardson, K. A. Marsh, J. S. Schoenwald, and J. F. Martin, "Acoustic Imaging of Objects in Air Using a Small Set of Transducers," *Proc. 1984 IEEE Ultrasonics*

Symposium.

[4] K. A. Marsh, J. M. Richardson, J. F. Martin, and J. S. Schoenwald, "Acoustic Imaging in Robotics Using a Small Set of Transducers," *Proc. of 4th Int'l Conf. on Robot Vision and Sensory Controls (1984).*

[5] H. P. Moravec and A. Elfes, "High Resolution Maps from Wide Angle Sonar," *Proc. 1985 IEEE Int'l Conf. on Robotics and Automation.*

[6] R. A. Jarvis, *Computer*, vol. 15, p. 8, 1982.

[7] J. S. Schoenwald and J. F. Martin, "Acoustic Scanning for Robotic Range Sensing and Object Pattern Recognition," *Proc. 1989 IEEE Ultrasonics Symposium.*

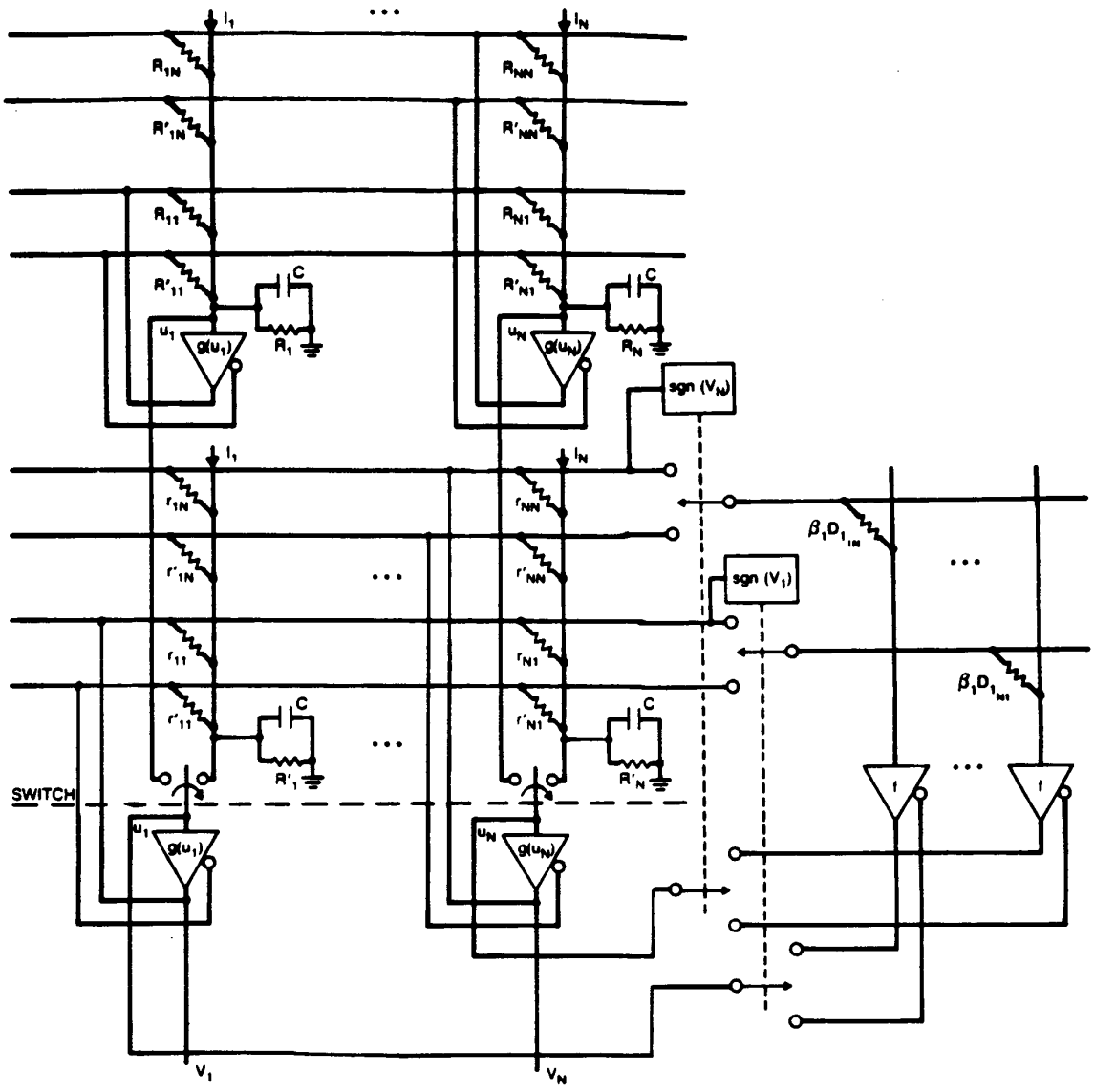

Figure 2 Schematic of the neural network.

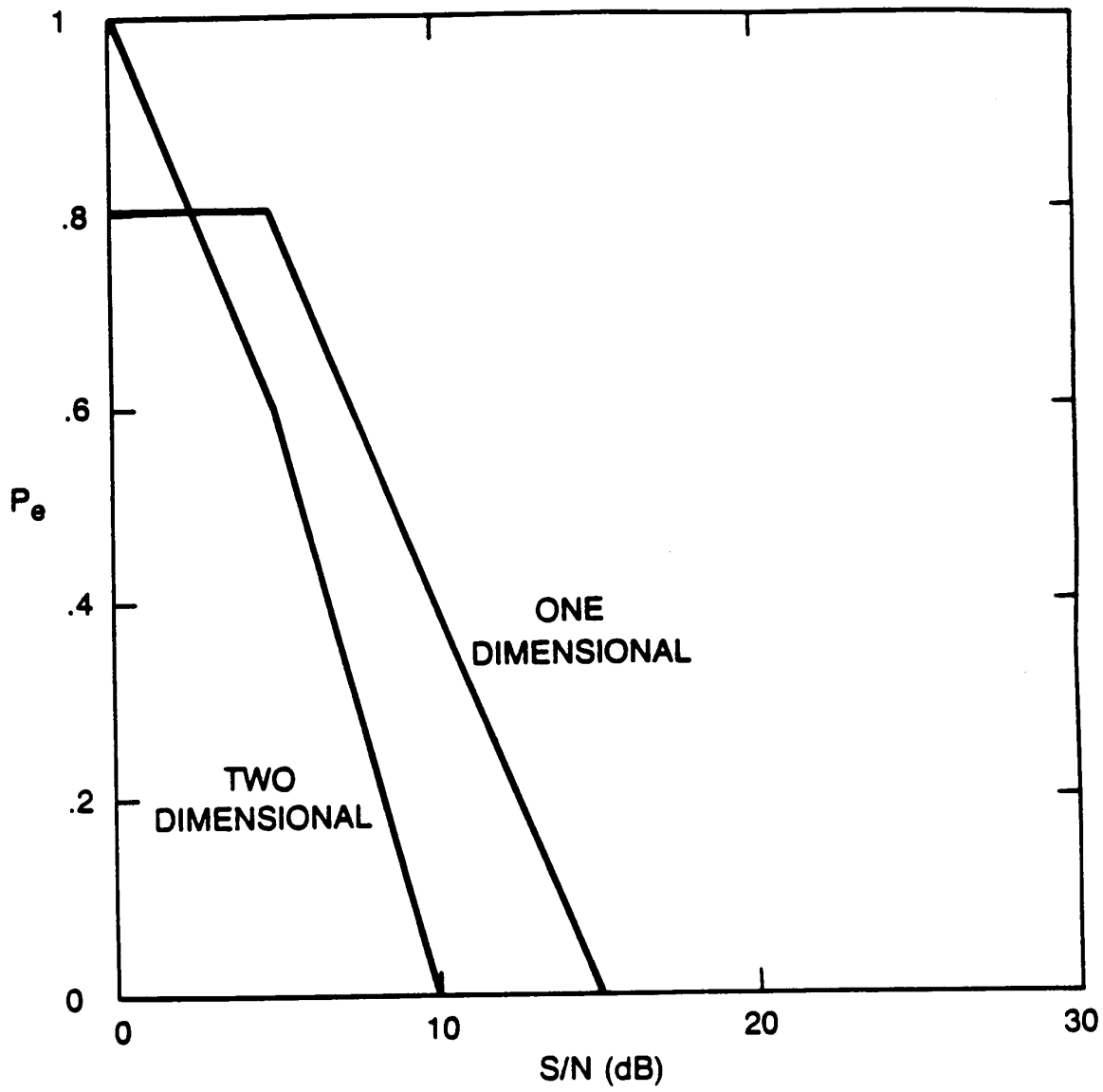

Figure 3 P_e versus S/N for the two step penalty function method implemented by a pair of neural networks.