Neural networks for image analysis

Chris M.W. Daft
GE Corporate Research and Development Center
Schenectady, New York 12301
daft@crd.ge.com

1 Abstract

Neural networks have become a popular tool in a broad range of disciplines in a short time [1]. They are fundamentally analog, non-programmed data processing structures. The networks are comprised of processing elements, each of which has set of inputs, a set of weights, and one output. Inputs are multiplied by their weights and summed: the output is computed as a nonlinear function of the summation. The outputs are used as inputs to other units. They offer fine-grained parallelism, and exhibit fault-tolerance. Three applications of neural networks to image analysis are presented. Current work on ultrasonic applications of neural networks are then cited.

2 Introduction

Why is there currently such a large degree of interest in neural networks? Early work with them, by Rosenblatt and others, took place in the 20 years before 1960. This was effectively halted by Minsky and Papert [2], who showed the limitations of the types of networks in use then. Several important mathematical results have caused them to rise, phoenix-like, from the ashes. First, there is a proof that a different kind of network can realize arbitrary nonlinear functions. A method for training these multi-layer networks has also been discovered. Yet another style of network has been found to perform well on difficult constrained minimization problems.

I would like to introduce three applications of neural networks to image processing: the silicon retina of Mead [3], a method of echo inversion by Farhat [4], and the fingerprint recognition of Engeler [5]. Carver Mead's work is significant because it forms a front-end for image analysis with a morphology modeled on the eye. One could then ask, why is it interesting to try to build computers which ape the methods and structure of the brain? Differences between the brain and "ordinary" computers are many and obvious. The computational metaphor of the brain knows nothing of bits, Boolean algebra, or linear system theory. Trivial questions of arithmetic are hard for the brain to compute accurately. But it has pattern recognition abilities far exceeding those of the most powerful supercomputers. Nabil Farhat's group at the University of Pennsylvania have done impressive work on the use of neural networks in image reconstruction. A description of some of this is included in this review because it concerns the regularization of an inverse problem, which is similar to many ultrasonic, medical and geophysical imaging applications. It also illustrates what many believe to be the future of neural network implementation: optics. Bill Engeler of GE Research been working on fingerprint recognition using a neural network. This is noteworthy not only as an example of neuromorphic image processing, but also because the representation of the data is inspired by measurements on the mammalian brain. In conclusion, I will indicate the current areas of research in ultrasonic applications of neural networks.

3 The silicon retina

The retina is the first stage in the visual system. It provides gain control and image enhancement, after transducing the light into electrical signals. It greatly reduces the signal bandwidth required to transmit the image to the rest of the brain. Irrelevant information, such as average illumination levels, are selectively rejected. We will see that the silicon retina succeeds in modeling the main operating features of its biological counterpart, to the extent that the ambiguities created by signal compression in the eye are mirrored in the artificial system.

Figure 1 shows the structure of a primate retina. The photoreceptors transduce the light into an electrical signal. This signal proceeds through the triad synapses to the bipolar cells. The bipolar cells are connected to the retinal ganglion cells, which form its output. As well as this, there are the outer and inner-plexiform layers. The outer plexiform layer is composed of the horizontal cells, just below the photoreceptors. Below these, in the inner plexiform layer, are the amacrine cells.

1051-0117/90/0000-1425 $1.00 © 1990 IEEE

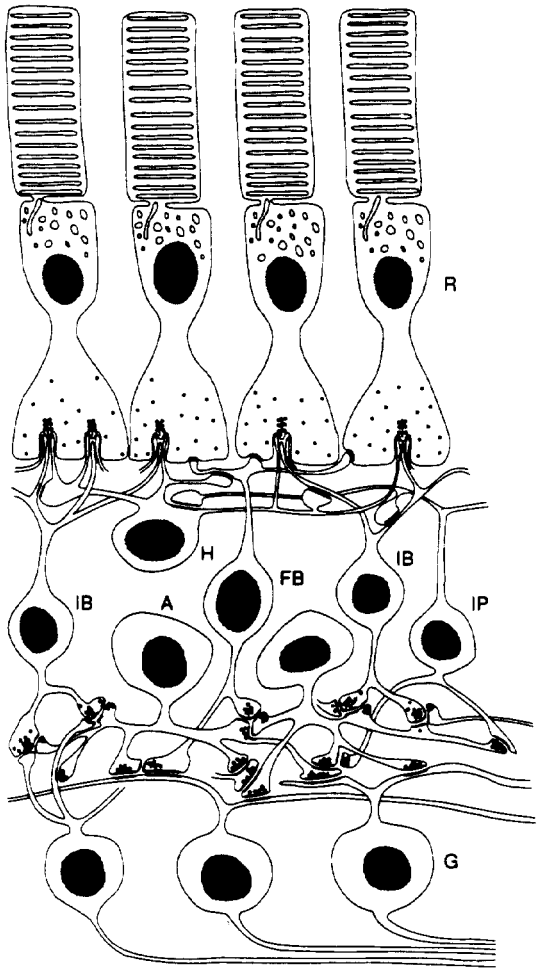

Figure 1: Conceptual drawing of a cross-section of a primate retina, indicating the primary cell types and signal pathways. The outer-plexiform layer is beneath the foot of the photoreceptors. The invagination into the foot of the photoreceptor is the site of the triad synapse. In the center of the invagination is a bipolar cell process, flanked by two horizontal cell processes. R: photoreceptor, H: horizontal cell, IB: invaginating bipolar cell, FB: flat bipolar cell, A: amacrine cell, IP: interplexiform cell, G: ganglion cell. Figures in this article are reprinted with kind permission of C.A. Mead, N.H. Farhat, W.E. Engeler, and Pergamon Press, Inc.

These cells show great diversity in their modes of operation: for example, the horizontal cells represent information with analog signals, while the retinal ganglion cells have dendrites and an axon, which produce outputs which are digital in amplitude but analog in time. The key processing in the outer-plexiform layer is the triad synapse, where the photoreceptor, bipolar and horizontal cells meet. The silicon retina emulates its behavior.

The retina functions in this way. The photoreceptor has a logarithmic transfer function. This allows the retina to process many orders of magnitude of light intensity. Also, differences in the output voltage depend on the contrast ratio of the observed light source. The horizontal cells are connected by gap junctions to form an electrically continuous network in which signals propagate. These cells can be thought of as interconnected passive cables. The network therefore computes a weighted sum of voltages surrounding a given photoreceptor, in which the weights decrease exponentially with distance. The processes of these cells are unmyelinated, so they possess a large capacitance. The voltage on a given horizontal cell is therefore a temporal as well as a spatial average. The receptive field of a bipolar cell is known as on-center, off-surround. The cell exhibits lateral inhibition: its center is stimulated by the photoreceptor, and the surround by the horizontal cells. The connection to the photoreceptor is excitatory; the other connections are inhibitory. Thus, the computation at the triad synapse is of a difference between the photoreceptor signal and the horizontal cells. This causes the output image perceived to be independent of the absolute illumination level. It also enhances spatial edges, like a Laplacian filter.

Figure 2 is a diagram of Mead's silicon retina. The photoreceptor is made of a vertical bipolar transistor. In it, photons with energies greater than the bandgap of silicon create electron-hole pairs as they are absorbed. These collect in the n-type base, lowering the energy barrier, and increasing the flow of holes from the emitter to the collector. This transistor is a byproduct of the CMOS fabrication process. The output current is fed to two diode-connected MOS transistors. They operate in the subthreshold region, and provide a logarithmic voltage output from the linear photoreceptor current, over four to five orders of magnitude of light.

The hexagonal, horizontal resistive layer computes a spatially weighted average. The resistors allow the weights for a given point to decrease approximately exponentially as the distance from that point increases. Parasitic capacitance on the chip provides temporal averaging. The computation at a node, corresponding to the triad synapse, includes two transconductance amplifiers. A follower (not shown) drives the resistive network

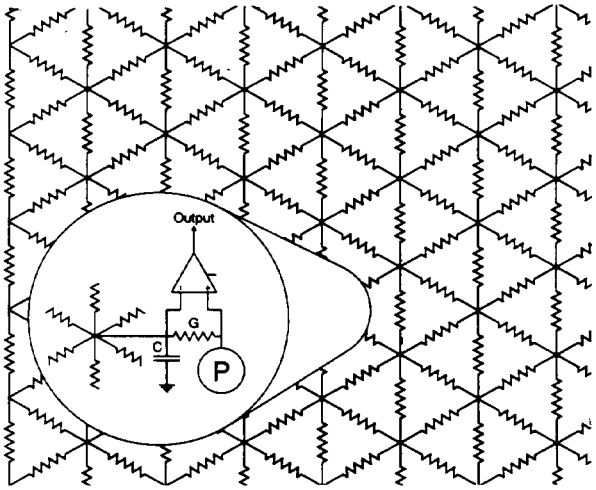

Figure 2: Diagram of the silicon retina showing the resistive network; a single pixel element is illustrated in the circular window. The silicon model of the triad synapse consists of the conductance G by which the photoreceptor P drives the resistive network, and the amplifier that takes the difference between the photoreceptor output and the voltage on the resistive network. The capacitor C represents the parasitic capacitance of the network.

with a current proportional to the voltage output of the photoreceptor. The differential amplifier (in figure 2) senses the difference in potential between the photoreceptor and the nearest node of the resistive network. The output of the "bipolar cell" equivalent is the difference between the center intensity and a weighted average of the surrounding points. The implementation of the retina as a single chip device has a 48x48 array of pixels, with a readout like a RAM. Pixels can be accessed sequentially, in order to display the output on a television screen.

The performance of Mead's retina shows striking similarities to the biological version. First, consider the sensitivity (figure 3). The output is prevented from saturating over an incredible range of illumination, due to the logarithmic response, and the lateral inhibition. The largest response is obtained when the photoreceptor's illumination is different from the spatial and temporal average of its neighborhood, i.e. when the image is changing rapidly in space or time. The step response (figure 4) shows an initial peak followed by a decay. When the stimulus is turned off, an overshoot occurs, and the output then returns to a baseline value.

Various visual illusions, such as Mach bands and the Craik-O'Brian-Cornsweet illusion, are illustrations of the mechanism of lateral inhibition in the retina. The

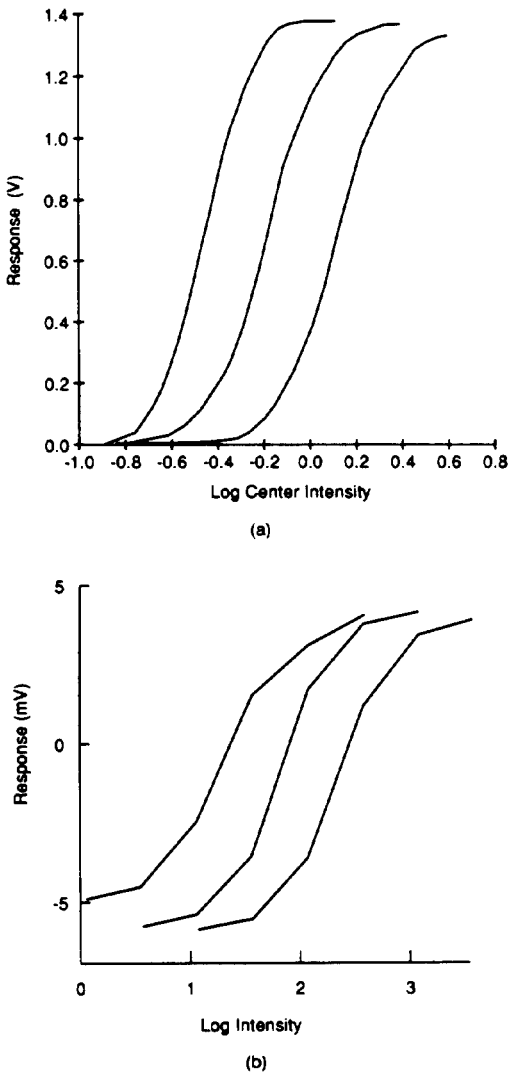

Figure 3: Intensity-response curves shift to higher intensities at higher background illuminations. (a) Intensity-response curves for a single pixel of the silicon retina. Curves plotted for three different background intensities. Stimulus was a small disk centered on the receptive field of the pixel. Steady state response is given. (b) Intensity-response curves for a depolarizing bipolar cell elicited by full field flashes. Test flashes were substituted for constant background illuminations. These curves are plotted from the peaks of bipolar response to substituted test flashes. Peak responses are plotted, and measured from the membrane potential just prior to response.

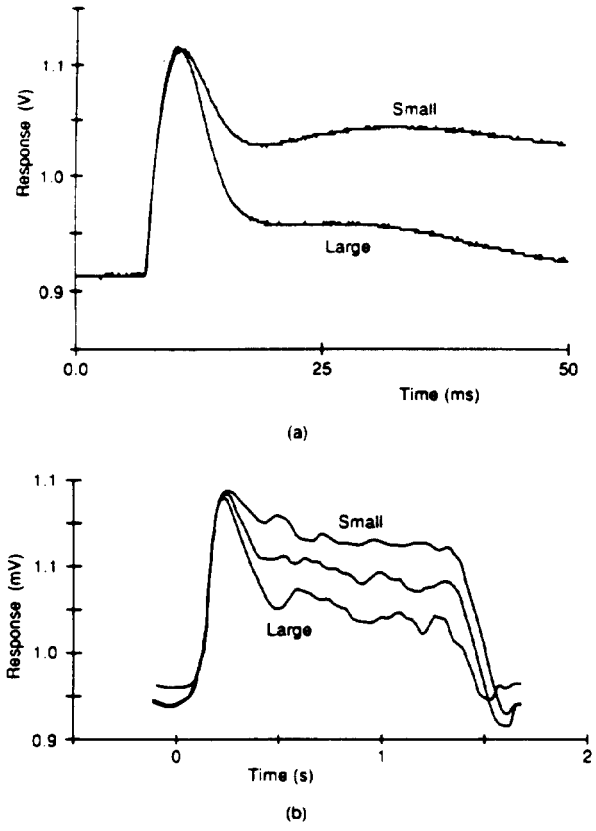

Figure 4: Temporal response of a bipolar cell of the mud puppy, *Necturus maculosus.* and of a pixel in the silicon retina, to different size test flashes. Test flashes of the same intensity but of different diameters were centered on the receptive field of the unit. (a) response of a pixel. Larger flashes increased the excitation of the surround. The surround response was delayed due to the capacitance of the resistive network. Because the surround level is subtracted from the center response, the output shows a delayed decrease for long times. This decrease is larger for larger flashes. (b) Response of *Necturus* bipolar cell.

silicon retina also has this center-surround response (figure 5).

4 Reconstructing radar images

Many problems in vision are ill-posed: for example, reconstructing a three-dimensional scene from a pair of two-dimensional images on a retina. The brain solves these problems with extraordinary ease. In radar imaging, ill-posedness arises from the fact that the microwave image data does not extend over an infinite frequency range, and the angle of observation is often limited. In microwave diversity radar imaging, the object of the is to extract size and shape information from scattering measurements.

Figure 6 shows the format of the data for a 2-D reconstruction. p_x and p_y correspond to the start and stop frequencies, and θ_1 and θ_2 are the start and stop viewing angles. Images of near-optical resolution can be obtained is the imaging aperture is large, and the frequency spread is wide. Often, providing a physical aperture of adequate size is expensive. In many applications, a synthetic aperture radar would require too much time to acquire an image.

A typical method to invert the data is to perform a 1-D inverse Fourier transform with respect to p, for a given θ. This is then called the range profile. The 2-D image is reconstructed by coherently summing the filtered, backprojected value of every range profile for all aspect angles. This works so long as the microwave wavelength is small in comparison to the object dimensions, and also if the Born approximation holds. It differs from medical tomography in that the object here is assumed to have zero impedance.

In practice, the frequency response is assumed to be zero outside the measured range. This results in a complex object function (which is obviously unphysical), and also to noise sensitivity. There is thus considerable motivation for robust reconstructions, satisfying *a priori* knowledge about the object. The neural network approach is formulated as an optimization problem, since neural nets are particularly suited to these [6]. Consider the one-dimensional task of reconstructing a range profile from frequency response data at a given aspect angle (the remainder of the standard reconstruction algorithm is well-posed). An energy function for the network is set up so that the Fourier transform of the range profile agrees with the frequency response over the measured frequency band. Regularization is used to help with the ill-posedness. The energy function is:

$$h(\mathbf{f}) = \|\mathbf{F}_d(p) - \mathbf{F}(p)\|^2 + \alpha R(\mathbf{f}) \qquad (1)$$

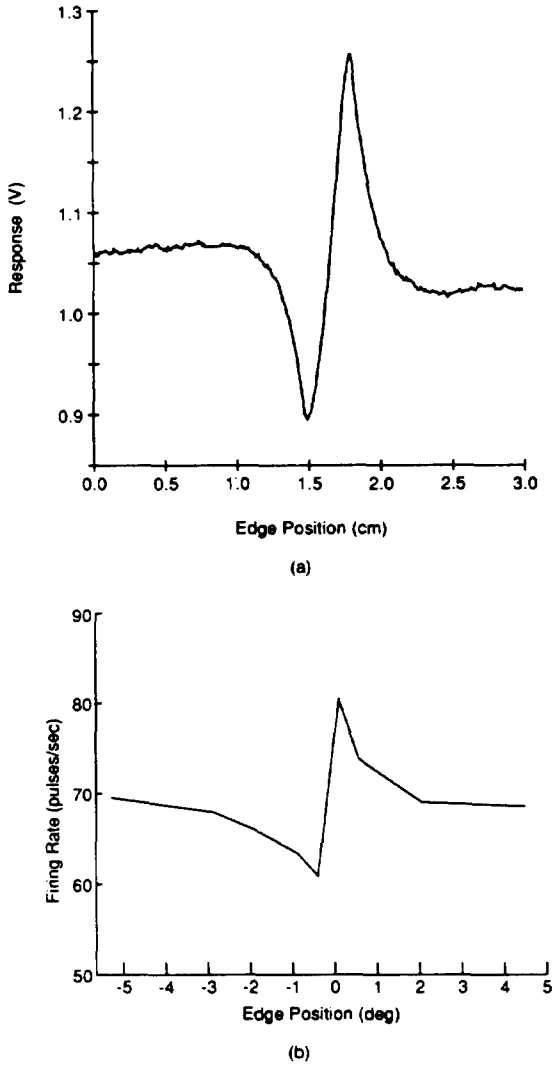

Figure 5: Spatial-derivative response of a retinal ganglion cell and a pixel to a contrast edge. The vertical edge was held stationary at different distances from the receptive field center. Contrast of the edge was 0.2 log units in both experiments. (a) Pixel output measured at steady state as edge was moved in increments of 0.01 cm at the image plane. Interpixel spacing corresponded to 0.11 cm at the image plane. (b) On-center C-cell of the cat. The contrast edge was alternately turned on and off. The average pulse density over the period 10-20 seconds after the introduction of the edge was measured for each edge position.

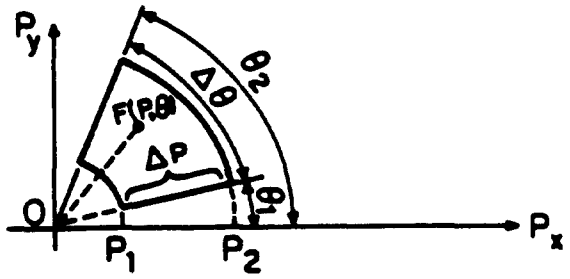

Figure 6: The p-space data collection in microwave imaging

where $\mathbf{f}$ is the object function (the state vector of the network), $\mathbf{F}_d(p)$ is the measured frequency response, $\mathbf{F}(p)$ is the Fourier transform of $\mathbf{f}$, and $R(\mathbf{f})$ is a regularization operator, chosen by *a priori* knowledge of the object to be reconstructed. Aerospace objects are generally smooth, so a regularization function of

$$R(\mathbf{f}) = \int \{f^2(r) + [\frac{df}{dr}]^2\} dr \qquad (2)$$

will keep the reconstruction smooth. The regularization parameter α controls the trade-off between fitness and smoothness of reconstruction.

Hopfield [6] introduced the idea of neural network processing being determined by a Hamiltonian, or energy function. The network here consists of a vector of processing elements, whose state of activation is $\mathbf{f}$. Each processing element is connected to all the others. We specify the connection weights between them to solve the problem at hand. In this network, the weights $T_{m,m'}$ do not change; they perform a real-valued Fourier transform on the current $\mathbf{f}$. A starting point on the energy surface is fixed by providing the partial frequency response data as the input to the network. This is the original activation $\mathbf{f}$ of the processing elements. The network then iterates to a solution, which will be the nearest minimum of the energy surface. Gradient descent is used to move towards this minimum. In order to decrease $H(\mathbf{f})$,

$$\Delta H = \frac{dH(\mathbf{f})}{df(m)}[f^{(j+1)}(m) - f^{(j)}(m)] < 0 \qquad (3)$$

The update from the jth to the $(j+1)$th value of the state vector is therefore chosen to be

$$\Delta f(m) = f^{(j+1)} - f^{(j)}(m) = -\lambda \frac{dH(\mathbf{f})}{df(m)} \qquad (4)$$

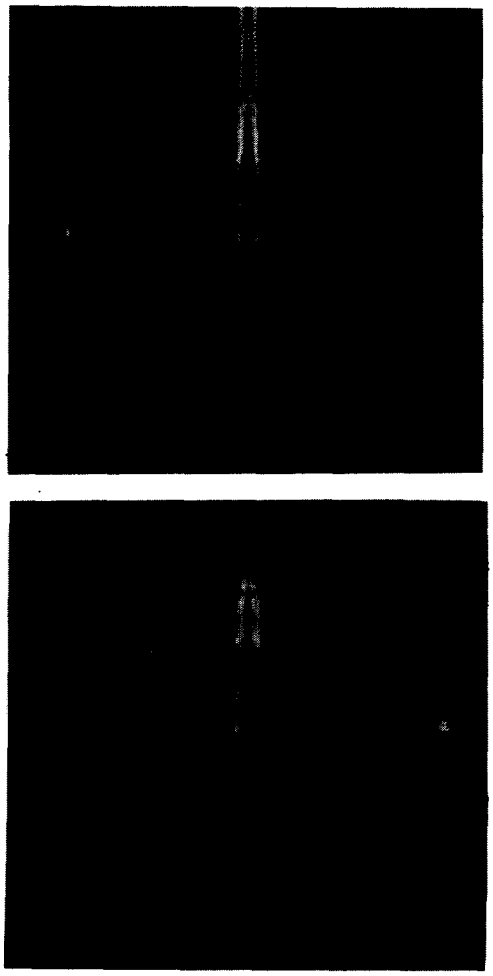

Figure 7: 2-D image reconstructed by FFT (top) and neural network processor (bottom)

Some algebra then leads to

$$\Delta f(m) = \lambda\{\sum_{m'=0}^{M} 2T_{m,m'} f^{(j)}(m') + I_m - S_m\} \quad (5)$$

where $T_{m,m'}$ is the matrix of connection weights that perform the underlying Fourier transform; I_m is the input data (transformed), and S_m is derived from the regularization function.

Results of this processing are shown in figure 7. In both the traditional and neural network cases, the computation to form the range profiles is followed by filtered backprojection. The neural net processor has improved signal-to-noise and resolution.

Two extensions to this are possible. First, the network can include a "neuromorphic" nonlinearity. As well as increasing the similarity to a biological network, this has the advantage of extending the regularization functions available, and further improves the image. Also, the image reconstruction can be made adaptive: as it proceeds, the regularization parameter α can be varied.

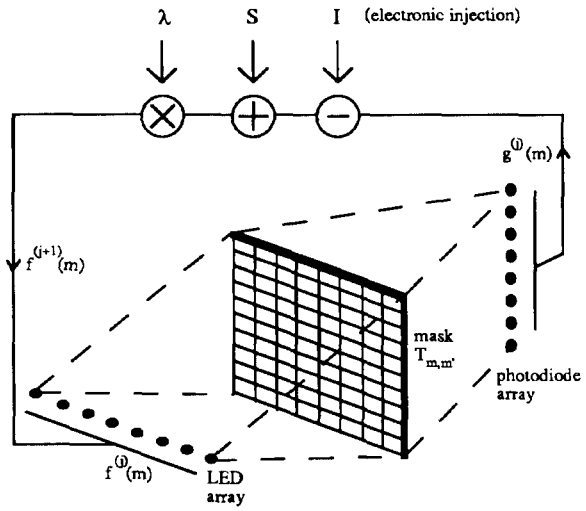

Figure 8: Opto-electronic implementation

The real-valued neural network computation lends itself to optical implementation. Figure 8 shows how this might be achieved. The neural state vector $\mathbf{f}$ is represented by the output of an LED array. The mask performs the matrix multiplication of $\mathbf{f}$ by $T_{m,m'}$. The result is received by a photodetector array. Feedback then occurs electrically, including the summation with S_m and I_m, and multiplication by λ. Farhat also describes a method for obtaining S_m optically.

5 Fingerprint recognition

This work tackles a hard pattern recognition problem using a neurally-inspired architecture. The data representation resembles that in the mammalian visual cortex, and the classifier is an adaptive network. It also illustrates many of the issues which must be overcome in order for neural networks to perform well. This is an example of the usefulness of Gabor filters for texture processing, and shows that neural networks can perform real-world feature extraction.

The image processing task is to extract the minutiae from noisy, grey-scale fingerprint images, as shown in figure 9. This is a harder task than classifying the 1-bit-

Figure 9: Grey scale fingerprint showing location of vertical minutiae

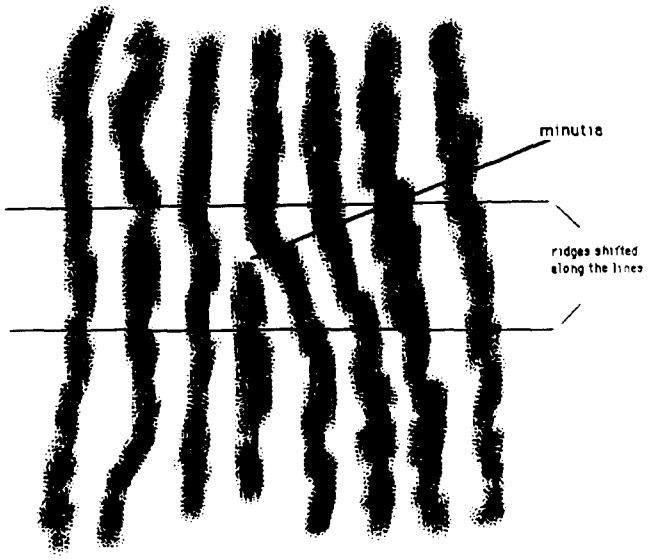

Figure 10: A small region of a fingerprint showing a vertical minutia. Phase changes between the two horizontal lines above and below the minutia are apparent.

deep nail to nail, rolled fingerprint commonly used by the police. These minutiae are one of the features used by humans in identifying people from their fingerprints. A minutia is a place where the fingerprint ridge ends or bifurcates (figure 10). It is necessary to extract the number, type, position and orientation of the minutiae.

These minutiae cause phase changes in a spatial frequency representation of the image. The obvious choice for a representation is a 2-D Fourier transform, but this would not preserve any spatial information. The Gabor transform is an optimal solution to the tradeoff between spatial and spatial-frequency resolution. The impulse response of the 2-D Gabor filter family is:

$$h(x,y) = g(x',y')exp(j2\pi Fx') \tag{6}$$

where (x',y') are rotated Cartesian coordinates $(x\cos\phi + y\sin\phi, -x\sin\phi + y\cos\phi)$, and

$$g(x,y) = ab\exp(-a^2x^2 - b^2y^2). \tag{7}$$

F is the spatial frequency and ϕ the orientation angle. a and b are scaling parameters. The 2-D Fourier transform of the filter is

$$H(u,v) = \pi\exp\{-\pi^2[\frac{(u-F)^2}{a^2} + \frac{v^2}{b^2}]\} \tag{8}$$

Daugman [7] has shown how a relaxation neural network similar to the previous example can be used to compute Gabor transforms efficiently. He has also demonstrated their power in texture recognition, since they can pick out orientation and frequency information with spatial resolution. Studies of the 2-D anisotropic receptive field profiles of neurons in the visual cortex show that this representation is used in the mammalian visual system.

In this application, the neural network is to act as a filter, operating on a kernel in the image. A window of 32 x 32 pixels would imply an input vector dimension of almost 1000. This is beyond the ability of current neural network algorithms to train. Therefore, a compressed representation is sought. First, consider the output of a Gabor filter (oriented in the vertical direction) to the image of figure 9. The area surrounding a minutia is composed of vertical lines which will produce large and equal signal amplitudes after filtering. The minutia itself does not match the convolution kernel, and will appear as a dark spot after filtering. There is also a phase change above and below the minutia, due to the introduction of an additional ridge. We therefore have two signatures for a minutia; these are shown in figures 11 and 12. The Gabor representation can be seen to preserve the characteristic features of the data.

The needed dimensionality reduction was obtained by demodulating the Gabor filtered output to baseband. This reduction in bandwidth permits some subsampling of the image, though care had to be taken in optimizing the subsampling strategy. Training and testing sets were made up from samples from minutiae in the vertical di-

Figure 11: Magnitude of 0 degree Gabor filtered fingerprint

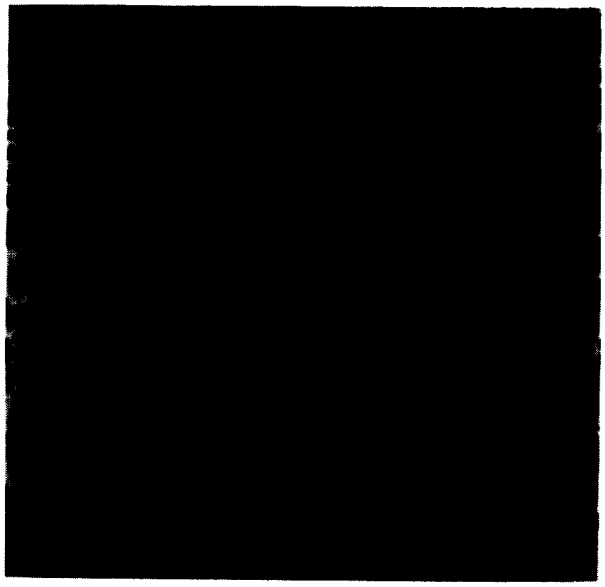

Figure 12: Phase of 0 degree Gabor filtered fingerprint

rection (50%), samples from minutiae in other directions (15%) and random samples (35%).

The in-phase and quadrature signals from the Gabor filter were fed into a three layer backpropagation network, and trained. Backpropagation is a method for training multi-layer perceptrons [8]. Poor generalization was found in this case. The reason is that the minutiae can be seen clearly in images of phase and magnitude, rather than I and Q. An additional layer in the network could have been trained to perform a transformation, but this would not have been an efficient choice. With the fledgeling neural network technology available, one should perform all processing which is possible algorithmically by standard methods, and produce the best possible representation for the network. It can then concentrate on the difficult cognitive task.

The network was therefore fed with baseband magnitude and phase directly, which resulted in a performance improvement. However, the false positive performance was still poor. Examination of the weights after training indicated that the network was only using the magnitude information. The next stage was to train two separate networks, one for magnitude and one for phase, logically ANDing the results. A different subsampling strategy had to be developed for the phase data (25 points rather than 5 points per kernel). Fine tuning of this kind further improved the results. The rest of the false positives were due to uneven lighting in the grey scale image. The d.c. components of the images were removed, and each kernel was scaled so that the deviation of the pixel values from the mean was the same. The results are shown in figure 13. When the two networks are combined, all the minutiae in the selected direction were identified. False positives were mainly caused by minutiae in nearby directions that were insufficiently attenuated by the Gabor filter. Comparisons with standard signal processing techniques showed that the method is effective. The image was processed using edge operators, threshold and thinned to extract the ridges. Minutiae were defined as places where the extracted ridges ended or split. The detection was found to be poorer, in both false positives and false negatives.

6 Conclusions

This paper has reviewed three applications of neural networks to image processing. The field is young, and somewhat prone to hyperbole at this stage. However, these examples indicate that good research is going on: these networks are solving interesting problems better than they had been solved before. Work has also been proceeding in ultrasonic applications. Winters [9] has used a Hopfield network to achieve superresolution with ul-

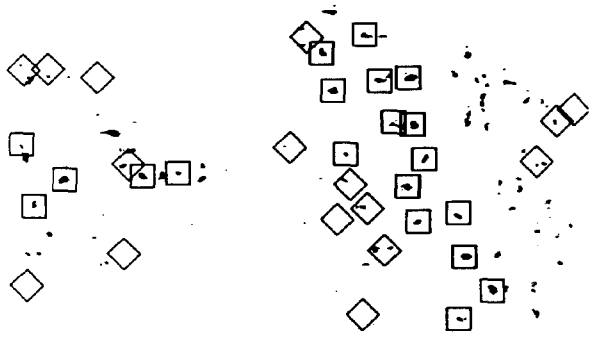

Figure 13: Final neural network scan of entire fingerprint. Black areas indicate that the neural network has detected a minutia; squares show the actual minutiae locations to be extracted. Diamonds are minutiae in adjacent directions, to which the neural network should be insensitive.

trasonic imaging in air. Yoneyama [10] has also worked on air imaging, training networks to recognize objects with robot vision in mind. Goldberg and Greenleaf [11] have used neural networks for texture analysis of mammographically indeterminant lesions in the breast. Nikoonahad [12] has applied neural networks to phase aberration correction, while the present author [13] has investigated the possibility of using them for ultrasonic tomography inversion. The neural network paradigm is so complementary to previous approaches to computing that we may soon see powerful hybrids of networks and algorithmic methods providing significantly more useful information processing.

7 References

1. R.P. Lippmann, "An introduction to computing with neural nets," *IEEE ASSP Magazine* April 1987, 4-22.

2. M. Minsky and S.A. Papert, "Perceptrons" (expanded edition). Cambridge, MA: MIT Press 1988.

3. C.A. Mead and M.A. Mahowald, "A silicon model of early visual processing," *Neural Networks* **1**, 91-98 (1988).

4. N.H. Farhat and B. Bai, "Echo Inversion and target shape estimation by neuromorphic processing," *Neural Networks* **2**, 117-125 (1989).

5. M-T. Leung, W.E. Engeler and P. Frank, "Fingerprint processing using back propagation neural networks," *International Joint Conference on Neural Networks*, **I**, 15-20 (IEEE publication number 90CH2879-5) (1990).

6. J.J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *Proc. Nat. Acad. Sci. USA* **79**, 2554-2558 (1982).

7. J.G. Daugman, "Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression," *IEEE Trans. Acoust. Speech Sig. Proc.* **36**(7), 1169-1179 (1988).

8. D.E. Rumelhart, G.E. Hinton and R.J. Williams, "Learning internal representations by error propagation," in *Parallel distributed processing: explorations in the microstructure of cognition* by D.E. Rumelhart, J.L. McClelland and the PDP Research Group. Cambridge, MA: MIT Press 1986.

9. J.H. Winters, "Superresolution for ultrasonic imaging in air using neural networks," *International Joint Conference on Neural Networks*, **I**, 609-615 (1988).

10. S. Watanabe and M. Yoneyama, "Ultrasonic robot eyes using neural networks," *IEEE Trans. Ultrason. Ferroelec. Freq. Control* **37**(3), 141-147 (1990).

11. V. Goldberg and J.F. Greenleaf, "Improvement in the specificity of US in the diagnosis of tumors in the breast by means of artificial intelligence," 76th Meeting of the Radiological Society of North America (1990).

12. M. Nikoonahad and D.C. Liu, "Medical ultrasound imaging using neural networks," *Electron. Lett.* **26**(8), 545-546 (1990).

13. B.C. Conrath, C.M.W. Daft and W.D. O'Brien, Jr. "Application of neural networks to diffraction tomography," *Proc. 1989 IEEE Ultrasonics Symposium*, 1007-1010 (IEEE publication number 89CH2791-2).