ARTIFICIAL NEURAL NETWORKS FOR AUTOMATIC TARGET RECOGNITION

Steven K. Rogers
Dennis W. Ruck
Matthew Kabrisky
Gregory L. Tarr
Department of Electrical and Computer Engineering
Air Force Institute of Technology
AFIT/ENG, Wright-Patterson AFB, OH 45433

Published in Proceedings of SPIE Conference on Applications of Artificial Neural Networks, Orlando, Florida April 1990 (Paper Number 1294-01)

Abstract

This paper will review recent advances in the applications of artificial neural network technology to problems in automatic target recognition. The application of feedforward networks for segmentation, feature extraction, and classification of targets in Forward Looking Infrared (FLIR) and laser radar range scenes will be presented. Biologically inspired Gabor functions will be shown to be a viable alternative to heuristic image processing techniques for segmentation. The use of local transforms, such as the Gabor transform, fed into a feedforward network is proposed as an architecture for neural based segmentation. Techniques for classification of segmented blobs will be reviewed along with neural network procedures for determining relevant features. A brief review of previous work on comparing neural network based classifiers to conventional Bayesian and K-nearest neighbor techniques will be presented. Results from testing several alternative learning algorithms for these neural network classifiers are presented. A technique for fusing information from multiple sensors using neural networks is presented and conclusions are made.

1 Introduction

Autonomous object recognition is an active area of interest for military and commercial applications. The Air Force Institute of Technology (AFIT) has been researching this area for the past twenty-five years. This paper reports on the recent work in the application of artificial neural network technology to problems in automatic target recognition (ATR). Given an input image from an infra-red or range sensor, the problem is to find interesting objects in those images and then classify those objects according to their type. For example, to find the tanks, trucks and jeeps as opposed to rocks, trees and hills. This is sometimes called the target/non-target problem. More specific classification problems, where targets types must be determined will also be considered. This problem is related to the general purpose robotic vision problem, but without the common constraints, such as controlled range, light or even aspect, available to those systems. Throughout this paper results are presented as found using a neural network development environment. This environment is the subject of a companion paper and was found to be critical to advances in artificial neural network applications to the target recognition problem. The system allows the display of the images and the workings of the networks to educate the users and allow creative insights to the workings of the system for subsequent modifications. The next section will address segmentation of images followed by a discussion of feature extraction. Classification alternatives are then presented followed by a neural network sensor fusion system. Experimental results and conclusions are presented. The authors do not necessarily feel that ATR systems must be all neural networks, but can in fact be hybrid systems with existing or other emerging technologies combined to make an automatic target recognizer. With this philosophy many of the neural network results will be shown wedged in between other technologies. The hope is to find the advantages/disadvantages of artificial neural networks in an ATR system.

Ex.1015 / Page 1 of 11
TESLA, INC.

2 Segmentation

Visual cortex, area 17 or sometimes called V1, can be modeled as a multiple image display screen which contains many images. In V1 there are the color images of the world, so called RGB images, texture images and also motion maps. Using these multiple cues, mammals can segment the world. We will ignore arguments that some mammals don't have color vision, such as dogs, since the argument is basically a matter of degree. A reasonable quality color system seems to have been retrofitted onto the lower mammal visual system for primates. It should be noted that many infra-primates have better color systems than we do. In either case the use of color, texture and motion to find lumps in the world is a cue that automatic target recognition systems should heed. For now let's restrict our attention to the texture analysis system. In the section on fusion, ideas that have an obvious relation to using multiple colors will be presented as well as ideas on how to combine these multiple views into a single or multiple competing hypotheses.

The texture map can be modeled as a Gabor function. Figure 1 shows some examples of Gabor functions. The reason some people believe that the cortical texture map can be modeled as a Gabor representation is because experiments have shown the receptive fields match to a high degree an appropriate parameterized Gabor function [4]. This work is not in complete disagreement with the earlier work of Hubel and Wiesel which proposed the existence of bar detectors [3]. At a recent neural network conference, IJCNN 90, Hubel still insisted that the appropriate model was bar-detectors and not smoothed bars as could be explained with Gabor functions. The use of Hough transform techniques which can be thought of as breaking the image up into its component bars is also being investigated and will be reported on separately in an article on processing synthetic aperture radar images. Let's return to the idea of a Gabor function as a model of how a neuron's receptive field is modulated. If these Gabor functions are correlated with an input image, the pixels in the resulting correlation image are a measure of the similarity of that part of the image with that Gabor function. If the output images from these correlations are then combined, a segmented image results. Figure 2 shows an example FLIR scene and the result of correlating four Gabor functions with that scene and then adding the resulting correlation images and thresholded. This is the simplest of the ideas of combining these correlation images. We have also tested a gating idea where only the maximum amplitude pixel from all of the correlations is maintained in the output image. The figure clearly shows this technique can segment blobs from the original FLIR image. Figure 3 shows a similar FLIR image segmented using conventional image processing techniques that work on the histogram of the pixel values. The quality of the segmentation is similar with the advantage going to the artificial neural network solution for ease of implementation and execution speed. Specifically, the Gabor transform can be implemented in an artificial neural network [1] or even in an optical system [8] which might be called an optical neural network.

As a side point many current researchers are developing front end vision systems that purport to model biological vision systems. It seems that a driving factor in this research is the ability of these models to reproduce common illusions. In the early 1960's Kabrisky proposed using low frequency Fourier transforms for understanding much of human visual information processing [5]. It was later shown by one of his students that common illusions could be explained with this Fourier model [2]. Recent work at AFIT has shown that the Gabor transform can also account for these common illusions [6]. Much of the rationale of these systems is to faithfully reproduce the type of processing that is going on in animals. For example, at a recent conference it was said that animal visual systems must be able to complete partially occluded boundaries because in the retina the light gets blocked by the blood vessels and interconnections of neurons prior to detection by the rods and the cones at the back of the eye. This logic shows a fundamental misunderstanding of how the visual system works. All static information is lost in the transmission of the image information back to the cortex. Anything frozen in the image, external or internal to the eye, is not mapped onto the cortex. Sacadic movements prevent external objects from being frozen in the image unless some perverted measures are taken by an experimental psychologist in order to study vision. The bottom line is even if such a phenomenon were going on in the visual system, why should an artificial system incorporate processing which by these researchers own arguments only exists to

Ex.1015 / Page 2 of 11
TESLA, INC.

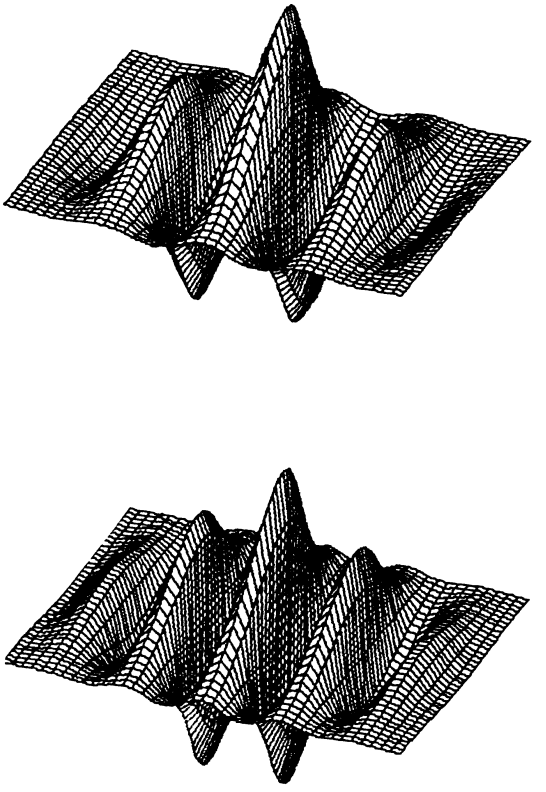


Figure 1: Gabor Function Examples

overcome the present implementation in the wetware?

Let's continue our design of the general purpose front end system. Implementing these local transforms in neural networks, either Gabors or Fourier or even Houghs, some means of determining whether the local area should pass a segmentation step must be accomplished. We propose a system that would take the transform information from little local windows and feed that information into a feedforward network. This network could be either an unsupervised or a supervised network. In the unsupervised case lots of images are passed through the system and the nodes organize to respond according to the density of the input in the transform spaces used. A subsequent calibration of the self-organized network could determine the classification of the nodes as representing regions that should pass the segmentation step. Currently we are testing our variations of the Kohonen Learning Vector Quantization techniques.

For the supervised network case, the local transform information is fed to a network along with a user classification of that region of the input image. By the supervisor feeding enough data through the network it should converge, learning to discriminate regions of interest from similar images. Techniques to determine which of the transforms are important to the segmentation process can be determined either after the training or during learning [9].

3 Feature Extraction

As the old adage goes, Good Features make Good Recongizers. This is true whether your recognizer is using an artificial network or whether it is using some statistical based decision mechanism. The purpose of this section is to demonstrate that from a set of features input to either a conventional or a neural

Ex.1015 / Page 3 of 11
TESLA, INC.

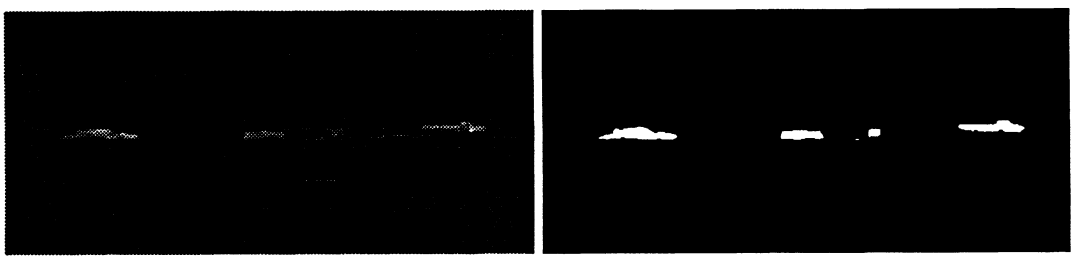


Figure 2: Original FLIR and Gabor Segmentation

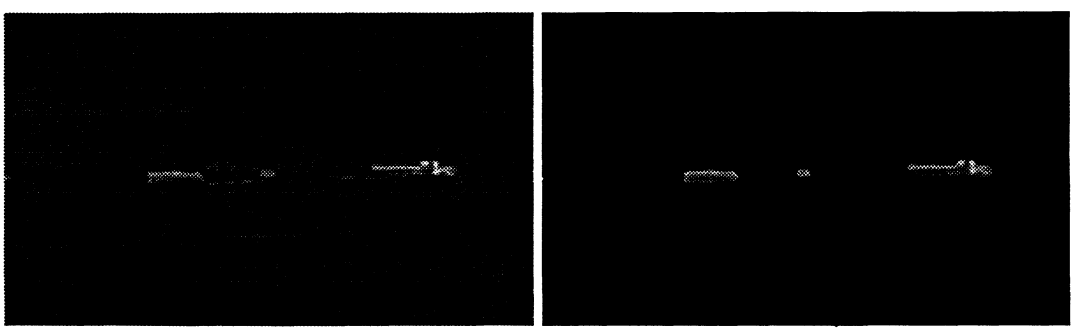


Figure 3: Conventional Heuristic Segmentation

Ex.1015 / Page 4 of 11
TESLA, INC.

Table 1: FLIR Features Evaluated

Feature	Description
Complexity	Ratio of border pixels to total object pixels
Length/Width	Ratio of object length to width
Mean Contrast	Contrast ratio of object's mean to local background mean
Maximum Brightness	Maximum brightness on object
Contrast Ratio	Contrast ratio of object's highest pixel to its lowest
Difference of Means	Difference of object and local background means
Standard Deviation	Standard deviation of pixel values on object
Ratio Bright Pixels/Total Pixels	Ratio of number of pixels on object within 10% of maximum brightness to total object pixels
Compactness	Ratio of number of pixels on object to number of pixels in rectangle which bounds object

Table 2: Ranking of FLIR Features

Feature	P_e Criterion Rank	Saliency Measure Rank
Complexity	8	8
Length/Width	7	7
Mean Contrast	6	3
Maximum Brightness	5	2
Contrast Ratio	4	4
Difference of Means	3	6
Standard Deviation	2	5
Ratio Bright Pixels/Total Pixels	1	0
Compactness	0	1

network based recognizer the same level of accuracies can be obtained and in fact the features are used in a similar manner.

The last statement should be qualified somewhat. If we use a conventional statistical based classification criterion, such as Bayesian, then some determination can be made on how important each of the input features are to reducing the probability of error. Specifically, using the set of features in Table 1, and ranking those features by individually testing the probability of error if only that feature is used, the features are ordered as shown. The question becomes when these same features are used by an artificial neural network for classification which are the important features. By using a simple definition for saliency, or importance, of a given feature we have shown that the artificial neural network finds basically the same importance of features [9]. The determination of the importance of a given feature turns out to be independent of the network starting conditions and consistent with the order of importance determined by the minimum probability of error used to rank importance by the Bayesian technique. Table 2 shows the features used for a FLIR classification problem and the order of importance as determined by the statistical approach and the artificial neural network classifier.

Further testing using only the most important features for classification demonstrated that these are

Ex.1015 / Page 5 of 11
TESLA, INC.

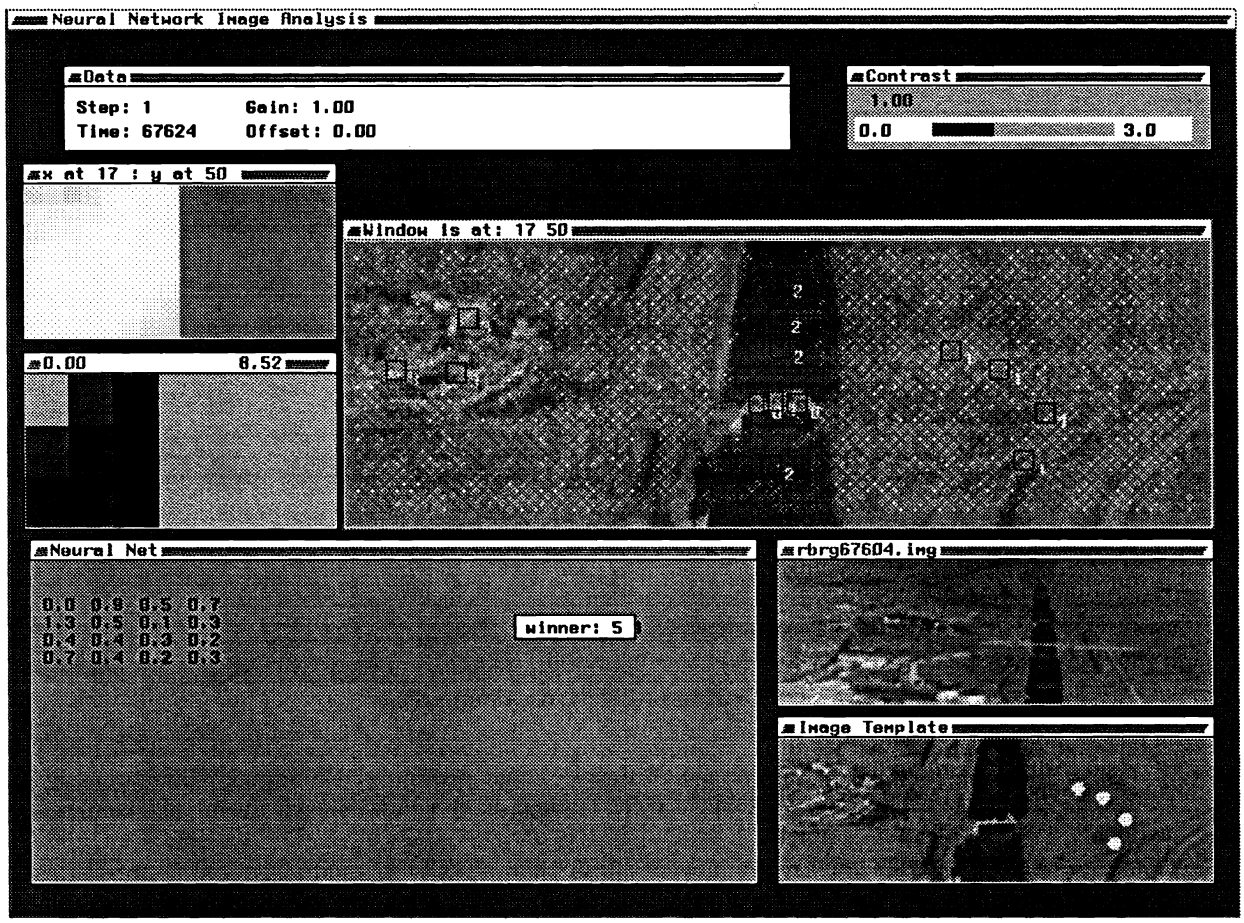


Figure 4: Neural Segmentation/Feature Extraction

in fact the important features. It also demonstrated that irrelevant of which variation of training rule that was used to learn, the same relative order of importance was found. Variations of the training rules is discussed in the next section. Even though the investigation that demonstrated the importance of the features to an artificial neural network was conducted after training, the technique could be incorporated during learning to automatically structure the network including the number of inputs.

The features shown above came from conventional techniques of identification of objects. These features were extracted with conventional image processing ,algorithms. Could a neural network be used on the front end to process the incoming image and automatically extract the features as well as segment? Since in the first section we proposed a neural based segmentation using the local transforms that could be implemented in artificial neural networks, could these coefficients then be used for classification by either a neural network or conventional classifier? Research into using the byproducts of the neural based segmentation, the local transform coefficients, can now be conducted since we can now determine which of the features, the coefficients, are important for classification via our saliency metric. An early version of that type of system is demonstrated below in Figure 4.

In Figure 4 a FLIR scene is shown with small boxes that represent areas where some teacher has denoted classification for segmentation. The system then computes local transforms, Hough/Fourier/Gabor, and a network determines which type of area is in that region of the picture. In the development system a color

Ex.1015 / Page 6 of 11
TESLA, INC.

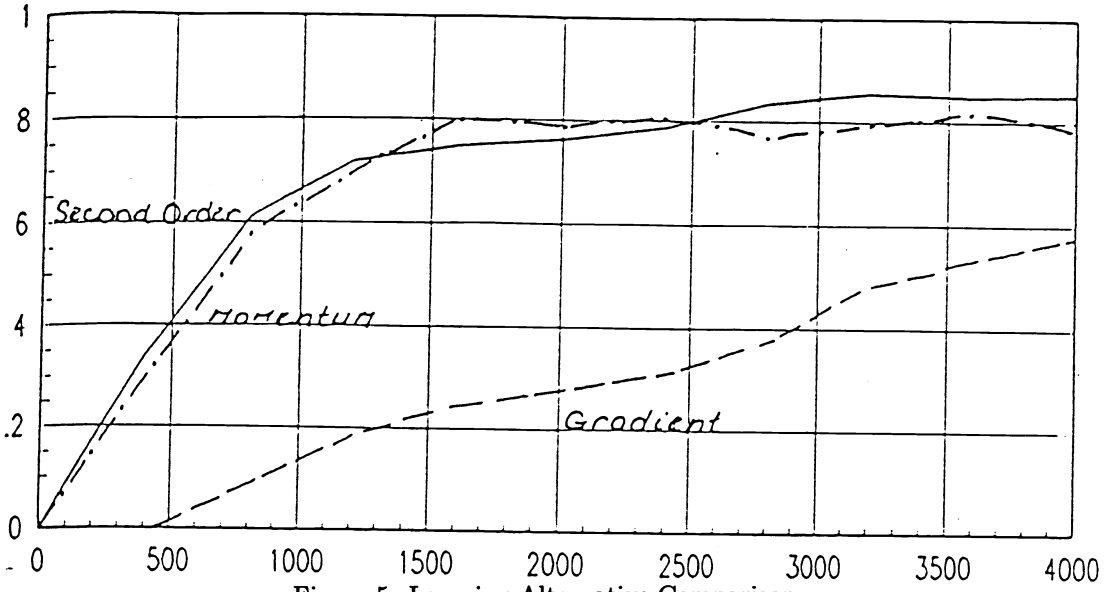


Figure 5: Learning Alternative Comparison

coded dot is used to show the operator the classification of that portion of the input. Unfortunately in this paper, the figure cannot be displayed, but a color transparency will be shown during the conference presentation.

We have now shown how artificial neural networks can be applied to segmentation and feature extraction. By far the most advanced application of artificial neural networks is in the area of classification. Since we have published several articles in this area, a brief review of this work will be provided emphasizing the recent efforts on alternatives to the common learning algorithms leading to the following section on sensor fusion.

4 Learning Alternatives

The most common technique for setting weights in a multilayer feedforward network is by the backpropagation algorithm. This algorithm was originally developed by Werbos as part of his PhD dissertation [12] and later rediscovered by others [7, 11]. This technique is a gradient search on an error surface. A variation of that technique is the use of second derivative information when updating weights. Another variation is the use of what is commonly called the momentum term. If all of these algorithms are tested on the same set of data how would the accuracies compare as functions of the number of presentations of the training data. Figure 5 shows the application of these variations on a FLIR target recognition problem.

Note in Figure 5 that both the second order and the momentum algorithms outperform the conventional backpropagation algorithm. This begs the question what are these algorithms and how do they all compare? Recently we were able to show a relationship between the common backpropagation algorithm and an extended Kalman filter being used to train the weights. What this work did was demonstrate that backpropagation is just a degenerate extended Kalman filter that has made assumptions about relations between weights [10]. Therefore if one tries both an extended Kalman filter and the conventional backpropagation on a set of data one would expect that the backpropagation would be bounded by the Kalman

Ex.1015 / Page 7 of 11
TESLA, INC.

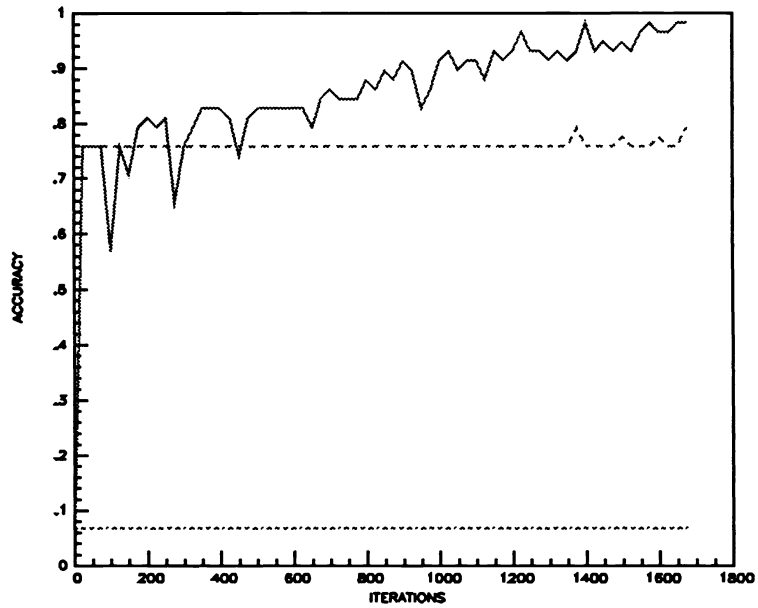


Figure 6: Tactical Data Training Classification Accuracy *vs* Iterations (Solid line: Kalman; Dashed line: Back propagation; Dotted line: Momentum method).

results. Figure 6 shows the application of both algorithms on a laser radar classification problem. All three methods asymptotically approach the same accuracy level. The Kalman method, however, achieves the limiting level with fewer iterations.

Before we jump to conclusions on the use of extended Kalman filtering for training weights, it must be noted that even though the Kalman algorithm learns faster in terms of number of presentations of the data it takes much longer in terms of numbers of computations. Second when both techniques were applied to the absolute range recognition problem the improvement is not observed, see Figure 7.

5 Sensor Fusion

In this section the problem of using artificial neural networks to fuse information from both the FLIR and the absolute range data sensor is discussed. Specifically, how well does the artificial neural network do compared to a Bayesian based fusion technique. The ability of an artificial neural network to fuse information from multiple sensors will be compared to k-nearest neighbor and non-parametric Bayesian techniques. The problem addressed is one of target versus non-target discrimination. A typical FLIR scene was shown before and a typical laser range image is shown in Figure 8.

Table 3 shows the features used for the fusion comparison. The way the fusion process works is by choosing a dominant sensor. This sensor extracts the features in the table below for blobs that pass its segmentation procedure. The dominant sensor then looks into the segmented image of the nondominant sensor to see if a corresponding blob exists. If it does then the features for that sensor are also extracted. If a segmented blob does not exists in the nondominant sensor, then those inputs are set to zero in the neural network for fusion. Table 4 shows the improvement achieved by fusing information from the multiple sensors. Also note the artificial neural network accuracy compared to conventional fusion algorithms.

Ex.1015 / Page 8 of 11
TESLA, INC.

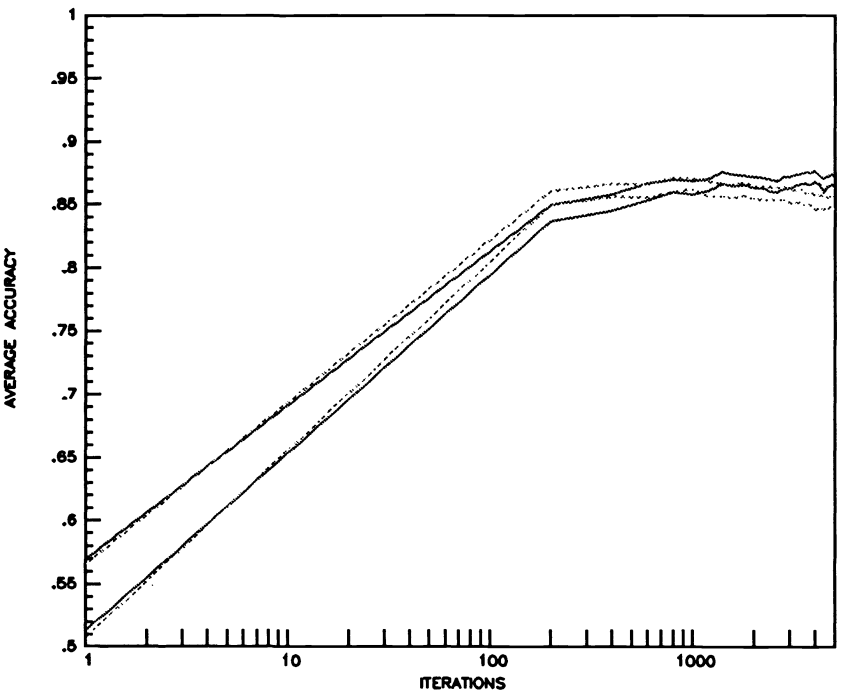


Figure 7: Average Testing Accuracy for Target/Non-target Problem on Absolute Range Data - 90% confidence (solid: back propagation; dotted: Kalman).

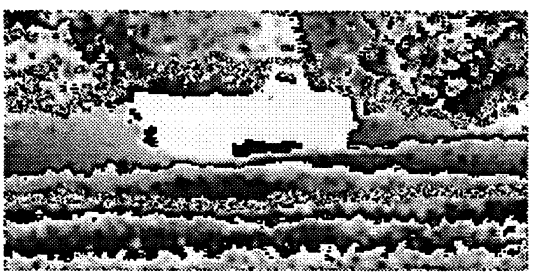


Figure 8: Absolute Range Image

Ex.1015 / Page 9 of 11
TESLA, INC.

Table 3: Features used for multisensor fusion

Feature	Description
FLIR Features	
Number of Pixels	Number of pixels on the blob
Background Std Dev	Standard deviation of local background pixels
Complexity	Ratio of border pixels to total object pixels
Absolute Range Features	
Complexity	Same as above
Pixel Standard Deviation	Standard deviation of pixel values for blob
Height	Height of blob

Table 4: Range Dominant Accuracy Results

Classifier	Range Only Accuracy	Range+FLIR Accuracy
	(95% Confidence Intervals)	
Multilayer Perceptron (6-5-2)	82%-88%	93%-97%
k Nearest-Neighbor	82%-88%	94%-98%
Non-parametric Bayesian	82%-90%	94%-98%

6 Conclusion

This paper has provided some information on the recent advances in applying artificial neural networks to problems in automatic target recognition. Ideas on use of this technology in segmentation, feature extraction and results in use for classification and sensor fusion were presented. Preliminary results in segmentation and feature extraction were presented and quantitative comparisons were made for learning alternatives and classification.

7 Acknowledgements

This work was supported by funding from the Rome Air Development Center, Griffiss AFB New York, RADC/IRRA, Lt Greg Alhquist.

References

[1] J.G. Daugman. Complete discrete two dimensional gabor transforms by neural networks for image analysis and compression. *IEEE Transactions on ASSP*, 36(7), July 1988.

[2] A. Ginsburg. Specifying relevant spatial information for image evaluation and display design: An explanation of how we see certain objects. In *Proceedings of the Society of Information Display*, volume 21.

[3] D. Hubel and T. Wiesel. Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology*, 195, 1968.

[4] J.P. Jones and L.A. Palmer. An evaluation of the two dimensional gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, 58(6), December 1987.

[5] M. Kabrisky. *A Proposed Model for Visual Information Processing in the Human Mind.* University of Illinois Press, 1966.

[6] R. Obendorf. Gabor transforms as a model of what we see. Master's thesis, AFIT, 1990.

[7] David B. Parker. Learning-logic. Invention Report S81-64, Stanford University, Stanford, CA, October 1982.

[8] S. K. Rogers, K. Ayer, M. Kabrisky, and M. Oxley. Forward looking infrared image segmentation and pattern recognition using gabor transform and joint transform correlation techniques. *Optical Engineering*, Winter 1990. Submitted for publication.

[9] D. W. Ruck, S.K. Rogers, and M. Kabrisky. Feature selection using a multilayer perceptron. *The Journal of Neural Network Computing*, 1990. Accepted for Publication to appear in Fall issue.

[10] D.W. Ruck, S.K. Rogers, P.S. Maybeck, and M. Kabrisky. Back propagation: A degenerate kalman filter? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1990. Submitted for Publication.

[11] David E. Rumelhart, James L. McClelland, and the PDP Research Group. *Parallel Distributed Processing*, volume 1: Foundations. MIT Press, Cambridge, Mass., 1986.

[12] P.J. Werbos. *Beyond Regression: New Tools for Prediction and Analysis in the Behovioral Sciences.* PhD thesis, Harvard University, 1974.

Ex.1015 / Page 11 of 11
TESLA, INC.