Neural Network for Analysis of Hyperspectral Imagery

Nikzad Benny Toomarian, Hamid Kohen
Jet Propulsion Laboratory
California Institute of Technology

Nahum Gat
Opto-Knowledge Systems Inc.

Abstract

Each material has unique spectral distribution that can be viewed as its "fingerprint". Advanced hyper and ultra spectral sensors will produce large amount of data. Therefore, advanced capabilities are required that can effectively deal with large "data cubes", possibly in real time. We have shown that unsupervised Kohonen type self-organized feature map can be used for hyperspectral image clustering and dramatic data reduction. All available spectral bands were used in the computation without any preprocess the data. We used AVIRIS data from Moffett Field, California to demonstrate convincing qualitative results. For our quantitative analysis we generated and used synthetic data sets. With the help of the ground truth, we demonstrated that all the networks tested identify all the classes correctly. However, in some cases, each class was represented by more than one neuron, a practical problem. A few possible avenues to overcome this problem are mentioned.

1. INTRODUCTION

Each material has unique spectral distribution that can be viewed as its "fingerprint", hence its importance in space exploration. Current sensors such as AVIRIS has only 224 channels of data. Future sensors will produce even higher data channels. Obviously, transmitting such amount of data to earth with current deep-space communication bandwidth is prohibitive. Processing of such vast amounts of data, especially in real time, are computer intensive. Therefore, there is a clear need for advanced image-processing software and hardware that can effectively deal with data sets often larger than 150 Mbytes. The general goal of our activities, has been to access the applicability of neural networks to hyperspectral image analysis and address the issues of training speed. The focus of this work is to provide an unsupervised neural network for classification of hyperspectral images. In the next section, we will review very briefly what hyperspectral data is. Section 3 will provide motivation for using neural networks by review some of the work in the are of neural network as applied to multi-spectral data analysis. Section 4 will summarize the self-organized feature map employed. We will discuss our simulation results in section 5. We conclude, in section 6, by outlining our ideas for improving the results.

2. WHAT IS HYPERSPECTRAL DATA?

Sunlight energy is continually reflected from our Earth's surface. Our eyes collect and our brains sense this visible reflected energy as colors such as green leaves, red crop or white snow. However, much of this reflected energy is invisible and called infrared. Sophisticated camera-like tools called imaging spectroradiometers collect the entire spectrum of reflected energy. Such tools are deployed in airplanes and, more recently, in satellites. Currently, data are available mainly from sensors mounted on aircraft like AVIRIS. Figure 1.0 represents a single pixel spectrum from AVIRIS. The x-axis is channel wavelength in micrometers. The y-axis is radiance, usually expressed in units of microwatts per square centimeter per nanometer per steradian.

The light curve of the Sun and the absorption features of the atmosphere dominate the general shape of an AVIRIS spectrum, see Figure 1.0 The Sun has a "blackbody" curve, which peaks in the green wavelengths and diminishes at higher and lower wavelengths. The atmosphere absorbs light at wavelengths that correspond to the absorption wavelengths of the components of the atmosphere: nitrogen, oxygen, carbon dioxide, water, and other elements. For example, the deep valleys that go down to near zero around 1.4 and 1.9 microns are due to water absorbing those wavelengths.

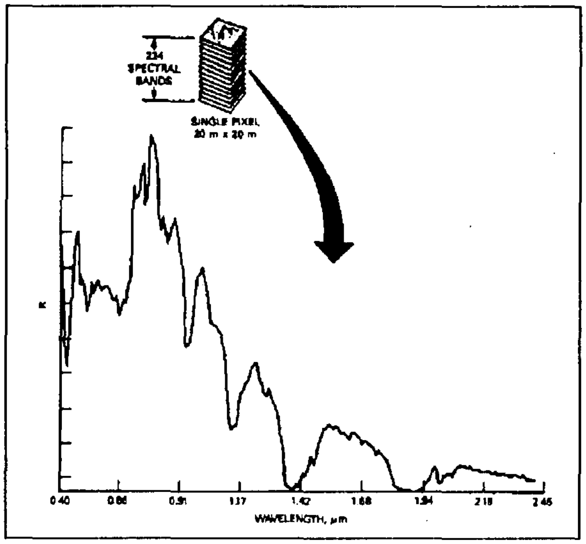

Figure 1.0 AVIRIS Single Pixel Spectrum

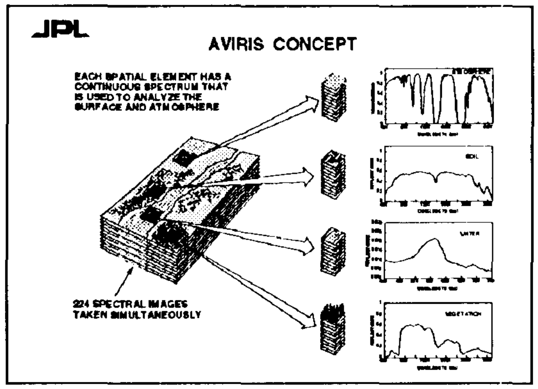

Figure 2.0 AVIRIS Concept

The peaks and valleys of a spectrum not due to the Sun or the atmospheres reveal information about the chemical composition of the pixel being examined. Every substance has its own spectrum, and one can look for those features from those spectra in the AVIRIS pixel spectra. Even living things have spectra. Green plants, for example, use chlorophyll to absorb the visible light from the sun, but reflect the infrared radiation. This manifests as a large jump in the spectra in the area where the red light (0.7 microns) merges into the infrared. The spectra presented above shows just such a "red edge", indicating that the pixel was showing vegetation. Figure 2.0 demonstrates the AVIRIS concept.

AVIRIS has only 224 channels of data. TRW's HyperSpectral Imager-HIS on-board failed Lewis & Clark spacecraft had 348 spectral bands. Future sensors that are based upon Fourier Transform Spectrometer, AKA ultraspectral, will produce even higher data dimensionality. Obviously, transmitting such amount of data to earth with current deep-space communication bandwidth is prohibitive. Processing of such vast amounts of data, especially in real time, are computer intensive. Therefore, there is a clear need for advanced image-processing software and hardware that can effectively deal with "data cubes", such as one shown in Figure 3.0, that are often larger than 150 Mbytes.

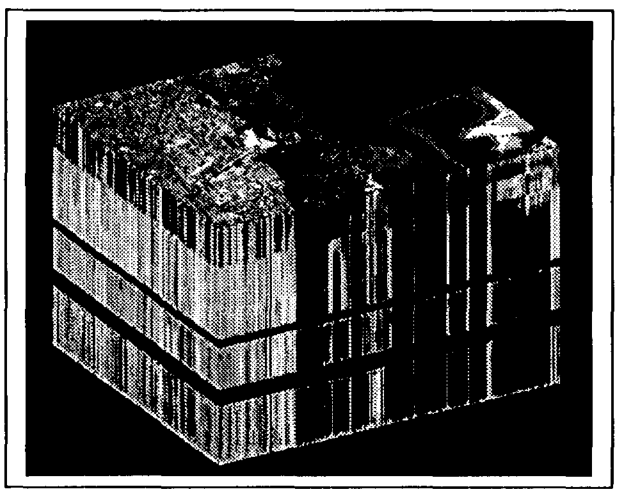

Figure 3.0 AVIRIS Moffett Field Image Cube

3. WHY NEURAL NETWORKS?

There are several premises that advocate neural networks as the computational framework for the analysis of hyperspectral data. Hertz, Krogh and Palmer (1991) have shown that many standard statistical classifiers are special case of neural networks. For instance, Yair and Gersho (1990) have pointed out that maximum a-posteriori classifiers (i.e., classifiers that choose the class with the highest aposteriori probability) are a special case of Bolzmann perecetron network. Yau and Manary (1990) have shown the equivalence between Gaussian classifiers and "sigma-pi-" networks. Ruck (1990) suggested that multilayer percepton networks provide an excellent approximation to a Bayes optimal discriminant function. In addition, Benediktsson (1990) has found neural networks are distribution free that can detect and exploit nonlinear data patterns with superior to statistical methods in terms of classification accuracy. This is an advantage over statistical methods, particularly when there is no knowledge of the statistical distribution functions of the data. Herman and Khazenie (1992), demonstrated that neural networks perform better or equal to conventional statistical classifiers on multispectral remote sensing data.

Hybrid inversion methodologies, whereby differences between measured and model-predicted reflectance are minimized, have shown very promising results. In these studies (Smith (1993) and Davis (1993)) neural networks were trained using a multiscatering model to accurately estimate parameters of interest from Landsat data. However, it is in the area of classification of multispectral images that neural networks have encountered their greatest success (Hermann (1990), and Bischof (1992)). In the remote sensing community, the question of how well neural network models perform as classifiers, has generated considerable interest (Benediktsson (1990), Hepner (1990), and Fitzgerald (1994)). For example, Bischof, Schneider and Pinz, (1992) have used feedforward backpropagation for remote sensing applications. They found during feedforward neural networks learning phase for large scale, real-data Hyperspectral image classification are slow but once training is completed, classification of new data is, quasi instantaneous. However, those who deal with large data volume, have practically given up on the use of neural networks due to their inherent long training time.

The general goal of our activities, in the past, has been to access the applicability of neural networks to hyperspectral image analysis and address the issues of training speed. In this vein, novel supervised neural networks classification techniques were developed, tested and compared to parametric statistical classification methods using specially developed performance benchmarking data sets. In general, neural networks have been demonstrated to match or outperform statistical methods. The following are some of the steps taken:

1. A Generalized Eigenvalue technique was developed for transforming hyperspectral data to domain in which maximum separability exist between signature classes. This will reduce the dimensionality of the data sets by more than an order of magnitude and improve classification performance.
2. A subnetwork technique was devised for performing hyperspectral image classification. Each subnet is trained to identify one class and reject all other existing classes. This procedure dramatically reduces the computing time and improves classification performance.
3. A recurrent neural network with enhanced learning algorithm has been adapted to address hyperspectral data analysis. This fully connected network, configured as a system of subnetworks was shown to perform very well on the benchmarking datasets.
4. A special hyperspectral classification benchmarking technique was developed, based upon synthetic signatures and various "noise" and "clutter" models. The use of synthetic "signatures" provides at most certainty in validating results.

Since our previous work was focused on supervised neural networks, the current work is emphasizing use of neural networks in an unsupervised mode. At this stage, we assume no spatial correlation between neighboring pixels. We further assume that each pixel is fully resolved (i.e., pixels with the same spectral distribution contain the same material). The goal of this work is to provide an unsupervised neural network for classification of hyperspectral images. No attempted was made to preprocess the data to extract any specific features or reduce the data size. We have used all 224 spectral bands although we know some bands provide very little, if any, extra information.

4. SELF ORGANIZING MAP

There are two different approaches to unsupervised learning.

- Grossberg, Carpenter and coworkers (1987-1992), base one method upon Adaptive Resonance Theory-ART. This approach, has the novel ability of performing controlled discovery of new clusters. ART while accommodating creation of new clusters it is not affecting the storage or recall capability of learned clusters.
- Kohonen (1995) base the second approach to unsupervised learning on self-organizing feature map.

The Kohonen self organized feature map, due to its simplicity, was chosen for this study. The principle philosophy of Kohonen algorithm is a mapping from the input data space $\mathbf{R}^n$ (in this case n is equal to number of spectral bands) onto a regular array of nodes that preserve the essential content of the information. This array of nodes though, in general, can be multidimensional, in practice, it is two-dimensional (**i=1,2,...,I & j=1,2,...,J**). Each input vector **x** member of $\mathbf{R}^n$ is fully connected to every node **(i,j)**, via a synaptic weight vector $\mathbf{w_{i,j}}$ member of $\mathbf{R}^n$. The grid of nodes has lateral connections and defines the topology of the network. The training can use any distance measure (for example Euclidean), to determine which node has a weight vector that is closer to the input vector. The node with closest weight vector is the winner. During the learning process, the weights of the winner node as well as those nodes that are topographically close to the winner, up to certain distance, will be modified. Thus the grid of nodes that can be square, hexagonal, cubic, or other geometrical shapes is organized into local neighborhoods that act as feature classifiers of the input data. The underlying principal here is to move the weight vectors so that they more closely align with the input vector. The algorithm used in this work has the following form:

1. Based upon the input data, the weight vectors, $\mathbf{w_{i,j}}$, were randomly initialized.
2. Randomly select and present an input vector **x** to the network.
3. Calculated the inner product of the input vector **x** with each weight vector. Select the node with the largest
4. Modify the value of the weights for the winning node as well as its neighbors as follows:

$$w_{ij}(t+1) = w_{ij}(t) + \eta(t)h(t)[x(t) - w_{ij}(t)]$$ **Equation 1**

Where **t** is the discrete-time coordinate, **h(t)** is the neighborhood kernel and η is the learning rate. The neighborhood kernel, **h(t),** is monotonically decreasing function of time and is usually defined in terms of distance from the winning node. One example neighborhood kernel that was used in the calculations is of the following Gaussian form:

$$h(t) = e^{(-r/2\sigma(2)^2}$$ **Equation 2**

Where **r** is the distance from the winning node to the neighboring ones. Hence, this distribution is centered on the winning node.

5. The width σ as well as the learning rate η are slowly reduced during the training.
6. If the changes in the weight vectors are significantly small, stop.
7. Increment the time, **t = t+1**, go back to step 2.

5. SIMULATION RESULTS

In our attempt to apply unsupervised neural networks to hyperspectral imagery, we will use a Kohonen based self-organizing feature map (http://nucleus.hut.fi/nnrc/nnrc-programs.html). This should partition an image cube into individual classes based upon the similarity among of the spectral bands of each pixel. Our classifier architecture takes all the spectral signature of a pixel, as input vector, and feed it to the network. This input vector is mapped onto a two-dimensional array of nodes as described in previous section. Figure 4.0 depicts the follow of the data from the Hyperspectral image cube into neural networks system for classification. Once the training is finished, the complete image is feed to the network for classification. The output is a two-dimensional image, with each pixel correspond to the same pixel from the input image and its content indicating the class number.

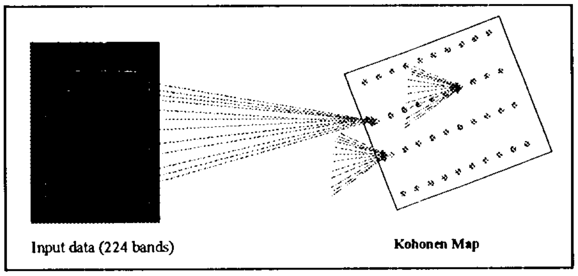

Figure 4.0 Neural Network Architecture

5.1 Moffett Field Data

To test the architecture outlined above, we have obtained and used the AVIRIS data from Moffett Field, California, Figure 3.0. Extensive parameter (e.g. number of classes and training time, training rate, output topology, neighborhood radius parameters, etc.) analysis was performed to obtain optimal parameters. Table 1.0 represents some of the cases in which Kohonen network parameters were manipulated. In all the cases, a hexagonal topology was chosen. The second column of Table 1.0, indicates the size of the network. Obviously, this will dictate the maximum number of classes one can get. For example, in scenario 9, assuming each neuron representing one class, we could have up to 75 classes. The first columns under the heading of Training Length, Learning Rate and Radius indicate the number of pixels and the initial learning rate and radius used in the first stage of the learning or "ordering", respectively. Ordering is customarily performed on relatively small (10% of the whole image) amount of the data with large values for learning rate and neighborhood radius. The second stage of learning is fine tuning the weights obtained from the ordering stage. Therefore, a large amount data point, small values of learning rate and radius are used. The second columns under the heading of Training Length, Learning Rate and Radius indicate the number of pixels, initial values for learning rate and radius, respectively.

Table 1.0 Moffett Field Image Cube 614 x 512 x 244

Scenario	Size of the Network	Training Length # of Pixels		Learning Rate		Radius		Quantified Error
2	5 x 5	10000	628736	0.05	0.02	10	1	2270
3	5 x 5	628736	3143680	0.05	0.02	10	1	1429
4	5 x 5	628736	3143680	0.05	0.02	3	0	1400
5	5 x 5	628736	3143680	0.05	0.01	13	5	1300
6	5 x 5	628736	0	0.05	0.01	13	5	2300
7	10 x 5	100000	314368	0.19	0.09	7	5	1000
8	15 x 5	10000	3143680	0.1	0.01	10	3	700
9	15 x 5	10000	3143680	0.1	0.009	10	5	1655
10	20 x 15	314368	7859200	.3	.01	40	1	569
11	8 x 32	314368	943104	0.3	0.01	40	1	595
12	20 x 15	314368	7859200	.3	.01	7	1	900
13	8 x 32	314368	943104	0.3	0.01	16	1	810
14	30 x 15	2000	0	.9	0	1	0	650

Let us focus on one of the cases in the Table 1.0, i.e., scenario 15 as a representative case. As indicated before, all 224 spectral bands of each pixel were used as input to the network. After the training, each input vector will be classified in one of the 75 possible classes and will be labeled with the class number. To

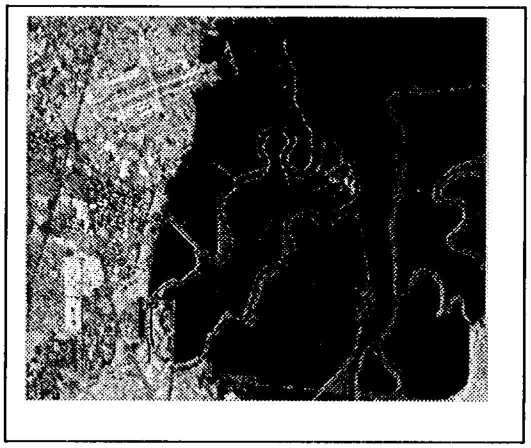

Figure 5.0 Visual Moffett Field Image (Color)

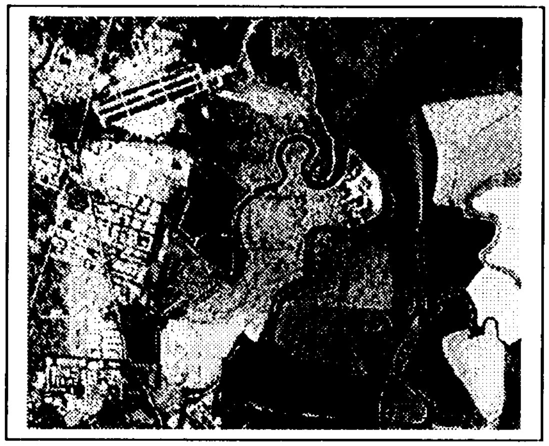

Figure 6.0 Classified Moffett Field Image (Gray)

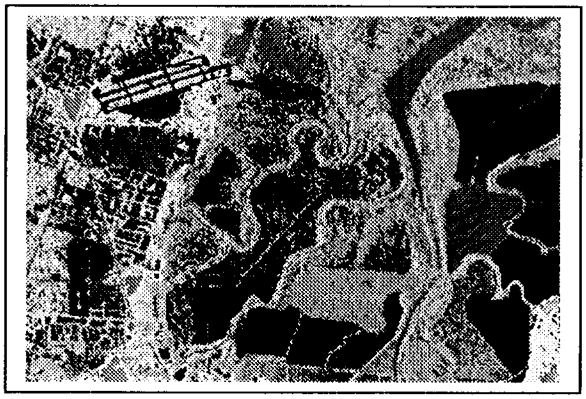

Figure 7.0 Classified Moffett Field Image (Color)

qualitatively demonstrate the results, each pixel will be assigned, based upon the class number it belong to, either a gray scale between 0 to 255 as in Figure 6.0, or alternatively an RGB color code as in Figure 7.0. The first visible spectral band of the image cube is repeated in Figure 5.0 for comparison. Visual inspection and comparison of Figures 6.0 and 7.0 with Figure 5.0, indicate a fairly good classification. However, not only this is a qualitative point of view, but also comparing only to one spectral band out of 224.

In order to perform any objective and quantitative evaluation of the methodology, one has to have ground truth data. Since there is no known ground truth data for Moffett field, it will hard if not impossible to determine the accuracy of the results. For example, no knowledge of actual number of classes that truly exists in Moffett field make it difficult to assess if the above-mentioned scenario over estimating or under estimating the number of classes and by how much. Based on observation of true color Moffett field image and for our test purposes, we have assumed a maximum of 75, 125, 256 and 456 different classes within the Moffett field data, as shown in Table 1.0.

5.2 Synthetic Data

To circumvent the ground truth problem we generated synthetic signatures for validation of the approach. We have introduced a small size known ground truth synthetic data sets. Presently, we have created 16 classes with 256×256 spatial pixels each, in 64 spectral bands. Each class represented by a spectral curve generated using an analytical mathematical expression. We have full control in selecting these expressions and in injecting any desired level [and distribution] of noise into the data for test purposes. I this study, however, we have added (either 2% or 5% of signature value at each band) normally distributed noise to the data.

The synthetic data set in addition to having ground truth is smaller. Therefore, it is much easier and faster to perform parametric analysis. In this spirit, we have perform extensive sensitivity analysis of the results with respect to different network parameters, such as learning rate, neighborhood size, network topology and size of training samples. We have generated Tables, similar to Table 1.0, for image cubes without noise, with 2% and with 5% noise. In the following, we will summarize the important aspect of our results.

Without noise

- The learning rate, the radius and number of pixels used for training did not have great impact on the number of classes correctly identified by the network.
- Network size had most important role in the performance of the network. The smaller the network (e.g. 3X5) or elongated rectangle (e.g. 1X16, or 2X16) the worst the results. Large networks (e.g., 6X8, 10X18) perform very well in identifying all 16 classes correctly.

With 2% & 5% noise

- The above conclusions of case without noise hold in this case as well.
- Knowing the ground truth help us to conclude that most of the networks tested identify all 16 classes correctly. However, each class was represented by more than one neuron. Results of one of our worst cases, in which a 32X8 size network was used, is depicted in Figures 8.0 and 9.0 for 2% and 5% noise, respectively. In this case, some classes occupy as little as two neurons, while others occupy as much as 30 neurons. This fact makes the task of identifying classes without knowing the ground truth or performing any post-processing impossible.

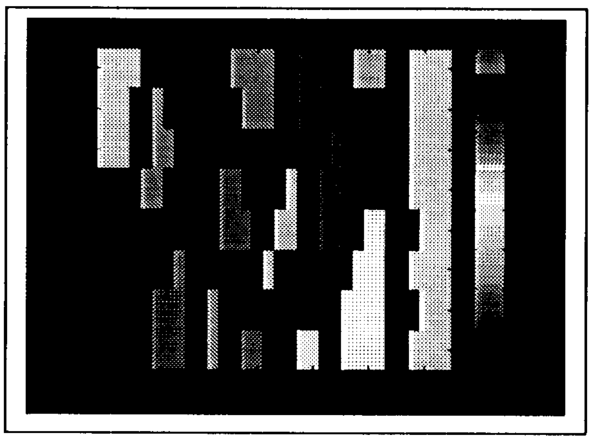

Figure 8.0 Classified Synthetic Image Cube with 2% Noise

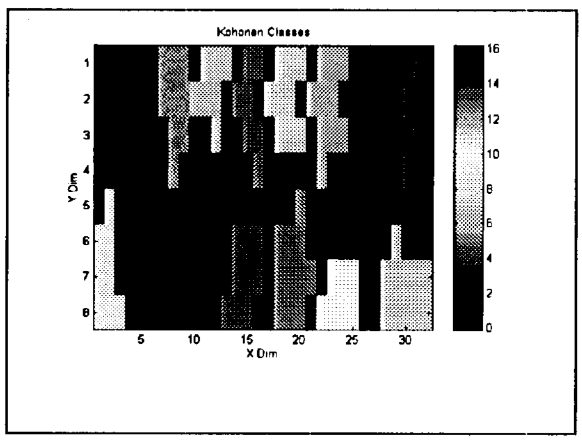

Figure 9.0 Classified Synthetic Image Cube with 5% Noise

6. CONCLUSION

Each material has unique spectral distribution that can be viewed as its "fingerprint", hence its importance in space exploration. AVIRIS has only 224 channels of data. Future sensors will produce even higher data channels. Obviously, transmitting such amount of data to earth with current deep-space communication bandwidth is prohibitive. Therefore, advanced capabilities are required that can effectively deal with large "data cubes", possibly in real time. The general goal of our activities, in the past, has been to access the applicability of neural networks to hyperspectral image analysis and address the issues of training speed. In this vein, novel supervised neural networks classification techniques were developed, tested and compared to parametric statistical classification methods using specially developed performance benchmarking data sets. In general, neural networks have been demonstrated to match or outperform statistical methods.

The goal of this work is to provide an unsupervised neural network for classification of hyperspectral images. The Kohonen self organized feature map, due to its simplicity, was chosen for this study. No attempted was made to preprocess the data to extract any specific features or reduce the data size. We have used all available spectral bands although we know some bands provide very little, if any, extra information. At this stage, we assume no spatial correlation between neighboring pixels. We further assume that each pixel is fully resolved.

Two type of data sets were used for our study.

- AVIRIS data from Moffett Field, California. Since, in this case, there is no ground truth data we were able to demonstrate only qualitatively our results. These results to a naked eye are very convincing.
- Synthetic generated data sets using analytical mathematical expressions. Extensive parametric sensitivity analysis was performed. The learning rate, the radius and number of pixels used for training had marginal effect on the number of classes correctly identified by the network. In cases where noise was added to the data, the ground truth help us to conclude that most of the networks tested identify all the classes correctly. However, each class was represented by more than one neuron. This fact makes the task of identifying classes without knowing the ground truth or performing any post-processing impossible

In conclusion, we have demonstrated that, in general, neural networks are ucfull toll for hyperspectral data analysis. In particular, we have shown that unsupervised, Kohonen type self-organized feature map, can be used for hyperspectral image clustering and dramatic data reduction. We further suggest that if one is interested in automated system, without human in the loop or supervised learning, one or a combination of the following actions be taken:

1. Perform preprocessing on the crude data. Application of principal component analysis, for example, will transform the hyperspectral data to domain in which maximum separability exist between signature classes. This will also reduce the dimensionality of the data sets and improve classification performance.
2. Post-processing of the data to eliminate the many neurons per class, or
3. Modify the Kohonen type algorithm from the ground up in such a way that prevents such possibilities.

All of these approaches are currently under investigation in our group. Results of these investigations are the subject of our future publication.

ACKNOWLEDGEMENTS

The research described in this paper was performed at the Center for Integrated Space Microsystems, Jet Propulsion Laboratory, California Institute of Technology. It was sponsored by Opto-Knowledge Systems Inc. as part of Small Business Technology Transfer- STTR program, Department of Energy and the National Aeronautics and Space Administration.

7. REFERENCES

Bischof, H., W. Schneider and Pinz, A.J. (1992), "Multispectral Classification of Landsat Images Using Neural Networks", IEEE Trans. Geosc. & Rem. Sens., 30 (3), 482-490.

Benediktsson, J.S., Swain, P.H. and Ersoy, O.K. (1990)," Neural Network Approaches Versus Statistical Methods in Classification Multisource Remote Sensing Data". IEEE Trans. Geosc. & Rem Ens., 28(4), 540-551.

Carpenter, G. A., Grossberg, S. (1987), "Computer Vision Grapghic Image Processing", 37:54-115.

Carpenter, G. A., Grossberg, S. (1987), "Applied Opt.", 26:4919-30.

Carpenter, G. A., Grossberg, S., Rosen D. B. (1991), "Neural Networks", 4:493-504.

Carpenter, G. A., Grossberg, S. (1990), "Neural Networks", 3:129-52.

Carpenter, G. A., Grossberg, S., Reynold, J. H. (1991), "Neural Networks", 4:565-88.

Carpenter, G. A., Grossberg, S., Rosen D. B. (1991), "Neural Networks", 4:759-71.

Carpenter, G. A., Grossberg, S., Masrkuzon N., Rosen D. B. (1992), IEEE Trans. Neural Networks, 3:698-713.

Davis, D.T. et al. (1993), "Retrieval of Snow Parameters by Iterative Inversion of a Neural network", IEEE Trans. Geosc. & Rem. Sens., 31 (4), 482-851.

Gersho, A. and Gray, R.M. (1992), Vector Quantization and Signal Compression, Boston: Kluwer Academic Publishers.

Hepner, G.F. et al. (1990), " Artificial Neural Network Classification Using a Minimal Training Set: Comparison to Conventional Supervised Classification", Photogram. Eng. & Rem. Sens., 56 (4), 469-473.

Hermann, P.D. and N. Khazenie (1992), " Classification of Multispectral Remote Sensing Data Using a Back-Propagation Neural Network", IEEE Trans. Geosc. & Rem. Sens., 30 (1), 80-88.

Hertz, J., A. Krogh and R. G. Palmer (1991), "Introduction to the Theory of Neural Computation, Addison-Wesley, Redwood City, CA.

Kohonen, T (1984), Self-Organization and Associative Memory, Berlin: Springer-Verlag.

Kohonen, T (1988), "Learning Vector Quantization," Neural Networks, 1 (suppl 1), 303.

Kohonen, T. (1995), Self-Organizing Maps, Berlin: Springer-Verlag.

Ruck, D.W. et al. (1990), "The Multilayer Preceptron as an Aproximation to Bayes Optimal Discriminant Function", IEEE Trans. Neuro. Netw., 1(4), 296-298.

Smith, J.A. (1993), "LAI Inversion Using Back-Propagation Neurral Network Trained with Multiple Scattering Model," IEEE Trans. Geoec. & Rem. Sens., 31 (5), 1102-1106.

Yair, E. and Gersho A. (1990), "Maximum a Posteriori Decision and Evaluation of Class Probabilities by Bolzmann Preceptron Classifiers", Procs. IEE, 78 (1), 1620-1628.

Yau, H.C. and Manary M.T. (1990), "Iterative Improvement of a Gaussian Classifier," Neural Networks, 3, 437-443.