Land-Cover Classification of Multispectral Imagery Using a Dynamic Learning Neural Network

K. S. Chen, Y. C. Tzeng, C. F. Chen, and W. L. Kao

Abstract

The results of the classification of SPOT high resolution visible multispectral imagery using a neural network are presented. The test site, located near Taoyuan in northern Taiwan, is in an agricultural area containing small ponds, bare and barren soils, vegetation, built-up land, and man-made buildings near the sea shore. The classifier is a dynamic learning neural network (DL) using the Kalman filter technique as its adaptation rule. The network's architecture consists of multilayer perceptrons, i.e., feed-forward nets with one or more layers between the input and output nodes. Selected data sets from 512- by 512-pixel three-band images were used to train the neural nets to classify the different types of land cover. Both simulated and real images were used to test classification performance. Results indicated that the DL substantially reduces the training time, compared to the commonly used back-propagation (BP) neural network whose slow training process prevents it from being used in certain practical applications. As for classification accuracy, the results were excellent. We concluded that the use of a dynamic learning network provides promising classification results in terms of training time and classification rate. In particular, the proposed network significantly improves the practicality of land-cover classification.

Introduction

The increased application of neural networks to remote sensing has emerged in recent years and continues to receive attention. This is probably due to the following characteristics of a neural network (Lippmann, 1987; Hush and Horne, 1993; Haykin *et al.*, 1991): (1) It has an intrinsic ability to generalize; (2) it makes weaker assumptions about the statistics of the input data than a parametric Bayes classifier; and (3) it is capable of forming highly nonlinear decision boundaries in the feature space and therefore has the potential of outperforming a parametric Bayes classifier when feature statistics deviate significantly from the assumed Gaussian statistics.

The comparison of traditional statistical methods (such as the *k*-nearest neighbor or the Gaussian classifier) made by Benediktsson *et al.* (1990) indicates that a neural network has the potential for processing multisource remote sensing data. Heermann and Khazenie (1992) used a back-propagation neural network to classify multispectral image data and concluded that a neural network is a feasible classifier for very large volume images. A similar study was also carried out by Bischof *et al.* (1992) and showed that the neural network outperforms the maximum-likelihood method. These studies most commonly used back-propagation learning scheme.

Nevertheless, all these applications have the drawbacks of the slow learning process that are inherently associated with a back-propagation training scheme. In this paper, we shall use a newly developed dynamic learning neural network to perform the classification of the SPOT HRV image. Two separate simulated images are constructed resembling a real image with known grey levels at each class; one is used for training the network and the other for testing. The methods for the generation and simulation results are presented in the following section. The next section illustrates the experimental result using a SPOT image where the ground truth was collected when it was imaged, to show the classifying capability of the presented network. Finally, concluding remarks are given.

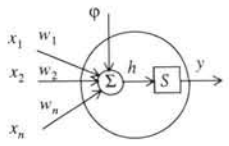

Figure 1. Modeling of a neuron.

Network Test

A Dynamic Learning Neural Network

The general configuration of a feed-forward multi-layer precepton (MLP) neural network is defined in this section. An MLP neural network consists of one input layer, one output layer, and one or more hidden layers in between. Each layer contains a number of nodes and a neuron, which is the basic element of a neural network. A neuron is modeled as shown in Figure 1.

The activation h and the output signal y are obtained by

$$h = \mathbf{w}^T \mathbf{x} + \varphi \quad (1)$$

Center for Space and Remote Sensing Research, National Central University, Chung-Li, Taiwan, Republic of China.

Photogrammetric Engineering & Remote Sensing, Vol. 61, No. 4, April 1995, pp. 403–408.

0099-1112/95/6104–403$3.00/0

$$y = S(h) \tag{2}$$

where the superscript T denotes the transpose operation, $\mathbf{x} = [x_1 x_2 \ldots x_n]^T$ is an n by 1 input vector, $\mathbf{w} = [w_1 w_2 \ldots w_n]^T$ is an n by 1 input vector, φ is an additive bias, and S is an activation function. The most commonly used activation function is the sigmoid function which is defined as $S(h) = 1 / (1+e^{-h})$.

A fully connected feed-forward MLP neural network is shown in Figure 2. It is assumed that the output signals from layer i can only be fed into layer $i+1$. There are n input nodes at the input layer, m output nodes at the output layer, and p layers of the hidden layer having n_i nodes at hidden layer i. The input of the neural network $\mathbf{x} = [x_1 x_2 \ldots x_n]^T$ is an n by 1 input vector and the output of the neural network $\mathbf{y} = [y_1 y_2 \ldots y_m]^T$ is an m by 1 output vector. To simplify the formulation, we let $n = n_0$ and $m = n_{p+1}$, so that we have $\mathbf{x} = S(\mathbf{h}_0)$ and $\mathbf{y} = S(\mathbf{h}_{p+1})$. The output from a set of input signals in each layer can be obtained according to the following recursive formula for $i=1,2, \ldots, p+1$:

$$\mathbf{h}_i = \mathbf{W}_i S(\mathbf{h}_{i-1}) + \mathbf{\Phi}_i \tag{3}$$

where $\mathbf{\Phi}_i = [\varphi_{i1} \varphi_{i2} \ldots \varphi_{in_i}]^T$ is an n_i by 1 bias vector, $\mathbf{h}_i = [h_{i1} h_{i2} \ldots h_{in_i}]^T$ is a n_i by 1 activation vector, $S(\mathbf{h}_i) = [S(h_{i1})\ S(h_{i2}) \ldots S(h_{in_i})]^T$ is an n_i by 1 signal vector, and $\mathbf{W}_i = [\mathbf{w}_{i1} \mathbf{w}_{i2} \ldots \mathbf{w}_{in_i}]^T$ is an n_i by n_{i-1} weight matrix. While the 1 by n_{i-1} weight vector in layer i is defined as $\mathbf{w}_{ij} = [w_{ij1} w_{ij2} \ldots w_{ijn_{i-1}}]$. The notations are defined as

- φ_{ij} : bias of the j-th node in layer i,
- h_{ij} : activation of the j-th node in layer i,
- $S(h_{ij})$: output signal of the j-th node in layer i, and
- w_{ijk} : weight connected between the j-th node at layer i and the k-th node in layer $i-1$.

After two modifications, (1) every node in the input layer and in all the hidden layers are fully connected to the output layer, and (2) the activation function is removed from each output node. The output of the modified network can be characterized as the weighted sum of the polynomial basis vectors as

$$\mathbf{y} = \underline{\mathbf{W}}\mathbf{x} \tag{4}$$

where $\underline{\mathbf{W}} = [\underline{\mathbf{w}}_1 \underline{\mathbf{w}}_2 \ldots \underline{\mathbf{w}}_m]^T$ is an m by M weight matrix. While the 1 by M long weight vector is defined as $\underline{\mathbf{w}}_k = [\mathbf{w}_{0k} | \mathbf{w}_{1k} | \ldots | \mathbf{w}_{ik} | \ldots | \mathbf{w}_{pk} | \varphi_{p+1,k}]$; the 1 by n_i vector is defined as $\underline{\mathbf{w}}_{ik} = [\underline{w}_{i1k} \underline{w}_{i2k} \ldots \underline{w}_{in_ik}]$; and $\underline{w}_{ijk}$ is the weight connected between the j-th node in layer i and the k-th node in the outer layer. It is also noted that the output of the MLP, Equation 4, is expressed as a linear function of the output weight vector, $\underline{\mathbf{W}}$, so that linear equations can be solved, thus minimizing the training error.

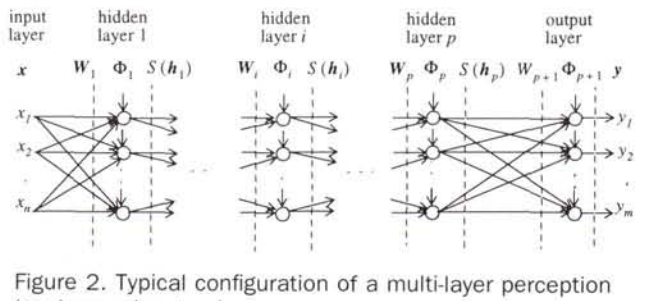

Figure 2. Typical configuration of a multi-layer perception (MLP) neural network.

Now that the network representation is complete, the training of the network employing the use of the Kalman filtering technique is in order. The Kalman filtering process is a recursive minimum mean-square estimation procedure (Brown and Hwang, 1983). Each updated estimate of the neural network weight is computed from the previous estimate and the new input data. The weight connected to each output node can be updated independently. The original problem can therefore be decomposed into m scalar problems for $k = 1, 2, \ldots, m$: i.e.,

$$y_k = \underline{\mathbf{w}}_k \underline{\mathbf{x}}. \tag{5}$$

By using the Kalman filtering technique, Equation 5 can be modeled in the form

$$y_k^j = \underline{\mathbf{w}}_k^j \underline{\mathbf{x}} + v_k^j \tag{6}$$

$$\underline{\mathbf{v}}_k^{j+1} = \underline{\mathbf{w}}_k^j \mathbf{A}^j + \mathbf{u}_k^j \mathbf{B}^j \tag{7}$$

where the superscript j denotes the added j-th training pattern, $\boldsymbol{A}^j$ is an M by M state transition matrix, $\boldsymbol{B}^j$ is an M by M diagonal matrix, $\boldsymbol{u}_k^j$ represents a 1 by M process error vector, and v_k^j is a scalar measurement error. Note that M is the number of hidden notes. The process error and the measurement error are assumed to be statistically independent and can be modeled as the zero mean, a white noise process whose correlations are given as

$$E[\mathbf{u}_k^{jT} \mathbf{u}_k^j] = \begin{cases} \mathbf{Q}_k^j & i = j \\ 0 & i \neq j \end{cases} \tag{8}$$

$$E[v_k^j v_k^j] = \begin{cases} r_k^j & i = j \\ 0 & i \neq j \end{cases} \tag{9}$$

$$E[\mathbf{u}_k^j v_k^j] = 0 \quad \text{for all } i \text{ and } j. \tag{10}$$

The update of the network weights is made according to the following recursions:

for $j = 1, 2, \ldots, N$

$$\hat{\underline{\mathbf{w}}}_k^j = \tilde{\underline{\mathbf{w}}}_k^j + \mathbf{g}_k^j [d_k^j - \tilde{\underline{\mathbf{w}}}_k^j \underline{\mathbf{x}}] \tag{11}$$

$$\tilde{\underline{\mathbf{w}}}_k^{j+1} = \hat{\underline{\mathbf{w}}}_k^j \mathbf{A}^j \tag{12}$$

where $\tilde{\underline{\mathbf{w}}}_k^j$ is the one step predicted estimate and $\hat{\underline{\mathbf{w}}}_k^j$ is the filter estimate of $\underline{\mathbf{w}}_k^j$, respectively, and $\mathbf{g}_k^j$ is the computed Kalman gain. The computed Kalman gain can be viewed as an adaptive learning rate and is computed according to the following steps:

$$\mathbf{g}_k^j = (\tilde{\mathbf{P}}_k^j \underline{\mathbf{x}})^T [\underline{\mathbf{x}}^T \tilde{\mathbf{P}}_k^j \underline{\mathbf{x}} - r_k^j]^{-1} \tag{13}$$

$$\hat{\mathbf{P}}_k^j = \tilde{\mathbf{P}}_k^j - \mathbf{g}_k^{jT} (\tilde{\mathbf{P}}_k^j \underline{\mathbf{x}})^T \tag{14}$$

$$\tilde{\mathbf{P}}_k^{j+1} = \mathbf{A}^{jT} \hat{\mathbf{P}}_k^j \mathbf{A}^j + \mathbf{B}^{jT} \mathbf{Q}_k^j \mathbf{B} \tag{15}$$

where $\tilde{\mathbf{P}}_k^j = E[(\underline{\mathbf{w}}_k^j - \tilde{\underline{\mathbf{w}}}_k^j)^T (\underline{\mathbf{w}}_k^j - \tilde{\underline{\mathbf{w}}}_k^j)]$ and $\hat{\mathbf{P}}_k^j = E[(\underline{\mathbf{w}}_k^j - \hat{\underline{\mathbf{w}}}_k^j)^T (\underline{\mathbf{w}}_k^j - \hat{\underline{\mathbf{w}}}_k^j)]$ are the one-step predicted and filter estimate error covariance matrices, respectively. The U-D factorization method (Brown and Hwang, 1983) is used to calculate the Kalman gain (Equation 13 through Equation 15) for reasons of stability.

To simplify the implementation of the Kalman filtering technique, we assume that $\mathbf{A}^j = \mathbf{B}^j = \mathbf{I}$ where $\mathbf{I}$ is a unit ma-

trix, r_k^i is a small positive number, $\mathbf{Q}_k^i = \sigma^2\mathbf{I}$ where σ^2 is an assigned variance of the process error, the initial network weights $\underline{\mathbf{w}}_k^1$ are set to be small random numbers and the initial one-setp predicted error covariance matrix is $\tilde{\mathbf{P}}_k^i = E[\underline{\mathbf{w}}_k^{1T}\underline{\mathbf{w}}_k^1]$.

Generation of Simulated Images

To test the performance of the dynamic learning (DL) neural network, we used a test image with known parameters, including the mean and variance of the grey levels for three bands in each class such that the classification accuracy can be more easily evaluated. It should be noted that the simulated images should be generated in such a way that they retain a geometrical complexity similar to the actual multispectral image. The test image was designed to rule out the possibility of ideal situations where the neural network may perform extremely well but have little practical value.

Two images were generated for the purpose of a network test. One is used for training, and the other for testing the classification. First, we seleclted a 512- by 512-pixel frame from an arbitrary SPOT image, and then we applied an ISODATA clustering algorithm (Ball and Hall, 1967) separating it into several classes (eight classes in this study). Next, we assumed that the distribution of the grey level was normal with predetermined mean and variance.

The assignments of the mean and variance in each class and each band are listed in Table 1. Notice that the variability of the grey levels depends on the choice of variance; the larger the variance, the larger the overlap. Then, the images were formed by randomly specifying a grey level for each class of each band according to the statistical distribution. In this study, we choose variances equal to 4 for all bands and classes.

TABLE 1. MEAN ASSIGNMENTS, VARIANCE = 4 IN ALL CASES

class	1	2	3	4	5	6	7	8
R	106	85	125	63	87	71	91	52
G	107	71	33	55	85	67	50	42
IR	99	73	48	61	83	69	59	51

Figure 3. Simulated SPOT-like image (Image I).

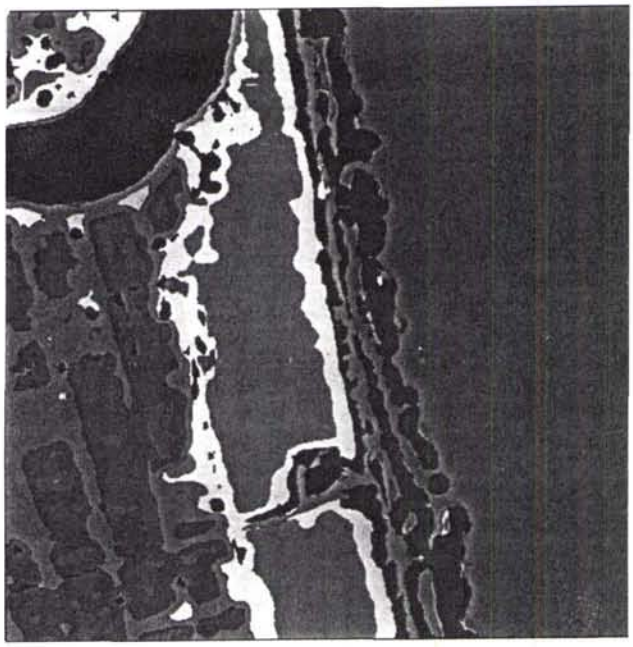

Figure 4. Simulated SPOT-like image (Image II).

The resulting image is shown in Figure 3. It is seen that the boundaries thus generated are rather complex, making classification using conventional statistical methods very difficult, if not impossible. The second image (Image II) is constructed from another SPOT image using the same procedure preserving similar properties but with different mean grey level values by one. Figure 4 shows the result, Image II, which apparently presents a different boundary pattern.

Test Results

To present the multispectral data to the network, we first encoded the input data sets in order to provide a means of discrimination of the similarities that diminish the accuracy of the neural network during the classification stage. A detailed discussion of the issue can be found in McCelland *et al.* (1986). The training cycle for the eight classes is now compared between DL and back-propagation using the delta rule as given in Figure 5. Also included for comparison is the fast learning (FL) network reported by Dawson *et al.* (1992). It is obvious that the BP fails to reach a predetermined error threshold which was set to be 10^{-3} in this study, even after 200 training cycles. The setting of the threshold value was determined experimentally, and was a trade-off between accuracy and training time. In general, a lower error threshold results in higher accuracy but requires more training time.

On the other hand, DL quickly achieves the required RMS error in just a couple of iterations. The performance of DL is

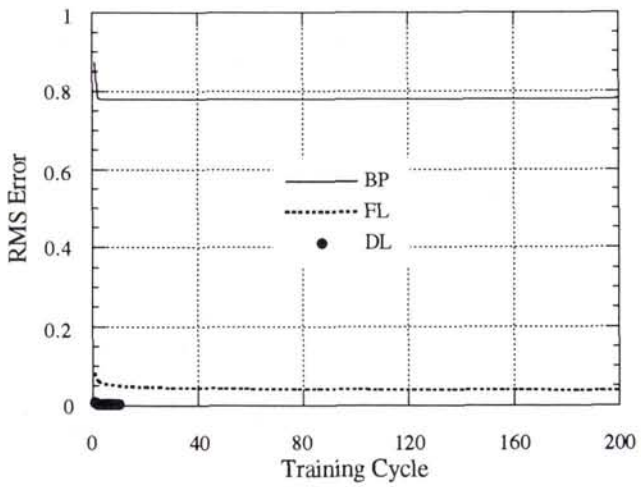

Figure 5. Comparison of the number of training cycles.

superior to that of FL. Once the network completes the training stage using the training patterns from Image I, it should generally be able to classify images with a statistically similar response without additional training.

To illustrate this statement, we applied the network directly to the classification of Image II. The resulting classification matrix is given in Table 2. A very high accuracy, well over 95 percent, was found for all classes. It should be remembered that the mean grey level values of Image II all deviated from those of Image I by one. The misclassification rate was found to be dependent on the correlation coefficient of the related classes. In other words, higher overlapped grey levels may degrade the classification accuracy, which is consistent with practical experience. To resolve this problem, other information, such as texture and radar response, is helpful but is beyond the scope of this study and will not be addressed further here.

Experimental Results

In the previous section, we demonstrated the excellent performance of the dynamic learning algorithm using two simulated images. Results indicate that the network training time is greatly reduced as compared to BP, giving a very high classification rate. The application of DL to land-cover classification of SPOT high resolution imaginary data is now in order.

TABLE 2. CLASSIFICATION MATRIX FOR IMAGE II

class	1	2	3	4	5	6	7	8	u.c.	rate,%
1	22260	0	0	0	0	0	0	0	793	96.6
2	0	22286	0	0	0	24	0	0	286	98.6
3	0	0	16482	0	0	0	0	0	432	97.4
4	0	0	0	95023	0	3	0	1	546	99.4
5	0	2	0	0	32592	7	0	0	193	99.4
6	0	2	0	14	0	34715	0	0	257	99.2
7	0	0	0	0	0	0	32071	0	76	99.8
8	0	0	0	0	0	44	0	3726	310	91.3

TABLE 3. LAND-COVER TYPES AND ASSOCIATED COLOR CODES FOR PLATE 1.

Type	Land-Cover Type	Color
1	Water	Blue
2	Vegetation	Garnet
3	Bare Soil	Bluish Yellow
4	Highly Reflective Roof	White
5	Built-Up Land	Calamine Blue
6	Barren Land	Grey
7	Sea Wake	Baby Blue

Test Site

The test site selected for the classification in this study was located near Taoyuan in northern Taiwan, an agriculture area containing small ponds, bare and barren soils, vegetation, built-up land, and man-made buildings near the sea shore. A total of seven categories based on the discriminating capability of the SPOT multispectral image were to be classified: water, vegetation, bare soil, a highly reflective roof, barren land, sea wake, and buildings, all listed in Table 3. Note that sea water and inland water were treated as one category. The standard color imagery is shown in Plate 1 where we can see that most areas were covered by short vegetation interspersed with small ponds. The small but very bright area represented a highly reflective building roof. The correlation between different bands was computed in order to understand the correlated properties of different types of land cover. The highly reflective roof was clearly the most distinguishable, while bare land and buildings are not well discriminated. Also indicated is the fact that all types of land cover are almost indistinguishable except the aforementioned roof or bare soil.

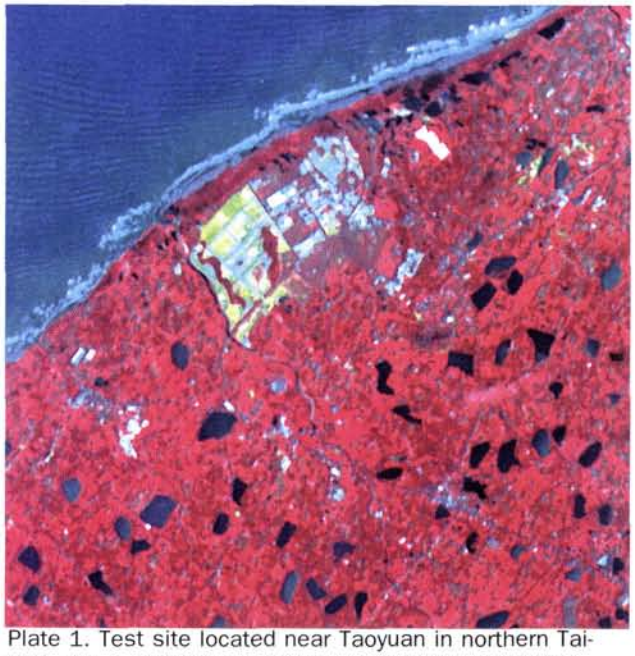

Plate 1. Test site located near Taoyuan in northern Taiwan with a pixel size of 512 by 512 (Copyright CNES).

Training Data Selection

The neural network was trained by presenting a training data set of known category assignments. For the purpose of this study, we generated a set of training data with the aid of visualization and some of the ground truth, along with available base maps. As a result, a total of 692 data points were selected as training data. The rest of the regions were then used for classification. The individual number of points for each category is given in Table 4. The estimated number of training data points over total data points is also listed in the Table. Among them, the highly reflective roof and barren land have the smallest percentage of training patterns due to the small areas they occupy in the overall space.

Classification Results

Plate 2 shows the classified image in which the major features of the seven categories of interest can be identified. To assess in more detail the performance of the DL neural network classifier, a set with a total 2000 pixels was randomly picked up from the image frame. An additional 1000 pixels was selected to aid the classification assessment in the area of the highly reflective roof and the built-up land which was more difficult to classify due to the more complex decision boundaries, smaller occupied regions, and less representative training data sets. A simple random sampling scheme (Congalton, 1988) was then applied to evaluate the classified map. Both 1:5000- and 1:25,000-scale base maps, and field works was used as auxiliary data to assist the performance evaluation. It was found that, overall, an accuracy of 92 percent was obtained.

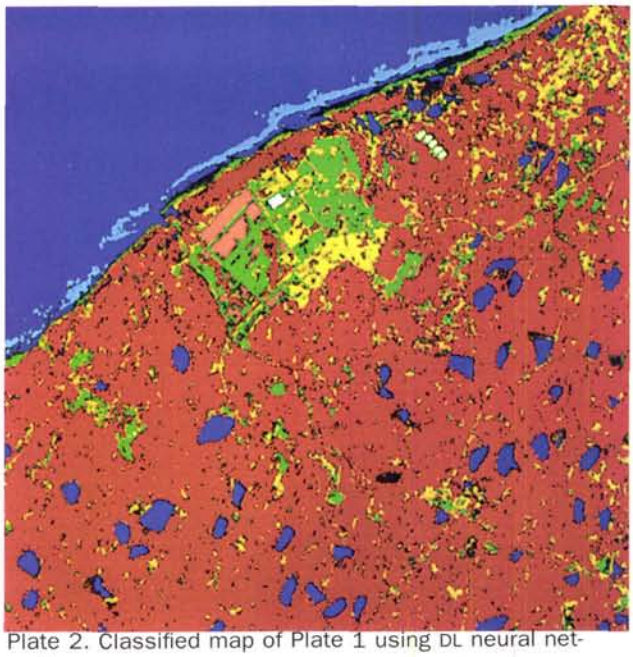

Plate 2. Classified map of Plate 1 using DL neural network.

Conclusions

The significant reduction of training time required and the high accuracy of the presented neural network when using a dynamic learning algorithm based on both simulated and real SPOT imagery results have been demonstrated. The capability of the neural network to resolve highly nonlinear and complex boundary problems, which are common for remotely sensed data, was also illustrated. In comparison, the presented network outperforms the back-propagation network in that the newly developed learning algorithm has enhanced the practical use of land-cover classification from multispectral imagery. It should be emphasized that the use of a Kalman filter enables seasonal change detection from satellite images because of the update equation which accounts for the change of states that can be incorporated into the network. This should be of interest for future investigation.

Acknowledgments

This work was supported by the National Science Council of Taiwan under grant NSC83-04240E-008-040. The authors are indebted to the anonymous reviewers for their penetrating and stimulating comments that substantially influenced this final presentation.

TABLE 4. TRAINING DATA SET SELECTION

Type	# of pixels	Est.%
water	242	34.97
vegetation	160	23.12
bare soil	71	10.26
hi. ref. roof	43	6.21
built-up land	76	10.98
barren land	39	5.64
sea wake	62	8.82

References

Ball, G. H., and D. J. Hall, 1967. A clustering technique for summarizing multivariate data, *Behavioral Science*, 12:153–155.

Barton, S. A., 1991. A matrix method for optimizing a neural network, *Neural Computation*, 3(3):450–459.

Benediktsson, J. A., P. H. Swain, and O. K. Ersoy, 1990. Neural network approaches versus statistical methods in classification of multisource remote sensing data, *IEEE Trans. Geosci and Remote Sensing*, 28(4):540–552.

Bischof, H., W. Schneider, and A. J. Pinz, 1992. Multispectral classification of Landsat images using neural networks, *IEEE Trans. Geosci and Remote Sensing*, 28:482–489.

Brown, R. G., and P. Y. C. Hwang, 1983. *Introduction to Random Signals and Applied Kalman Filtering*, 2nd Ed., Wiley, New York.

Chen, M. S., and M. T. Manry, 1991. Back-propagation representation theorem using power series, *Proc IJCNN*, pp. 295–300.

Congalton, R. G., 1988. A comparison of sampling schemes used in generating error matrices for assessing the accuracy of maps generated from remotely sensed data, *Photogrammetric Engineering & Remote Sensing*, 54, No. (5):pp. 593–600.

Dawson, M. S., J. Olvera, A. K. Fung, and M. T. Manry, 1992. Inversion of surface parameters using fast learning neural networks, *Proc. IGARRS'93*, Houston, Texas, pp. 910–912, Houston, Tx.

Haykin, S., W. Stehwien, C. Deng, P. Weber, and R. Mann, 1991.

Classification of radar clutter in an air traffic control environment, *Proc. IEEE*, 79(6):742–772.

Heermann, P. D., and N. Khazene, 1992. Classification of multispectral remote sensing data using a back-propagation neural network, *IEEE Trans. Geosci and Remote Sensing*, 30:81–88.

Hush, D. R., and B. G. Horne, 1993. Progress in supervised neural networks, *IEEE Signal Processing Magazine*, 10(1):8–39.

Lippmann, R. P., 1987. An introduction to computing with neural nets, *IEEE Acoustic, Speech and Signal Processing Magazine*, 4(2):4–22.

McCelland, J. L., D. E. Rumelhart, and G. E. Hinton, 1986. The appeal of parallel distributed processing, *Parallel Distributed Processing*, Vol I (Rumelhart and McCelland, editors), MIT Press, Cambridge, Massachusetts.

Tzeng, Y. C., and K. S. Chen, 1993. A dynamic learning neural network using Kalman filtering technique, submitted paper.

(Received 23 June 1993; revised and accepted 1 February 1994)

K. S. Chen

K. S. Chen graduated from National Taipei Institute of Technology, Taipei, Taiwan and received the Ph.D. degree in electrical engineering from the University of Texas at Arlington in 1990. From 1985 to 1990 he was with the Wave Scattering Research Center at the University of Texas at Arlington. He is now an Associate Professor at the Center for Space and Remote Sensing Research, National Central University, Taiwan. His major research has been in the areas of wave scattering and propagation from terrains and sea, radar signal and image simulation and analysis, and their applications to remote sensing. He is member of IEEE and was recipient of 1993 Young Scientist Award from the International Union of Radio Science (URSI).

Yu-Chang Tzeng

Yu-Chang Tzeng received the B.S. degree from the National Taiwan Institute of Technology, Taiwan, ROC, in 1985, the M.S. degree from the National Taiwan University, Taiwan, ROC, and the Ph.D. degree from the University of Texas at Arlington as a research assistant. He is now an associate professor with the Department of Electronics, Lien-Ho Junior College of Technology and Commerce, Taiwan, ROC. His research interest has been in the areas of microwave remote sensing with an emphasis on volume scattering and its applications to snow and sea ice.

Chi-Farn Chen

Chi-Farn Chen was born in Taipei, Taiwan. He received the B.S. degree in Atmospheric Physics from the National Central University, Taiwan. He obtained the B.S. degree in Meteorology and Ph.D. degree in Civil and Environmental Engineering from the University of Wisconsin-Madison. He has been an Associate Professor at the Center for Space and Remote Sensing Research, National Central University, Taiwan, since 1989. His research interests include digital image processing, neural networks, and integration of remote sensing and GIS for environmental problems.

Wen-Liang Kao

Wen-Liang Kao received the B.S. degree in physics from National Kaohsiung Normal University, Kaohsiung, Taiwan, ROC. He is currently enrolled as a graduate student at the Institute of Space Science, National Central University. Since 1992, he has been with Center for Space and Remote Sensing Research of NCU where his main research interests include remote sensing image processing and analysis, neural network, and rough surface scattering.