Classification of Land Cover Using Optimized Neural Nets on SPOT Data

Abstract

Often papers, which compare the performance of neural net classifiers with traditional classifiers, tend to de-emphasize the importance of the network configuration. This paper presents the results of experiments with the use of neural nets for land-cover classification in a SPOT satellite image. Segments in the image are described by textural features calculated from gray-level difference statistics. The segments are found using an algorithm based on iterative use of adaptive noise filtering and region growing. An artificial neural net, a backpropagation net with two hidden layers (often referred to as a three-layer perceptron), is used to classify the segments. A variety of segment sizes are generated by the region growing procedure, and some of these are obviously too small for a textural description. Therefore, three nets are created to process different sets of features describing segments in accordance with the segment size. To improve the classification, the networks are optimized with respect to the size of the hidden layers, by removing the least significant neurons. Furthermore, least significant input neurons are removed as a means of feature selection.

Introduction

In recent years a number of papers have been published introducing the use of artificial neural nets for classification of remotely sensed images. In many cases, however, only a limited part of the power of the neural nets is actually being utilized. Standard backpropagation networks (Rummelhart and McClelland, 1986) are used as handy alternatives to traditional Gaussian classifiers. In some cases the nets are plainly used to process pixelwise, multispectral data. Textural features as described by Van Gool et al. (1985) have been used in a number of studies (see, for example, Møller-Jensen (1990)), but the choice of textural features seems somewhat arbitrary. Finally, the optimal size of the layers in the network are in most cases found by comparing the classification results obtained with a number of discrete configurations.

This paper introduces an optimized neural net, processing spectral signatures and a number of textural features based on the gray-level difference statistics (Weszka *et al.*, 1976), to classify land cover on a multispectral SPOT image. The aim of the classification is to assign each pixel in the image to one of nine land-cover types listed in Table 1. For evaluation, the nine classes were merged into four major clusters (water, artificial/urban, fields, and forest).

The optimization algorithm proposed by Mozer and Smolensky (1989) is used to determine the size of the input layer, that is, which of the input variables to use, as well as the size of the hidden layer in the neural net. The textural features are calculated in segments generated by region growing in an image which has been processed iteratively with an edge enhancing adaptive filter.

The raw satellite image used as the test scene is shown in Figure 1. The image (500 by 500 pixels) was obtained during spring and covers a 10- by 10-km region just south of an urban center. It contains a mixture of both densely built-up areas, suburbs, rural land, and waterbodies.

Segmentation

The literature is comprised of an immense amount of studies on image segmentation. A survey of the major techniques is found in Haralick and Shapiro (1985). The method presented here is basically a single linkage region growing, in which the similarity criteria is the Euclidian distance, Δ, between the spectral signatures. This is obviously a very simple method to use for the segmentation of an image as complex as a satellite image. The reason for applying this simple method is that, prior to the region growing procedure, the image has been filtered to reduce the "noise" level in the image, in order to avoid occasional merging of adjacent, distinct segments. Noise in this context may be present both due to the detector and, more important, due to pixels covering several types of land cover. In this study the noise is reduced through iterative use of an adaptive, edge preserving, contrast enhancing filter (Dreyer, 1991). The filter is basically a modified K-Nearest Connected Neighbor Mean (Lønnestad, 1988) where K, the number of pixels used to calculate the mean, is variable. Iterative use of the filter generates images, which appear segmented, without loss of fine structure. Besides, as seen in Figure 2, iterative use of the filter leads to relaxation; the effect of the filter becomes negligible after a limited number of iterations.

The aspect of processing time was not an essential issue in this study, so few efforts were spent on improving the effiency of the algorithms. In its present form each iteration of multispectral filtering in three bands takes approximately 15 minutes and the region growing is then done within a few minutes on a PC 486.

The average size of the segments generated by the method depends on the choice of Δ, as can be seen in Figure 3. The actual size of Δ needs some consideration. On the one hand, if the number is small the segments will be small, containing too few pixels to calculate textural features with statistical significance. On the other hand, if Δ is large, the possibility of accidental merging of two distinct types of landcover into one segment becomes high.

The actually used value, $\Delta^2 = 30$, was chosen from visual inspection of the segmentation results for a number of different values. For values larger than the chosen, visual inspection reveals segments which cover areas of several types of landcover.

Photogrammetric Engineering & Remote Sensing,
Vol. 59, No. 5, May 1993, pp. 617–621.

0099-1112/93/5905–617$03.00/0

Peter Dreyer
Operations Research Group, Jutland Telephone, Sletvej 30, DK-8310 Tranbjerg-Aarhus, Denmark

TABLE 1. LAND-COVER CLASSES.

Landcover	Cluster
Deep water Shallow water	Water
Densely built-up Loosely built-up Roads Railroads	Artificial cover
Fields (grass, permanent crops) Fields	Fields
Forest	Forest

Features

Because the region growing is carried out in filtered, contrast enhanced images, where texture is surpressed, it is important to note here that the texture calculated for the generated segments is calculated on data from the raw satellite image. Six features are used to describe texture. The features were selected among a multitude of measures listed by Haralick *et al.* (1973). The chosen features have been used in previous studies (Lee *et al.*, 1990; Franklin and Peddle, 1989; Weszka, 1976). These are (in italics to distinguish from ordinary statistical measures)

Mean
Scatter
Contrast
Entropy
Local homogeneity
Second Angular Moment

derived from the gray-level difference statistics as given in Weszka *et al.* (1976). For each image band, each feature is calculated for four displacement vectors (1,0), (0,1), (2,0), and (0,2). These rather small displacements were chosen in the light of the similarly small average size of the segments. From the mean of the first two, the feature for a displacement distance of 1 is derived, and from the latter two, the feature for a displacement distance of 2 is obtained. This is done in order to avoid distinguishing texture due to orientation.

With six textural features, each calculated for two displacement distances in each of three image bands, 36 textural measures are obtained. Furthermore, the mean and standard deviation in spectral signature inside the segment is calculated, giving a total of 42 measures to describe each segment.

The Neural Network

Over the last couple of years, a number of papers have analyzed the use of neural nets for image classification. Hepner (1990) shows that, for land-cover classification of TM images based on spectral signature in a 3- by 3-pixel window (in order to consider texture), the performance of a neural net is superior to the conventional maximum-likelihood method. Lee *et al.* (1990) have published the results of cloud classification in visible, single channel MSS images. They obtain remarkable results

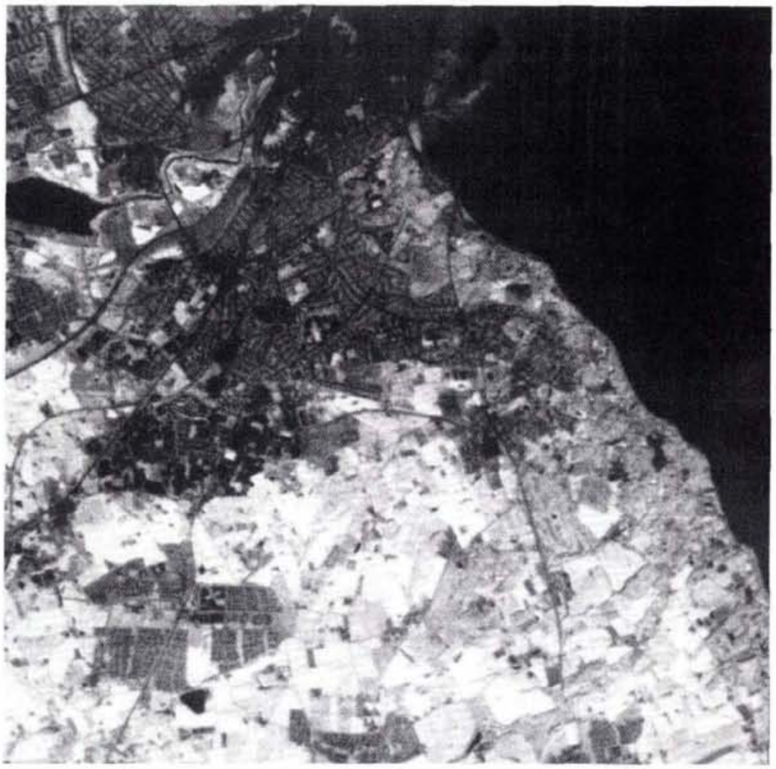

Figure 1. Test scene. 500 by 500 pixels raw IR-band of multispectral SPOT satellite image covering region just south of the city of Aarhus, Denmark. (© SPOT Image 1988 CNES.)

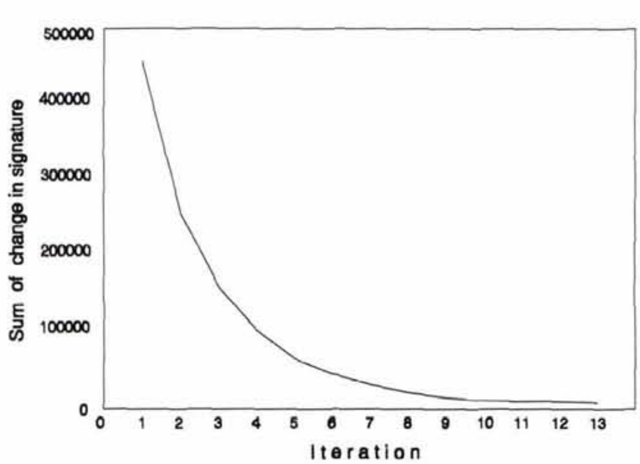

Figure 2. The sum of the absolute change in pixel values over the image in all three spectral bands, caused by a single iteration, as a function of iteration number.

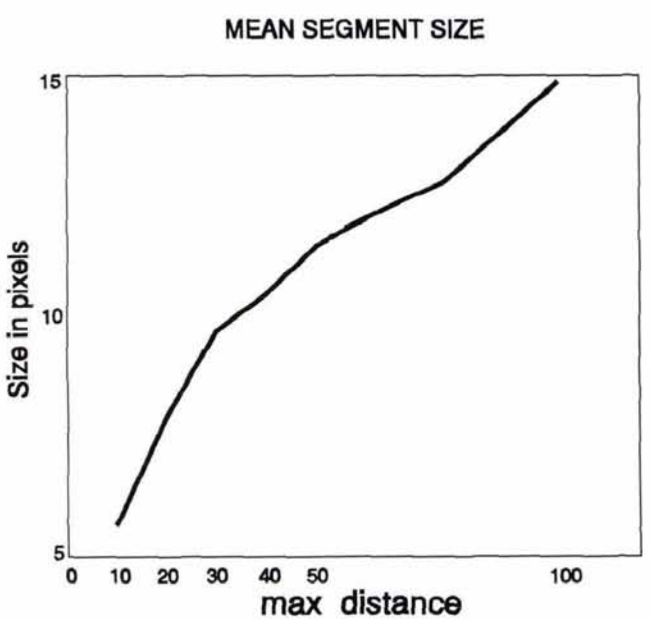

Figure 3. Average segment size in image segmented with the above mentioned method, as function of the square of the maximum Euclidian distance allowed by region growing.

based on textural features calculated in rather large windows (512 by 512 pixels). In both of the above cases, backpropagation networks were used for the classification.

Also in this present study a standard backpropagation net with sigmoid transfer functions and momentum term is used with further emphasis on the optimization of the neural net.

Because the method of segmentation generates segments consisting of just one or a very few pixels, as well as larger segments, it is obvious that texture features can not be calculated or have statistical significance for all segments. Therefore, any further analysis of the segments must depend on the actual size of the segment. According to size, the segments are thus divided into three groups:

A: Segment size $< N_1$, $N_1 = 5$
B: $N_1 \leq$ Segment size $< N_2$, $N_2 = 20$
C: Segment size $\geq N_2$

The threshold sizes were selected rather arbitrarily. Segments in group A are classified with a neural net with only three input neurons representing mean of the spectral signature. Segments in group B are classified using a net with an input layer with six neurons, representing the mean as well as the standard deviation of the spectral signature inside the segment. Finally, segments in group C are classified with a net with 42 input neurons, utilizing all measures calculated in the segment. In each case the output layer has nine neurons, each representing one of the land-cover classes.

The number of hidden layers in the backpropagation net determines the complexity of the mapping of the input domain onto the output domain. As explained by Lippmann (1987), any kind of mapping is possible when using two hidden layers. Because, *a priori*, the complexity of the mapping required to produce the classification was unknown, both nets with a single and with two hidden layers were tested. It appeared to be impossible to train nets with a single hidden layer to the same degree as nets with two hidden layers. Hence, nets with two hidden layers were selected for further study.

Initially, the A and B nets had 20 neurons/hidden layer and the C net had 15 neurons/hidden layer. However, it is a well established fact, that the size of the hidden layers has a strong influence on the performance of the neural net. For a given task, this size may be optimized with respect to a tradeoff between the ability of the network to generalize and the ability to learn the learning samples. A number of papers address this problem. In some cases, as in Mozer and Smolensky (1989), the optimization is done by removing hidden neurons, while others (Le Cun *et al.*, 1990) suggest removal of interconnecting weights based on second-order terms in the error function. In any case, the underlying philosophy, also known as "Occam's Razor" is that, for a set of networks solving a given problem, the net with fewer neurons gives the best generalization (Thodberg, 1990).

Here Mozer and Smolensky's algorithm is applied to simplify the neural nets. By calculating the relevance of each neuron in the input and hidden layer, it is possible to find the least important neurons. It is now possible to remove a number of specific neurons and leave the rest of the network unchanged. By doing so, the reconfigurated net is "born" with the weights from the previous configuration, representing some degree of learning. There is a risk of ending up in a local minimum of the error function but, as noted by Mozer and Smolensky, tests indicate that it is more probable that a network of the same size, starting with randomized weights, will be stuck in a local minimum. In this study, for nets with an optimal size obtained by iterative thining, it appeared to be impossible to train these nets in one step, that is, starting with random weights as opposed to weights representing some degree of learning.

In Mozer and Smolensky's approach, the relevance is defined as a difference between error functions. It is estimated as a first-order partial derivative of the error function. Because the transfer function of the neuron is highly non-linear, this estimate will in some cases be a very poor approximation, in which case higher order terms from the Taylor expansion might have to be used. The skeletonization

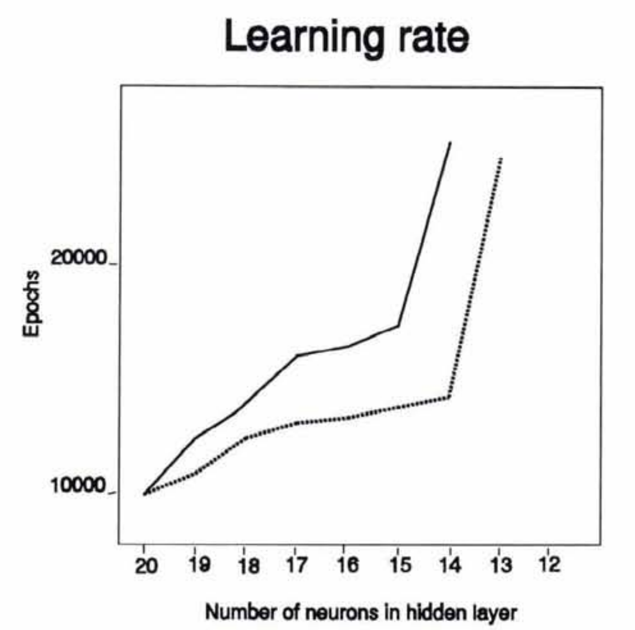

Figure 4. Number of epochs needed to retrain a re-configurated net having removed the least significant hidden neuron (dotted line), compared to random removal of neurons.

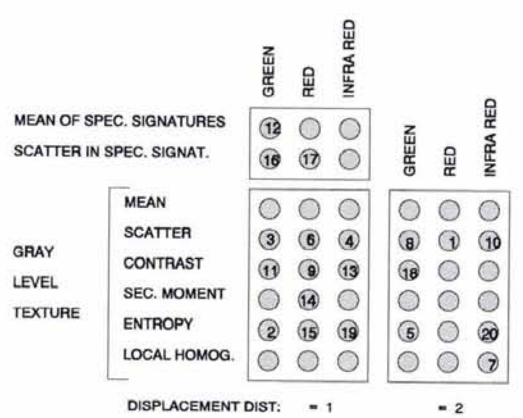

Figure 5. Order of removal of least significant neurons in input layer, originally processing 42 input features.

derived from Mozer and Smolensky, as compared to the skeletonization resulting from the exact calculation of the relevance, should be subject to further study.

The question of when to remove a neuron still remains: How low should the relevance be to imply removal? There seems to be no straightforward relationship between the significance of a removed neuron (compared to the significance of the remaining neurons) and the amount of retraining needed to restore the level of learning. This makes it difficult to use the value of the significance to decide whether or not to remove the neuron. Instead, the algorithm used here is simply "give it a try." As seen in Figure 4, the number of epochs needed for retraining is rather low, until a certain point, where the number increases dramatically. Removing any neuron beyond this point makes retraining impossible.

It is important to note that in the input layer the relevance has further meaning; if a neuron has little importance the corresponding feature has little importance and may be left out of the analysis.

Results

Using the previously described method, the net with three input neurons was skeletonized from 20 neurons in each of the hidden layers to only 13 and 12 neurons respectively. For the net with six input neurons, each of the hidden layers was thinned from 20 to eight neurons. In both cases the input layer was kept unchanged, even though the significance in the input layer reflected the fact that the red and green bands of the SPOT satellite have a high degree of correlation and a low dynamic range in the scene.

For the net processing all 42 features, neurons in the input layer were removed after the optimization of the hidden layers (from 15 in each layer to seven in each). Figure 5 shows the order in which the input neurons were removed. It is no surprise that the *scatter* of the gray-level statistics were removed at an early stage; because the *means* are close to zero, the *scatter* and the *contrast* are highly correlated. However, it is surprising and contrary to the results of Lee *et al.* (1990) that the *means* have relatively high significance. This may be due to the fact that string-like segments in the image, such as roads, often display a slow, monotonous increase or decrease in pixel values along the length of the segment. This again occurs because roads cover only a limited part of the area inside a pixel.

At this point it may be of value to give an interpretation of the individual measures. When looking at the definitions, it appears to be possible to divide the measures into two groups: those describing the homogeneity (or lack of) of the pixel values (the amplitude of the texture) and those describing the regularity of the texture. It is, for example, possible to have very ordered texture as the pattern on a checkboard; however, the homogeneity of the segment with such a pattern depends on the contrast between the bright and the dark squares. To the former group belong *mean*, *contrast*, and *local homogeneity*, while *scatter*, *entropy*, and *second angular momentum* belong to the latter group.

Based on the first thinning of the input layer, a new net was created in which all input neurons processing features describing *scatter*, *contrast*, and *entropy* were removed in accordance with Figure 5. In the light of the interpretation of the texture measures, it is interesting that *scatter* and *entropy*, both indicators of the regularity of the texture, seems to have little significance. Further studies are needed to draw any conclusions with respect to this effect.

Following the training procedure, the neural nets were finally used to classify the entire satellite image. The performance of the original and optimized nets is given in Table

Table 2. Classification Results for Various Methods. The Values Indicate Percentage of Correctly Classified Segments among 300 Test Segments, Prior to and After Optimization. (For the Analysis of the Larger Segments, the Test Sample Consisted of only 180 Segments.)

Classification method	Water	Urban	Field	Forest
1: Neural Net 3 inp	80/75	93/94	71/76	95/98
2: Neural Net 6 inp	85/90	91/93	68/73	91/93
3: Neural net 42 inp	58/73	73/92	90/93	66/88

2. The performance is evaluated on a number of test segments, which were selected and classified by visual inspection and compared with the results of the network classification. (It is worth noting that the visual classification may introduce some uncertainty in the ground truth. This is partly due to errors in segmentation and partly due to the fact that the classes tend to cluster in the spectral feature space.)

Conclusion

From Table 2 it seems that the best performance is obtained with a classifier which considers only spectral signature. *Visual* inspection seems to suggest that taking scatter into account gives slightly better results. However, for some of the classes the classification using textural features is clearly of lower quality. This is in particular the case for the non-optimized net (probably due to the fact that the net is so large that it learns the limited number of training samples "by heart"). In the case of forest, texture is very marked in the raw satellite images and filtering does not remove this completely. Therefore, in forest areas the segmentation is, to some extent, based on the texture. This implies that the original large forest segments are broken down into small segments, where the texture at random deviates from real, descriptive texture. The lack of improvement in accuracy from the use of texture is in good agreement with Franklin and Peddle (1989). Note, however, that use of textural features leads to a significant improvement in detection of fields. Distinction between urban areas and areas with shallow water does not appear to be helped with the use of texture.

The most striking effect in Table 2 is the improvement in performance following an optimization of the nets. In some cases the improvement is rather small, probably within the uncertainty of the figures, but the trend is obvious. Even if the improvement in accuracy is minor, optimization still results in highly improved efficiency. In all cases the number of connections is reduced by approximately a factor of 4, resulting in the same increase in computing speed.

The results obtained at present prove the applicability of the described method, consisting of segmentation and texture based classification. The results seem to suggest that the choice of texture measures is important, but also that texture may actually be inopportune for some classes of land cover. There is no comparison of the neural net classifier with other traditional classifiers in this paper, but from looking at neural nets only, it seems obvious that optimization of the net configuration is important for the performance. Further testing might lead to some improvement due to optimization of free parameters such as similarity criteria for segmentation and the threshold segment sizes for the different net sizes.

References

Dreyer, P. 1991. Pre- and post-processing in classification of land-cover in digital multispectral satellite images, *Proceedings of 7th SCIA 1991, Ålborg*, p. 1019.

Gool, L. Van, P. Dewaele, and A. Oosterlinck, 1985. Texture Analysis Anno 1983, *Computer Vision, Graphics, and Image Processing*, (29):336.

Franklin, S. E., and D. R. Peddle, 1989. Spectral texture for improved class discrimination in complex terrain, *Internat. Jour. Remote Sensing*, 10(8):1437.

Haralick, R. M., K. Shanmugan, and I. Dinstein, 1973. Textural Features for Image Classification, *IEEE Trans. Systems, Man and Cybernetics*, 3(6):610.

Haralick, R. M. and L. G. Shapiro, 1985. Image Segmentation Techniques, *Computer Vision, Graphics and Image Processing*, (29):100.

Hepner, G. F., T. Logan, N. Ritter, and N. Bryant, 1990. Artificial Neural Network Classification Using Minimal Training Set, *Photogrammetric Engineering & Remote Sensing*, 56(4):469.

Le Cun, Y., 1990. Optimal Brain Damage, *Advances in Neural Information Processing Systems II*, Morgan Kaufmann Publ.

Lee, J., R. C. Weger, S. K. Sengupta, and R. M. Welch, 1990. A Neural Network Approach to Cloud Classification, *IEEE Transactions on Geoscience and Remote Sensing*, 28(5):846.

Lippmann, R. P., 1987. An Introduction to computing with Neural Nets", *IEEE ASSP Magazine*, April, p. 4.

Lønnestad, T., 1988. Et Filter for Kantbevarende Støyfjerning ved Klassifica sjon, *Proceedings from NOBIM conference 1988*, Oslo, p. 169.

McClelland, J. L., E. R. Rumelhart, and the PDP Research Group, 1986. *Parallel Distributed Processing*, Vol. I, The MIT Press, Massachusetts.

Mozer, M. C., and P. Smolensky, 1989. Using Relevance to Reduce Network Size Automatically, *Connection Science*, 1(1):3.

Møller-Jensen, L., 1990. Knowledge-Based Classification of an Urban Area Using Texture and Context Information in a Landsat-TM Imagery, *Photogrammetric Engineering & Remote Sensing*, 56(6):899.

Thodberg, H. H., 1990. Occams razor—improving generalisation of neural networks, *Preprint from Danish Meat Research Institute*.

Weszka, J. S., C. R. Dyer, and A. Rosenfeld, 1976. A Comparative Study of Texture Measures for Terrain Classification, *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-6(4):269.

(Received 7 May 1992; revised and accepted 8 December 1992)