032596

LARGE AREA CLASSIFICATION FOR HYPERSPECTRAL DATA SETS

Xiuping Jia
Department of Electrical Engineering
University College, The University of New South Wales
Australian Defence Force Academy,
Campbell, ACT, 2600, Australia
x-jia@adfa.oz.au

David Landgrebe
Professor & Acting Head
School of Electrical & Computer Engineering
Purdue University,
West Lafayette, IN 47907-1285, USA
landgreb@ecn.purdue.edu

ABSTRACT

With the popularisation of remote sensing techniques and applications, large area classification will become inevitable. Conventionally, when using supervised classification methodologies for the analysis of multispectral data, class training fields must be selected from the whole area to be classified in order to create reliable statistics for each class. This tends to be labour-intensive, and the costs increase significantly when a large area is to be analysed.

In this paper, an investigation on how to improve the generalisation of training results from a local area to fit a larger area is presented. Problems which must be dealt with and strategies which might be used are outlined and illustrated, including some newer algorithms which have only recently appeared in the literature. Experimental results using an AVIRIS data set showed that the training results obtained from a small area was able to be adjusted to improve the classification accuracy on a neighbouring area nine times the size of the region from which training was drawn.

INTRODUCTION

Conventionally, classifier training is required to be done using well distributed training pixels in the image to be classified in order to ensure that the training pixels are representative and therefore, class statistics are reliable (Richards, 1993). Training pixel labeling, however, is an expensive process, which include field visits, reference map production or collection and photointerpretation of multispectral images, etc. Moreover, the ground truth information must be approximately contemporary with the multispectral data to be recorded and must be applied to the correctly registered image. For a large area image classification, the high cost incurred in the class training becomes a great concern.

With this problem in mind, the intension of this work is to investigate the possibility of using limited ground truth information for a local area to classify other areas in the vicinity. Theoretically, hyperspectral data have higher separability power than multispectral data. However, as the dimensionality goes up, the number of training samples needed to adequately define the shape of the class distribution in N-dimensional space goes up quite rapidly (Landgrebe, 1995). Therefore, a key difficulty is that the conflict between the requirement of larger numbers of spectral bands for class separability and the limited number of training pixels available.

In this paper, problems encountered for this work are demonstrated and analysed, followed by general consideration and experimental results.

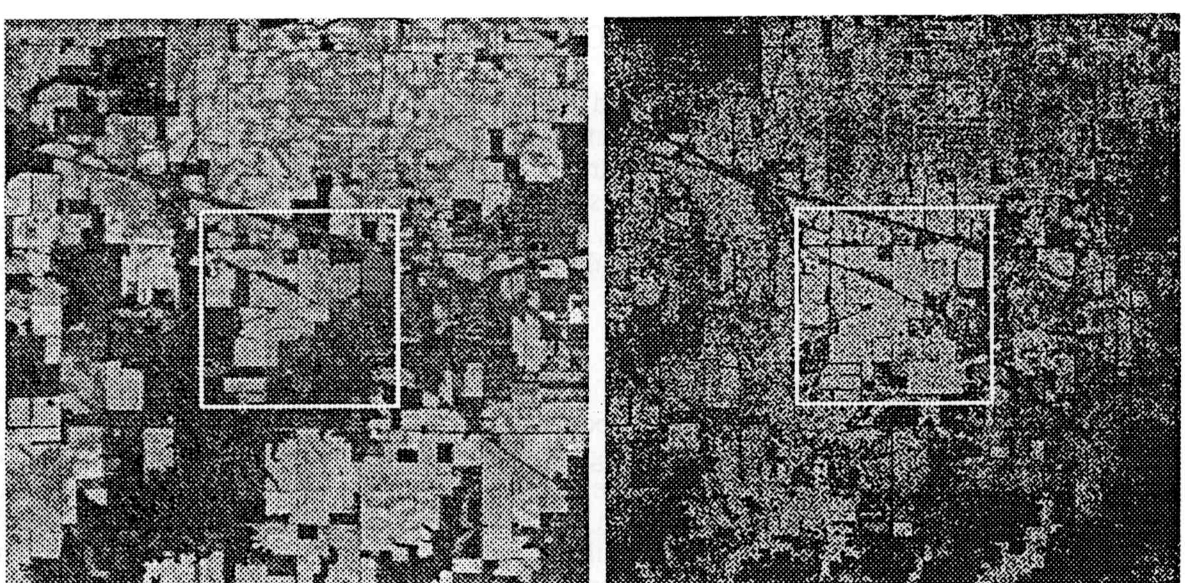
DEMONSTRATION OF THE PROBLEM

The image data used for this work is from an area of mixed agriculture and forestry in Northwestern, Indiana, USA. The data was recorded in June 1992 by AVIRIS with 210 bands. The centre area, 145 by 145 pixels, of about 8.4 km^2, is used as training site (Fig. 1(a)). Experiments on using the training results obtained from this local area to label the surrounded area, which is 9 times of the centre area, have been studied and analysed.

Firstly, the class training was done in the centre area. 16 classes were defined and their statistics were estimated using the training pixels selected from this region based on the ground truth. (Landgrebe, 1994) demonstrated that the classification accuracy was 75.5% for the training data using 9 AVIRIS bands which occur in the middle of the 6 Thematic Mapper reflective bands plus 3 additional ones. But, 98% was reached using the first 50 new features obtained after the original 210 bands were transformed using Decision Boundary Feature Extraction (DBFE) (Lee and Landgrebe, 1993). The function of DBFE is to find an optimal linear transformation to enhance class separability, based directly upon the training samples. The new features obtained are in order of their importance relative to discriminating between the classes defined. This showed the higher separation power of the hyperspectral data.

The training results were then applied to the extended area as shown in Fig. 1 (a) and the probability map is showed in Fig. 1 (b). A probability map indicates the degree of membership of the class to be labelled for each pixel (Landgrebe and Biehl, 1995). Using Gaussian maximum likelihood classification, the decision rule is: $\mathbf{x} \in \omega_i$ if $g_i(\mathbf{x}) > g_j(\mathbf{x})$ *for all* $j \neq i$

where $g_i(\mathbf{x}) = -\ln|\Sigma_i| - (\mathbf{x} - \mathbf{m}_i)^t \Sigma_i^{-1} (\mathbf{x} - \mathbf{m}_i)$, $i = 1, 2, \ldots, M$, where $\mathbf{x}$ is a pixel brightness vector, $\mathbf{m}_i$ is the mean brightness vector for class i, and Σ_i is its covariance matrix of size $N \times N$, and N is the total number of spectral bands. M is the number of classes available for labeling the pixel. The brightness in a probability map is direct proportional to the maximum likelihood value, $Maximum(g_i(\mathbf{x}))$.



(a) (b)

Fig. 1. (a) Image segment selected for large area classification exercise. The centre area is used as training site. (b) Classification probability map using the training results obtained from the centre area.

The average probability is 46.3% in the training site, 18.7% on the extended area. The classification accuracy on the test data of the extended area was 35% only. This results show that the classification is reliable in the centre area, but not in the extended area and demonstrate the problem of classifier generalisation.

There are two main physical reasons for the poor classification on the extended area. One is that the radiometric error resulting from atmospheric and other effects varies from nadir and the extremities of the swath for the systems, particularly those with a wide fields of view, because of the appreciable difference in atmospheric path length, the differences in illumination and view angles, and other effects. Secondly, there are some slight differences in spectral response from place to place for an information class. For example, the different moisture contents and different types of soils may present variations to the class spectral response.

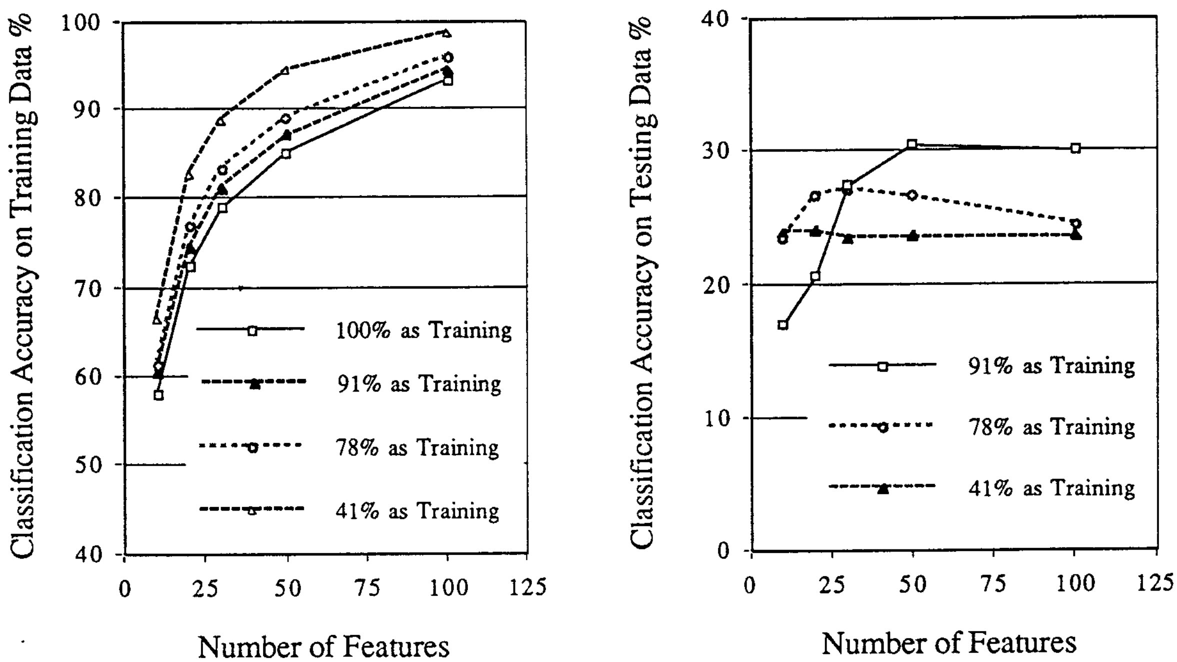
Six classes have been selected from the image for detailed investigation. Since the data were collected in the early part of the growing season, soybean and corn canopies presented only about a 5% ground cover. These two classes were further divided into three subcategories depending upon the different tillage practice used on a given field: no-till, minimum-till, and clean-till. The no-till fields would have a substantial amount of residue from the crop of the previous year, the minimum-till field would have only a moderate amount, and the clean-till would show only the species canopy on a soil background. The separation of these six classes, two species in each of three conditions, represent a very challenging classification problem, one that represents a very poor S/N ratio where the response from the plant leaves represent signal and that from the soil/residue background represents noise. They are used for the following tests.

An experiment was made to examine the effect of using a pixel as training data or as testing data against the number of features used. In the real case as the project setup, training pixels are only available in the centre area, totalling 6554. Testing pixels were selected from the extended area totalling 9434. The classification results on the testing data were not satisfactory as suggested above. Three simulations were conducted for classifier performance comparison by using part of testing data (Simulations 1 and 2) or all the testing data as training data (in Simulation 3) (Table 1). The classification accuracy on the training data and testing data for each case is plotted against the number of DBFE features used as shown in Figs. 2 (a) and 2 (b), respectively.

Table 1 Number of Training and Testing Pixels in Each Test

	No. of Training Pixels	No. of Testing Pixels	Percentage of Training Pixels
Real Case	6554	9434	41%
Simulation 1	12519	3467	78%
Simulation 2	14586	420	91%
Simulation 3	15980	0	100%

From Fig. 2 (a), it can seen that the classification accuracy on the training data increases with the increasing number of features used. The number of training pixels has a significant effect on the performance. Using the least training pixels, 41% of the total labelled pixels, the shape of the class distributions fits the training samples the best, while using all the labelled pixels as training data give the lowest classification accuracy on those data itself. This results are opposite to the testing data. For the testing data, the more training data, the better the classification results achieved as would be expected. The Hughes phenomenon (Hughes, 1968) is observed in Fig. 2 (b) where the classification accuracy does not continue to increase with the number of features used. Rather, it starts to decrease after more than a certain number of features used. A conclusion can be drawn from the above tests: High dimensional feature space provides the possibility to separate even arbitrary classes. However, enormous numbers of training samples are needed to generate reliable class statistics. The training data can be classified very well as long as enough features are available. But, when the number of training samples are fixed, the accuracy first increases, then drops with increasing features. These points are significant for finding a suitable way for classifier generalisation discussed below.



(a) (b)

Fig. 2. Classification accuracy (a) on the training data; (b) on the testing data, using different percentage of labelled pixels as training data

PROBLEM SOLUTION

There are several problems to deal with for classifier generalisation:

Reliable and separable class statistics estimation It is essential that the class spectral response be unique and reasonably separable from class to class. Attention must be paid when one seeks strong class separability by increasing the number of features to be used. With higher dimensionality, the ratio of the number of training pixels to the number of features decreases and therefore the robustness of the class spectral response drops. High performance obtained on the training data may be a misleading. Since there are only a limited number of training pixels for each class in the remote sensing situation, it is vital to use the smallest number of features which achieve adequate separability. As a result, feature extraction will be the first step to compress the rich information contained in a hyperspectral data such as AVIRIS data with its more than 200 bands, into, say 10 to 30 bands (Landgrebe, 1995). Basically, the number of training samples per class is a parameter to limit how many features can be used.

New classes detected in the extended area This can be done using classification with a threshold or finding them directly from the probability map. Regions with low likelihood are candidates for the new spectral classes and should be trained separately. When we have no ground truth information for them, we can simply regard them as unknown classes. But, it is important to include them in the class list, in order that there is a logical class to which to assign every pixel.

Class spectral response modification The class statistics estimated from the training pixels selected from a limited area need to be modified to accommodate the variations in the extended area. The key point is how to make use of extra information provided in the extended area to achieve this. Unsupervised clustering techniques may be employed on the extended area for subclass detection. Alternatively, more training pixels may be picked up using iterative classification and class statistics can be recalculated after each iteration with some new training pixels added. However, these methods can be sensitive to the degree of separability between the classes of interest and may be suitable for main ground cover types mapping only. Alternatively, the statistics enhancement algorithm (Shahshahani and Landgrebe, 1994) can be employed for general class spectral response modification. Statistics enhancement focuses on modifying the original class statistics estimated from the local training pixels to fit the global image statistics, i.e. the probability density function of the entire data set is modelled as a mixture of Gaussian class densities. By doing this, unlabelled pixels can play a significant role in class spectral response formation. A requirement for running statistics enhancement is the list of classes must be exhaustive, which is required for a proper classifier training as well.

EXPERIMENT RESULTS

The above procedure were applied to the image data discussed previously. After careful examination, nine classes were defined. DBFE was applied to the 210 bands based on these classes of interest and the first 15 features were kept. Using the original statistics, the highest accuracy on the testing data was 70.7% when the first 10 features were used. Statistics enhancement was conducted after 22 new clusters were defined and taken into account with the original 9 classes. The highest accuracy on the testing data using the enhanced statistics increased to 73.1% when 10 features were used. More comparisons using different number of features are plotted in Fig. 3.

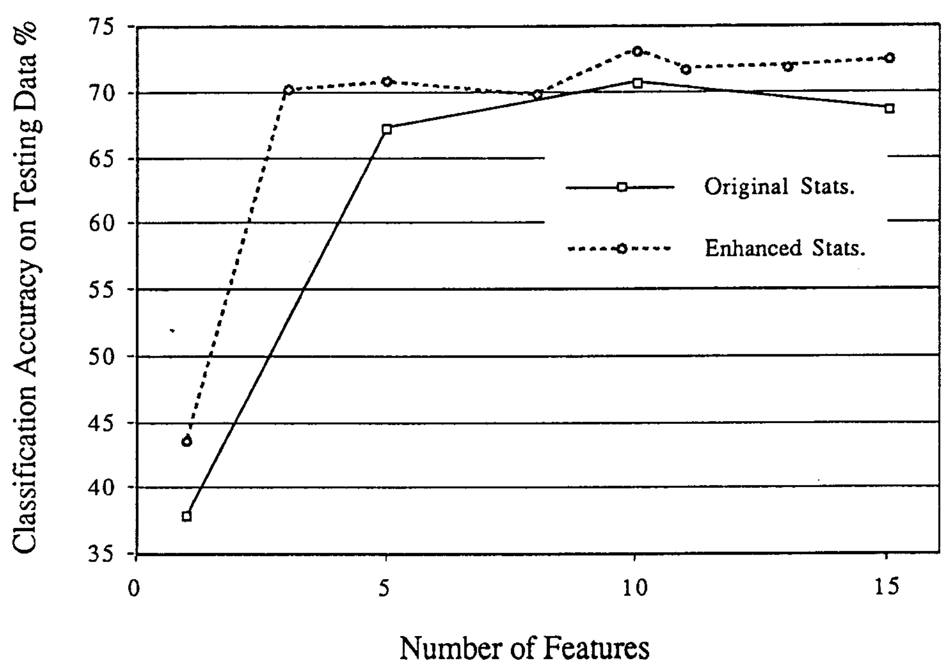


Fig. 3. Classification accuracy comparison using the original class statistics and the enhanced statistics.

CONCLUSIONS

Hyperspectral data provide stronger power in class discrimination and possibility of classifier generalisation. However, to make class statistics reliable, it is important to carry out feature extraction to reduce the dimensionality. **Class spectral response modification is then a critical step for classifier generalisation** to accommodate variations in the areas other than the training site. Nevertheless, better class modelling needs to be developed to a imitate human being's ability in ignoring noise and detecting signal only.

All of the algorithms used in this work are contained in MultiSpec©, a software system with a substantial amount of documentation which may be downloaded without cost from the World Wide Web at the following URL:

http://dynamo.ecn.purdue.edu/~biehl/MultiSpec/

The current version of this system runs under the Macintosh operating system. A Windows version of MultiSpec is under construction, and a partially complete version of it is also downloadable, but it does not yet include all of the algorithms used herein. This Windows version is expected to be complete within the next several months.

REFERENCES

Hughes, G.F. (1968) 'On the mean accuracy of statistical pattern recognizers,' IEEE Trans. Infor. Theory, Vol.IT-14, No. 1, pp. 55-63.

Landgrebe, D.A. (1994) 'Multispectral data analysis: A signal theory perspective,' © 1994 David Landgrebe, A Tutorial on Multispectral Image Data Analysis, 43 pages. Available via the WWW at the following URL: http://dynamo.ecn.purdue.edu/~biehl/MultiSpec/

Landgrebe, D.A. (1995) 'Land satellite information extraction: past, present, and future', The American Society of Photogrammetry and remote sensing symposium, Vienna, VA, September, 25-28.

Lee, C. and Landgrebe, D.A. (1993) 'Feature extraction based on decision boundaries', *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 15, No.4.

Richards, J.A. (1993) *Remote Sensing Digital Image Analysis: An Introduction.* 2nd ed. Berlin: Springer-Verlag, 1993.

Shahahahani, B.M. and Landgrebe, D.A. (1994) 'The effect of unlabelled samples in reducing the small size problem and mitigating the Hughes phenomenon', *IEEE Trans. Geosci. Remote Sensing*, Vol. 32, No.5.