# Class Modeling and Remote Sensing Image Classification Using Selected Spectral and Spatial Features

*Xiuping Jia; School of Information Technology and Electrical Engineering, University College, UNSW, Australian Defence Force Academy; Canberra ACT 2600 Australia*

## Abstract

*To improve the quality of the thematic map generated from spectral data, spatial information can be taken into account. Texture features, such as contrast, correlation, energy and homogeneity, can be combined with the original spectral data so that more detailed presentation of each class' characteristics is obtained. On the other hand, we are facing a very high dimensional data set to handle. In this paper, classification using a large number of features is addressed with strong emphasis on the need of feature selection and feature extraction. Texture features and associate parameters are examined in terms of correlations between them and their contributions toward class separability. Class data normality test is introduced using skewness and kurtosis before the Bhattacharyya distance is applied for effective feature selection. A progressive two-class decision classifier is adopted for a flexible multistage classification. Experimental results show a significant improvement by the pair-wise classification using properly selected features.*

## Introduction

Multispectral or hyperspectral data record the reflectance of various earth cover types at different wavelengths. This spectral information has been widely used in pixel labeling [1]. However, such pixel-based classification may introduce 'pepper and salt' appearance on the generated thematic map. To overcome this problem, spatial information can be taken into account.

Classification using texture features have been studied in recent years [2-5]. It is expected the texture measures, such as contrast, correlation, energy and homogeneity, can be combined with the original spectral data so that more detailed presentation of each class' characteristics is obtained. On the other hand, we are facing a very high dimensional data set to handle. A number of texture measures with various spatial distances in different directions can be generated from each spectral band. The resulting high dimensional data can experience the Hughes phenomenon [6] or the curse of dimensionality. Feature reduction is required to reduce the computational cost and, more importantly, to improve the robustness of the class models established.

In this paper, texture features and associate parameters are examined in terms of correlations between them and their contributions toward class separability. Class data normality test is introduced using skewness and kurtosis before the Bhattacharyya distance is applied for effective class separability based feature selection. A binary classification scheme called progressive two-class classification is adopted for pixel labeling. Experimental results show a significant improvement by the pair-wise classification using properly selected features.

## Test Data Set

A subset of Landsat Enhanced Thematic Mapper Plus (ETM+) data recorded over Canberra region, Australia, on 25 April 2001 has been used for demonstrating the developed procedure. The area consists of a lake, city centre and surrounding suburbs, and bushland. The ETM+ sensor records data in six spectral bands in the visible and middle infrared range with the spatial resolution of 30x30 $m^2$. It also has an image in the thermal infrared wavelength with lower spatial resolution. This channel is not used in this work. The radiometric resolution is 8 bits.

## Texture Feature Generation

Texture in an image is determined by the pattern of the brightness values within a neighborhood. The Gray Level Co-occurrence Matrix (GLCM) [7] has been used widely to present the spatial relationships of pixels. It calculates how often a pixel with the gray value $i$ occurs in a neighboring pixel with the value $j$. The neighbors can be on its immediate right, above, or in other directions. They also can be two or more pixels away. This parameter is referred as spatial distance in this paper. The element $glcm(i,j)$ is simply the sum of the number of times that the pixel with value $i$ occurred in the neighbor with value $j$ in the input image. Several image texture statistics can be derived from the normalized $glcm(i,j)$, ie.

$$\sum_{i,j} glcm(i,j) = 1,$$

as detailed below [8].

***Contrast***: It is also called variance and inertia. It measures the local variations in the image.

***Correlation:*** It measures the linear dependency of brightness of neighboring pixels at the spatial distance defined.

***Energy:*** It is also known as uniformity or the angular second moment.

***Homogeneity:*** It measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal.

Above texture measures can be generated for each pixel of an original spectral band using the $glcm(i,j)$ calculated within a local window. The new bands/features can then be stacked togather with the original spectral bands. The window size selection needs to consider the spatial resolution of the image and the distance of neighborhood defined in generating GLCM. With the testing data set, since it is prohibitive to calculate a GLCM for the full dynamic range of 0 to 255, they are scaled to 32. 192 texture bands were generated with following parameters.

| | |
|---|---|
| No. of original spectral bands: | 6 |
| No. of texture measures: | 4 |
| Spatial Distances of neighbors: | 1 with window size of 5x5 |
| | 2 with window size of 7x7 |
| Directions of the neighbors: | 4 ($0^0$, $45^0$, $90^0$ and $135^0$) |

The outputs of the texture measures were then scaled to 0 to 255 to match the spectral data range. Together with original 6 spectral bands, there are 198 features available. The list of the features is given in Table 1. However, it is inappropriate to use all of them directly for classification. Obviously, not all the features are useful for individual projects. The useless features damage the reliability of class training. It often leads to high classification accuracy on training data but low on testing data, and therefore the rest of the image. In other words, the class model has poor generalization. Feature reduction is a must step before classification can be carried out.

**Table 1: The list of spectral and spatial features**

| Feature Numbers | Features |
|---|---|
| 1 - 6 | ETM+ Bands: B1 to B6 |
| 7- 30 | Contrast in 4 directions at distance 1 for B1 to B6 |
| 31 - 54 | Correlation in 4 directions at distance 1 for B1 to B6 |
| 55 - 78 | Energy in 4 directions at distance 1 for B1 to B6 |
| 79 - 102 | Homogeneity in 4 directions at distance 1 for B1 to B6 |
| 103 - 198 | Same as 7 - 102 but at distance 2 |

## Feature Extraction and Selection

It is expected some texture features are highly correlated. After examining the correlation properties among all the features shown in Fig. 1, it was found that the Homogeneity features for the 6 spectral bands in all the directions with the same spatial distance are highly correlated. The Energy features are insensitive to directions as well. Each of this highly correlated group of features has been compressed by principal component transform and replaced by the first component, respectively, leaving 116 features for further consideration.

Further feature selection can be performed based on class separability provided by a subset of data. The separability is related to mean vector difference and/or covariance matrix difference. Since the wide variation of texture for most of ground cover types, the normal distribution based separability measure, Bhattacharyya distance (B-dis), is preferred. B-dis between class $i$ and class $j$ with mean vector $\mathbf{m}_i$ and $\mathbf{m}_j$, covariance matrix $\Sigma_i$ and $\Sigma_j$, respectively, is defined as [1]:

$$\mathrm{B-dis} = \frac{1}{8}(\mathbf{m}_i - \mathbf{m}_j)^t \left\{\frac{\Sigma_i + \Sigma_j}{2}\right\}^{-1} (\mathbf{m}_i - \mathbf{m}_j) + \frac{1}{2}\ln\left\{\frac{\left|(\Sigma_i + \Sigma_j)/2\right|}{\left|\Sigma_i\right|^{1/2}\left|\Sigma_j\right|^{1/2}}\right\} \tag{1}$$

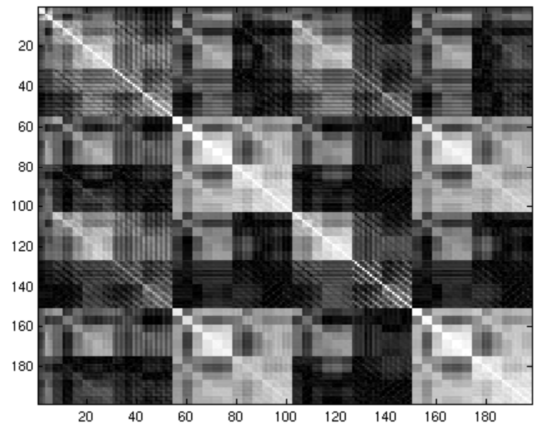


***Figure 1**. Correlation matrix for the 198 features in the order given in Table 1. (white = 1 or -1; black = 0.)*

However, if the class data is far away from normal distribution, as we can see in the example given in Fig. 2, the feature selection results using B-dis will not be reliable and will not lead to high classification accuracy. Another issue to consider is that the separability measure is class pairwise calculation. With more than two classes, the highest average B-dis is normally used as the criterion for subset selection. This may not be effective since different class pairs can be separated most by different subsets of features. To address these problems, normality test for each class pair is introduced. The features which pass the test will be considered further using B-dis for final feature selection.

The classical statistics of skewness, $s$, and kurtosis, $k$, can be used to test normality [9]. They are defined as

$$s = \frac{\sum_{i=1}^{n}(x_i - m)^3}{(n-1)\sigma^{3/2}} \tag{2}$$

$$k = \frac{\sum_{i=1}^{n}(x_i - m)^4}{(n-1)\sigma^2} \tag{3}$$

where $x_i$ is the sample brightness; $n$ is the number of samples; $m$ is the samples mean; and $\sigma$ is the sample variance. If $x$ has a normal distribution, $s$ is expected to be 0 and $k$ is expected to be 3. For a given class pair, the features that provide $s$ value relatively close to 0 and $k$ value relatively close to 3 are selected.

Four difficult-to-separate classes, Nature Reserve, Residential Area, Grassland, and Pine Plantation, were selected in this study. The details of the classes are given in Table 2. Table 3 lists the feature selection outcomes.

## Classification

A progressive two-class decision classifier (pTCDC [10]) (or a Directed Acyclic Graph (DAG) [11]) was adopted for classification. With this scheme, a multiclass problem is converted into $M(M-1)/2$ ($M$ is the number of classes) independent two-class separations. At each decision node, only one class pair is considered. It provides the flexibility in using the decision rules and features required by individual class pairs.

A Gaussian maximum likelihood classification using the features listed in Table 3 was implemented in the pTCDC. The classification accuracy received was compared with the cases of using spectral data only and using all the spectral and spatial features, respectively. They are given in Fig. 3. It can be seen that using selected features for each class pair provides the highest overall classification accuracy.

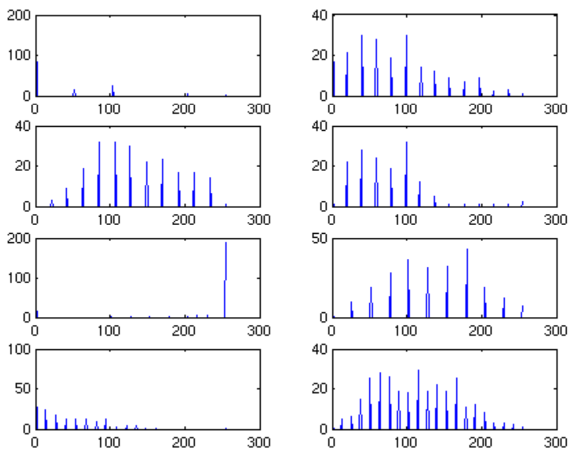

***Figure 2.*** *Histograms for 8 different classes ranging from water body, forest to buildings on the Contrast feature derived from band 1 with distance 1 in the horizontal direction.*

**Table 2: Number of training and testing pixels used in the experiment**

| Class No. | Class Name | No. of Training Data | No. of Testing Data |
|---|---|---|---|
| 1 | Nature Reserve | 220 | 143 |
| 2 | Residential Area | 240 | 168 |
| 3 | Grassland | 173 | 123 |
| 4 | Pine Plantation | 220 | 112 |

**Table 3: Feature selection results**

| Class Pair | No. of Features Selected | | Features Selected | B - dis |
|---|---|---|---|---|
| | By Normality | By B-dis | | |
| 1 & 2 | 14 | 3 | 4,6,116 | 3.48 |
| 1 & 3 | 19 | 1 | 4 | 10.63 |
| 1 & 4 | 11 | 6 | 4,8,19,64,110,114 | 3.31 |
| 2 & 3 | 18 | 2 | 3,4 | 6.12 |
| 2 & 4 | 28 | 2 | 67,116 | 4.44 |
| 3 & 4 | 15 | 2 | 4,115 | 8.34 |

## Discussion and Conclusion

Spatial features can add new information and used together with spectral data. However, not all the features are useful for individual mapping applications. Feature selection is important preprocessing.

For each class pair separation, a few features are often adequate. Two-class classification scheme is effective and easy to implement.

Skewness and kurtosis were introduced for normality test. Other methods are available, such as the Jarque-Bera test and the Lilliefors test. Further study will be on alternative normality testing.

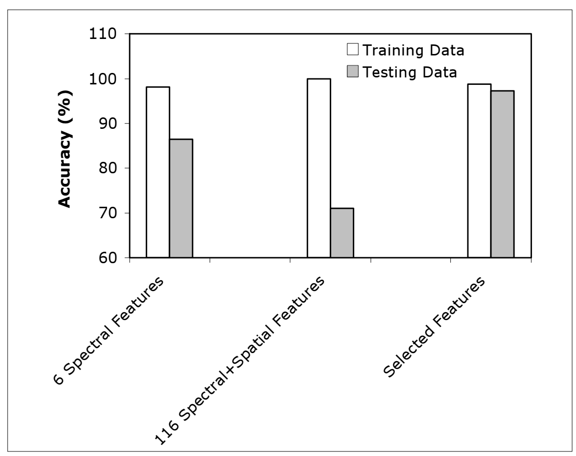


***Figure 3.*** *Classification accuracy comparison.*

## References

[1] John A. Richards and Xiuping Jia, Remote Sensing Digital Image Analysis, 3rd Ed., (Springer Verlag, 1999).

[2] X. Zhang, N.H. Younan and C.G. O'Hara, Wavelet Domain Statistical Hyperspectral Soil Texture Classification, IEEE Trans. On Geosci. and Remote Sensing, 43, 3, pg. 615-618. (2005).

[3] M. Torma and K. Luojus, Effect of Texture Measures to Separability of Land Cover Classes Using ERS SAR Images, IGARSS04, pg. 2684-2687.

[4] Y-W. Wang, Y-F. Wang, Y. Xue and W. Gao, A New Algorithm for Remotely Sensed Image Texture Classification and Segmentation, IGARSS'03, pg. 3534-3536.

[5] J. Mao and A.K. Jain, Texture Classification and Segmentation Using Multi-Resolution Simultaneous Autoregressive Models, Pattern Recognition, 25, 2, pg. 173-188. (1992).

[6] G.F. Hughes, On the Mean Accuracy of Statistical Pattern Recognizers, IEEE Trans. Infor. Theory, 14, pg. 55-63. (1968).
[8] Matlab, The MathWorks, Inc. 1984-2005.
[7] R. Haralick, K. Shanmugam and I. Dinstein, Textural Features for Image Classification, IEEE Trans. On Systems, Man, and Cybernetics, 3, pg. 610-621. (1973).
[9] K.V. Mardia, Tests of Univariate and Multivariate Normality. Handbook of Statistics, P.R. Krishnaiah (ed.), 1, pg. 279-320. (1980).
[10] X. Jia and J.A. Richards, Progressive Two-Class Decision Classifier for Optimization Class Discriminations, Remote Sensing of Environment, 63, 3, pg. 289-297. (1998).
[11] J. C. Platt, N. Cristianini and J. Shawe-Taylor, Large Margin DAGs for Multiclass Classification, Advances in Neural Information Processing Systems, 12, S.A. Solla, T.K. Leen and K.-R. Muller (eds.), MIT Press, pg. 547-553. (2000).

## Author Biography

*Xiuping Jia received the B. Eng. degree from Beijing University of Posts and Telecommunications, China, in 1982 and the Ph.D. degree in electrical engineering from The University of New South Wales, Australia, in 1996. She joined the University College, The University of New South Wales, Australian Defence Force Academy, Canberra, in 1988. Her research area is remote sensing and imaging spectrometry. She is the co-author of the remote sensing textbook, Remote Sensing Digital Image Analysis, (Berlin, Germany: Springer-Verlag, 3rd ed., 1999).*

.