Geoinformation for European-wide Integration, Benes (ed.)

The use of InSAR texture features in land use classification

Mahesh Pal & Paul M. Mather
School of Geography, University of Nottingham, Nottingham NG7 2RD, United Kingdom
pal@geography.nottingham.ac.uk , paul.mather@nottingham.ac.uk

Keywords: InSAR, coherence image, texture, decision tree

ABSTRACT: The planning of agricultural policy at European and national levels makes use of crop statistics derived from remote sensing data. The main aim of using these data is to obtain regular and reliable crop statistics during the growing season via a monthly crop status bulletin. Optical data are being used on an operational basis to provide agricultural statistics as early as possible in the year because, with the use of satellite imagery, large areas can be covered under the same atmospheric conditions at a very limited cost. But cloud cover poses the greatest restriction on the acquisition of data that may be required at different intervals. This limitation has been somewhat alleviated by the use of Synthetic Aperture Radars (SARs) which are essentially all-weather systems. A number of works have already reported the use of SAR intensity images for land use classification. Another SAR-based technique, called SAR interferometry, uses two single look complex (or SLC) SAR images which contain both amplitude and phase information. Coherence images can be produced from two SLC images, often as part of the interferogram generation process.
This study was carried to study the impact of the use of texture features derived from coherence and intensity images on land use classification accuracy, in combination with coherence and intensity images. For this study, five SLC images covering a period from 02/05/1996 to 20/09/1996 were selected. The classification problem involved the identification of seven land cover types; namely, wheat, potato, sugar beet, onion, peas, barley and carrots that cover the bulk of the area of interest. Interferometric processing was carried out using all five SLC images. Five intensity and three coherence images were obtained and it was found that only the tandem pair (02/05/1996 and 03/05/1996) provided good quality coherence images. Texture features based on the GLCM, Markov random fields, local statistics, and fractals were derived from all intensity and the coherence images for further use in image classification. As the number of input features become large for classification, a feature selection based on a statistical measure called Hotelling's T^2 was carried out to select the three best features for each intensity and coherence image. Five different data sets (1) five median filtered intensity images, (2) median filtered coherence and five intensity images, (3) median filtered coherence and three texture features of coherence images and five intensity images, (4) data set (1) with texture features and, (5) data set (2) with texture features were used in this study. Three different classifiers (ML, neural and decision tree) were used.
Results suggests that the classification based on data set (2) is about 9% more accurate than data set (1) with all the classification system used thus, suggesting that the incorporation of coherence information can be very useful for land cover classification, in combination with intensity images. By using the texture features of all the images (data set (5)) an overall classification accuracy of more than 82% was achieved, which suggests that the use of texture features derived from InSAR data can results in acceptable levels of accuracy being attained for land cover classification. This study also suggests that a decision tree classifier can be used very effectively for land cover classification studies involving radar data.

1 INTRODUCTION

Interferometric Synthetic Aperture Radar (InSAR) is an airborne or spaceborne system that acquires microwave images of a given area on the ground from two different positions. Microwave images have two components. The first is the intensity of the backscatter. Intensity images are widely used in remote sensing. Often, several intensity images are averaged

to produce a multi-look intensity image, which has less speckle noise than a single look image, though resolution is degraded. The second component of a microwave image is the phase of the signal. Comparison of phase differences on a pixel-by pixel basis for two microwave images taken from two different points provides information about the surface elevation of the pixel. In fact, phase measures the altitude of the platform carrying the SAR instrument, and surface elevation is computed from a combination of orbital parameters that gives the position of the spacecraft and phase information recorded by SAR.

Synthetic Aperture Radar (SAR) has a number of advantages for land observations over visible/infrared systems due to its all-weather, day and night image acquisition capabilities. A SAR based technique called interferometry has attracted interest due to its use in the generation of DEMs, detection of terrain displacement, monitoring of ice sheet motion and measuring ocean surface currents, etc. SAR interferometry uses a pair of complex SAR images (called single look complex, or SLC images) of the same area taken from two different viewpoints. The image thus obtained contains information on the phase and amplitude of the backscatter and can be used to generate interferograms. The phase component of the interferogram (responsible for the interferometric fringes) is the phase difference between two SAR images.

Coherence, which gives the degree of correlation of the phase component of the two images, and which is used to characterise the quality of interferograms, can be calculated on a pixel by pixel basis to produce a coherence image as a by-product of interferometry. These images are generally used for verifying the phase coherence of an interferogram. Over the past few years, a number of works reported in the literature show the potential uses of coherence images in mapping change detection, forest mapping, geological and ice mapping and land use classification.

This paper begins with a brief review of the operation of an InSAR system and the calculation of the coherence between successive SAR images is described. Various textures measures, feature selection as well as the classifiers used in this study are discussed in brief and finally the results of this study are discussed.

2 INTERFEROMETRIC SYNTHETIC APERTURE RADAR

Interferometric Synthetic Aperture Radar (InSAR) mapping was introduced by Graham (1974). Radar interferometry is a technique for extracting three-dimensional information about the earth's surface by using the relative phase difference of two coherent synthetic aperture radar images obtained by two receivers separated by a cross-track distance (termed the baseline) to derive an estimate of the earth surface height. The horizontal resolution (range and azimuth resolution) of the system is dictated by the SAR bandwidth (frequency range contained in a signal) and antenna length. These parameters can be selected so as to satisfy the topographic resolution requirements. The vertical accuracy of the system is ultimately limited by the wavelength used by the SAR, which, for microwaves, is on the centimetric scale. In addition, InSAR has the capability to provide all-weather performance, in contrast to optical sensors.

The phase difference ϕ between the signals received from the same surface element at the two antenna positions is

$$\Phi = 4\pi(r2 - r1)/\lambda \qquad (1)$$

Where λ is the SAR wavelength (which must be the same for the two observations), H is the altitude of platform, r1 and r2 are the distances between the radar antenna and the scatterers for platform positions. The parameter θ is the local incidence angle. The height of the point N can be determined from

$$Z = H - r1\cos\theta \qquad (2)$$

There are three main ways to acquire SAR interferometric data. These are along-track, across-track and repeat-track (multi-pass) interferometery. In both across-track and along-track interferometry two SAR antenna systems are mounted on a single platform, which is generally an aircraft because it has not yet been possible to mount two SAR systems on a satellites. Repeat-pass interferometry requires only one antenna mounted on a satellite. The data used in this study (table 1) were acquired by the ERS-1 and ERS-2 satellites, which carry identical SAR instruments.

Table 1. SLC (single look complex) used in the study

Sensor	date
ERS-1	02/05/1996
ERS-2	03/05/1996
ERS-2	07/06/1996
ERS-2	16/08/1996
ERS-2	20/09/1996

3 COHERENCE

In SAR interferometery, coherence is defined as a measure of the degree of resemblance of radar phases of two SAR images acquired from two different positions. The degree of correlation that exists between two SAR images is called the complex degree of coherence. The value of this coefficient var-

ies between zero and one, where a zero value means no interference, which implies that there will be no fringes on an interferogram derived from the area.

3.1 *Coherence Magnitude Estimation*

The complex coherence of two zero-mean complex signals g_1 and g_2 for stationary random processes is defined as (Born and Wolf, 1980):

$$\gamma = \frac{E\left[g_1 g_2^{*}\right]}{\sqrt{E\left[\left|g_1\right|^2\right]E\left[\left|g_2\right|^2\right]}} \qquad (3)$$

where g_2^{*} is the complex conjugate value of signal g_2 and E[ ] denotes the expectation value. The magnitude of γ ($|\gamma|$) is called the degree of coherence and the phase of γ is called interferometric phase. Under the assumption that the processes involved in the above equation are ergodic in mean[1] ensemble average is found by coherently averaging the complex values of N single look pixels. The coherence is then defined as:

$$\bar{\gamma} = \frac{\left|\sum_{i=1}^{N} g_{1,i} g^{*}_{2,i}\right|}{\sqrt{\sum_{i=1}^{N}\left|g_{1,i}\right|^2 \sum_{i=1}^{N}\left|g_{2,i}\right|^2}} \qquad (4)$$

An accurate estimate of coherence cannot be determined directly from this equation when we have a topographically-induced phase component in the data, which is the case for InSAR images. So a revised coherence estimate is given by:

$$\bar{\gamma} = \frac{\left|\sum_{i=1}^{N} g_{1,i} g^{*}_{2,i}\, e^{-j\phi_i}\right|}{\sqrt{\sum_{i=1}^{N}\left|g_{1,i}\right|^2 \sum_{i=1}^{N}\left|g_{2,i}\right|^2}} \qquad (5)$$

Where $e^{-j\phi_i}$ is a phase factor for the local imaging geometry.

In this study, three coherence images at one, thirty five and seventy day interval were generated. Due to the large base line length (>1000 m) and the time interval, the phase coherence obtained at 35 days and 70 days was poor. Hence, in the remainder of the study, only the 1-day coherence image and five intensity images produced during the interferometric processing of the single look complex images were used for land use classification. A median filter with a window size of 5×5 was used for speckle reduction in all intensity and coherence images.

[1] There are two way to calculate the mean of a random variable: 1.Time average: by integrating a particular member function over all time or 2. Ensemble average: average together the values of all member functions evaluated at some particular point in time. A random variable is ergodic if and only if (1) the time average of all member functions are equal, (2) the ensemble average is constant in time, and (3) the time average and the ensemble average are numerically equal. Thus, for ergodic random variables, time average and ensemble average are interchangeable

4 CLASSIFICATION

Three type of classifiers are used. These are the maximum-likelihood, multi-layer backpropagation neural network and the decision tree classifier. A number of studies have reported the use of neural classifier in remote sensing, but very few works have so far reported the use of decision tree methods in land use classification, especially its performance compared to neural network classifiers. The use of decision tree classifier in remote sensing is discussed by Swain et al. (1977), Friedl *et al*. (1996), Defries et al. (1998), Hansen et al. (1996, 2000), and Muchoney et al. (2000). This section provides a brief summary of the characteristics of the three classifiers used in this study.

4.1 *Maximum-Likelihood Classifier*

The Maximum likelihood Classifier (MLC) is based on the assumption that the classes are normally distributed in attribute space. MLC is a pixel-based method and can be defined as follows: a pixel with an associated observed feature vector X is assigned to class c_j or

$$X \in c_j \text{ if } g_j(X) \rhd g_k(X) \text{ for all } j \neq k, j,$$

$$k = 1, \ldots\ldots, N$$

For multivariate Gaussian distributions $g_k(X)$ is given by:

$$g_k(X) = \ln(p(c_j)) - \frac{1}{2}\ln\left|\sum_k\right| - \frac{1}{2}(X - M_k)^t \sum_k^{-1}(X - M_k) \qquad (6)$$

where M_k and $\sum_k$ are the mean vector and covariance matrix, and g_k is the discriminating function.

Implementation of the MLC involves the estimation of class mean vectors and covariance matrix using training patterns chosen from the images and based on the ground reference image of the study area.

4.2 *Artificial Neural Network Classifier*

For this study, a feed-forward artificial neural network (ANN) is used. This is the most widely used neural network models, and its design consists of one input layer, at least one hidden layer, and one output layer. Each layer is made up of nonlinear processing units called neurones or units, and the connections between neurons in successive layers carry associated weights. Connections are directed and allowed only in the forward direction, e.g. from

input to hidden, or from hidden layer to a subsequent hidden or output layer. Non-linear processing is performed by applying an activation function to the summed inputs to a unit. Back-propagation is a gradient-descent algorithm minimising the error between the output of the teaching input/output pairs and the actual network outputs. Therefore, a set of input/output pairs is repeatedly presented and the error is propagated from the output back to the input layer. The weights on the path back through the network are updated according to an update rule and a learning rate. ANNs are not specified by the characteristics of their processing units and the training or learning rule only. The network topology, i.e. the number of hidden layers, the number of units, and their interconnections, also has an influence on classifier performance. In this study we use the network architecture and number of patterns used for training as suggested by Kavzoglu (2001).

4.3 *Decision Tree Classifier*

In the usual approach to the classification, a common set of features is used jointly in a single decision step. A different approach is to use a multistage or sequential hierarchical decision scheme. The basic idea involved in any multistage approach is to break up a complex decision into a union of several simpler decisions, hoping that the final solution obtained this way would resemble the intended or desired solution. Hierarchical classifiers are a special type of multistage classifier that allows rejection of class labels at intermediate stages.

Classification trees offer an effective implementation of such hierarchical classifiers. Indeed, classification trees have become increasingly important

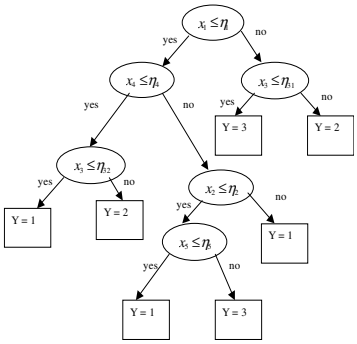

Figure 1. A classification tree for a five dimensional feature space and three classes. The x_i`s are the feature values, the $\eta_i$`s are the thresholds, and y is the class label.

due to their conceptual simplicity and their computational efficiency. A decision tree classifier has a simple form which can be compactly stored and that efficiently classifies new data. A decision tree classifier perform automatic feature selection and complexity reduction, while the tree structure gives easily understood and interpreted information regarding the predictive or generalisation ability of the data.

The decision tree is constructed by recursively partitioning a data set into purer, more homogenous, subsets on the basis of a set of tests applied to one or more attribute values at each branch or node in the tree. This procedure involves three steps: splitting nodes, determining which nodes are terminal nodes, and assigning class label to terminal nodes. The assignment of class labels to terminal nodes is straightforward: labels are assigned based on a majority or a weighted vote when it is assumed that certain classes are more likely than others.

A tree is composed of a root node (containing all the data), a set of internal nodes (splits), and a set of terminal nodes (leaves). Each node in a decision tree has only one parent node and two or more descendent node (figure 1). An observation vector is classified by moving down the tree and sequentially subdividing it according to the decision framework defined by the tree until a leaf is reached.

In this study, *See5.0* software (a univariate decision tree algorithm using the information gain ratio as an attribute selection measure) is used. The most important element of a decision tree estimation algorithm is the method used to estimate splits at each internal node of the tree and to reduce the problem of overfitting the training data, thus reducing the generalisation capability of the classifier. To reduce this problem, the original tree is pruned to reduce the classification errors when the data outside of the training set are to be classified. To address the problem of overfitting of the training data. *See5.0* uses error-based pruning which prunes the tree using a bottom-up approach, and by using the training data itself.

5 FEATURE EXTRACTION AND SELECTION

A number of methods have been developed to deal with spectral and spatial information, in order to achieve improved classification performance. In comparison with tonal measures, the definition of texture features appears more difficult. The main difficulty faced by the researcher is to define a set of meaningful features to characterisetexture properties.

Based on the exture descriptors available in the literature, four approaches are used in this study. The first approach uses the Grey-Level Co-occurrence Matrix (GLCM) (Haralick et al., 1973). The second approach uses the features derived from local statistics. The third approach is based on the fractal geometry of the image and the fourth approach is based on the Multiplicative Autoregressive Random (MAR) field model. Hotelling's T^2 statis-

tics is used to find the best four features out of the ten features of each intensity and coherence image, including the intensity and coherence image values.

5.1 *Texture features based on GLCM*

In this study, five indices are used as texture measures obtained from both coherence and intensity images. In what follows, g(i, j) denotes the (i, j)th entry in a normalised GLCM, and N_g denotes number of distinct grey levels in the quantised image.

1. Angular Second Moment (ASM):

$$ASM = \sum_i \sum_j [g(i,j)]^2 \tag{7}$$

2. Contrast (Con):

$$Con = \sum_{n=0}^{N_g-1} n^2 \left\{ \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} g(i,j) \right\} \tag{8}$$

3. Correlation (Cor):

$$Cor = \frac{\sum_i \sum_j (i,j) g(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y} \tag{9}$$

4. Inverse Different Moment (IDM):

$$IDM = \sum_i \sum_j \frac{1}{1+(i-j)^2} g(i,j) \tag{10}$$

5. Entropy (Ent):

$$Ent = -\sum_i \sum_j g(i,j) \log(g(i,j)) \tag{11}$$

5.2 *Feature based on local statistics*

Only one feature, "variance", is calculated based on local statistics. It can be calculated from the following formula using a moving window:

$$Variance = \frac{\sum (DN_{ij} - \mu)^2}{n-1} \tag{12}$$

where DN_{ij} represents the DN value of pixel at position (i, j), n is the number of pixels in a moving window and μ represents the mean of the moving window, which is calculated from:

$$\mu = \frac{\sum DN_{ij}}{n}.$$

5.3 *Textures based on fractal dimension*

The concept of fractal dimension can be useful in the measurement, analysis, and classification of shape and texture. A number of approaches exists in literature to estimate the fractal dimension (D). In this study the method based on box-counting proposed by Sarkar and Chaudhri (1994) is used.

This method is described as follows: consider that the image of size M×M pixels has been divided into grids of size s×s, where $M/2 \geq s > 1$ and L is an integer. Consider the image as a 3D surface with (x, y) denoting a 2-D position and the third co-ordinate (z) being grey level. The (x, y) space is divided into grids of size s×s. Thus for each grid there is a column of boxes of size $s \times s \times s'$, where s can be a multiple of the side length of a pixel in (x, y) and s' can be a multiple of the grey level in z-direction. If the total number of grey level is G then s' is calculated by

$$\left\lfloor \frac{G}{s'} \right\rfloor = \left\lfloor \frac{M}{s} \right\rfloor \tag{13}$$

Assign the numbers 1, 2, …., n in turn to each box in the column from bottom to the top. Let the minimum and maximum grey level of the image on the (i, j)th grid fall in boxes number p and k, respectively. Then the number of boxes needed to cover the surface on the (i, j)th grid is

$$n_r(i,j) = = p - k + 1$$

where

$$r = s/M$$

After taking contributions from all grids, the total number of boxes needed to cover the whole image with box size $s \times s \times s'$ is

$$N_r = \sum_{i,j} n_r(i,j) \tag{14}$$

N_r is counted for different values of s. The fractal dimension D is therefore estimated from the least square linear fit of log (N_r) against log (1/r). This method is computationally efficient and gives a good approximation to the boxes intersecting the image intensity surface. Counting N_r in this manner gives a better approximation to the boxes intersecting the image intensity surface, when there is sharp grey level variation in neighbouring pixels in the images.

5.4 *Estimation of texture parameters in the MAR model*

Frankot and Chellapa (1987) proposed the gaussian autoregressive random field models for the logarithm of radar image intensity in two-dimensions, which they called the lognormal multiplicative autoregressive (MAR) model. Initially the MAR concept was originally used to model image data, and the parameters of the model have been found to be highly correlated with the spatial distribution of the data. For this reason, they can be used as a texture descriptors for image classification (Solberg and jain., 1997). Three parameters of the MAR model are generally estimated and used for texture descriptors. These are the parameters θ, variance σ_w^2, and the mean value δ_q of the stationary random process q respectively. The parameter estimation method em-

ployed is least square estimation as suggested by Kashyap and Chellappa (1983) while the neighborhood defined by N = {(0,-1), (-1,-1), (-1,0)}and θ_r is an exponent weighting factor for neighborhood r, used to compute these parameters.

Let an image p(s), $s \in \Omega$, be represented by the following white-noise-driven multiplicative system:

$$p(s) = \prod_{r \in N} [p(s+r)]^{\theta_r} \cdot v(s) \quad (15)$$

Where Ω = {0, 1, ..., M-1}×{0, 1, ..., M-1}. N is the neighbourhood set defining model support, v(s) is a lognormal white-noise process referred to as the driving process, and s=(m, n), a 2-D index to an image. The random field p(s) is said to obey a lognormal MAR model if q(s) = ln p(s) and w(s) = ln v(s). Then equation 15 becomes:

$$q(s) = \sum_{r \in N} \theta_r \cdot q(s+r) + w(s) \quad (16)$$

For MAR models, the least squares estimates based on p(s) are as:

Covariance:

$$\sigma_w^2 = \frac{1}{M^2} \sum_{s \in \Omega} \left[q(s) - m_q - \theta^T z(s) \right]^2 \quad (17)$$

where

$$\theta = \left[\sum_{s \in \Omega} z(s) \cdot z^T(s) \right]^{-1} \left[\sum_{s \in \Omega} z(s) \left(q(s) - m_q \right) \right]$$

and mean

$$m_q = \frac{1}{M^2} \sum_{s \in \Omega} q(s) \quad (18)$$

where

$$z(s) = q(s+r) - m_q, \quad r \in N$$

Due to the variations in texture, different images will generally show different value of means, noise covariance, and parameter θ, and so it possible to use these three parameters to describe the texture of image being used for classification.

5.5 *Feature selection using Hotelling's T^2 statistic*

Several multivariate statistical test techniques can be used to determine the degree of discrimination between the classes in a given dataset, by using the means and co-variance matrices of the classes. Hotelling's T^2 statistic is a well known technique that can be used in a descriptive statistical test to estimate the discriminating power of a feature or relative importance of a feature.

Hotelling's T^2 statistic is used to test the null hypothesis that the (population) multivariate means of the two groups under study do not differ significantly. It provides a multivariate generalisation of the Student's t test and is related to the problem of how best to discriminate between two groups. T^2 is calculated from:

$$T^2 = \frac{n_1 n_2}{n_1 + n_2} . D^2 \quad (19)$$

$$D^2 = (m_1 - m_2)' \sum^{-1} (m_1 - m_2)$$

where D^2 is the measure known as Mahalanobis' D-squared, which measures the overall similarity between the two groups. $\sum^{-1}$ is the inverse matrix of the pooled co-variance matrix Σ, and m_1 and m_2 are the mean vectors for the groups, which contain n_1 and n_2 individuals, respectively.

The value of Hotelling's T^2 increases as interclass separation increases. The statistical significance of T^2 can be evaluated using a transformation to the F distribution. It should be noted that the number of observations need not be the same for the two samples, but the number of features must be the same.

Hotelling's T^2 is used to identify only three best texture features for each image (five intensity images and one coherence image) to be used for land cover classification. For intensity and coherence images, nine texture features including five grey-level co-occurance features two features from the MAR model (markov mean and covariance), one feature calculated from fractal geometry and one feature from first order statistics (variance) are used for feature selection. With intensity images the four best features selected and used for further classification are fractal dimension, the intensity image itself, contrast from GLCM and variance from first order statistics. For the coherence image, the four features selected are coherence, markov mean, correlation and entropy from GLCM (Table 2).

6 CLASSIFICATION RESULTS AND DISCUSSION

6.1 *Datasets used and Results*

For this study the different data sets used for classification are:

1. All five median filtered intensity images.
2. combination of filtered coherence image with all five intensity images.
3. A combination of coherence image and its three best texture features with all five median-filtered intensity images.
4. All five intensity images in combination with the texture features of all these five intensity images.
5. The coherence image, all five intensity images, and the three best texture features for each of the coherence and intensity images.

Table 2.Various features obtained by applying Hotelling's T^2 feature selection method and used in final classification process.

Image	Features selected	Hotelling's T^2 value
Coherence image	Correlation and entropy from GLCM and MAR mean	709.15
Intensity images	Variance from local statistics, contrast from GLCM and fractal dimension	476.18

Three different classifiers (section 4) were used. The user-defined parameters for the neural network are set according to the recommendations made by Kavzuglu (2001). Random sampling was used to collect the pixels with all data sets. The sampled pixels were divided into two parts, one of which was used for training and one for testing the classifiers, so as to remove the bias in using the same set of pixels during training and testing. Overall classification accuracy (in %) and Kappa coefficients are shown in table 3, using maximum likelihood, neural and decision tree classifiers (figure 2 and figure 3 gives classified images with data set 2 and 5 respectively).

Table 3.Classification accuracies and Kappa values (in bracket) with all five different datasets with three different classifiers

Data sets used	Overall classification accuracy (%) and Kappa value		
	Classifiers used		
	ML	NN	DT
1	59.2 (0.52)	60.8 (0.57)	69.9 (0.65)
2	68.7 (0.64)	69.9 (0.66)	77.8 (0.74)
3	70.6 (0.66)	73.7 (0.71)	77.9 (0.74)
4	68.7 (0.63)	73.1 (0.70)	78.6 (0.75)
5	77.1 (0.73)	82.9 (0.81)	82.7 (0.80)

6.2 *Discussion*

Comparison of the results shown in table 3 (data sets 1 and 2) suggests that the inclusion of coherence information with the intensity images results in an improvement in overall classification accuracy for all of the three classifiers used. The increase in accuracy with the maximum likelihood classifier is from 59.2% to 68.7% (Kappa value 0.524 to 0.635), the neural classifier shows a rise from 60.8% to 69.9% (Kappa value 0.572 to 0.664), while results from the decision tree classifier vary from 69.9% to 77.8% (Kappa value 0.65 to 0.741). The result shows that there is an increase of about 8-9% in accuracy when coherence information is added, irrespective of the classifier used. Evaluation of results presented so far in this study suggests that the decision tree classifier performs better than either the maximum likelihood or the neural classifiers. It gives a significant increase in classification accuracy of about 10% as compared to both of the other classification systems with data sets 1 and 2.

The study further suggests that the inclusion of texture measures helps to improve classification accuracy. For this study, the three best features are selected for each intensity and coherence image. Initially, data set 3, which includes texture features obtained from the coherence image, was used for classification. The results suggests that an improvement of 4% in classification accuracy is achieved by using the neural network classifier as compared to data set 2, while use of the maximum likelihood and the decision tree classifiers suggests no major improvement in classification accuracy.

When all five intensity images with their associated texture features (data set 5) are used, a significant improvement in classification accuracy is shown as compared to the data set 1. Classification accuracy increases by an amount ranging from 8.7% to 12.3% depending on the classifier used, thus suggesting the importance of texture features with InSAR intensity images in land use classification. Further studies were carried out after adding the coherence and its texture features with dataset 5 (data set 6). An increase of between 4 and 9% in classification accuracy as compared to data set 6 suggests that the coherence image provides discriminating information about the land surface and can be used effectively for land use classification in combination with intensity images obtained from interferometric SAR data. The highest accuracy obtained by this combination (data set 6) is 82.9% with a neural classifier, which is slightly higher than the decision tree classifier (82.7%).

7 CONCLUSIONS

A number of combinations of data sets derived from interferometric SAR intensity and coherence images have been evaluated for crop discrimination. The results obtained from five different datasets (table 3) shows that coherence image in combination with the intensity images provides additional discriminative power in land use classification studies. The highest accuracy obtained is about 82.9% while using 24 features, justifying the importance of the texture features, as suggested by earlier work by Dutra and Huber (1999) in land cover classification using interferometric SAR data. However, in practice, one often encounters the so-called dimensionality problem, i.e., with a fixed and relatively small sample size, the classification accuracy may actually decrease when the number of features is increased (Hughes, 1968). This means that the use of a larger number of features requires a corresponding increase in the number of training samples, so that the results obtained are reliable.

It has also been found that decision tree classifier achieves better results than the statistical and neural classifiers in almost all cases, except with dataset 5, where its performance is slightly inferior to that of the neural classifier. The same number of training data were used for the three classifiers for dataset 5, thus indicating the limitations of univariate decision tree with limited training data size as the number of features increases.

ACKNOWLEDGEMENTS

Mahesh Pal thanks the Association of Commonwealth Universities (ACU), London, for providing scholarship for this study. The authors would like to thank Dr. Taskin Kavzoglu for allowing the use of his feature selection programs. Computing facilities were provided by the School of Geography, University of Nottingham.

REFERENCES

Born, M. and Wolf, E. 1980. P*rinciple of Optics*. Oxford: Pergamon Press.

Defries, R. S., Hansen, M., Townshend, J. R. G., and Sohlberg, R. 1998**.** Global land cover classification at 8 km spatial resolution: the use of training data derived from Landsat imagery in decision tree classifiers. *International Journal of Remote Sensing,* 19, 3141-3168.

Dutra, L. V. and Huber, R. 1999. Feature extraction and selection for ERS-1/2 InSAR classification. *International Journal of Remote Sensing*, 20, 993-1016.

Friedl, M. A., Brodley, C. E., and Strahler, A. H. (1999) Maximizing land cover classification accuracies produced by decision tree at continantal to global scales. *IEEE Transactions on Geoscience and Remote Sensing*. 37, 969-977.

Frankot, R. T. and Chellappa, R. 1987. Log Normal random field models and their applications to radar image synthesis. *IEEE Transactions on Geoscience and Remote Sensing*, 25, 195-207.

Graham, L.C. 1974. Synthetic Interferometer Radar for topographic mapping. *Proceedings of IEEE*, 62, 763-768.

Hansen, M., Dubayah, R., and Defries, R. 1996. Classification trees: an alternative to traditional land cover classifiers. *International Journal of Remote Sensing,* 17, 1075-1081.

Hansen, M., Defries, R., Townshend, J. R. G., and Sohlberg, R. 2000. Global land cover classification at 1 km spatial resolution using a classification tree approach. *International Journal of Remote Sensing,* 21, 1331-1364.

Harlick, R.M., Shanmugam, K. and Dinstein, I. 1973. Texture features for image classification. *IEEE Transactions on System, Man, and Cybernetics*, 3(6), 610-621.

Kashyap, R. L. and Chellappa, R. 1983. Estimation and choice of neighbours in spatial-interaction models of images. *IEEE Transactions on Information Theory*, IT-29, 60-72.

Kavzoglu, T. 2001. *An Investigation of The Design and Use of Feed-forward Artificial Neural Networks in the Classification of Remotely Sensed Images*. PhD thesis. University of Nottingham, Nottingham, UK.

Muchoney, D., Borak, J., Chi, H., Friedl, M., Gopal, S., Hodges, J., Morrow, N., and Strahler, A. 2000. Application of MODIS global supervised classification model to vegetation and land cover mapping of Central America. *International Journal of Remote Sensing*, 21, 1115-1138.

Sarkar, N. and Chaudhuri, B. B. 1994. An efficient differential box-counting approach to compute the fractal dimension of images*. IEEE Transactions on System, Man, and Cybernetics*, 24, 115-120.

Solberg, A. H. S., and Jain, A. K. 1997. texture fusion and feature selection applied to SAR imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 35, 475-479.

Swain, P. H., and Hauska, H. 1977. The decision tree classifier: design and potential. *IEEE Transactions on Geoscience Electronics*, 3, 142-147.