A NON-PARAMETRIC CLASSIFICATION STRATEGY FOR REMOTELY SENSED IMAGES USING BOTH SPECTRAL AND TEXTURAL INFORMATION

Mukesh Kumar[1] and Douglas A. Miller[2]
[1] Dept. of Civil and Environmental Engineering, [2] EMS Environment Institute
The Pennsylvania State University, USA
muk139@psu.edu, miller@essc.psu.edu

ABSTRACT
A classification strategy which does not require *a priori* assumptions about the statistical distribution of training pixels in each class is proposed. This method uses an indicator kriging approach in feature space to classify remotely sensed images incorporating both spectral and textural information of bands. Texture information is used only in cases where spectral information is not sufficient to resolve the assignment of the pixel to a class. Application of the proposed methodology on a remotely sensed natural scene shows an improvement in the overall classification accuracy with respect to the case when the scenes are classified by the traditional supervised Gaussian maximum likelihood classification (GMLC) method using either spectral band only or using both spectral and textural bands. A marked improvement in classification accuracy is obtained particularly for the classes for which the GMLC's assumption of multivariate normal distribution of training pixels in a class fails miserably.

KEY WORDS
Remote sensing, Image Classification, Indicator Kriging.

1. Introduction

Remotely sensed data are widely used for thematic map generation by performing multispectral digital classification. Traditional image classification methods use only spectral information (pixel values) as a basis to analyze and classify remotely sensed images (also termed *per-pixel* classification methods). However, if pixels belonging to two different classes possess similar brightness values, they can be distinguished based on their textural information [1]. Several efforts have been made to incorporate texture into the classification process with varying degrees of success [2], [3]. The reason for varied successes stems from the fact that whenever only spectral information is sufficient to assign a pixel to a particular class, added texture information is either trivial or contributes to confusion in classification thus decreasing its accuracy. So it is imperative to use texture data intelligently such that it is used only as an information source rather than noise.

More often than not, the classification techniques used are parametric in nature. For example GMLC, which is one of the most common supervised classification methods, assumes that the probability distribution function (PDF) of pixel values in each training class is Gaussian in nature, i.e. they follow normal distribution. However, the assumption of normality may not be valid in several practical applications. For instance when the PDFs of the classes may be skewed or have a double peak (in the case of heterogeneous classes). It may also be very difficult to determine the probability distribution of a class based on a limited number of training pixels. The existence of cases where the assumption for parametric classification schemes is not valid has led to the continuing search for a classification strategy which is free of *a priori* assumptions about the PDFs of training pixels in a class.

Accounting for the shortcomings of traditional classifiers viz. *a priori* parametric assumptions and an improper use of texture information – as highlighted in previous two paragraphs – this paper presents an attractive alternative classification strategy to conventional statistical classifiers which does not require any assumptions about distribution and uses both spectral and textural information *intelligently*.

The paper is organized in five sections. Section 2 presents, sequentially, the steps performed to implement the proposed methodology. The respective subsections describe the algorithmic details of corresponding steps. Section 3 discusses the application of the proposed methodology on a test data set. In Section 4 the result obtained from the proposed methodology has been compared to that obtained using the GMLC method. Section 5 is a summary of this research effort.

2. Approach

A flowchart depicting the sequence of steps performed in the proposed methodology is shown in Fig. 1. The following subsections describe the steps and algorithmic details of the proposed method.

2.1 Texture Band Extraction

At a simple level texture can be thought of as the variability in tone within a neighborhood, or the pattern of

spatial relationships among the gray levels of neighboring pixels [4]. Texture

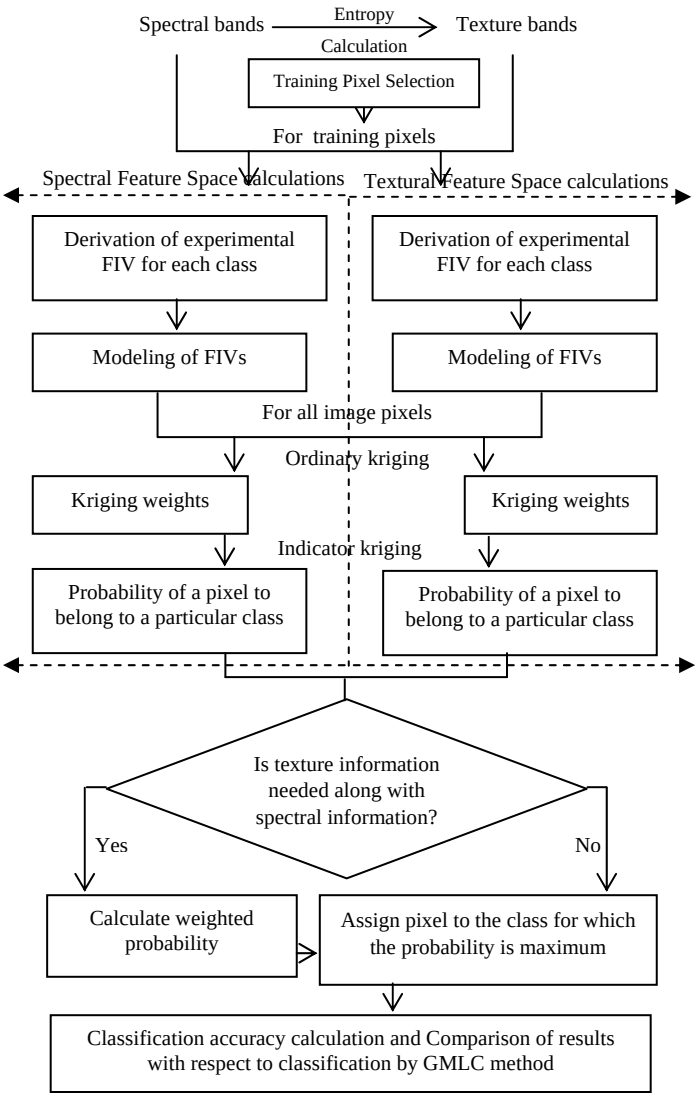

Fig. 1: Procedural steps in proposed classification strategy

can be modeled as a grey level function using one of the numerous statistical approaches like gray level difference histogram (GLDH) [5], gray level co-occurrence matrix (GLCM) [1] and wavelet transform [6]. Each of these methods has some inherent advantages and disadvantages in terms of their computational efficiency and classification of texture patterns. In this paper, GLCM method is used to calculate texture as it performs best in terms of higher texture discrimination accuracy [7]. Texture is quantified by entropy which is a measure of the degree of disorder or heterogeneity in an image [1] and is calculated as

$$entropy = -\sum_{i}\sum_{j} p(i,j)Log(p(i,j)) \qquad (1)$$

where, $p(i,j)$ is the average of normalized frequency taken on all directions.

2.2 Classification Algorithm

The proposed method is a supervised classification approach. Basic steps in the proposed methodology include: training pixel selection, defining an experimental frequency-in-interval variogram (FIV), modeling a FIV for each class, kriging and calculation of probability of assignment of a pixel to a particular class and finally, listing and evaluating the criteria when texture will be used along with spectral information in resolving the assignment of a pixel to a particular class. The following subsections describe the steps and algorithmic details of the classification algorithm.

- **Training Pixel Selection:** Specific sites that represe-t areas of known classes *a priori* on the ground are located on the image. These sites are called training sites and pixels belonging to these specific locations are called training pixels. As suggested by [8], random pixels are chosen from within a training area rather than using contiguous blocks. Once the training sites corresponding to different classes are located, their spectral and textural properties are used to classify the image. In the proposed methodology, all subsequent analyses are performed in feature space. A synthetic but representative N_b (= 3) dimensional feature space with total number of classes, N_c = 4, is shown in Fig. 2. It is obvious from Fig. 2 that pixels which belong to the same class are closer and those from different classes are farther separated. This implies that if two pixels are near to each other in feature space, then it is more probable that they will belong to the same class. On the contrary, if they are further apart then they are expected to belong to different classes. Based on the previous argument, the distribution of training pixels of each class in feature space is modeled. The modeled

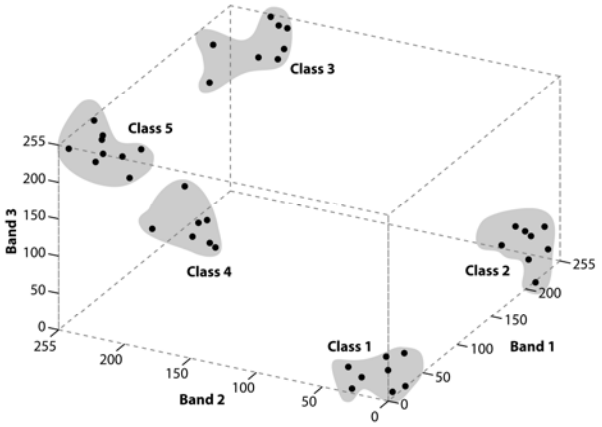

Fig. 2: Schematic Diagram of a feature space

distribution is eventually used to calculate an assignment probability for a particular pixel to the respective classes. Training pixel distribution in feature space is captured using the FIV metric. FIV is obtained for each class in both spectral and textural feature space separately.

- **Fraction-In-Interval Variogram:** The FIV metric is

FIV metric is similar to a traditional semivariogram in the sense that both quantify the spatial autocorrelation in the data, showing that the variation in data is small within a small region and increases with distance until it stabilizes at a distance called the range of the variogram [9]. However, though a semivariogram is generally defined in a geographical or image space, the FIVs define spatial autocorrelation in the data in a feature space. In earlier efforts that used kriging for image classification, the variogram was calculated in the image space within a rectangular window for each training land-cover type to perform texture classification [10], [11]. The analysis was based on the premise that pixels are spatially autocorrelated in image space and are expressed as unique semivariograms for each class. However, computation of the variogram function is difficult in the image space for highly heterogeneous areas and for classes, like a winding river, which are only a few pixels wide and which often change direction [12]. Also a large number of observations are needed in image space to take into account the anisotropy of classes. More so, these strategies are essentially capturing only the texture information using variograms in image space. On the whole, classification strategies which use variogram calculations in image space have limited flexibility.

In the proposed methodology, a modified form of the traditional variogram called FIV is derived in spectral and textural feature space instead of in image space. FIV, $\gamma_c(h)$ is calculated in a feature space as

$$\gamma_c(h) = 1 - \frac{nc(h)}{n(h)}$$

$$h_{(+/-)} = \frac{\max DistP - \min DistP}{Nh} \text{ x } i_{+/-} \qquad (2)$$

where, nc(h) is the number of distinct training pixel pairs such that both belong to class c and are $> h\_$ and $<= h_+$ apart in feature space and n(h) is the number of distinct training pixel pairs such that at least one training pixel in the pair belongs to class c and are $> h\_$ and $<= h_+$

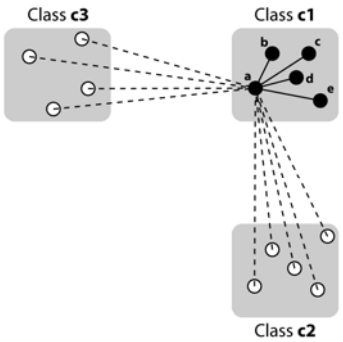

Fig. 3: Calculation strategy for $\gamma_c(h)$ in a representative two dimensional space

apart. h_+ and $h\_$ are the either extremes of the lag *h*. In the calculation of lag *h*, maxDistP is the maximum distance between any two pixels in a feature space and minDistP is the minimum distance and i varies from 1 to Nh where Nh is the number of lag intervals. Fig. 3 depicts a representative situation using synthetic data for a two dimensional feature space. nc(h) is the number of lines that connect pixels which belong to class c1 (these are shown as solid lines in Fig. 3) and have a length $> h\_$ and $<= h_+$. Similarly, n(h) is the number of lines connecting a pixel which belongs to class c1 to any other pixel (these include both the solid and dotted lines shown in Fig 3) and has a length $> h\_$ and $<= h_+$. $\gamma_c(h)$ is calculated using equation (2). Note that for better clarity and interpretability of Fig. 3, only the lines emanating from pixel **a** are shown. In fact, lines emanating from every pixel in c1 (i.e. from pixel **a**,**b**,**c**,**d** and **e**) are considered in the calculation of nc(h) and n(h). In this example, class c1 has 5 training pixels and if there are 14 training pixels in all (belonging to class c1, c2 and c3), then nc(h) corresponding to class c1 and lag h will be equal to the number of pixel pairs out of the total 10 ($= {}^5C_2$) pairs (these pairs have both pixels belonging to class c1) such that the distance between pixels in the pair is $> h\_$ and $<= h_+$. Similarly n(h) will be equal to number of pixel pairs out of the total of 55 ($= {}^5C_2 + 5*(14-5)$) pairs (these pairs have at least one pixel belonging to class c1) such that the distance between pixels in the pair is $> h\_$ and $<= h_+$. Obviously, if h is small, all pixel pairs with distance between pixels in the pair being $> h\_$ and $<= h_+$ will belong to same class. In that case $n(h) \approx nc(h)$ which implies $\gamma_c(h) \approx 0$. Conversely for a large h, all pixel pairs with distance between pixels in the pair being $> h\_$ and $<= h_+$ will belong to a different class. In that case $nc(h) \approx 0$ which implies $\gamma_c(h) \approx 1$. $\gamma_c(h)$ is therefore a spatially distributed bounded structure in feature space associated with a given land cover class c. The value of $\gamma_c(h)$ varies between 0 and 1. For a particular class c, $\gamma_c(h)$ is calculated for all possible Nh lags. The value of Nh is selected large enough to smoothly model $\gamma_c(h)$. The ordered set of values $\gamma_c(h1), \gamma_c(h2) \ldots\ldots \gamma_c(Nh)$ are calculated at $h1$, $h2 \ldots Nh$ respectively and this summarizes the spatial relation in training data. The calculated FIVs (also called an experimental FIV henceforth) have more or less an erratic appearance due to sampling fluctuations. To get FIV values at unsampled lags, a smooth modeled variogram representing regional variation is obtained. Functions used for modeling a variogram are such that the negative of $\gamma_c(h)$ is a positive definite function [13]. Some generally accepted model functions used for fitting experimental variograms are called authorized models viz. spherical, linear, Gaussian, exponential etc. [9]. More often than not, experimental FIVs do not consist of a single spatial structure. In such cases, different authorized model equations are added to yield the FIV model that

best fits the calculated value of $\gamma_c(h)$. Each particular structure in such a nested model can consist of different range and structural parameters representing varied scales of spatial dependence.

- **Kriging and Classification Probability:** Utilizing the feature space spatial correlation structure inherent in FIVs, kriging provides a best linear unbiased estimate of unmeasured values calculated using the weighted values measured in a local neighborhood. Based on a limited number of observations, the utility of entire the analysis rests upon the derivation of a probability distribution [14]. The kriged estimate at x_0 using the observed values at x_1, $x_2 \ldots x_n$ is evaluated using the following equations

$$\sum_{i=1}^{i=n} \lambda_i \gamma_c(x_i x_0) + \psi(x_0) = \gamma_c(x_i x_0)$$

where $\sum_{i=1}^{n} \lambda_i = 1$ (3)

$\gamma_c(x_i x_0)$ in (3) is the FIV magnitude at h = $|x_i - x_o|$. Note that the reliability of this entire analysis hinges on the appropriateness of variograms and so proper modeling of FIV as discussed in the previous section is quite crucial. n+1 equations are solved for all MxN image pixels and corresponding to N_c FIV configurations which uniquely define each class in feature space. A linear system of equations in n+1 independent variables for all the (MxN) pixels present in the image can be assembled in form of a matrix algebra computation as shown below in equation (4)

$$\underbrace{\begin{bmatrix} \gamma_c(x_1,x_1) & \gamma_c(x_1,x_2) & \ldots & \gamma_c(x_1,x_n) & 1 \\ \gamma_c(x_2,x_1) & \gamma_c(x_2,x_2) & \ldots & \gamma_c(x_2,x_n) & 1 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \gamma_c(x_n,x_1) & \gamma_c(x_n,x_2) & & \gamma_c(x_n,x_n) & 1 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix}}_{\mathbf{A}} * \underbrace{\begin{bmatrix} \lambda_{1,0} & \lambda_{1,1} & \ldots & \lambda_{1,MxN-1} \\ \lambda_{2,0} & \lambda_{2,1} & \ldots & \lambda_{2,MxN-1} \\ \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots \\ \lambda_{n,0} & \lambda_{n,1} & \ldots & \lambda_{n,MxN-1} \\ \psi(x_0) & \psi(x_1) & \ldots & \psi(x_{MxN-1}) \end{bmatrix}}_{\lambda_c} = \underbrace{\begin{bmatrix} \gamma(x_1,x_{img[0]}) & \gamma(x_1,x_{img[1]}) & \ldots & \gamma(x_1,x_{img[MxN-1]}) \\ \gamma(x_2,x_{img[0]}) & \gamma(x_2,x_{img[1]}) & \ldots & \gamma(x_2,x_{img[MxN-1]}) \\ \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots \\ \gamma(x_n,x_{img[0]}) & \gamma(x_n,x_{img[1]}) & \ldots & \gamma(x_n,x_{img[MxN-1]}) \\ 1 & 1 & \ldots & 1 \end{bmatrix}}_{\mathbf{b}} \quad (4)$$

x_i in (4) is the position vector of i^{th} training pixel in N_b dimensional feature space. Similarly, $x_{img[i]}$ denotes the position vector of the i^{th} image pixel. Note that x_is and $x_{img[i]}$s are vectors of size N_b*1. Matrix algebra performed in (4) is repeated for each class. Since, there are different FIVs for each class (details in section II-B2), the values of $\gamma_c(h)$ terms in (4) are different for different classes even when the value of h is the same. The solution to (4) is a two dimensional vector λ_c of size n*(M*N). Each column of λ_c, of size n*1, is the vector of weights on a particular pixel exerted by all the training pixels. It signifies the degree of influence of the FIV on that particular pixel. For each pixel, N_c vectors of weights are obtained. If the pixel is closest to the center of training pixel cluster c, the column vector belonging to vector λ_c corresponding to a particular image pixel is chosen for further calculations.

Note that kriging does not assure positive weights. Negative weights are assigned during kriging when the value of the data to be interpolated has a larger magnitude than the rest of the data points being used in the interpolation. Since FIV is bounded and can only vary between 0 and 1, negative weights are generally not encountered. However, due to deviations in the modeled FIV from the experimental FIV, the modeled function can sometimes produce FIV values exceeding 1 at higher lags. This may lead to assignment of very small negative weights. These weights are set to zero.

Indicator kriging uses the ordinary kriging weights λ_c and class information of each training pixel (TP) to estimate the probability of assignment of an image pixel to a particular class. Indicator Kriging makes no assumption about the theoretical distribution of the data [13]. Defining an indicator variable as

For (j=1,2,….,N_c)

For (i=1,2,…..,n)

$$I_{j,i} = \begin{cases} 1 & \text{if } TP \in c_j \\ 0 & \text{if } TP \notin c_j \end{cases} \quad (5)$$

The weighted average of values of indicator variables is an estimate of the probability that a pixel belongs to class j [11] and is given by

For (j=1,2,…..,N_c)

$$P_j = \sum_{i=1}^{n} \lambda_i I_{j,i} \quad (6)$$

Once the probability of an image pixel belonging to a particular class has been obtained in both spectral and textural feature space, a metric is needed that would finally assign an image pixel to a class based both on spectral and textural information.

- **Critereon for use of texture:** The metric of concern is given by

$$\text{Robustness} = R = \sum_{i=1}^{N_c} (\max P - P_i)^2 \qquad (7)$$

where maxP is the largest of the probabilities, P_i, of a particular pixel to belong to any class. Robustness can be calculated for each pixel corresponding to both spectral and textural feature space. The assignment of a pixel to a class j is done purely on the basis of P_j(spectral) i.e. probability calculated in spectral feature space, whenever spectralRobustness (specR) > texturalRobustness (texR). In that case if (P_j(spectral) = maxP) >> $\forall$ P_i(spectral) for i $\neq$ j, then the pixel is assigned to class j. However in cases of small spectralRobustness, i.e. whenever it is difficult to distinctively assign a pixel to a class entirely on the basis of a probability calculation performed in spectral feature space, the pixel is assigned to the class j for which the weighted probability is a maximum. The weighted probability, wP_j is calculated by

$$wP_j = \frac{P_j(spec.)*\sqrt{specR} + P_j(tex.)*\sqrt{texR}}{\sqrt{specR} + \sqrt{texR}} \qquad (8)$$

where $P_j(spec.)$ and $P_j(tex.)$ are the probability of assignment of a pixel to a particular class j, in spectral and textural feature space respectively. Small spectralRobustness can either be due to all the P_is being small or because of P_is (for all i $\neq$ j) is close to P_j = maxP. The robustness strategy is depicted in Fig. 4. Note that in

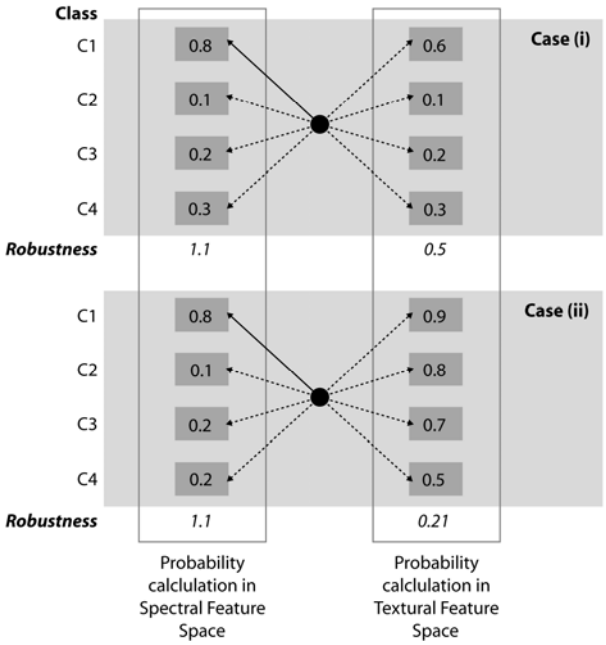

Fig. 4: Robustness calculation (i) maxP in spectral space > maxP in textural space (b) maxP in spectral space < maxP in textural space, however the robustness of spectral space is higher. Note that in both the cases, image pixel (denoted by a circle in this figure) is assigned to class 1.

Fig 4-case(i), the larger the maximum probability of assignment of a particular pixel to a class in a feature space, the greater is the robustness. However, Fig 4-case(ii) shows that in some cases, just the comparison between maximum probabilities of the assignment of the pixel to any class in two feature spaces is not a sufficient measure in classification. It is obvious from Fig 4-case(ii) that though in the textural feature space, a pixel has larger maximum probability, but it also faces more confusion because of a greater probability that it will get assigned to other classes. This reduces its robustness. In contrast, the maximum probability of TP to get assigned to a class in blue feature space is larger relative to its probability of assignment to other classes resulting in higher robustness. To sum up, both spectral and texture information is used to produce the classified map but texture is used selectively depending on the robustness criteria. Classification accuracy is calculated to ascertain the performance of the proposed methodology.

2.3 Accuracy Analysis

To check the classification accuracy two sources of information are needed. The first is the classified image and the second is the reference image having training and test pixels. Corresponding to each test pixel position, class value is gathered from the derived classified image and matched using an error or confusion matrix [15]. The accuracy of classification is estimated from all the available samples – both training and test sets – using the Khat index k. This index uses all the elements of the confusion matrix to measure how well the remotely sensed classification agrees with the reference data [16]. The overall Khat statistic is defined as

$$k = \frac{(P_o - P_c)}{1 - P_c} \qquad (9)$$

where P_o and P_c are overall accuracy and chance agreement respectively. In the present study, classification of the AOI has been performed using the proposed method while considering both spectral and textural information. The results are then compared to classification obtained using traditional GMLC algorithm [17], [16] with only spectral bands and also with both spectral and textural bands. For overall or individual class agreements, pairwise statistical tests using a Z-statistic [16] is performed to assess the significance of any difference observed between classifications performed using different techniques. At the 95% confidence level, a classification "a" is considered to be significantly better (/worse) than classification "b" if Z_{ab} > 1.96 (/< -1.96). If the absolute Z_{ab} value is below 1.96, then classifications "a" and "b" are considered statistically similar. If one wishes to choose the confidence interval of 99%, the corresponding critical Z_{ab} value is 2.58.

As stated earlier, one of the key points of this methodology is no *a priori* assumption about the distribution of training data. It still needs to be verified – how the proposed methodology performs when the distribution of training pixels of a class doesn't follow multivariate normal distribution (as assumed by GMLC

classifier). In order to verify the aforesaid hypotheses, it is imperative to first test the multivariate normality of training classes.

2.4 Test of Multivariate Normality

Mardia's test [18] of multivariate skewness and kurtosis assesses the degree to which multivariate data deviates from multinormality. Assuming $\mathbf{z}_i$ to be a vector representing the i-th training pixel of class c. The dimension of vector $\mathbf{z}_i$ will be N_b*1. Let $\bar{z}$ be the vector of sample means, and $\mathbf{S_z}$ be the sample covariance matrix. If there are m training pixels in class c then the multivariate skewness (*Sk*) and kurtosis (*Ku*) are estimated by

$$Sk = b_{1,N_b} = \frac{1}{m^2}\sum_{i-1}^{m}\sum_{j=1}^{m}((z_i - \bar{z})^T S_z^{-1}(z_j - \bar{z}))^3$$

$$Ku = b_{2,N_b} = \frac{1}{m}\sum_{i=1}^{m}((z_i - \bar{z})^T S_z^{-1}(z_i - \bar{z}))^2 \qquad (10)$$

[18] also discusses different estimates of *Sk* and *Ku* for the case when m is large.

3. The Experiment

As a test application of the proposed strategy, a remotely sensed scene of a mixed agriculture/forestry landscape in the Indian Pine Test Site in Northwestern Indiana imaged by Airborne Visible-Infrared Imaging spectrometer

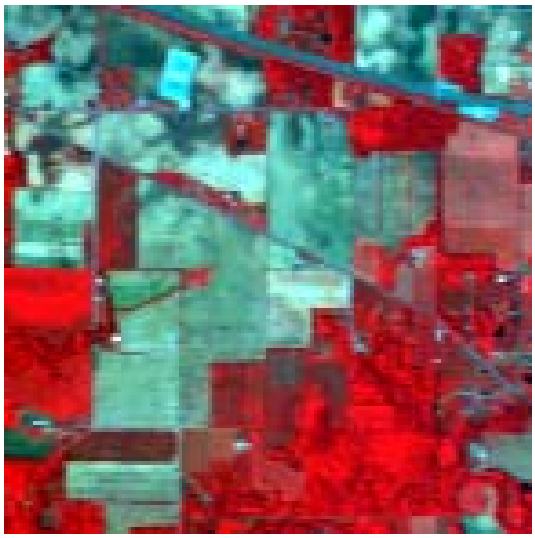

Fig. 5: False color composite for AVIRIS acquired scene (using bands 6, 3 and 1)

(AVIRIS) [19] is classified and analyzed. Fig. 5 shows

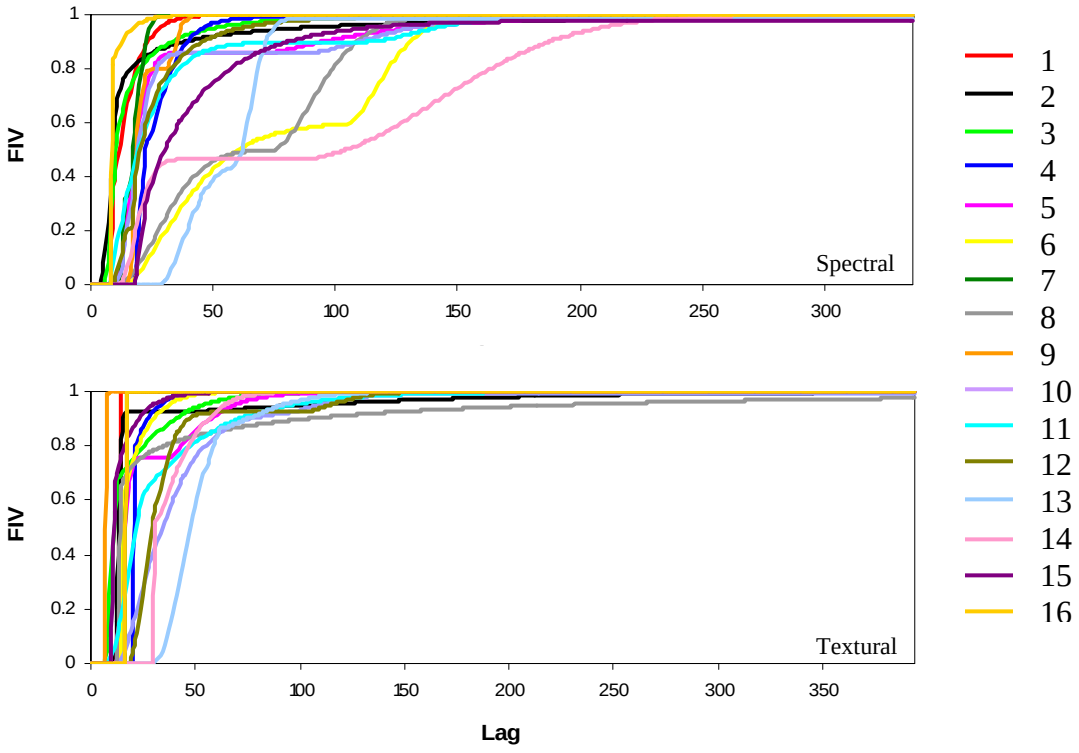

Fig 6: Modeled FIVs (a) Spectral feature space (b) Textural feature space. Note that actual Lag is equal to Lag shown in the graph divided by 450

false color composite (FCC) of the scene. Note that though AVIRIS has 224 bands for each scene, only 9 bands equally distributed over the entire spectral region are used in the present analysis as this simplifies the logistics while still being sufficient to demonstrate efficacy of the method. Though several efforts have focused on selection of the best combination of bands that lead to higher classification accuracy, in this paper attention is focused on a new classification algorithm. We refer the reader to [20] for more details on feature selection with AVIRIS data. The image contains 145 rows by 145 columns of pixels. Overall 2603 and 1301 test and training are selected respectively. FIV is obtained for all N_c classes in both spectral and textural feature space using eqn. (2). In accordance with the discussion in section 2.2, combinations of authorized models are used to model these experimental FIVs. Modeled FIVs are shown in Fig. 6. These modeled functions are used for the calculation of kriging weights using eqn (3). The calculated kriging weights are eventually used to calculate the probability of assignment of each pixel to a class using eqn (6).

4. Results

The classified image obtained by implementing the proposed methodology on the remotely sensed image is shown in Fig. 7. Scene is classified into 16 classes viz. Alfalfa (1), Corn-notill (2), Corn-min (3), Corn (4), Grass/ Pasture (5), Grass/ Trees (6), Grass/ Pasture-mowed (7), Hay-windrowed (8), Oats (9), Soy-notill (10), Soy-mintill (11), Soy-clean (12), Wheat (13), Woods (14), Bldg-grass-trees-drives (15) and Stone-steel towers (16). A similar classification map is also derived using (a) the GMLC classifier with only spectral bands and b) GMLC classifier with both spectral and textural bands. The same training and test pixel set is used for each experiment.

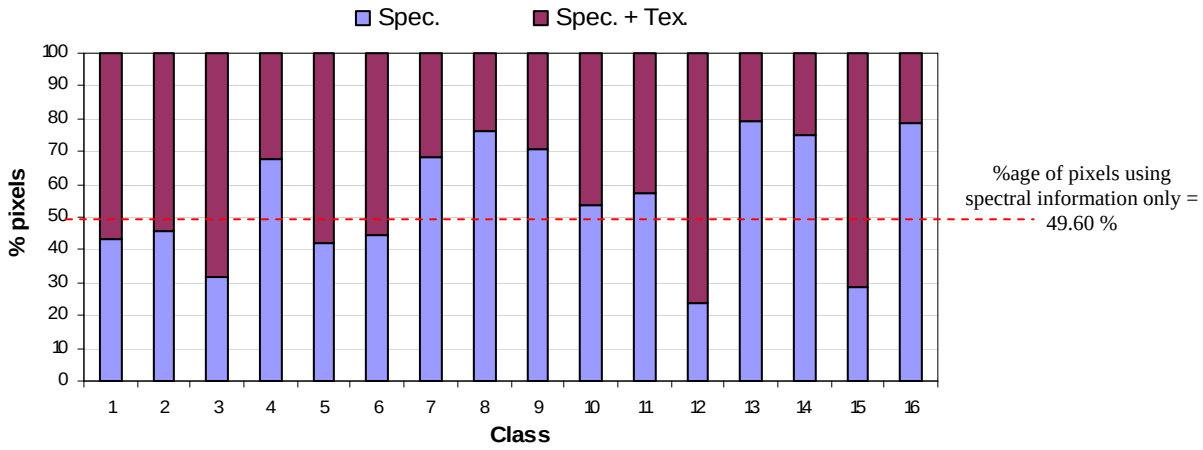

Fig. 8: Percentage of times only spectral information is used to assign a pixel to a class

Accuracy analysis of the respective classified maps is performed based on the discussion in Section 2.3. Kappa accuracy is calculated for each class and also for the overall classified image. Calculated values are tabulated in Table I. Relative comparison of the performances of various classification strategies are performed by using the Z–statistic. A five percent increase in overall accuracy of test pixels is obtained by using GMLC(S+T) with respect to GMLC (S). The proposed method further increases the accuracy by 2 percent. Also note that, the first column in Table I lists the group of which the training pixels of the corresponding classes are members to. A class falls in group 1 if it's training pixels follow a multivariate normal distribution and they fall in group 2 if they drastically violate the multivariate distributional assumption of GMLC. Identification of classes into two groups was performed based on the analysis in Section 2.4. Four classes are found to have statistically significant Z-statistics for classification accuracy. Interestingly these are the classes for which the normality distributional assumption for training pixels is violated. For some classes, like class 7, 9 and 13, GMLC(S) outperformed GMLC(S+T) and the proposed method. Fig. 8 shows the percentage of pixels which are assigned to a particular class based just on spectral information (i.e. the case where spectralRobustness > texturalRobustness). The percentage of the total pixels in the scene that are assigned a particular class based just on spectral information only is around 49.6%. This means that around half of the pixels in the image don't need texture information for getting classified to a particular class. Obviously, this percentage will be dependent on the heterogeneity of the classes. Expectedly, for highly homogenous classes like Hay, Woods, Wheat and Stone-steel towers, a large percentage of pixels get classified based just on spectral information.

Table I: Comparison of proposed methodology (PM) to GMLC strategy

		Training Pixels					Test Pixels				
Group	Class	GMLC(S) (a)	GMLC(S+T) (b)	PM (c)	Z_{ab}	Z_{ac}	GMLC(S) (a)	GMLC(S+T) (b)	PM (c)	Z_{ab}	Z_{ac}
1	1	0.843	0.859	0.859	0.854	0.85	0.68	0.813	0.846	1.424	2.06
2	2	0.834	0.85	0.860	0.68	0.921	0.606	0.791	0.829	3.739	5.36
1	3	0.862	0.878	0.949	0.159	0.489	0.672	0.822	0.888	0.691	1.078
1	4	0.936	0.9539	0.962	0.359	0.351	0.894	0.941	0.933	0.645	0.772
1	5	0.81	0.825	0.829	0.133	0.190	0.698	0.782	0.785	0.343	0.453
2	6	0.963	0.981	0.994	0.768	1.238	0.902	0.948	0.955	1.220	1.7
1	7	0.996	0.95	0.95	-1.49	-1.49	0.966	0.901	0.949	-1.664	-1.151
2	8	0.992	0.961	0.973	-0.122	-0.123	0.924	0.949	0.958	0.437	0.67
1	9	0.947	0.965	0.965	0.345	0.611	0.928	0.923	0.927	-0.813	-0.509
2	10	0.817	0.832	0.843	0.891	1.253	0.692	0.811	0.858	1.429	2.168
2	11	0.886	0.863	0.874	-0.668	-0.6	0.814	0.772	0.900	-0.622	1.159
1	12	0.88	0.896	0.894	0.190	0.228	0.625	0.917	0.944	0.867	1.168
1	13	0.948	0.906	0.906	-0.343	-0.440	0.941	0.912	0.921	-0.311	-0.343
2	14	0.941	0.94	0.952	-0.341	1.264	0.878	0.84	0.903	0.556	-0.588
1	15	0.864	0.880	0.877	0.279	0.325	0.673	0.732	0.816	0.604	1.206
2	16	0.915	0.932	0.944	0.933	1.564	0.899	0.945	0.951	1.426	1.963
Khat Accuracy		**0.8596**	**0.8711**	**0.888**			**0.789**	**0.830**	**0.852**		

5. Conclusions

Indicator kriging using FIV provides a good tool to estimate the probability of occurrence of a class at a particular pixel location. The classification methodology described in this paper leads to increased classification accuracy. These results may be explained by the fact that no assumption is made about the distribution of pixels in a class and the additional benefit of using texture information only when spectral information is not enough to classify a pixel. However, the proposed methodology still requires expert input on the number of lags to be used.

Acknowledgements

We thank anonymous reviewer for a number of helpful suggestions that greatly improved the presentation of paper. Thanks are also due to Dr. C. J. Duffy, College of Engineering and EMS Environment Institute, The Pennsylvania State University for funding this work.

References

[1] R. M. Haralick, Statistical and Structural Approaches to Texture, *Proceedings of the IEEE*, Vol. 67, No. 5, 1979, 786-803.

[2] M. Herold, X. H. Liu, & K. C. Clarke, Spatial Metrics and Image Texture for Mapping Urban Land Use, *Photogrammetric Engineering and Remote Sensing*, vol. 69, no. 9, 2003, 991-1002.

[3] J. L. Cushnie, The Interactive Effect of Spatial Resolution and Degree of Internal Variability within Land Cover Types on Spatial Classification Accuracies, *International Journal of Remote Sensing*, 8(1), 1987, 15-29.

[4] E. H. Shih, & R. A. Schowengerdt, Classification of and Geomorphic Surfaces using Landsat spectral and textural features. *Photogrammetric Engineering and Remote Sensing*, 49, 1983, 337-347.

[5] R.W. Connersand & C.A. Harlow, A Theoretical Comparison of Texture Algorithms, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 1980, 204-222.

[6] C. Zhu & X. Yang, Study of Remote Sensing Image Texture Analysis and Classification Using Wavelet," *International Journal of Remote Sensing*, vol. 19, 1998, 3197-3203.

[7] B. Tso, & P. M. Mather, *Classification Methods for Remotely Sensed Data*, London: Taylor and Francis, 2001

[8] J. B. Campbell, Spatial Correlation Effects Upon Accuracy of Supervised Classification of Land Cover. *Photogrammetric Engineering and Remote sensing*, Vol. 47, 1981, 335-357

[9] R. Webster, & M. A .Oliver, *Geostatistics for Environmental Scientist*, John Wiley & Sons Ltd, England, 2001, 111.

[10] M. Chica-Olmo & F. Abarca-Hernández, Computing Geostatistical Image Texture for Remotely Sensed Data Classification, *Computer & Geosciences*, 26, 2000, 373-383.

[11] P. Goovaerts, Geostatistical Incorporation of Spatial Coordinates into Supervised Classification of Hyperspectral Data, *Journal of Geographical Systems*, 4, 2002, 99-111.

[12] P. M. Atkinson & P. Lewis, Geostatistical Classification from Remote Sensing: An Introduction, *Computer and Geosciences*, 26, 2000, 261-371.

[13] A.G. Journel, Non-parametric Estimation of Spatial Distributions, *Mathematical Geology*, 15, 1983, 445-468.

[14] G. Krige, & C. Camisani, Early South African Geostatistical Techniques in Today's Perspective, *Geostatistics*, Kluwer, 1989, 1-19.
[15] Y. N. M. Bishop, S. E. Fienberg and P. W. Holland, *Discrete Multivariate Analysis: Theory and Practice*, MIT Press, Cambridge, Massachusetts, 1975.
[16] R. G. Congalton and R. A. Mead, A Quantitative Method to Test for Consistency and Correctness in Photo-Interpretation, *Photogrammetric Engineering and Remote Sensing*, 49(1), 1983, 69-74.
[17] J.A. Richards, *An Introduction to Remote sensing digital image analysis*, Springer Verlag, 1986.
[18] K.V. Mardia, Tests of Univariate and Multivariate Normality. In P.R. Krishnaiah (ed.), *Handbook of Statistics*. vol. 1; 279-320. Amsterdam: North Holland, 1980.
[19] LARS,(2005, May).*Multispec* [Online]. Available: ww.ece.purdue.edu/~biehl/MultiSpec/
[20] L. Bruzzone and S.B. Serpico, A Technique for Feature Selection in Multiclass Cases," *International Journal of Remote Sensing*, vol. 21, 2000, 549-563.