Hyperspectral Image Analysis with Projection Pursuit and MRF Segmentation Approach – Unsupervised and Supervised.

Anjan Sarkar*, Ashish Vulimiri, Suman Paul, Md Jawaid Iqbal, Avishek Banerjee and Shibendu S Ray

***Abstract*—This work proposes methods for hyperspectral image analysis in both situations viz., (i) when concurrent ground-truth is unavailable and (ii) when available. The method adopts a projection pursuit (PP) procedure with entropy index to reduce the dimensionality followed by Markov Random Field (MRF) model based segmentation. Ordinal optimization approach to PP determines a set of "good enough projections" with high probability, the best among which is chosen with the help of MRF model based segmentation. When ground-truth is absent, the segmented output obtained is labeled (tuned) with the desired number of classes so that it resembles the natural scene closely. With available ground-truth, some special reflectance characteristics based on the ground-truth data of the study area in question are incorporated vis-a-vis the MRF model based segmentation stage while minimizing the energy function in the image space. In our crop study, some biophysical parameters of crops are appropriately incorporated in MRF model based segmentation stage. Subsequently segments are validated with training samples so as to yield a classified image with respect to varieties and stages of crops. Three illustrations are presented with (i) EO-1 archived data for an unsupervised case, (ii) EO-1 image data with concurrent groundtruth and (iii) AVIRIS-92AV3C, Indian Pine test site image with concurrent groundtruth. The classification accuracies of some nonparametric approaches and that of proposed methodology are provided for the illustrations (ii) and (iii) along with the computational time.**

***Index Terms*—Projection pursuit, Entropy Index, Markov Random Field and Ordinal optimization, classification accuracy**

I. Introduction

THE Hyperion Sensor aboard the EO-1 spacecraft provides high quality data with 30 meter spatial resolution over more than two hundred channels for analyzing the complex variability of landscape surface cover on earth. Investigations have been reported for matching image pixel spectra available from the hyperspectral data and the standard library comprising reflectance characteristics for a large class of materials with some success [21]. In classifying hyperspectral image data, the use of conventional techniques which are suitable for multispectral data have not yielded satisfactory results. This is because of mathematical and practical limitations such as Hughes Phenomenon [9] and others. A solution to this is to reduce the high dimensionality (corresponding to the number of channels) to a moderate dimensionality to allow analysis by computers as well as by users. Principal Component Analysis (PCA), an exploratory data analysis technique, is the most common and conventional data reduction method, generally used in the absence of any prior knowledge of a scene. PCA is performed with the idea of capturing largest variations in projection which are thought of as structured but which often fails in practice. To carry out projection, first a scalar projection score x for a data vector $\mathbf{y}$ along a direction $\mathbf{w}$ is defined as an inner product of $\mathbf{w}$ and $\mathbf{y}$, $x = \mathbf{w^T y}$, which gives a reasonably good measure of the performance of the projection direction $\mathbf{w}$ for a pixel vector $\mathbf{y}$. A natural approach for projecting the data set in an appropriate way is to evaluate the performance of the projection by optimization based on the projection scores of some index of "interestingness", a parameter which is task relevant and depends on the objective. This approach leads to the "projection pursuit"(PP) technique, a term introduced by Friedman and Tukey [6], that seeks out a linear projection of the multivariate data onto a lower dimensional space by means of optimization of the index of "interestingness", referred to henceforth as the "projection index". In order to find the projection optimal for our objective, we must first select an appropriate projection index. In PCA, which is a special case of PP, this index is the variance of the projection scores. Thus PCA seeks the projection which maximizes the signal sources.

PP, as defined above, is the process of making such selections (signals of interest) by local optimization over projection directions of some index of "interestingness". The index of interestingness is designed purposely to reveal clustering characteristics hidden in the multivariate high dimensional data. When hyperspectral image data are reduced to a moderate dimensionality with PP, one can go for segmentation approach to classification to determine the complex variability of landscape surface cover.

Some interesting projection indices available in literature are

A. Sarkar (anjan@maths.iitkgp.ernet.in) is with the Department of Mathematics and Computing at the Indian Institute of Technology, Kharagpur, India.

A. Vulimiri was with the Department of Computer Science and Engineering at the Indian Institute of Technology, Kharagpur, India. He is currently pursuing his Ph.D. in the University of Illinois-Urbana Champaign.

S. Paul is with the Department of Electronics at the Indian Institute of Technology, Kharagpur, India.

Md. Jawaid Iqbal is with the Department of Computer Science and Engineering at the Indian Institute of Technology, Kharagpur, India.

A. Banerjee is with the Department of Computer Science and Engineering at the Indian Institute of Technology, Kharagpur, India.

S. S. Ray is with the Space Applications Centre, Ahmedabad, India, of the Indian Space Research Organization (ISRO).

EDICS Code: ELI-COL

due to Friedman and Tukey [6], Huber [8], Jones and Sibson [11], Chiang and Chang [2], Jimenez and Landgrebe [12], and Ifarraeguerri and Chang [10] among others. The projection indices the above authors have employed are of three types, viz.

1) class distance measures (e.g., Bhattacharyya distance [12], The Friedman-Tukey Index [6]),
2) entropy indices or information divergence indices (Ifarraeguerri and Chang [10]), and
3) moment indices (Jones and Sibson [11] and Chiang and Chang [2]).

The concept of using entropy index emerges from the analysis of the statistical behavior of very high dimensional data sets in lower dimensions due to Renyi [19] and Huber [8]. Projection is a convolution, which due to Central Limit Theorem tends the data sets towards normality and thus their inherited structures are lost. Among the distributions with mean zero and unit variance, the usual order-1 entropy measure $\int -flogf$ is minimized when f has standard normal density. Therefore the entropy measure based on this can be used to calculate the tendency towards normality of the data set. By employing this entropy index we shall be able to identify "uninterestingness" as tendency towards normality and deviation from it as otherwise. The analytical form of the index as defined by Ifarraeguerri and Chang [10] is given by

$$J(f,g) = \int_{-\infty}^{\infty} glog\frac{g}{f} + \int_{-\infty}^{\infty} flog\frac{f}{g} \tag{1}$$

or,

$$J(f,g) = \sum_{-\infty}^{\infty} g_i log\frac{g_i}{f_i} + \sum_{-\infty}^{\infty} f_i log\frac{f_i}{g_i} \tag{2}$$

where g is equated with standard normal density and f is the distribution estimated from the data set. Following such a projection index, Ifarreguerri and Chang [10] have reduced the dimensionality from 210 to a meager 11 for a 256×256 HYDICE sensor scene. The component images resulted from PCA and PP as exhibited in [10] show the potential of the methodology. But the computation involved in such a methodology is highly intensive. It appears that the computation time required with such a methodology for a complete hyperspectral scene, which can be fairly large – EO-1 images, for instance, consist of 242 channels, each of size 256×3128 – would be unrealistically high.

Due to the potential of Ifarraeguerri and Chang's [10] information divergence index for PP, we adapt it to reduce the dimensionality "almost as good as optimally". Further, we do so with a technique that leads to faster computational time, followed by Markov Random Field (MRF) segmentation with the resulting component images to identify the best projected data for subsequent use for classification. Instead of determining best projection with the whole data set, we replace "best for sure" with "good enough with high probability"as per ordinal optimization suggested by Ho [7]. This "softening of the goal" eases the computational burden in our problem and it is much easier to find the best projection within, say, the top 100 than to get the absolute best projection so that some meaningful solution can be obtained in a reasonable time. To do this, we reduce the search space from $|W| = 8 \times 10^5 (= 256 \times 3128)$ to a 1000 random samples $(= N,$ say) from the whole data set. For smaller size image we may consider 500 samples or so. Then, the probability that at least one sample will yield one of the best projections on the top $100 (= k)$ is approximately 0.1175 which is computed from the expression $1-(1-\frac{k}{|W|})^N$ [7]. This implies that we need $\frac{1}{0.1175} \simeq 8$ iterations of the procedure (on the average) to guarantee that our projection is among the 100 best projections. Likewise, if we desire to get our best projection within top 10 of all best projection then we need to perform 80 iterations on the average. The set of 1000 samples that gives "good enough projection with high probability" should unfold good enough separated clusters from the high dimensional cloud, which in turn would yield the least energy in the MRF segmentation scheme as described in Sarkar et al. [20].

The MRF procedure we adopt here yields a segmented output which subsequently can be clustered into a desired number of classes in the absence of groundtruth so as to make the output closely resemble the natural scene of the area under study. The output so obtained only reveals different unidentified classes.

When groundtruth samples are provided we may exploit the information available to us in order to further classify and label the image. We demonstrate how this can be done by an example of how crop-classes, upto a level of 4 sub-classes of a given crop-type, might be identified in agricultural imagery. Here we show how to exploit the significant amount of information with respect to the crop's biophysical characteristics available in the spectral reflectance curve of each pixel in hyperspectral images. Other features of the groundtruth data would be utilized in other kinds of imagery.

The biophysical characteristics that we use, such as normalized difference vegetation index(NDVI), calculated using the radiances at the 803nm and 671nm wavelengths, and red edge points (REP), calculated using the radiance curve in the wavelength region 680nm to 780nm, are respectively measures of the condition and growth-stage of a plant. During the segmentation stage, while minimizing the energy function associated with the MRF model, all the above measures, such as NDVI values and red edge points, are suitably incorporated to obtain the segmented image. After obtaining the optimal segmented image a cluster validation scheme is carried out. A schematic representation of the methodology is exhibited in Fig.1. The portion demarcated by the dotted line describes the supervised scheme with the classified image as the end result.

The originality of the paper lies in underlining how PP can be exploited in a reasonable time frame with ordinal optimization scheme in an MRF model based segmentation procedure for analyzing hyperspectral images. The methodology also takes into account how the significant amount of information in the feature space (radiance curve for different wavelengths) available in hyperspectral images are analyzed to understand the plant's biophysical parameters and integrated in the segmentation process in a meaningful way. Although the paper is based on two previous works [10] and [20] but for ordinal optimization it develops a novel technique in which the two works are appropriately combined resulting in a methodology

which is computationally faster. The analytical form of the entropy index in [10] is easy to implement besides its potential as stated earlier and how well the different projections with different set of sample data segregate the classes in the feature space is clearly brought out by the relative value of the energy function associated with the MRF in [20]. Further, the energy function associated with the MRF segmentation procedure can be made as a two pronged one by considering a specific variable of the energy function in [20] as a binary one. As such other significant information that are available for biophysical characteristics of plants or other aspects of the study area of the hyperspectral image can also be easily integrated in this segmentation process through this binary variable. This way, whatever the hyperspectral image data have to tell us regarding the ground complexity of the land cover as well as plants' biophysical condition (in our case), everything has been embodied in this process of analysis. Thus the proposed methodolgy analyse the hyperspectral image in totality and is expected to produce high classification accuracy in a faster manner. Also, when concurrent groundtruth information is not available this work shows a way for meaningful qualitative comparison with the natural scene.

The paper is outlined as follows. Section II describes briefly the projection pursuit method with the ordinal optimization scheme. Section III presents measures of crop biophysical parameters that determine the stage of crops. The MRF model based segmentation scheme is presented in Section IV. In Section V we illustrate experimental results with discussion followed by the conclusion in Section VI.

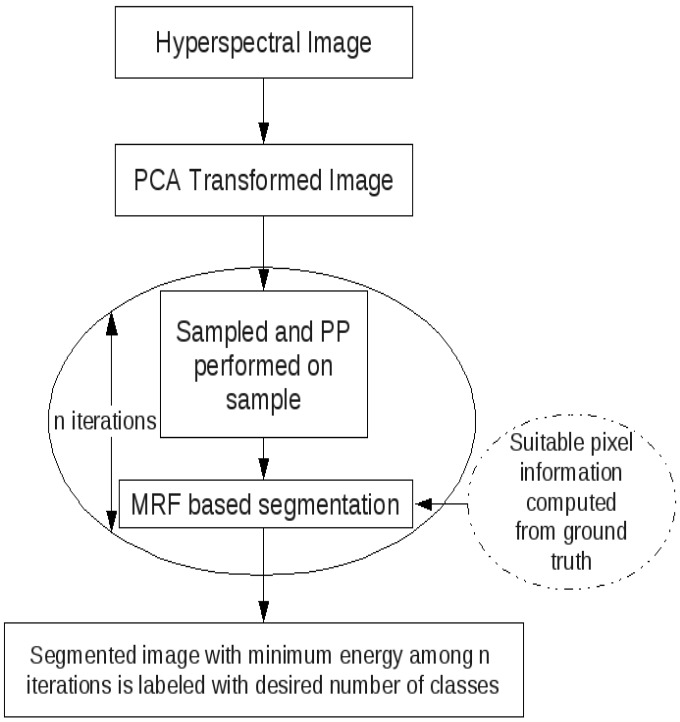


Fig. 1. **Workflow of the methodolgy**

II. Projection Pursuit

The projection pursuit procedure that we shall follow has been proposed by Ifarraguerri and Chang [10], but for the completeness of the work, the methodology and the way we use it is briefly described in this paragraph. This approach relies on the belief that the "interesting" projection vectors are located in or near the point cloud and they can be approximated by the pixel spectrum nearest to it in terms of the inner product defined above as "projection score". So, to find the best projection, the whole data set is simply projected along each pixel vector, their divergence from normality is calculated in terms of the value of the projection index, and the one that corresponds to the highest value of the projection index (the least normality) is chosen as the desired direction. After finding the direction as described, other projections are searched at directions orthogonal to it. The pre-reduction of dimensionality is done by using PCA and thus the dimension of the space is fixed in which the projection is sought so that we need to search only that number of different orthogonal directions as specified by significant principal components.

The hyperspectral image data, consisting of $r \times c = h$ pixels over d bands, is arranged in a matrix $\mathbf{X}_{d\times h}$. As an initial reduction of dimension, PCA is performed using the well known "covariance-free method" [23] and the first k eigenvectors corresponding to k significant principal components are stored in $\mathbf{E_{d\times k}}$. Then the principal component transformed data matrix is obtained as

$$\mathbf{Z}_{k\times h} = \mathbf{E}^T\mathbf{X}, \tag{3}$$

a copy of which is saved as $\mathbf{Z}_{orig}$. A random sample consisting of N pixels from k component bands has been selected and as described in [10], the projection index is calculated as

$$J^{(l)}(\mathbf{p},\mathbf{q}) = \sum p_i log\frac{p_i}{q_i} + \sum q_i log\frac{q_i}{p_i} \tag{4}$$

This procedure is carried out for all N pixels. The l for which $J^{(l)}(\mathbf{p},\mathbf{q})$ is maximum is then searched:

$$arg(max(J^{(l)}(\mathbf{p},\mathbf{q}))) = l_{best} \tag{5}$$

Then the vector $\mathbf{z_{l_{best}}}$ corresponding to the l^{th}_{best} pixel spectrum is termed as the optimal projection vector. $\mathbf{z_{l_{best}}}$ is orthonormalized and is appended as a column $\mathbf{w_r}$ of a matrix $\mathbf{W}$:

$$\mathbf{w_r} = \frac{\mathbf{z_{l_{best}}}}{||\mathbf{z_{l_{best}}}||}, \tag{6}$$

Subsequently, to search for the directions orthogonal to the columns $\mathbf{w_r}$ of $\mathbf{W}$, at each step $\mathbf{Z}$ is projected orthogonal to $\mathbf{W}$ by the following:

$$\mathbf{Z}^{(new)} = [\mathbf{I} - \mathbf{W}(\mathbf{W^T W})^{-1}\mathbf{W^T}]\mathbf{Z} \tag{7}$$

$\mathbf{Z}^{(new)}$ is the new $\mathbf{Z}$ for the next iteration. This whole procedure is carried out k times to obtain a $k \times k$ matrix $\mathbf{W}$. For the optimally projected data set, we perform

$$\mathbf{Z}_{opt} = \mathbf{W}^T\mathbf{Z}_{orig} \tag{8}$$

Again another random sample is taken as $\mathbf{X}$ and we repeat the procedure for a specified number of times which depends on the size of the sample and the degree of proximity (with the true solution) we want to attain by the formula cited above, and by the same reasoning, this is certain to produce one of the best solutions. The best solution (projection) among all these iterations is identified in the MRF based segmentation scheme described in section IV. The criterion by which we identify the best solution among the number of iterations considered is the minimum segmentation energy associated with the MRF procedure.

III. Measures of crop biophysical parameters

The hyperspectral image that we consider under our study of supervised knowledge comes from crop area and as such we may take note of some of the key biophysical parameters such as leaf area index (LAI), NDVI values and red edge points (REP) etc that play an important role in explaining crop growth and condition. These biophysical characteristics can be measured from the spectral reflectance curve for each pixel and may be exploited in the segmentation process. However incorporating these values would sensibly improve the classified image only if there are concurrent ground truth values for these parameters. In practice it may not be easy to measure such values, particularly LAI, and in fact in our study area, ground truth data are collected in conformity with different stages of the crops only. We thus use NDVI and REP values from spectral information of each pixel for determining crop condition and stage discrimination respectively. If the study area concern is agriculture, these values viz., NDVI and REP will surely help in distinguishing similar agricultural subclasses yielding improve classified image.

NDVI is calculated with NIR=803nm and RED=671nm wavelengths. The REP for each crop pixel are calculated with radiance curve in the wavelength region 680nm to 780nm by polynomial fitting of the curve [18]. Based on the ground truth data four stages of growth are identified, viz (i) emergence (ii) tillering/booting (iii) milking and grain filling and (iv) maturity. The corresponding REP range of each stage is noted for each pixel from ground truth samples and subsequently these are used in segmentation stage along with NDVI values to determine the classified image at level-4 subclass.

IV. MRF model based segmentation scheme

We follow the scheme of Sarkar et al. [20] in defining the MRF on a region adjacency graph(RAG) of initial over-segmented regions - the details are available in [20]. Our discussion here is directed in formulating the energy function used in the MRF based segmentation approach so as to show how the contribution of biophysical features of plants available in the image can be incorporated in the segmentation (classification) process. Minimizing this energy function will result in a maximum a posterior probability (MAP) estimate of the optimal segmented image. We impose two constraints as per our notion of optimal segmentation.

1) An optimal segmented image region R_i should be uniform with respect to the measured characteristics as obtained from the projection pursuit component images. This implies that better the revelation of the hidden clustering characteristics of the high dimensional point cloud in lower dimension, more uniform the segmented region.
2) Two distinct adjacent regions R_i and R_j should be as dissimilar as possible with respect to the measured characteristics as evident from the selected projection pursuit component images as well as distinguishing bio-physical features such as REP etc over the region pairs. These bio-physical features over the regions are based on majority criterion or the average value of those pixels comprising the region.

First, a tonality based initial segmentation procedure [14] is carried out on the first projection pursuit component image, and then the same regions $R_i(1), i = 1, 2, ...Q$ are grown in each of the remaining projection pursuit component images. Thus the multi-component image is initially over-segmented into a set of Q disjoint regions denoted by $R_1 = R_1(p), R_2 = R_2(p), ..., R_Q = R_Q(p), p = 1, 2, ..., P$, where P is the number of component channels. Representing each region R_i as a node with multi-component channel information, a RAG, $\Gamma = (R, E)$ is defined, where $R = \{R_i; 1 \leq i \leq Q\}$ is a set of nodes and E is a set of edges connecting them. A MRF is defined on RAG with neighborhood system $\eta = \eta(R_i) : 1 \leq i \leq Q$. Here $\eta(R_i)$ is a set of regions in R which are neighbors of R_i. Before we can tackle segmentation problem in the MRF framework we need an expression for posterior distribution given the realization $\mathbf{y}$ [16].

Let the events $\{\mathbf{X} = \mathbf{x}\}$ and $\{\mathbf{Y} = \mathbf{y}\}$ represent respectively a specific labeling configuration and a specific realization. The Bayesian estimate of the configuration is the one which maximizes the posterior distribution. The posterior distribution also follows a Gibbs distribution under the assumption of conditional independence of the likelihoods [4]

$$P(\mathbf{X} = \mathbf{x}|\mathbf{Y} = \mathbf{y}) = \frac{e^{-U_{ps}(\mathbf{x}|\mathbf{y})}}{Z_{ps}} \tag{9}$$

where

$$\begin{aligned} U_{ps}(\mathbf{x}|\mathbf{y}) &= -ln\Big[\prod_{i=1}^{Q} P(Y_i = y_i|X_i = x_i)P(\mathbf{X} = \mathbf{x})\Big] \\ &= -ln\Big[\prod_{i=1}^{Q} P(Y_i(1) = y_i(1), ..., Y_i(P) = y_i(P) \\ &\qquad \Big|X_i = x_i)P(\mathbf{X} = \mathbf{x})\Big] \end{aligned}$$

using the relation $x = e^{lnx}$. This enables us to express the posterior energy function as a sum of the potentials at different points in the image lattice [4], [5]. The term $Z_{ps} = \sum_{x \epsilon \Omega} e^{-U_{ps}(\mathbf{x}|\mathbf{y})}$ is the normalizing constant. The maximum aposteriori estimate $\hat{x}$ of $\mathbf{x}$ is obtained by minimizing the posterior energy function $U_p(\mathbf{x}|\mathbf{y})$. Since the energy function $U_{ps}(\mathbf{x}|\mathbf{y})$ is a sum of the clique potentials $V_c(\mathbf{x}|\mathbf{y})$, it is necessary to select appropriate cliques and clique potentials to achieve the desired objective. For the cliques and clique

functions, only the set of adjacent two-region pairs each of which is directly connected in the RAG are considered here.

Incorporating the above two constraints we define the clique potential (energy) function [20] by two processes, Region Process H and Edge Process B as

$$V_c(\mathbf{x}|H,B) = V_c(\mathbf{x}|\mathbf{y}) = \eta_{ij}\mathbf{W}_{i,j} + \theta_{ij}(1-\eta_{ij})\mathbf{B}_{ij} \quad (10)$$

Here,

$$\mathbf{B}_{ij} = \frac{n_i n_j}{n_i + n_j}(\mathbf{M}_i - \mathbf{M}_j) * (\mathbf{M}_i - \mathbf{M}_j)^{'},$$

and

$$\mathbf{W}_{ij} = \frac{1}{\nu_{ij}}\left[\sum_{k=1}^{n_i}(\mathbf{Y}_{ik} - \mathbf{M}_i) * (\mathbf{Y}_{ik} - \mathbf{M}_i)^{'} + \sum_{k=1}^{n_j}(\mathbf{Y}_{jk} - \mathbf{M}_j) * (\mathbf{Y}_{jk} - \mathbf{M}_j)^{'}\right]$$

$\mathbf{M_i}$ being the mean vector of region R_i consisting of n_i number of pixels, $\nu_{ij} = n_i + n_j - 2$ and η_{ij} a binary variable taking values 0 and 1. ν_{ij} takes the value 1 when the following two conditions are together satisfied depending on the values of NDVI. With NDVI greater than zero, we identify the crop regions. In case of crop regions, the first condition is that the REP values of R_i and R_j follow the same pattern. The second is that the regions (in clique potential) are homogeneous with respect to the multi-component channel pixel intensity values. If the NDVI are not greater than 0 the first condition is ignored. If any of these conditions are violated, η_{ij} takes the value zero. It may be noted that $\eta_{ij} = 1$ indicates that $x_i = x_j$. With this variable (η_{ij}) the biophysical condition of the crops is coupled with data level fusion in the energy minimization process.

We also note that in the unsupervised scheme only the second condition needs to be satisfied.

The parameter θ_{ij} controls the weight to be given to the two processes for regions involved in the clique c.

A suitable comparative criterion among the elements of these two matrices $\mathbf{B}_{ij}$ and $\mathbf{W}_{ij}$ is necessary for deciding the merging of two adjacent regions of the over-segmented image. Since the ratio of $\mathbf{B}_{ij}$ and $\mathbf{W}_{ij}$ can be expressed as

$$T^2 = (\mathbf{M}_i - \mathbf{M}_j)'[(1/n_i + 1/n_j)\mathbf{s}_{pooled}]^{-1}(\mathbf{M}_i - \mathbf{M}_j) \quad (11)$$

where $\mathbf{s}_{pooled} = \frac{\mathbf{S}_i + \mathbf{S}_j}{\nu_{ij}} = \mathbf{W}_{ij}$, $\mathbf{S}_i$ being the sum of squares and cross product matrix of the region R_i, the comparative criterion needed here is based on Hotelling's T^2 statistic.

Therefore, the regions R_i and R_j in the clique should be merged if $T^2 < F_\alpha$ and the regions should not be merged if $T^2 \geq F_\alpha$, where $P[T^2 > F_\alpha] = \alpha$, the level of significance.

The segmented image is obtained by minimizing the energy function $U_{ps}(\mathbf{x}|H,B) = \sum_{c\epsilon C} V_c(\mathbf{x}|H,B) = \sum_{c\epsilon C} V_c(\mathbf{x}|\mathbf{y})$ as described above.

It is obvious that the more the hidden clustering characteristics of the high dimensional data set revealed by PP in lower dimensions the better is the segmented output with lesser and lesser energy. Corresponding to every thousand random samples chosen from the hyperspectral image, PP is performed and the projected data yields a number of component images as per the order of the chosen lower dimension. Subsequently segmentation is performed for the desired number (say, P) of component images for each of the iterations. For n number of such iterations (each with 1000 random sample) n sets of MRF based segmented image are determined. The segmented image that yields the least energy is the best projected data set for the given hyperspectral image. As said earlier we choose our search space 1000 random sample instead of the full hyperspectral image of size 256×3128 or so. With eight$(= n)$ iterations only, each iteration taking a small duration of time, we ensure that the segmented output (with minimum energy) is among the best 100 cases.

This segmented output generates directly an image that is almost the classified image when supervised knowledge (training samples) is incorporated in the segmentation stage as shown in Fig.1 in the dotted area. It is almost the classified image in the sense that only the labels of the segments (clusters) are absent. We may note here that if the complete exhaustive ground truth knowledge is available, one may use per pixel labels following maximum likelihood technique (with projection pursuit component images). Subsequently these per pixel labels may be combined in the segmentation process like REP values of the regions to obtain a classified image.

In the absence of any supervised knowledge, our MRF model based segmented output yields hundreds of different regions as per similar pixel intensities in the component images. For example, a large forest area comprising different types of clusters of trees of the same species would unfold in the segmented output as a number of regions corresponding to clusters of trees of the same species which have similar intensity. But as the intensities of the clusters of trees of different species are not expected to be grossly different the fragmented segments can be tuned (grouped) into a single class or validate with similar label and identified as forest cover. In the same way, crop areas, fallow land etc also comprise fragmented regions as per similar pixel intensities which may be combined. This implies that our MRF model based segmented output may be tuned into lesser number of segments (classes) in the absence of any ground truth knowledge. This is a kind of unsupervised cluster validation technique. However, this tuned output that we propose corresponds to a labeling of the segments to the desired number of classes and remains unidentified with the set of natural landcover classes of the scene in the absence of ground truth. But it renders a means of comparison in a crude way between a classified image (without natural class label) of an unsupervised classifier and an expected natural scene of the area under study.

The simple technique that we follow for such a tuning comprises two steps and is carried out as follows:

1) Determine the R_{c_2} number of distances of T^2 values between all pairs of R number of regions in the segmented image.
2) Assign the label i of R_i to R_j of those two regions for $i < j$, whose T^2 value among all R_{c_2} distances is minimum.

Follow steps 1 and 2 till $R_0 - N_c$ number of times where R_0 and N_c are respectively the intial number of regions in the segmented image and the desired number of classes. We note that at each itaration, the number of region decreases by one and thus after $R_0 - N_c$ iterations we get the desired number of regions(classes).

The tuned segmented output thus exhibits the different regions in as many gray shades as N_c number of classes. In order to make the regions (classes) of the images more vivid and distinct we subsequently exhibit the image in false colour composite using the technique of Jinxiu et al [13] to transform the image from gray to pseudocolor.

Cluster Validation Scheme

The unsupervised segmentation of projection pursuit component images results, in general, in a large number of clusters (segments) say, R_j , $j = 1, 2, ...L$ with unique cluster numbers being assigned to each pixel. We validate those clusters following the scheme of Sarkar et al [20] with some modification. Unlike the case in [20], where ground-truth was collected exhaustively for all classes in the study area, here we may have ground-truth samples only for some specific type of classes, say only crop classes. We note here that in actual practice a scene comprises variety of classes viz., water, urban, forest etc. In our case, let there be K such crop classes whose ground-truth data are available. First we identify those clusters in the segmented image that have ground-truth samples (see details in [20]). In this stage only K crop classes would be identified with number of segments greater than or equal to K. Next we pick up all those unlabeled clusters R_j, one at a time and compare with each of the K labeled clusters with respect to their means. In this comparison we actually attempt to recognize each of the unlabeled segments with any of these labeled clusters (segments). These labeled segments can each be thought of as the prototype clusters corresponding to each of the ground-truth classes $i = 1, 2, .., K$. Now we set up the null hypothesis H_0 : the given unlabeled segment R_j is identical to the ith labeled segment R_i against the alternative H_1 : the given unlabeled segment R_j is different from R_i. Under the null hypothesis, Hotelling's T^2 statistic, eq(11) may be used. Therefore, the segments R_i and R_j should have the same label if $T^2 < F_\alpha$ and otherwise, if $T^2 \geq F_\alpha$, where $P[T^2 > F_\alpha] = \alpha$. We note $S_{pooled} = \frac{S_i + S_j}{\nu_{ij}}$ and S_i represents the sum of squares and cross product matrix. We note that for a given R_j, the above condition may be satisfied for more than one prototype class. In such a case we recognize R_j with that prototype class R_i for which the computed T^2 is found to be minimum. An unlabelled segment will remain unclassified at the level of significance α if $T^2 \geq F_\alpha$ for each of the prototype classes.

V. Experimental Results

Three illustrations of hyperspectral image analysis have been presented in this paper – the first example has no concurrent ground truth and the second has the concurrent ground truth. Both were acquired from NASA, EO-1 satellite (Hyperion sensor). The third example deals with AVIRIS-92AV3C image which has been collected from the website [1] along with its goundtruth. Comparison tables of classification accuracies and Kappa coefficients of some nonparametric methods along with the propsed methodology are provided for the images whose groundtruths are available. Besides determining classification accuracy to judge the merit of the proposed methodology, an experiment has been carried out in this example to evaluate quantitatively how good is the unsupervised classified image.

Example-1

An archive image data over the area specified by the co-ordinates UL(23.5628N, 87.4528E), UR(23.5478N, 87.5259E), LL(22.7796N, 87.2651E), and LR(22.7647N, 87.3377E) of the southern part of West Bengal, India has been considered. The area was imaged on 17 January 2003. The hyperspectral image for this example is of size 256×3128 with 220 bands covering the portion of the spectrum of wavelength 400-2500 nm with an window size of 10nm. The archive image data had been preprocessed for atmospheric correction with *Geomatica*. Those bands which had a DN of zero or which were corrupted with noise have been removed and finally 143 bands have been used as an input to our methodology for unsupervised labeling. The bands used are ([12-52, 77-120, 143-164 ,189-224])

In step 1, PCA has been carried out with whole set of image data, that is, with 256×3128 pixels each having 143 bands. From PCA, the first 15 component images, that take into account 99.9% of the total variation, have been considered for further processing. In the next step, as described in Section II, PP has been performed with a random sample of 1000 pixels taken from the 15 component images. As described earlier, such an iteration (of carrying out PP with 1000 randomly chosen sample) has been executed 8 times to determine one among the hundred best solutions, that is, a "good enough projection with high probability". To determine the best projection among these eight cases, the MRF model based segmentation procedure with the projected data (component image) has been carried out for each iteration. MRF based segmentation procedure on each of the eight cases yields a homogeneity (uniformity) measure (value of the energy function associated with the underlying MRF). In fact we may rank each of the projections from best to worst by listing them in the ascending order of the values of the energy function associated with MRF segmentation stage. The one which yields least energy (implying the best partitioning) in the segmentation stage gives us the desired solution. After carrying out PP on all of the 15 component images derived from PCA, we note that the last few (five in the present case) component images (orthogonal to one another) carry no significant information. Thus MRF model based segmentation has been carried out on 10 component images derived from PP in each iteration.

As there are no groundtruth samples, tuning of the segmented output is carried out as described at the end of Section IV with desired number of landcover classes and is exhibited in false color composite. The desired number of classes has

been considered as 30 odd classes from a post date survey of the area.

Fig. 2(a) exhibits the full (256×3128) original image displayed in false color composite with three bands $R(803.3nm)$, $G(681.2nm)$ and $B(548.9nm)$. Fig. 2(b) and Fig. 2(c) exhibit the segmented image based on, respectively, the PCA component images and the PP projected component images corresponding to the minimum energy case (for the 8 iterations). In order to compare the two cases in detail we consider a similar subset of Fig. 2(b) and Fig. 2(c) as shown in Fig. 3(a) and Fig. 3(b). Comparison of these two Figs (Fig. 3(a) and Fig. 3(b)) would reveal which of the two cases does better separation of segments (classes). The total processing time for 8 iterations including PCA for this example with an image size of 256×3128 was around 3000 seconds on an Intel Pentium D 3.00GHz CPU.

Example-2

This image was acquired on 12 March 2005 over the area specified by the co-ordinates UL(29.558780N, 77.763235E), UR(29.543599N, 77.840277E), LL(28.716203, 77.548268E), and LR(28.701110N, 77.624694E) of western Uttar Pradesh (Modipuram), India. The image is geocoded, slanted to the right and of size 961×3521 with 240 bands covering the same portion of the spectrum with the same window size as in example-1. We consider here a subscene of size 200×710 for our analysis with 122 bands having meaningful DN values. The consideration of this subscene comes from the fact that the ground truth data collected was found to be scattered in this area. The classwise break up of this ground truth as collected from this site are as follows:

1) wheat grain filling(61)
2) wheat maturity(26)
3) sugarcane mature(22)
4) wheat milking(48)
5) sugarcane-ratoon maturity(14)
6) wheat booting(15), and
7) wheat emergence(14)

The number in braces indicates the total number of ground truth pixels in the corresponding class. As the total number of pixels in some classes are very small, it is difficult to deal with a separate set for testing the classification accuracy. In view of this, we randomly set aside 20% of the available samples from each of the classes for determining classification accuracy. As in example-1, we follow an analogous procedure upto the segmentation stage. While minimizing the energy function associated with the MRF we also combine NDVI and REP information for different stages of each crop as measured from the spectral reflectance curve for each pixel. The REP values are calculated from the radiance curve in the wave length region 680nm to 780nm by polynomial fitting of the curve. For wheat crop, four distinct wavelength regions are observed for the REP values corresponding to the four different stages. Because of the first condition, as stated in Section IV, the optimal segmented regions have been determined with additional knowledge viz., the supervised knolwedge. We now carry out the cluster validation scheme as described at the end of Section IV and subsequently classification accuracy is evaluated.

Since the search space in this example is $|W| = 200 \times 710$, with eight iterations of the procedure we would be able to determine one of the twenty best solutions. The original image is exhibited in Fig. 4(a). The classified image of the proposed methodolgy is exhibited in Fig. 4(b). Table-1 provides the classification accuracies of some nonparametric techniques, viz., SVM (Support Vector Machines)-linear, SVM-polynomial, SVM-RBF(Radial Basis Function) and the proposed methodology. This Table also exhibits corresponding Kappa coefficients of each technique along with execution time.

TABLE I
TABLECOMPARISON OF ACCURACY AND TIME DURATION OF OUR APPROACH WITH SOME OTHER APPROACHES IN INTEL PENTIUM D 3.00GHZ CPU.

Approach	Parameters	Accuracy[%](κ)	Time[s]
SVM-Linear	C=100	76.47(0.684)	30
SVM-polynomial	degree=4,C=100,γ=1	70.58(0.609)	20
SVM-RBF	C=100 γ=1	70.60(0.612)	20
K-nn classifier	K=7	70.0(.631)	30
Proposed methodology	α=.01	89.29(0.904)	985

Example-3

In this example we consider the AVIRIS-92AV3C, (Indian Pine test site) image [1] of size 145×145 with 220 bands and compare the classification accuracy of the proposed methodology with the nonparametric approaches as in Example-2. The fifty percent groundtruth samples(4757) have been used for training and remaining (4588) samples of the total 9345 have been used for testing [see Table-I in [17]]. The groundtruth comprises sixteen classes, seven classes are discarded because of insufficient number of samples as mentioned in the procedure suggested in [17] and remaining nine classes have been considered. The REP and NDVI values are calculated as in example-2 and incorporated in the segmentation process. The original image along with the classified image have been exhibited in Fig.5(a) and Fig.5(b). Fig.5(c) exhibits the tuned image with nine classes. The coloring of the segments of Fig.5(c) has been done as per the color of the legend of the classified image so as to have a visual comparison. The search space of the image is $|W| = 145 \times 145$. With 500 random samples and five iterations for 5 pca components (4 pp components) images that accounts for 93% of total variation, we determine one of the best 10 solutions. Table-II compares our findings for the proposed methodology with respect to accuracies and computation time with some nonparametric approaches viz., SVM, SVM-RBF and K-nn classifier. In order to compare the findings as in [17] for this image data we also report [Table-II, of [17]] of the approaches, executed in SUN ultra system in Table-III. We now compare our unsupervised classified image quantitatively. This is just to verify how accurately the segments are labelled by this unsupervised approach. To do this we first carry out a mapping of the 9 tuned classes with the groundtruth classes followed by overlaying

the testing samples on these tuned classes to determine the classification accuracy. For mapping the tuned classes we also overlay the groundtruth samples on tuned segments. The tuned segments will have the same label as that of the groundtruth class label whose maximum samples fall in that segment. In some cases there may be some segments where no ground truth falls. Such a segment will remain as a unclassified class with respect to groundtruth class labels.

TABLE II
COMPARISON OF ACCURACY AND TIME DURATION OF OUR APPROACH WITH SOME OTHER APPROACHES IN INTEL PENTIUM D 3.00GHZ CPU.

Approach	Classification (Overall) Accuracy[%](κ)	Time[s]
SVM-Linear	87.64 (0.854)	295
SVM-polynimial	86.42 (0.839)	130
SVM-RBF	82.63 (0.795)	159
Proposed methodology	94.01(0.931)	175

TABLE III
COMPARISON OF ACCURACY AND TIME DURATION AS IN[17] (IN SUN ULTRA 80 WORKSTATION)

Approach	Classification (Overall) Accuracy[%]	Time[s]
SVM-Linear	87.10	40342
SVM-RBF	93.42	2702
K-nn classifier	83.94	2618
RBF classifier	86.99	4743

Discussion

In Example-1, while comparing Fig. 2(b) and 2(c) we note that there are thirty different shades of colour corresponding to thirty odd classes, although some of these shades appear very similar. But the courses of all these rivers are distinctly visible in Fig 2(c) while in Fig 2(b) it is not so. Now we compare Fig.3(a) and Fig.3(b) which are similar subsets of Fig.2(b) and Fig.2(c) and remembering that both these are tuned in the same way, we note the following points. Firstly, Fig. 3(a) exhibits 1218 number of regions compared to Fig. 3(b) which has 1469 regions. Secondly the segments displayed in Fig. 3(b) are more vivid as compared to that of Fig 3(a). In particular, the different crop regions (red colored regions in Fig. 2(a)), the canal and the sublane towards the bridge over the river are best identified in Fig.3(b). This figure (Fig. 3(b)) is the segmented output that yields the minimum energy for the number of iterations considered and hence is one of the best projections onto the lower dimension. It thus brings out the potential of our methodology, as the proposed methodology outperforms PCA with respect to class separation.

In the original image (Fig. 4(a)) of Example-2, the red colored region corresponds to vegetation whereas the blue region corresponds to scattered barren land surface. For this example, the output of the proposed methodology yields a classified image (Fig. 4(b)) comprising regions of two types of crops of different stages and unclassified regions (for barren land and other vegetation areas). Different stages of wheat and sugarcane are identified with different colors as exhibited in Fig. 4(b) along with black color for the unclassified regions (classes). The two deep blue vertical lines featuring in the original image on the western part and a diagonal line like feature across the image in the central part are also visible as unclassified part in Fig. 4(b). Further we note that almost all the sky blue regions that correspond to barren land in original image have been appropriately identified. For example one can compare the reverted L type blue region in the central part of the original image and the corresponding black region in Fig. 4(b). In order to examine whether the class labeling is appropriate we check the classification accuracy of the crop classes. Since the ground-truth samples for some classes is very small we examine the accuracy by setting aside 25% of the total ground-truth samples as test set for each of the classes. That is to say, with 75% of the samples for each of the classes we perform cluster validation scheme and test the overall classification accuracy with the remaining 25% samples. This 25% selection of the test set has been carried out randomly a number of times and the accuracy is computed in each case. This repetition of accuracy evaluation has been done to mitigate the effect of less number of groundtruth samples for some classes. It has been observed that overall accuracy does not differ widely in these repetitions, instead remains in a range 87.3%- 91.2 % with mean 89.29%. For the sake of comparison, some non-parametric approaches have been run and the results have been reported in Table I. As can be seen from the value of κ coefficient (Table-I), the accuracy of the proposed methodolgy is significantly higher. The extra time that this methodolgy takes in comparison with other methods may be accommodated for this high accuracy.

In example -3, the overall accuracy and κ coefiicient are found to be 94.01% and 0.931 respectively. The latter value corroborates the fact that it is significantly higher with respect to other methods. This high classification accuracy compared to other approaches (see Table-II) as well as the exhibited classified image (Fig.5(b)) that reveals different classes very close to original image(Fig.5(a)), demonstrate the potential of the methodology. The black regions that are observed in Fig.5(b) are unclassified regions and correspond to those areas whose ground truth samples have not been considered in training set. The roads and the railway track that are visible in original image from NW to SE are found here as unclassified classes in Fig.5(b). Further, the computation time of 175 seconds that this methodology takes to analyse this hyperspectral image in comparison with other nonparametric approaches is quite fast (see Table-II). The accurcies in Table-III explain that only SVM-RBF approach is comparable with our methodology.

Second part of the experiment with Example-3 deals with quantitative evaluation of classification accuracy for the unsupervised image analysis (tuned image) as stated at the end of section IV. The overall classification accuracy is found to be 51% for this unsupervised case (Fig.5(c)).

This classification accuracy, although not high, but the image Fig.5(c) suggests that the technique adopted for un-

supervised analysis of the hyperspectral image could be a way to determine the expected natural scene in the absence of groundtruth information.

VI. Conclusion

PP procedure with information divergence index identifies some collection of projections based on randomly chosen samples. When coupled with MRF model based segmentation procedure it determines the "good enough projection" with high probability through ordinal optimization. The proposed methodology appears to have a potential for hyperspectral image data analysis as is evident from the experimental results. By tuning the segmented output to a desired number of classes a visual comparison is possible with the original image in the absence of any groundtruth knowledge. The experimental results suggest that the complex variability of the landscape surface cover over earth has been potentially analyzed with the proposed methodology. When detailed groundtruth is available, the classified image is determined by means of a cluster validation technique. The classification accuracy as observed shows that the methodology has merit. In essence, the methodology shows a way out how to analyze hyperspectral images reducing the "curse of dimensionality". The methodology is not computationally intensive if we desire "good enough with high probability" instead of determining "best for sure". When we intend to have our best projection within top k of all best projection, the order of computational complexity is $O(I * (nlogn + W * N * m^2))$, where I = $O(\frac{W}{N*k})$ is the number of iterations, W = size of search space, nlogn (n being the number of initial segmented regions) is the complexity of MRF segmentation, m = dimensionality of the hyperspectral image, N = number of samples chosen in each iteration of PP.

Thus this method is adaptable to trade off its one of the best solution among k best solution against intensive computation.

Acknowledgment This work was supported by ISRO under the grant for the "Classification of hyperspectral remote sensing data to discriminate between crop condition, variety and stage",
Ref. ISRO/RES/4/535/2006-07, March 29, 2007.

References

[1] AVIRIS NW Indiana's Indian Pines 1992 data set[Online]. Available: ftp://ftp.ecn.purdue.edu/biehl/Mutispec/92AV3C (original files) and ftp://ftp.ecn.purdue.edu/biehl/$PC_/$MultiSpec/ThyFiles.zip (ground truth).

[2] S. S. Chiang and C.I. Chang, "Unsupervised target detection in hyperspectral Images using projection pursuit," *IEEE Trans. Geosci. Remote Sens.*, vol.39, no.7, pp. 1380-1392, 2001.

[3] K. Crammer and Y. Singer." On the Algorithmic Implementation of Multi-class SVMs", JMLR, vol. 2, pp. 265-292, 2001.

[4] R. C. Dubes and A. K. Jain, " Randiom field models in image analysis ", *J. Appl. Statist.*,vol. 16, pp 131-164, 1989.

[5] R. C. Dubes, A. K. Jain, S. G. Nadabar, and C. C. Chen, "MRF model based algorithms for image segmentation", in *Proc. Int. Conf. on Pattern Recognition*, 1990, pp. 808-814.

[6] J. H. Friedman and J. W. Tukey, "A projection pursuit Algorithm for exploratory projection pursuit," *IEEE Trans. Comput.*, vol. C-23, pp. 881-889, 1974.

[7] Y. C. Ho, "The No-freelunch theorem and the human-machine interface," *IEEE Control System Magazine*, pp. 8-11, 1999.

[8] P. J. Huber, "Projection pursuit", *Ann. Staist.*, vol.13, no.2, pp. 435-475, 1985.

[9] G. F. Huges, "On the mean accuracy of statistical pattern recognizers", *IEEE Trans. Inform. Theory*, vol.IT-14, no.1, pp. 55-63, 1968.

[10] A.Ifarraeguerri and C. Chang, " Unsupervised Hyperspectral Image Analysis with Projection pursuit", *IEEE Trans. Geosci. Remote Sens.*, vol.38, no. 6, pp. 2529-2538, 2000.

[11] M. C. Jones and R. Sibson, "What is projection pursuit?", *J. R. Statist.Soc. A*, vol.150, no.1, pp.1-36, 1987.

[12] L. O. Jimnez and D. A Landgrebe, "Hyperspectral data analysis and supervised feature reduction via projection pursuit", *IEEE Trans Geosci. Remote Sens.*, vol.37, no. 6, pp 2653-2667, 1999.

[13] Li Jinxiu, Li Junli and Wei Ping, "Pseudo colour coding of Medical Images Based on Gradient", *Ist Int. Conf. on Bioinf. and Biomed. Eng.*, ICBBE, pp. 932-935, July 2007.

[14] B Kartikeyan and A Sarkar," A unified approach for image segmentation using exact statistic," *Comput.Vis. Graph. Image Process.*, vol.48, pp.217-229, 1989.

[15] V. Kecman, *Learning Soft computing Support vector Machines, Neural Networks, and Fuzzy logic Models*, MIT press, Cambridge, 2001.

[16] S.Z.Li, *Markov Random Field Modeling in computer vision*, New York:Springer-Verlag, 1995.

[17] F Melgani, L.Bruzzone,"Classification of Hyperspectral RemoteSensing Images With support Vector Machines" *IEEE Trans. Geosci. Remote Sensing*, vol.42, pp. 1778-1790, 2004.

[18] A. C. Moses, A. K. Skidmore, "A new technique for extracting the red edge position from hyperspectral data: The linear extrapolation method", *Remote Sensing of Environment*, Vol. 101, Issue 2, pp. 181-193, 2006.

[19] A. Renyi, *Probability theory*, Amsterdam, North -Holland, 1970.

[20] A. Sarkar, M. K. Biswas, B. Kartikeyan, V. Kumar, K. L. Majundar and D. K. Pal, "A MRF Model Based Segmentation Approach to Classification for Multispectral Imagery", *IEEE Trans. Geosci. Remote Sensing*, vol.40, pp. 1102-1113, 2002.

[21] F. D. Van der Meer, S. M. De Jong and W. Bakker, " Analytical techniques in spectrometry", *Imaging Spectrometry: Basic principles and prospective applications*, Kulwer Aca. Pub., Amsterdam, 2003.

[22] V. N. Vapnik, "*Statistical Learning theory*", Wiley, New York, 1998

[23] J. Weng, Y. Zhang and W. S. Hwang, "Candid Covariance-Free Incremental Principal Component Analysis",*IEEE Trans. Pattern Ana Mach Int*, vol.25, No. 8, pp. 1034-1041, Aug. 2003.

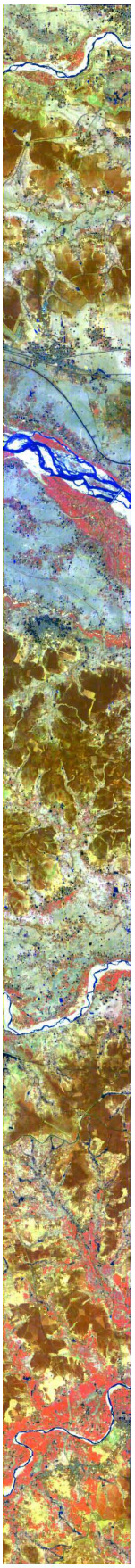


Fig 2(a): Original Image

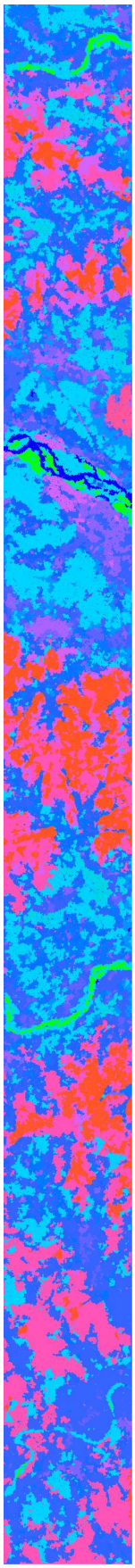


Fig 2(b): PCA Segmented Image

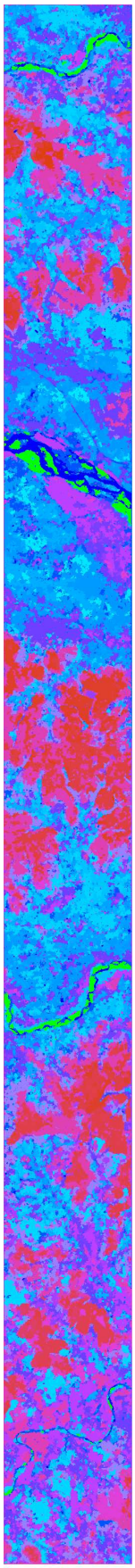


Fig 2(c): Segmented Image - Min Energy

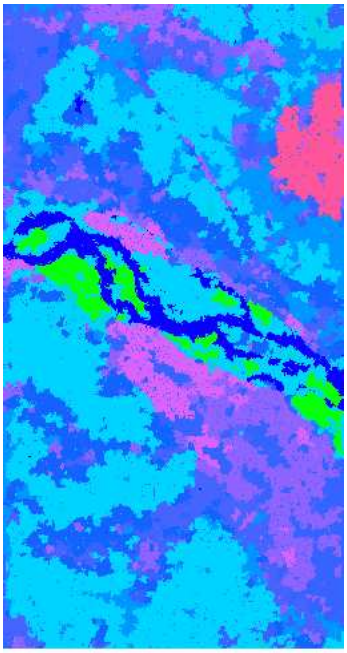


Fig 3(a): Subscene of Fig 2(b)

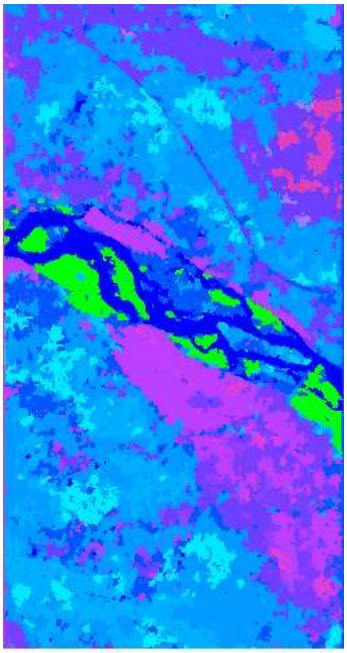


Fig 3(b): Subscene of Fig 2(c)

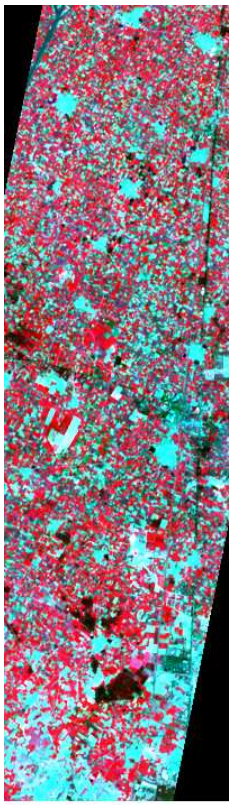


Fig 4(a): Original Image

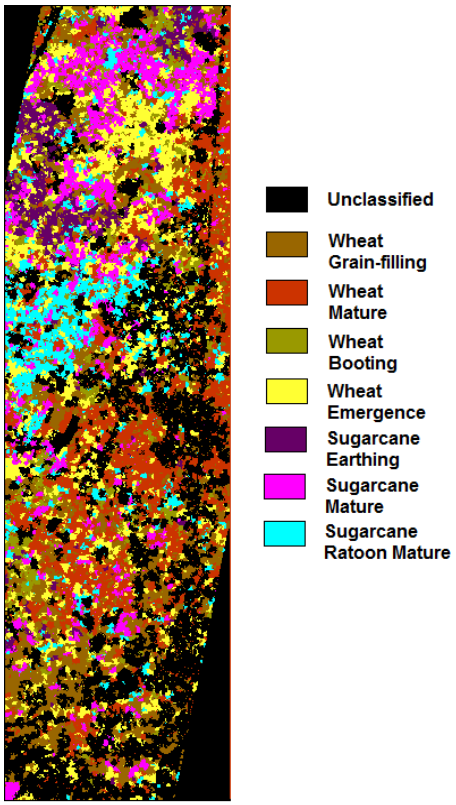


Fig 4(b): Classified Image

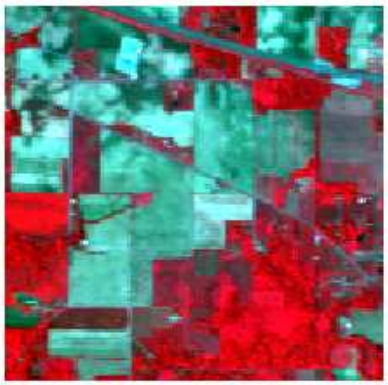


Fig 5(a): Original Image

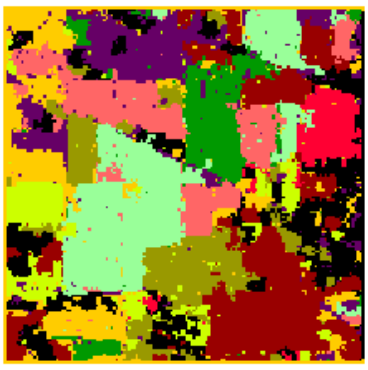
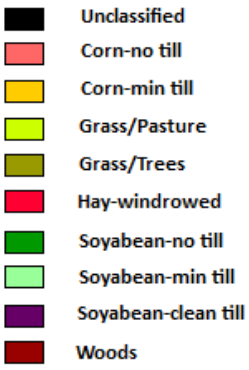


Fig 5(b): Classified Image



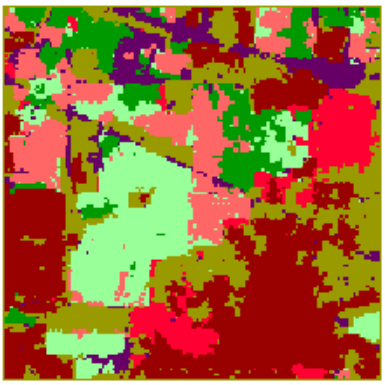


Fig 5(c): Tuned Image