

Research Article

Unsupervised and Supervised classification of Hyperspectral Imagery data using Projection Pursuit and Markov Random Field Segmentation.

A Sarkar ^{*}, A Vulimiri [†], S Paul [•], Md. Jawaid Iqbal [†], A. Banerjee [†], R. Chatterjee[†] and S S Ray [‡]

^{*} Department of Mathematics, IIT Kharagpur,
India.(anjan@maths.iitkgp.ernet.in)

[†] Department of Computer Science and Engineering, IIT Kharagpur, India.

[•] Department of Electronics and Communication Engineering, IIT Kharagpur,
India.

[‡] Space Applications Center, Ahmedabad, India

(Received 00 Month 200x; in final form 00 Month 200x)

This work presents a classification technique for hyperspectral image analysis when concurrent ground-truth is unavailable and available. The method adapts a principal component analysis based projection pursuit (PP) procedure with an entropy index to reduce the dimensionality followed by the Markov Random Field (MRF) model based segmentation. An 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 with the desired number of classes so that it resembles the natural scene closely. When the landcover classes are in detailed level, some special reflectance characteristics based on the classes of the study area in question are determined. These are later incorporated in MRF model based segmentation stage while minimizing the energy function in the image space. Segments are evaluated with training samples so as to yield a classified image with respect to the type of ground-truth data. Two illustrations are presented with (i) EO-1 Hyperion sensor image with concurrent groundtruth at detailed level classes and (ii) AVIRIS-92AV3C image with concurrent groundtruth - for supervised cases. Comparison of classification accuracies and computational times of some nonparametric approaches with that of the proposed methodology are provided for the illustrations. Experimental results demonstrate that the method provides high classification accuracy and is computationally faster compared to other methods.

1. Introduction

In this paper we address the problem of classifying very high dimensional image (hyperspectral) data (Hyperion Sensor, 2000, AVIRIS, 1992) . AVIRIS stands for Airborne Visible Infrared Imaging Spectrometer. Such imagery affords one all discriminating details needed for fine delineation of many agriculture classes or material classes. Our aim is to build a classification technique for hyperspectral images that yields high accuracy and is computationally faster.

Investigations have 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 (Van der Meer *et al.*, 2003). Mention

This work was supported by Indian Space Research Organization(ISRO) under the grant for “Classification of hyperspectral remote sensing data to discriminate between crop condition, variety and stage”, Ref. ISRO/RES/4/535/2006-07, March 29, 2007.

may be made of a semi-supervised graph-based scheme (Camp-Valls *et al.*, 2007) and supervised classification for AVIRIS data (Benendiktsson, 1995), among others. An overview on the recent advances in the techniques for hyperspectral image analysis is available in Plaza *et al.* (Plaza *et al.*, 2009).

The classification techniques that have been successfully applied to multispectral data viz., Gaussian maximum likelihood (ML) etc., are not effective for hyperspectral image data. This is because of mathematical and practical limitations such as Hughes Phenomenon (Huges, 1968) and others. The main reason is related to small ratio between the number of available training samples and the number of features used in a classifier (parametric) in the context of supervised classification. The classification accuracy decreases with increasing number of features in the classifier after some optimum value, which depends on the number of training samples and the kind of classifier adopted. To circumvent this problem, some valuable methods are introduced in the literature. Covariance matrix regularization technique that can be applied to estimate covariance matrix appropriately for ML classifier have been addressed by (Tadjudin and Landgrebe, 1999, Hoffbeck and Landgrebe, 1999) for limited training sample data.

Some interesting approaches for hyperspectral image analysis are due to (Shashahni and Landgrebe, 1994, Jackson and Landgrebe, 2002), etc. among others. The former work is the first that demonstrates the advantage of using unlabeled data to improve the performance of the classifier with some success but when those samples come from outliers, boundary pixels and unknown classes the performance deteriorates. The main focus of this work has been the use of unlabeled data via expectation maximization(EM) to obtain better estimates of the class specific parameters. Subsequent extention to EM approach have been carried out by (Jackson and Landgrebe, 2002, Dundar and Landgrebe, 2004) with semilabeled data in EM iterations. In these approaches the available labeled data are first used to train a supervised classifier to obtain a tentative labels for the unlabeled data. The semilabeled data thus obtained are then used to modify the classifier iteratively. Investigation due to (Li, *et al.*, 2010) of using labeled and unlabeled training samples to improve classification accuracy shows a greater potentiality as it includes spatial information also.

On the other hand, nonparametric approaches viz., Artificial Numeral Network(ANN), SVM and etc., may be applied in some cases but the computation cost of such approaches is often much higher than that of parametric classifiers (Fakunaga, 1990). The larger the number of training samples or the number of classes, the greater is the computation time in these supervised nonparametric approaches. The methodology of considering spatial information along with spectral information to improve classification accuracy of hyperspectral image has been seen as a recent trend. The approaches due to (Fauvel, *et al.*, 2008, Bolton and Gader, 2009, Chen *et al.*, 2009B, Velasaloco-Forero and Manian 2009) show some degree of success. We attempt our investigation in similar lines by combining spectral and spatial information.

To develop a classification technique that yields high classification accuracy vis-a-vis computationally faster we may carry out two-stage processing. In the first stage- we adapt a method that allow reduction of data dimensionality with small loss of information. In the second stage- we carry out a suitable classification methodology with reduced dimensionality which is not computationally intensive and at the same time yields high classification accuracy.

Principal Component Analysis (PCA), an exploratory data analysis technique, is the most common and conventional data reduction method. This is generally used in the absence of any prior knowledge of a scene. PCA is performed for capturing

largest variations in projection and is optimal when the background associated with signal source (the clutter) is Gaussian. As said in (Ifarraquerri and Chang, 2000) information content in hyperspectral images does not always coincide with such projection. In hyperspectral images the clutter includes contribution from interference sources such as natural background signatures as well as nonrandom noise (Chang and Du, 1999). Furthermore maximum variance is not the only information content for data reduction that we are interested in. In fact we seek such projections that should have a cluster-detecting ability.

Some interesting projection indices available in literature are due to Friedman and Tukey, Huber, Jones and Sibson, Chiang and Chang, Jimenez and Landgrebe, and Ifarraquerri and Chang among others. The projection indices the above authors have employed are of three types, viz.

- (i) class distance measures (e.g., Bhattacharyya distance (Jimenez and Landgrebe, 1999), The Friedman-Tukey Index (Friedman and Tukey, 1974)),
- (ii) entropy indices or information divergence indices (Ifarraquerri and Chang, 2000), and
- (iii) moment indices (Jones and Sibson, 1987, Chiang and Chang, 2001).

Friedman criticized the moment index (Jones and Sibson, 1987), that it was strongly attracted to projections which contained outliers. Jones and Sibson again emphasized that Friedman and Tukey's index searches for departures from parabolic form of the projected density rather than detecting clusters. The concept of using entropy index emerges from the analysis of the statistical behavior of very high dimensional data sets in lower dimensions (Renyi, 1970, Huber, 1985). With such a projection index, Ifarraquerri and Chang have reduced the dimensionality from 210 to 11 for a 256×256 HYDICE sensor scene. The component images resulted from PCA and PP as exhibited in (Ifarraquerri and Chang, 2000) show the cluster-detecting ability of the methodology. But the computation involved in such a methodology is intense. 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, each of size 256×3128 – would be unrealistically high.

As seen above, the first task to reduce the dimensionality can be carried out with projection index as described in (Ifarraquerri and Chang, 2000) in a manner that reduces computation time. Therefore instead of determining the absolute best projection with the whole data set, we seek to obtain a projection which is good enough i.e., which has a high probability as per the ordinal optimization suggested by (Ho, 1999). This softening of the goal eases the computational burden in our problem and it is much easier to find a projection which falls within the top τ such that some meaningful solution can be obtained in a faster manner.

The next step is to carry out classification with the resulting component images to attain high classification accuracy. To do so we adapt an MRF based segmentation approach to classification due to (Sarkar *et al.*, 2002). In this MRF approach, the PP component with highest projection index value is initially segmented with a tonal-region characteristics based segmentation technique(Kartikeyan and Sarkar, 1989). The use of such an initial segmentation technique on a hyperspectral image that contains dominant textural property will result in oversegmented regions. An MRF is defined on the region adjacency graph(RAG) of these oversegmented regions. Information of other PP components are suitably incorporated in the RAG. Due to the energy minimization process associated with the underlying MRF, oversegmented regions are expected to be merged yielding an optimal segmentation.

With the value of the energy function associated with the MRF model, each projection data in terms of its cluster detecting ability may be ranked. The value

of the energy function that yields the minimum value in the segmentation process determines the best segmented output among all the iterations (projection) based on the random samples and thereby determining in a faster manner, a segmented output which is *good enough* among such projections.

For a hyperspectral image we may note that apart from its image space there exists a significant amount of information from the spectral reflectance curve of each pixel with respect to the different material characteristics. In case of the crop area under study, the area may comprise detailed level classes viz., the different stages of a crop. In such situations the crop's biophysical characteristics with regard to stages of a crop may be extracted from the spectral reflectance curve of each pixel and can be exploited in our segmentation approach to classification methodology. This type of fusion of image space and per pixel spectral reflectance information may help further in determining such segments so as to yield an improved classified image. The biophysical parameters that may be used, viz., 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 REP as per the stages of crop-growth may be suitably incorporated to obtain the segmented image. In general depending on the detailed level of classes, suitable parameters may be evaluated accordingly and if integrated with the image space in the segmentation process, it would yield a classified image with improved accuracy.

In the absence of groundtruth this segmented output can be further clustered into a desired number of classes 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. We term such an output as a tuned image. When groundtruth samples are provided a cluster validation scheme is carried out on the segmented image to obtain a classified image. 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 an ordinal optimization scheme in an MRF model based segmentation procedure for analyzing hyperspectral images. This methodology also takes into account how the significant amount of information in the feature space(radiance curve for different wavelengths) available in the hyperspectral images is analyzed for detailed level classes and integrated in the segmentation process in a meaningful way. Although the paper is based on two previous works (Ifarraeguerri and Chang, 2000) and (Sarkar *et al.*, 2002) it develops a novel technique for the ordinal optimization in which the two works are appropriately combined resulting in a methodology which is computationally faster and expected to yield high classification accuracy. The proposed methodology has been tested with one EO-1 Hyperion sensor image and one AVIRIS image data. The experimental results demonstrate that the proposed methodology is effective for achieving high classification accuracy in a faster manner.

The paper is outlined as follows. Section 2 and Section 3 describe briefly the data reduction method with the ordinal optimization scheme and measures of crop biophysical parameters respectively. The MRF model based segmentation scheme is presented in Section 4. Section 5 describes the classification procedure -unsupervised and supervised. In Section 6 we illustrate two experimental results with discussions followed by the conclusion in Section 7.

insert figure 1 here

2. Data Reduction

Projection Pursuit (PP):

For the completeness of the work, the methodology of Ifarreguerri and Chang and the way we use it are briefly described in this section.

To find the best projection, in general 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 which is usually applied to data before the application of projection pursuit. In PCA all the components are given equal weights by scaling them with their corresponding eigenvalue and the observed variables are transformed into uncorrelated variables with unit variance. In other words, suppose $\mathbf{Z}_{d \times h}$ is a sample of h vectors of d dimension with mean $E[\mathbf{Z}]$ and covariance $\Sigma = [\sigma_{ij}]$. The spectral decomposition of Σ is $\Sigma = \mathbf{U}\Lambda\mathbf{U}^T$ with \mathbf{U} orthogonal, whose column contains eigenvectors and Λ , the diagonal eigenvalue matrix. The vector \mathbf{Z} is transformed by $\mathbf{Z}' = \Sigma^{-\frac{1}{2}}(\mathbf{Z} - E[\mathbf{Z}]) \Rightarrow E[\mathbf{Z}'] = 0$ and $\Sigma' = \mathbf{I}$, where $\Sigma^{-\frac{1}{2}} = \mathbf{U}\Lambda^{-\frac{1}{2}}\mathbf{U}^T$. The data set is thus *sphered* by using a linear transform to cause the transformed data to have zero mean and unit variance matrix and which in turn simplifies the design of projection pursuit (Nason, 1995).

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 the significant principal components. The hyperspectral image data, consisting of $r \times c = h$ pixels over d bands, is arranged in a matrix $\mathbf{Z}_{d \times h}$. PCA is performed using the covariance-free method" (Weng *et al.*, 2003) 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

$$\mathbf{Y}_{k \times h} = \mathbf{E}^T \mathbf{Z}, \quad (1)$$

a copy of which is saved as \mathbf{Y}_{orig} . A random sample consisting of N pixels from k component bands has been selected and the projection index as described in (Ifarraeguerri and Chang, 2000) is calculated as

$$J^{(l)}(\mathbf{p}, \mathbf{q}) = \sum_i p_i \log \frac{p_i}{q_i} + \sum_i q_i \log \frac{q_i}{p_i}, \quad l = 1, 2, \dots, k \quad (2)$$

where $\mathbf{p} = \mathbf{Y}^T Y_l$ is the projection score vector (p_i being the i th normalized projection score in l th direction) and $\mathbf{q} = [q_i]$ is a $(N+1) \times 1$ vector (considering floor of $\frac{N}{2}$ as a substitute of $\frac{N}{2}$ when N is odd), $q_i = \int_{(i-1)\nu}^{(i+1)\nu} f dx$, i running from $\frac{-N}{2}$ to $\frac{N}{2}$ and f and ν are respectively the normal density and bin size.

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

$$l_{best} = \arg \max_l (J^{(l)}(\mathbf{p}, \mathbf{q})) \quad (3)$$

Then the vector $\mathbf{y}_{l_{best}}$ corresponding to the l_{best}^{th} pixel spectrum is termed as the optimal projection vector. $\mathbf{y}_{l_{best}}$ is normalized and is appended as a column \mathbf{w}_r of a matrix \mathbf{W} :

$$\mathbf{w}_r = \frac{\mathbf{y}_{l_{best}}}{\|\mathbf{y}_{l_{best}}\|}, \quad (4)$$

Subsequently, to search for the directions orthogonal to the columns \mathbf{w}_r of \mathbf{W} , at each step \mathbf{Y} is projected orthogonal to \mathbf{W} by the following:

$$\mathbf{Y}^{(new)} = [\mathbf{I} - \mathbf{W}(\mathbf{W}^T \mathbf{W})^{-1} \mathbf{W}^T] \mathbf{Y} \quad (5)$$

$\mathbf{Y}^{(new)}$ is the new \mathbf{Y} which allows us to search for new orthogonal projections. This procedure (eq(2) through eq(5)) is carried out k times to obtain a $k \times k$ matrix \mathbf{W} . For the optimally projected data set, we perform

$$\mathbf{Y}_{opt} = \mathbf{W}^T \mathbf{Y}_{orig} \quad (6)$$

The entire procedure described above (eq(1) through eq(6)) is repeated n times. Each iteration is performed by taking a random sample \mathbf{Y} from the data matrix given in eq(1) followed by the steps as described in eq(2) through eq(6) so as to determine \mathbf{W} . Subsequently \mathbf{Y}_{opt} , the optimally projected data is determined with the MRF based segmentation scheme which is described in the next section. The number of iterations n depends on the size of the random sample and the degree of proximity (to the true solution) we desire to attain as per ordinal optimization due to (Ho, 1999). The criterion by which we identify the best solutions among the number of iterations considered is the minimum segmentation energy (top rank information) associated with the MRF procedure.

Under the assumption that the optimum direction (of projection) lies in the vicinity of the point cloud and can be approximated by the point nearest to it in terms of the projection score, a search with a random sample of size N for a number of iterations is expected to yield the desired projection. The MRF segmentation approach adopted here determines the optimum one. However if groundtruth data is available some samples belonging to each class may be accommodated in this set. In this way we optimize the projections and in this sense, our method differs from the approach of Ifarraguerri and Chang. Their approach suggests that the projections are determined either with a fixed 1000 pixels sampled from the image at uniform intervals or with the whole set of image data. In the former case, the projections may not be *good enough* and in the latter case the procedure is too computational expensive.

3. Measures of crop biophysical parameters

If the hyperspectral image under study comes from an agricultural area where there are detailed level crop classes viz., emergence, milking and mature etc. as different stages of crop growth, we may take note of some of the key biophysical parameters such as NDVI values and REP etc. These parameters, NDVI and REP play an important role in explaining the plant condition and growth stages respectively. Shifts in the REP in longer or shorter wavelengths have been used as a means to estimate changes in foliar chlorophyll content and are also indicators of vegetation stress (Smit *et al.*, 2004)). Based on this notion we may investigate REP values corresponding to different stages of a crop and use it as an indicator of the different stages of a crop. We briefly discuss how the measures viz., NDVI and REP values can be determined from the spectral reflectance curve. NDVI is usually calculated with reflectance on the wavebands NIR=803nm and RED=671nm , using the index, $NDVI = \frac{IR(803nm) - R(671nm)}{IR(803nm) + R(671nm)}$.

The REP for each crop pixel are calculated with radiance curve in the wavelength region 680nm to 780nm. First, the derivative curve of the spectral reflectance spectrum in the above range is determined and smoothed by a polynomial fitting of the curve (Moses and Skidmore, 2006). Two straight lines are fitted on the derivative curve in each flank 680 and 700, and 725 and 760. One at two points of the far-red (680 and 700) and another at two points of the NIR (725 and 760) spectrum. The REP is defined by the wavelength value at the intersection of these two extrapolated lines. At the intersection, the two lines have equal λ (wavelength) and the first derivative reflectance values. REP is the λ at the intersection and is computed by the equation, $REP = -\frac{c_1 - c_2}{m_1 - m_2}$, where c_1 and c_2 , and m_1 and m_2 represent the intercepts and slopes of the far-red and NIR lines respectively. Far-red wavebands at 680 and 690nm in combination with NIR wavebands 730.50 and 760nm have been considered as the optimum combination to identify the stages of the crop.

With positive NDVI values, a pixel is identified as a crop pixel. The REP value of that pixel is then determined. All those pixels comprising a particular crop of a specific stage over the homogeneous region yields an average REP value corresponding to that stage of the crop for this region. Thus every homogeneous region, coming from a particular crop area is characterized as per its stage with regard to the REP values. These values for different stages of crops for groundtruth samples have been found to be distinct. We note that the REP values shifts towards higher wavelength in the prescribed range as mentioned earlier as the stages grow to maturity stage. After maturity it shifts towards shorter wavelength in the prescribed range. In the segmentation stage, this information of biophysical parameters may be taken into account as one of the criteria along with the image space to determine the value of the energy function associated with the underlying MRF model.

4. MRF model based segmentation scheme

We follow the scheme of (Sarkar *et al.*, 2002) in defining the MRF on a RAG of initial oversegmented regions - the details are available in (Sarkar *et al.*, 2002).

First, a tonality based initial segmentation procedure (Kartikeyan and Sarkar, 1989) is carried out on the projection pursuit component image with highest projection index value. Subsequently the same regions $R_i(1), i = 1, 2, \dots, Q$ 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 $R_1 = R_1(p), R_2 = R_2(p), \dots, R_Q = R_Q(p), p = 1, 2, \dots, P$, 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. An MRF is defined on the 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 . Through an MRF model, one can incorporate the desired statistical properties among the nodes of the RAG by choosing appropriate cliques and clique potential functions. We note that clique c is defined as a fully connected subgraph of the RAG, Γ unlike the (Dubes and Jain, 1989, Jackson and Landgrebe, 2002) where a clique c is defined in a usual way as a subset of points in image lattice \mathbf{S} in which every member is a neighbor of the other. The clique in (Li, et al., 2010) is defined on the set of points of first and second-order neighborhoods with respect to distance between horizontal and vertical neighbors.

Our discussion here is primarily directed in formulating the energy function used in the MRF based segmentation approach. Minimizing this energy function will result in a maximum *a posteriori* probability (MAP) estimate of the optimal segmented image. We impose two constraints as per our notion of optimal segmentation.

- (i) An optimal segmented image region R_i should be uniform with respect to the measured characteristics as obtained from the projection pursuit component images.
- (ii) 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. Furthermore the region pairs should be distinct with respect to ancillary information, if any, from the radiance curve of the hyperspectral image.

Let $\mathbf{X} = \{X_i, 1 \leq i \leq Q\}$ denote a family of random variables representing the labels of the regions $R_i = R_i(p), p = 1, 2, \dots, P$ in R . Here each X_i is a discrete random variable describing the label for all P component channels, taking values from a finite set $\omega = \{\omega_1, \omega_2, \dots, \omega_q\}$. Let $\Theta = \{(x_1, x_2, \dots, x_Q) : x_i \in \omega, 1 \leq i \leq Q\}$ be the set of all configurations. Further, let $\mathbf{Y} = [Y_1, Y_2, \dots, Y_p]^t$ be a multicomponent random variable representing pixel intensities of the component images and \mathbf{y} be a specific realization. We can now deal with the segmentation problem in the MRF framework by the posterior distribution given the image \mathbf{y} (Li, 1995).

The Bayesian estimate of the configuration $\{\mathbf{X} = \mathbf{x}\}$ maximizes the posterior distribution of \mathbf{X} , given the realization $\{\mathbf{Y} = \mathbf{y}\}$. The posterior distribution also follows a Gibbs distribution under the assumption of conditional independence of the likelihoods (Dubes and Jain, 1989)

$$P(\mathbf{X} = \mathbf{x} | \mathbf{Y} = \mathbf{y}) = \frac{e^{-U_{ps}(\mathbf{x}|\mathbf{y})}}{\Pi_{ps}} \quad (7)$$

where

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

using the relation $x = e^{ln(x)}$. This enables us to express the posterior energy function as a sum of the potentials at different points in the image lattice (Dubes and Jain, 1989, Dubes *et al.*, 1990). The term $\Pi_{ps} = \sum_{x \in \Omega} e^{-U_{ps}(\mathbf{x}|\mathbf{y})}$ is the normalizing constant. The maximum *a posteriori* estimate $\hat{\mathbf{x}}$ of \mathbf{x} is obtained by minimizing the posterior energy function $U_{ps}(\mathbf{x}|\mathbf{y})$. 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 by two processes, Region Process H and Edge Process B as

$$V_c(\mathbf{x}|H, B) = V_c(\mathbf{x}|\mathbf{y}) = \eta_{ij} \mathbf{W}_{i,j} + \theta_{ij}(1 - \eta_{ij}) \mathbf{B}_{ij} \quad (8)$$

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

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

\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. η_{ij} takes the value 1 when the following two conditions are together satisfied. 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 the clique potential) are homogeneous with respect to the multi-component channel pixel intensity values. If the NDVI values are not greater than 0 the first condition is ignored. If any of these conditions are violated, η_{ij} takes the value zero. The case $\eta_{ij} = 1$ indicates that $x_i = x_j$, that is, both the regions have the same label.

It appears that through this variable η_{ij} , integration of pixel based spectral reflectance information taken as an average over each of the regions that brings about the homogeneity of the pair of regions at detailed level can be carried out. If the study area concern is agriculture, the information on REP will certainly help further in distinguishing similar agricultural subclasses(different stages of a crop) yielding an improved classified image. We also note that, if such parameters values are not tractable from the classes only the second condition needs to be satisfied.

The parameter θ_{ij} in eq(8) controls the weight to be given to the two processes(Region H and Edge B) 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

$$T^2 = (\mathbf{M}_i - \mathbf{M}_j)'[(1/n_i + 1/n_j)S_{pooled}]^{-1}(\mathbf{M}_i - \mathbf{M}_j) \quad (9)$$

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 $U_{ps}(\mathbf{x}|H,B) = \sum_{c \in C} V_c(\mathbf{x}|H,B) = \sum_{c \in C} V_c(\mathbf{x}|\mathbf{y})$ as described above.

With an objective of separation of classes, the better the PP carried out with a random sample, the better would be the MRF segmented output with lesser value of the energy function. Corresponding to every random samples of size N chosen from PCA component images, PP is performed and the projected data yields a number of component images. Subsequently segmentation is performed for the desired number ($P \leq k$) of component images at each iteration. For n such iterations 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.

5. Classification

Unsupervised scheme

The unsupervised segmentation of projection pursuit component images results, in general, in a large number of clusters (segments) say, R_j , $j = 1, 2, \dots, L$ with unique cluster numbers being assigned to each pixel. We may tune(group) these regions into lesser number of segments(classes) in the absence of any ground truth knowledge with the following two steps.

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

Follow steps 1 and 2 until $R^o - N_c$ number of times where R^o and N_c are respectively the initial number of regions in the segmented image and the desired number of classes. We note that at each iteration, the number of regions decreases by one and thus after $R^o - N_c$ iterations we get the desired number of regions(classes).

The tuned segmented output thus exhibits the different regions in as many gray shades as the number of classes N_c . To make the regions (classes) of the images clearer and distinct a suitable technique may be applied to transform the image from gray to pseudocolor. The technique due to Jinxiu et al (Jinxiu *et al.*, 2007)) performs the task to some extent.

Supervised Scheme

We now validate those clusters. Unlike the case in (Sarkar *et al.*, 2002), 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, viz., different crop classes. We note here that in actual practice a scene contains a variety of classes viz., water, urban, forest etc. In our case, let there be C such crop classes whose ground-truth data are available. First, the ground truth samples of all the C classes are first overlayed on the segmented image. The segment that has groundtruth samples will have a label as per the groundtruth class. In this stage only C crop classes would be identified with the number of segments greater than or equal to C . Next we pick up all those unlabeled clusters R_j , one at a time and compare with each of the C labeled clusters with respect to their means. In this comparison we actually attempt to reconcile 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 $i = 1, 2, \dots, C$. Now we set up the null hypothesis H_0 : the given unlabeled segment R_j is identical to the i th 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(9) may be used. Therefore, the segments R_i and R_j should have the same label if $T_{ij}^2 < F_\alpha$ and otherwise, if $T_{ij}^2 \geq F_\alpha$, where $P[T_{ij}^2 > F_\alpha] = \alpha$, $\mathbf{S}_{pooled} = \frac{\mathbf{S}_i + \mathbf{S}_j}{\nu_{ij}}$ and \mathbf{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_{ij}^2 is found to be minimum. An unlabeled segment will remain unclassified at the level of significance α if $T^2 \geq F_\alpha$ for each of the prototype classes.

A point to be noted is that at the first stage as stated above, while identifying the prototype clusters(segments) the ground truth samples of all the C classes are first overlain on the segmented image. The number of ground truth samples of a particular class may either fall in a number of fragmented segments each with a small number of pixels that are supposed to belong to the same class or in a single segment of a large number of pixels. All these fragmented segments or the large segment will now have a specific groundtruth class label. If the total number of pixels of these labeled segments or that of the single large labeled segment is at least thirty, then the estimate of the parameters viz., \mathbf{M}_i and \mathbf{S}_{pooled} in eq(8) with few PP components are based on large samples. Such a value of sample size is usually considered large on the principle of central limit theorem. This means that even with a moderate number of training samples for each of the class this methodology in some cases would be able to carry out the cluster validation appropriately.

6. Experimental Results

Two illustrations of hyperspectral image analysis are presented. The first example deals with an AVIRIS-92AV3C image which was collected from the website (AVIRIS, 1992) along with its groundtruth. As it has sufficient groundtruth data,

rigorous comparison of some nonparametric and a parametric(ML) method with the proposed methodology are provided. Subsequently the quantitative evaluation of the technique of unsupervised classification as described in section 5 has been performed for this example to demonstrate its performance. The second example deals with an image acquired from the Hyperion sensor aboard the EO-1 space-craft. This has very limited concurrent ground truth data on detailed classes, viz., on different stages on the same crop and as such cluster validation has been carried out carefully to examine the potential of the proposed methodology with regard to its attainment of high classification accuracy.

Example-1

For this illustration we consider the AVIRIS-92AV3C, (Indian Pine test site) image (AVIRIS, 1992) of size 145×145 for 220 -bands with enough groundtruth samples and compare the classification accuracy of the proposed methodology with a nonparametric method. As this image (Fig.2(a)) has been extensively studied by Melgani and Bruzzone in (Melgani and Buzzon, 2004, Chen et al., 2009B) we follow the same procedure as in (Melgani and Buzzon, 2004, Chen et al., 2009B) for using groundtruth data. First, the fifty percent groundtruth samples(4758) have been used for training and the remaining (4587) samples of the total 9345 have been used for testing [see Table-I in (Melgani and Buzzon, 2004)]. The groundtruth comprises sixteen classes, seven classes are discarded because of insufficient number of samples as mentioned in the procedure suggested in (Melgani and Buzzon, 2004) and the remaining nine classes have been considered. For this scene the crop classes are not in detailed level and as such NDVI values and REP etc are ignored as these will not play any role to enhance classification accuracy.

The search space of the image is $|W| = 145 \times 145$. The method used for PCA computation of d-dimensional space is based on covariance-free technique (Weng *et al.*, 2003). The number of significant PCA components(VD) has been found to be 11, obtained by Malinowski's method. Although this method overestimates the VD in real image data to an extent, it is used for the ease of computation as it is based on only eigenvalue findings of d-dimensional space. Other methods are based on computationally intensive procedures. We now determine the probability π that at least one sample of $500 (= N)$ from the search space $|W| (= 145 \times 145)$ will yield one of the best projections on the top $20 (= \tau)$ with the expression, $1 - (1 - \frac{\tau}{|W|})^N$ (Ho, 1999) and thereby that of the number of iterations as $\frac{1}{\pi} \simeq 3$. With 500 random samples, we desire to carry out PP and MRF with 11 PCA component images for 3 iterations. The MRF segmentation has been carried out with 11 PP component images in each of the 3 iterations. Subsequently classification accuracy has been determined with the training and test set data for the segmented image having minimum energy in these three iterations.

For the proposed methodolgy we note that the number of sample size to carry ut PP is one of the important criterion to make the procedure computationally fast vis-a-vis to determine the high classification accuracy. As such we have performed the same exercise for two other cases, each with sample size 300 and 400 for three iterations in each case. Table-1 summarises the classification accuracies and the computation time for different cases. The value of τ that the solution ranks as one among top τ with corresponding sample size has also been provided in Table-1. The confusion matrix for PP-MRF with 500 samples (case(v)) has been presented in Table-2.

Malgani and Bruzzone (Melgani and Buzzon, 2004) have considered different multiclass strategies one against all (OAA), One Against One (OAO), Binary Hi-

erarchical Tree-Balanced Branches (BHT-BB) and Binary Hierarchical Tree-One Against All (BHB-OAA) schemes with these data set. Since OAO multiclass strategy gives the highest classification accuracy among all the multiclass strategies (Melgani and Buzzon, 2004) we have compared the classification accuracy and the computation time of the proposed methodology with that of OAO strategy in the same system. The classification accuracy and computation time for One Against One (OAO) multiclass strategy(Case(i)) using nonlinear SVM based on Gaussian kernel function(SVM-RBF) is given in Table-1 The confusion matrix for this approach has been presented in Table-3.

The significance test of the difference between the two accuracies viz., for the proposed methodology and of SVM RBF has been done using McNemar's test (Agresti, 1996, Bradley, 1968). The test is based on the value of the standardized normal test staistic viz.,

$$z = \frac{f_{12} - f_{21}}{\sqrt{f_{12} + f_{21}}},$$

f_{ij} being the frequency of sites lying in 2×2 confusion matrix element i, j (Foody, 2004). The elements f_{ij} are derived from the confusion matrices of the two classifiers whose accuracies are being compared. For case(i) and case(v), this statistic yields the value 4.66 (see Table-4), implying that they are significantly different. Likewise, other cases are compared similarly, and the findings are reported in Table-1. The (*) in Table-1 indicates that the PP-MRF classifier is significantly different from SVM-RBF with respect to the overall accuracy.

insert table 1
insert table 2
insert table 3
insert table 4

Table-5 presents class by class accuracies of SVM-RBF OAO multiclass strategy along with the 3 cases of the proposed methodology. From these values it appears that the proposed methodology (PP-MRF) produces consistent performance for all the classes for different sample sizes. It is interesting to note that for the class ω_6 , the accuracy is relatively low for all the cases presented in Table-5.

Further, the conventional nonparametric approaches viz. SVM-linear, SVM-polynomial and K-nn classifier yield classification accuracies 87.64%, 86.42% and 82.63% respectively as reported also in (Melgani and Buzzon, 2004). The simple parametric(ML) technique has also been used for this dataset with 11 PCA component images as well as with 220 bands. The accuracies obtained are found to be 66.1% and 53.1% respectively. Such low values of accuracies can be attributed to the reason that this method does not consider contextual information of pixels which is an essential criterion for a technique of achieving high clasification accuracy. Furthermore we have also compared our results with the classical approach(7) of (Tadjudin and Landgrebe, 1999) with the groundtruth samples comprising the 7 crop classes. Twenty percent of the total 6541 has been used for training and remaining 5333 have been used for testing. The proposed methodology(PP-MRF 500,11,11) yields the accuracy of 94.01%. It may be noted that the accuracy reported in (Tadjudin and Landgrebe, 1999) is 89.10% for these seven crop classes.

insert Figure 2(a), 2(b)

Now we compare our result with the work of (Chen et al., 2009B) following 11 classes groundtruth data as considered by them. The approach is based on SVM and uses stacked generalization along with complementary information of magnitude and shape feature spaces. Among several variation in using the stacking method with respect to the feature spaces, the stacking with decision values of binary SVMs trained in the magnitude feature space, called SSVM-V-M yields superior results in their first experiment. Next experiment has been carried out to show improved

discriminant capability by using both stacked generalization and combined feature space. The method has been termed as SSVM-V-MS. Subsequently to reduce the computational complexity, this method has been considered with meta level feature selection based on Bhattacharyya distance termed as SSVM-BD. The latter approach does infact reduces the computation time than that of the SSVM-V-M approach but with decreased accuracy.

The proposed methodology(PP-MRF with 500 sample) has been compared with the results of the three approaches viz., SSVM-V-M, SSVM-V-MS and SSVM-BD on AVIRIS data for the 11 ground truth classes. The user's accuarcy and producer's accuracy have been computed with the PP-MRF approach in Intel Pentium D 2.80 GHz CPU with 1-GB RAM. The findings along with OA and computational time have been reported in Table-6. As shown in (Chen et al., 2009B) the OA due to SSVM-V-MS approach is significantly different than that of the other two aproaches viz., SSVM-V-M and SSVM-BD and so is OA of the PP-MRF method. The significance test of the difference between OA's of the PP-MRF and SSVM-V-MS by McNemar's test yields the value of $z= 1.60$. This suggests that the OAs are not significantly different but the computation time of PP-MRF method is one fifth of the time expended by SSVM-V-MS method.

insert table 5

insert table 6

This high classification accuracy as well as the classified image, Fig.2(b) suggest that the methodology has a potential. The black regions that are observed in Fig.2(b) are unclassified regions and correspond to those areas whose ground truth samples have not been considered in the training set. The roads and the railway track that are visible (as thick lines) in the original image from NW to SE are found here as unclassified classes in Fig.2(b).

Finally we compare our unsupervised classified image quantitatively with this data. This is just to verify how accurately the segments are labeled with groundtruth classes. 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 an unclassified class with respect to the groundtruth class labels.

insert Figure 2(c)

The overall classification accuracy is found to be 59% for this unsupervised case (Fig.2(c)). This classification accuracy, although not high, but the image Fig.2(c) suggests that the technique adopted for unsupervised analysis of the hyperspectral image could be a way to determine the expected natural scene in the absence of groundtruth information.

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 portion of the spectrum of wavelength 400-2500nm with an window size of 10nm. We consider here a subscene of size 200×710 (Fig.2(a)) for our analysis with 122 bands having meaningful DN values.

The consideration of this subscene comes from the fact that the limited ground truth data collected was found to be scattered in this area that covers mainly two types of crops viz., wheat and sugarcane. The classes of this ground truth as collected from this site with the total number of ground truth pixels in parentheses are as follows:

- (i) Wheat grain filling(61)- this stage is before the ripening stage.
- (ii) Wheat maturity(26)- the beginning stage of matured wheat, i.e., the state of being mature.
- (iii) Sugarcane mature(22)- fully grown stage.
- (iv) Wheat milking(48)-this is the initial stage of grain formation- i.e., when the grain is squeezed a milky solution is apparent.
- (v) Sugarcane-ratoon maturity(14)- this is the second crop of sugarcane which is grown from parts left in the field and at the state of being mature.
- (vi) Wheat booting(15)- this stage starts with the development of head within sheath of the leaf and ends with the tip of the head emerging out of the sheath of the leaf before milking stage.
- (vii) Wheat emergence(14)-in this stage, seedlings start coming out of the soil.

At first PCA has been carried out with the whole set of image data, that is, with 200×710 pixels each having 122 bands. As in example-1, the number of significant PCA components, referred to as virtual dimensionality (VD)(Chang and Du, 2004) for hyperspectral imagery has been identified with Malinowski's method (Malinowski, 1997, Chang and Du, 2004). The estimated VD has been found to be 15 for this image data. To obtain one of the best soultion in top 150 ($= \tau$) with 500 samples in a search space of $|W| (= 200 \times 710)$ we need to carryout two iterations.

In the next step, as described in Section 2, PP has been performed with a random sample of 500 pixels taken from the 15 PCA component images. After carrying out PP with all the 15 component images derived from PCA the MRF model based segmentation has been carried out. The NDVI and REP values for different stages have been computed pixelwise as described in Section -3 and subsequently the average value of REP corresponding to each oversegmented crop region are maintained. Table-7 provides the REP values for different stages of wheat crop. In the segmentation process these are suitably incorporated as outlined in the end of section 4. For next iteration, PP-MRF segmentation scheme has been similarly run and the one which yields least energy in the segmentation stage between these two iterations gives us the desired solution. Subsequently with this segmented output the cluster validation has been carried out as described under the subsection *Supervised scheme* to have the classified image.

insert table 7

The original image and the classified image of the proposed methodology have been displayed in Fig. 3(a) and Fig. 3(b) respectively. Fig. 3(a) exhibits 200×710 original image displayed in false color composite (FCC) with three bands $R(803.3nm)$, $G(681.2nm)$ and $B(548.9nm)$. As vegetation has high reflectance in NIR, the crop classes appears reddish in FCC. Other than crop classes there are some barren agriculture areas, these areas appear gray to brown normally. However, with the presence of moisture in these barren area it obtains a bluish tinge.

Table-8 provides the classification accuracies of some conventional nonparametric techniques viz., SVM-linear, SVM-polynomial, SVM-RBF(Radial Basis function-one against one (OAO) multiclass strategy), K-nn classifier and the proposed methodology. These non-parametric approaches have been run with 15 PCA component images as well as with full data set (122 bands). The classification accu-

racies of these methods have been found to be very low in both cases. The values in Table-8 exhibits overall classification accuracy with 15 PCA component images. To obtain the best parameter values of SVM techniques, different experiments have been carried out by varying the values of gamma (γ) and penalty (C) with the training data set and the one corresponding to maximum accuracy have been found and reported in Table-8. Since the OAO multiclass strategy usually yields higher classification accuracy (Chen et al., 2009A) and has lesser computational complexity among all other multiclass strategies of SVM-RBF, the result for this strategy has been reported in Table-8. This table also exhibits Kappa coefficient of each technique along with execution time.

insert table 8

insert Figure 3(a), 3(b)

As seen in Fig 3(b) different stages of wheat and sugarcane are identified with different colors. Unclassified regions (classes) are shown in black. The two deep blue vertical lines featuring in Fig. 3(a) on the western part are visible as an unclassified area in Fig. 3(b) and so is the diagonal line like feature across the image in the central part of Fig. 3(a) (see similar rectangles in the two images in Fig.3(a) and Fig. 3(b)). Further we note that almost all the sky blue regions that correspond to barren land in the original image have been appropriately labeled as unclassified areas. For example one can compare the reverted *L* shaped blue region in the central part of the original image and the corresponding black region in Fig. 3(b), marked with rectangles in both figures.

Since the ground-truth samples for some classes is very small we examine the accuracy by setting aside 30% of the total ground-truth samples as test set for each of the classes. That is to say, with 70% of the groundtruth samples for each of the classes we perform the cluster validation scheme and test the overall classification accuracy with the remaining 30% of the groundtruth samples. Such a random selection of 30% of the total groundtruth samples has been carried out a number of times and accuracy has been determined at each of the such 30% test set. It has been observed that the overall accuracy lies in the range 84.43%- 89.71 % in these repetitions for the two sample sizes 500 and 300, with mean 87.07%. For comparing accuracy of the proposed methodology with that of other approaches we have used McNemar's test(Agresti, (1996), Bradley, (1968)). This is because the groundtruth sample size is very small in this case. For comparing the two classifiers corresponding to case(iii) and case(v) as given in table-8, the value of *z*-statistic has been found to be 2.8 without resampling and that of 24.34 with resampling using 9999 permutations. This shows that these are significantly different. As can be seen from Table-8, the accuracy of the proposed methodolgy is significantly higher than that of other (nonparametric) methods. It may be noted that for other methods, computation time is primarily proportional to the groundtruth samples and as such computation times for those methods are quite fast due to very small number of training and test data but the accuracy yielded have been found to be poor.

7. Conclusion

In this hyperspectral classification methodology, PP procedure identifies some collection of projections from randomly chosen samples based on the information divergence index. When coupled with MRF model based segmentation procedure it determines the "good enough projection" with high probability through ordinal optimization. The classification accuracies presented in Table-1 suggest that the proposed methodolgy (PP-MRF) yields significantly higher accuracy than that of

the most effective OAO(SVM-RBF) method tested for 9 class groundtruth data set vis-a-vis with lesser computation time with reasonable groundtruth sample size. As demonstrated in Table -6 with 11 class ground truth data set, the computation time for this method is far superior with comparable accuracy. In accordance with an ordinal optimization procedure, we have focused our attention to perform the classification task in a faster manner.

Fast computation is expected to make the methodology very useful for remote sensing practitioners. Computation time for the PP-MRF methodology with random samples in the range of 300 to 500 would be very less for moderate size imagery and is expected to yield a high classification accuracy as shown in example-1. If groundtruth sample size is very small, computation time for nonparametric methods may be less but the classification accuracy of the proposed methodology is expected to be much higher as is evident in example-2.

In our approach the selection of the number of significant(VD) PCA components has been carried out with Malinowski's method for faster computation. As this method is known to overestimate the VD for real image we have tried MRF segmentation with reduced number (< the number of PCA) of PP component images to circumvent the problem. Few PP component images having relatively low projection index values are dropped while carrying out MRF segmentation followed by classification. It is observed that the classification accuracy does not significantly increase but the computation time reduces. If some quantitative methods are developed for selecting the appropriate number of PP component images based on the corresponding projection index values, it will make the procedure further useful.

When we intend to have our best projection within top τ of all best projections, the order of computational complexity is $O(I \times (n\log n + W \times N \times m^2))$, where $I = O(\frac{W}{N \times \tau})$ is the number of iterations, W = size of search space, $n\log n$ (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.

The key features of the methodology are as follows:

(i) This method offers a pragmatic solution to trade off, from computational point of view, a comparably inexpensive method for obtaining one among top τ best solutions, against the resource intensive approaches for computation of the absolute best solution.

(ii) The proposed methodology has a provision to include ancillary information such as NDVI, REP values etc when crop classes are of detailed level along with image space and is expected to produce high classification accuracy. With Example-2 we have demonstrated how to integrate ancillary information in the segmentation process when landcover comprises detailed level classes to obtain improved classification accuracy in spite of the pitfalls for limited number of groundtruth samples.

(iii) When concurrent groundtruth information is not available this work shows a way for meaningful qualitative comparison with the natural scene.

As there are sufficient groundtruth samples for Example-1, the results of this example corroborate the fact that high classification accuracy has been achieved in a faster manner which has been the objective of this PP-MRF based classification methodology. Example-2 has been illustrated to demonstrate how the spatial and spectral information are used together within classification -although this example has a very limited groundtruth data. For projection, we have chosen the entropy index because of its cluster detecting ability. When projection has been carried out with other types of index as stated in section-1, classification accuracy has been achieved to a maximum of 93% only for example-1.

References

- AGRESTI,A. 1996, *An introduction to categorical Data Analysis*, Wiley New York, N.Y.
- BOLTON, J. AND GADER, P. 2009, "Random set framewrk for context based cl;assification with hyperspectral Imagery", *IEEE Transactions on Geoscience and Remote Sensing*, **47(9)**, pp.3810-3821.
- BRADLEY, J. V.,1968, *Distribution-free staistical tests*, *Printice Hall*, Englewood Cliffs, New Jersey.
- AVIRIS NW Indiana's Indian Pines 1992 data set[Incline]. Available: ftp: ftp.ecn.purdue.edu biehl Mutispec 92AV3C (original files) and ftp: ftp.ecn.purdue.edu biehl PC MultiSpec ThyFiles.zip (ground truth).
- BENEDIKTSSON, J. A., SVEINSSON, J. R. AND ARNASON, K. 1995, Classification and Feature extraction of AVIRIS data. *IEEE Transactions on Geoscience and Remote Sensing*, **33(5)**, pp.1194-1205.
- CAMPS-VALLS,G., MARSHEVA,T. V. B., AND ZHOU,D., 2007, "Semi-Supervised Graph-Based Hyperspectral Image Classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol.45, no.10, pp. 3044-3054.
- CHANG, C-I., AND DU, Q,1999, "Interference and noise-adjusted Principal components analysis" *IEEE Transactions on Geoscience and Remote Sensing*, vol.37, Sept(5), part-2, pp.-2387-23960.
- ., 2004, "Estimation of Number of Spectrally Distinct Signal Sources in Hyperspectral Imagery" *IEEE Transactions on Geoscience and Remote Sensing*, vol.42, no.3, pp. 608-619.
- CHEN, J., WANG, C. AND WANG, R., 2009, Adaptive Binary Tree for fast SVM Multiclass classification, *Neurocomputing*, vol.72, pp. 3370-3375
- ., 2009,"Using stacked generalization to combine SVMs in magnitude and shape Feature spaces for classification of hyperspectral data" *IEEE Transactions on Geoscience and Remote Sensing*, vol.47, no.7, pp. 2193-2205.
- CHIANG, S. S. AND CHANG, C.I., 2001 "Unsupervised target detection in hyperspectral Images using projection pursuit," *IEEE Transactions on Geoscience and Remote Sensing*, vol.39, no.7, pp. 1380-1392.
- CRAMMER,K. AND SINGER, Y., 2001." On the Algorithmic Implementation of Multi-class SVMs", *Journal of Machine Learning Research*, vol. 2, pp. 265-292.
- DUBES, R. C. AND JAIN, A. K., 1989, " Random field models in image analysis ", *Journal of Applied Statistics*,vol. 16, pp 131-164.
- DUBES, R.C., JAIN,A.K., NADABAR,S.G. AND CHEN, C.C., 1990, "MRF model based algorithms for image segmentation", in *Proceedings 10th International Conference on Pattern Recognition*, pp. 808-814.
- FAUVEL, M., BENEDIKTSSON, J. A., CHANUSST, J. AND SVEINSSON, J R., 2008, " Spectral and spatial Classification of hyperspectral data using SVMs and Morphological profiles",*IEEE Transactions on Geoscience and Remote Sensing*, vol.46, no.11, pp. 3804-3814.
- DUNDAR, M.M AND LANDGREBE, D. A.,2004, "A cost-effective Semispervised Classifier Approach with Kernels" *IEEE Transactions on Geoscience and Remote Sensing*, vol.42, no.7, pp. 264-270.
- FRIEDMAN, J. H. AND TUKEY, J. W., 1974, "A projection pursuit Algorithm for exploratory projection pursuit," *IEEE Transaction Computers*, vol. C-23, pp. 881-889.
- FRIEDMAN,J. H., 1987 "Exploratory projection pursuit" *J. Am. Stat. Ass*, 82(397), pp. 249-266.
- FUKUNAGA, K. 1990 *Introduction to Statistical Pattern Recognition* , 2nd ed., New york, Academic.
- FOODY, G.M, 2004,"Thematic Map Comparison: Evaluating the Statistical Significance of Difference in classification accuracy" *Photogrammetric Engineering and Remote Sensing*, Vol.70., No.5, 2004, pp.627-633.
- HO, Y.C., 1999, "The No-free lunch theorem and the human-machine interface," *IEEE Control System Magazine*, pp. 8-11.
- HOFFBECK, J.P AND LANDGREBE, D. A., 1999, "Covariance Matrix Estimation and

- Classification with limited Traing data" *IEEE Transaction Pattern Ana Mach Intel.* vol.18, no. 7, pp.763-767.
- HUBER, P. J., 1985, "Projection pursuit", *Annals of Statistics*, vol.13, no.2, pp. 435-475.
- HUGES, G.F., 1968, "On the mean accuracy of statistical pattern recognizers", *IEEE Transaction Information Theory*, vol.IT-14, no.1, pp. 55-63.
- HYPERION SENSOR, 2000, <http://eo1.usgs.gov/sensors.php>
- IFARRAEGUERRI, A. AND CHANG, C., 2000, "Unsupervised Hyperspectral Image Analysis with Projection pursuit", *IEEE Transactions on Geoscience and Remote Sensing*, vol.38, no. 6, pp. 2529-2538.
- JACKSON, Q. LANDGREBE, D. A., 2002, "Adaptive Bayesian Contextual Classification Based on Markov Random Fields," *IEEE Transactions on Geoscience and Remote Sensing*, vol.40, no. 11, pp. 2454-246..
- JONES, M.C. AND SIBSON, R., 1987, "What is projection pursuit?", *Journal of Royal Staistical Society A*, vol.150, no.1, pp.1-36.
- JIMENEZ, L. O. AND LANDGREBE, D. A., 1999, "Hyperspectral data analysis and supervised feature reduction via projection pursuit", *IEEE Transactions on Geoscience and Remote Sensing*, vol.37, no. 6, pp 2653-2667.
- JINXIU,L., JUNLI, L. AND PING,W., 2007, "Pseudo colour coding of Medical Images Based on Gradient", *First International Conference on Bioinformatics and Biomedical Engineering*, ICBBE, pp. 932-935.
- KARTIKEYAN, B. AND SARKAR, A., 1989," A unified approach for image segmentation using exact statistic," *Computer Vision Graphics and Image Processing*, vol.48, pp.217-229.
- KECMAN, V., 2001, *Learning Soft computing Support vector Machines, Neural Networks, and Fuzzy logic Models*, MIT press, Cambridge.
- LI, J., BIOUCAS-DIAS,J., AND PLAZA, A., 2010, "Semi-supervised Hyperspectral Image Segmentation using multinomial logistics regression with active learning", *IEEE Transactions on Geoscience and Remote Sensing*, vol.48, pp. 4085-4098.
- LI, S. Z., 1995, *Markov Random Field Modeling in computer vision*, New York:Springer-Verlag.
- MALINOWSKI, E. R., 1997,"Determination of the number of factors and experimental error in a data matrix," *Analytical Chemistry*, vol.49, pp.612-617.
- MCKENZI, D.P., MACKINON,A.J., PELADEAU,N., ONGHENNA,P., BRUCE,P.C., CLARKE, D.M., HAARIGAN, S., AND McGORRY,P.D., 1996," Comparing correlated kappas by resampling: Is one level of aggrement significantly different from another?", *Journalof Pschiatric Research*, vol.30,pp.483-492.
- MELGANI, F. AND BRUZZONE, L., 2004,"Classification of Hyperspectral RemoteSensing Images With support Vector Machines" *IEEE Transactions on Geoscience and Remote Sensing*, vol.42, pp. 1778-1790.
- MOSES, A.C. AND SKIDMORE, A. K., 2006, "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.
- NASON, G. P., 1995, "Three dimension projection persuit", *Applied Statistics*, vol.44, No. 4, pp. 441-430.
- PLAZA, A., BENEDIKTSSON,J. A., BOARDMAN, J. W., BRAZILE, J., BRUZZONE, L., CAMPS-VALLS,G., CHANUSSOT,J., FAUVEL,M., GAMBA,P, GUALTIERI, A., 2009, "Recent advances in techniques for hyperspectral image processing", *Remote Sensing of Environment*, vol.113, pp. S110-S122.
- RENYI, A., 1970, *Probability theory*, Amsterdam, North -Holland.
- SARKAR,A., BISWAS,M.K., KARTIKEYAN,B., KUMAR,V., MAJUNDAR,K.L. AND PAL,D.K., 2002"A MRF Model Based Segmentation Approach to Classification for Multispectral Imagery", *IEEE Transactions on Geoscience and Remote Sensing*, vol.40, pp. 1102-1113.
- SHASHAHNI, B.M., LANDGREBE, D. A., 1994," The effect of Unlabeled samples in reducing the small samples size problem and mitigating the Hughes phenomenon", *IEEE Transactions on Geoscience and Remote Sensing*, vol.32, pp. 1087-1095.
- SMITH,K.L., STEVEN, M.D. AND COLLS, J.J., 2004,"Use of Hyperspectral Derivative

- Ratios in Red Edge Region to Identify Plant Stress Responses to Gas Leak., *Remote Sensing of Environment*, Vol. 92, Issue 2, pp. 207-217.
- TADJUDIN, S., LANDGREBE, D. A., 1999,"Covariance estimation with limited samples" *IEEE Transactions on Geoscience and Remote Sensing*, vol.37, pp. 2113-2118.
- VAN DER MEER,F. D., DE JONG, S. M., AND BAKKER, W., " Analytical techniques in spectrometry", *Imaging Spectrometry: Basic principles and prospective applications*, Kulwer Aca. Pub., Amsterdem.
- VAPNIK,V. 1998 "Statistical Learning theory", Wiley, New York.
- VELASCO-FORERO,S. AND MANIAN, V.,"Improving hyperspectral image classification using spatial preprocessing", *IEEE Transaction on Geoscience and Remote Sensing Letters*, vol.6, No. 2, pp. 297-301.
- WENG,J., ZHANG, Y. AND HWANG,W.S., 2003, "Candid Covariance-Free Incremental Principal Component Analysis",*IEEE Transaction Pattern Analysis and Machine Intelligence*, vol.25, No. 8, pp. 1034-1041.

Table 1. Comparison of PP-MRF with SVM-RBF(Intel Pentium D 2.80GHz CPU)

cases	Approach	OA[%]	τ	No. of channels	Time[s]
(i)	SVM-RBF OAO	93.96	-	220 channels	455
(ii)	SVM-RBF OAO	73.36	-	11 PCA	210
(iii)	PP-MRF(300)	95.28*	25	11, 11	138
(iv)	PP-MRF(400)	95.23*	20	11, 11	210
(v)	PP-MRF(500)	95.01*	20	11, 11	355

Table 2. Confusion matrix for case(viii) of PP-MRF

. To Class No from class	1	2	3	4	5	6	7	8	9	total samples	Accu
	-	-	-	-	-	-	-	-	-	-	-
1	652	15	0	3	0	7	15	0	0	692	0.942
2	7	385	0	0	0	0	0	0	0	392	0.982
3	0	8	219	6	0	0	0	0	4	237	0.924
4	1	5	2	341	0	0	6	1	2	358	0.952
5	0	0	0	1	247	0	0	0	0	248	0.996
6	30	5	0	1	0	432	0	13	0	481	0.895
7	32	7	0	0	0	19	1152	2	10	1222	0.942
8	4	12	1	0	0	0	0	292	0	309	0.945
9	1	0	0	0	0	0	6	3	638	648	0.985

Table 3. Confusion matrix for case(i) of SVM-RBF

. To Class No from class	1	2	3	4	5	6	7	8	9	total samples	Accu
	-	-	-	-	-	-	-	-	-	-	-
1	625	23	2	1	0	9	32	0	0	692	0.903
2	4	351	7	0	0	6	16	8	0	392	0.895
3	2	6	224	2	0	0	3	0	0	237	0.945
4	0	0	1	357	0	0	0	0	0	358	0.997
5	0	0	0	0	248	0	0	0	0	248	1.000
6	25	1	5	0	0	425	21	4	0	481	0.884
7	31	9	5	1	0	29	1143	4	0	1222	0.935
8	0	3	9	0	0	0	4	293	0	309	0.948
9	1	0	0	1	0	2	0	0	644	648	0.994

Table 4. 2×2 Confusion matrix derived from Table-1 and Table-2

		SVM-RBF			
		Correct	Incorrect	Total	
PP-MRF	Correct	4281	77	4358	
	Incorrect	29	200	229	
		4310	277	4587	

Table 5. Comparison of Class by Class Accuracy of OAO multiclass strategy of SVM-RBF and the best case (as per sample size) for the proposed methodology

Methods	Sample size	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6	ω_7	ω_8	ω_9	Overall Accu
OAO	-	90.3	89.5	94.5	99.7	100.0	88.4	93.5	94.8	99.4	93.96
PP-MRF	300	90.5	96.9	98.3	97.5	100.0	92.5	93.0	94.8	99.2	95.87
PP-MRF	400	94.8	97.4	97.9	95.0	100.0	89.4	97.1	95.1	99.4	96.24
PP-MRF	500	94.2	98.2	92.4	95.2	99.6	89.5	94.2	94.5	98.5	95.01

Table 6. User's and Producer's accuracy along with overall accuracy and computational time achieved by different Methods

Classes Methods	O.A.	Time		ω_1	ω_2	ω_3	ω_4	ω_5	ω_6	ω_7	ω_8	ω_9	ω_{10}	ω_{11}
PP-MRF	94.73	355	User's	96.69	92.05	90.75	92.62	90.32	92.99	86.76	95.80	94.10	99.51	100
			Producer's	99.15	87.34	89.85	94.16	99.45	100	95.27	92.69	92.16	97.78	98.03
SSVM-V-M	93.36	645.3	User's	78.29	92.42	92.31	96.37	98.94	99.19	92.26	91.05	88.30	98.13	99.53
SSVM-V-MS	94.43	1725.1	Producer's	86.32	87.77	90.38	95.98	99.47	99.59	93.16	92.38	88.84	99.06	99.07
SSVM-BD	93.81	1529.4	User's	84.68	94.66	93.20	96.77	99.20	100	93.61	91.90	90.16	98.11	99.69
			Producer's	89.74	89.21	91.77	96.39	99.47	99.59	95.44	92.87	92.77	98.11	99.07
			User's	77.44	93.11	92.85	96.40	99.20	99.59	92.63	91.83	89.41	97.22	99.53
			Producer's	88.03	87.53	90.52	96.79	99.20	99.59	94.14	92.95	90.70	99.06	99.07

Table 7. Values of REP for wheat stages.

emergence	booting	milking	maturity	grain filling
693.63	700.62	701.57	706.52	699.7

Table 8. Comparison of OA and Time Duration of PP-MRF with other approaches in Intel Pentium D 3.00GHz CPU.

Case	Approach	Parameters	OA[%]	Time[s]
(i)	SVM-linear	C=100	76.47	30
(ii)	SVM-polynomial	degree=4,C=100, $\gamma=1$	70.58	20
(iii)	SVM-RBF- OAO	C=5000 $\gamma=0.01$	73.2	20
(iv)	K-nn classifier	K=7	70.0	30
(v)	PP-MRF(15, 15, 500)	$\alpha=.01$	89.71*	700
(vii)	PP-MRF(15,15, 300)	$\alpha=.01$	84.43*	430

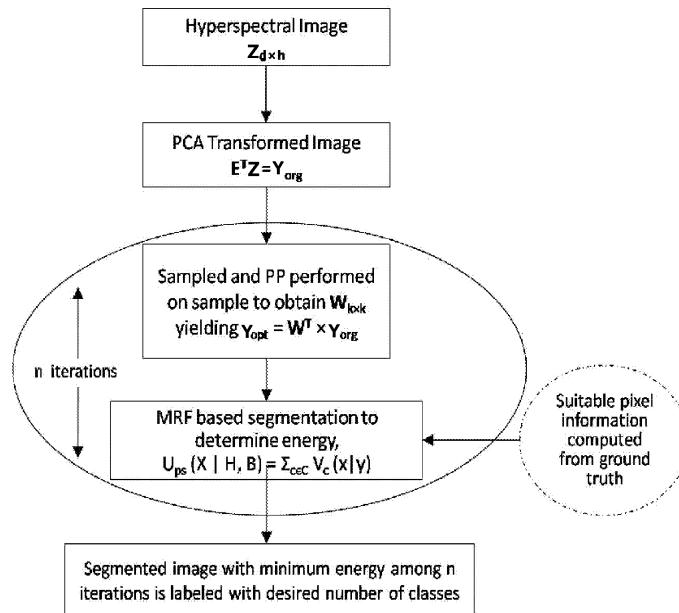


Fig 1: Iteration based PP-MRF Segmentation Approach to Classification

Figure 1. Iteration based PP-MRF Segmentation Approach to Classification.



Fig 2(a): Original Image

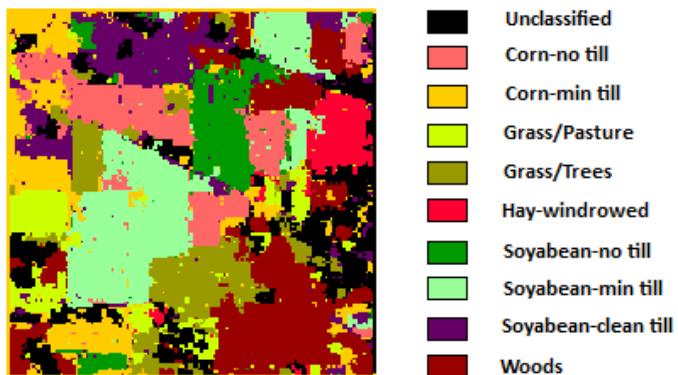


Fig 2(b): Classified Image

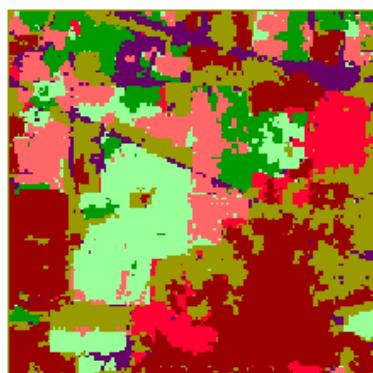


Fig 2(c): Tuned Image

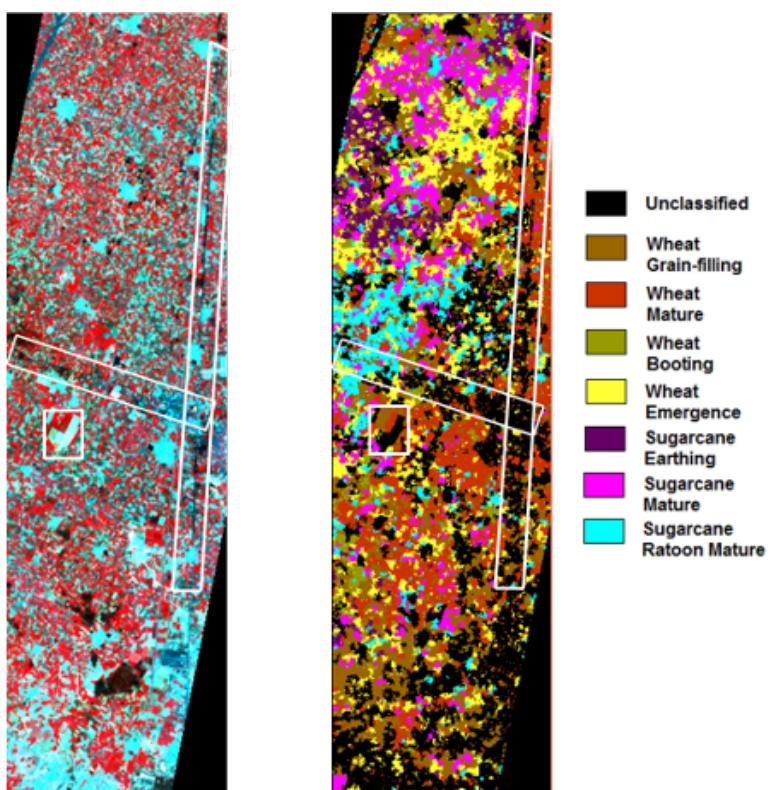


Fig3(a): Original Image

Fig 3(b): Classified Image

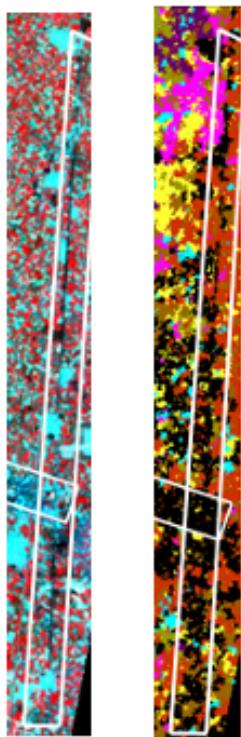


Fig 3(c): Zoom area of
original image(left) and
classified image(right)