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

A Sarkar ⋆, A Vulimiri ⋆⋆, S Bose ⋆, S Paul • 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

Abstract- *This work deals with hyperspectral image analysis in the absence of ground-truth. 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. The segmented output so obtained is labeled with desired number of landcover classes in the absence of ground-truth. While comparing with original hyperspectral image the methodology outperforms principal component analysis with respect to class separation as exhibited in the illustration of an archive EO-1 hyperspectral image. The technique is not a computational intensive as is usually the case in hyperspectral image analysis. When training samples are available, the segmented regions yields a classified image with any cluster validation technique viz.[12].*

Index Terms: Projection pursuit, Entropy Index, Markov Random Field and Ordinal optimization.

1 Introduction.

The Hyperion Sensor aboard the EO-1 spacecraft provides high quality data with 30 meter spectral resolution over more than two hundred channels for analyzing complex variability of landscape surface cover on earth. Investigation have been reported for matching image pixel spectra available from the hyperspectral data and the standard library comprising reflectance characteristics for large class of materials with some success [13]. 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 Huges Phenomenon [5] 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 projections 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 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 "projection persuit"(PP) technique, a term introduced by Freidman and Tukey [2] that seeks out a linear projection of the multivariate data onto a lower dimensional space by means of optimization of the index of "interestingness", defined as 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 projection that maximizes the signal sources.

PP as defined above, is the process of making such selections (signals of interest) by local optimisation over projection directions of some index of "interestingness". The index of interestingness has been 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 index available in literature are due to Freidmen and Tukey [2], Huber [4], Jones and Sibson [7], Chiang and Chang [1], Jimenez and Landgrebe [8], and Ifarraeguerri and Chang [6] among others. The Projection indices the above authors have employed are of three types, viz.
1) class distance measure (e.g., Bhattacharya distance [8],

The Friedman-Tukey Index [2]),
2) entropy index or information divergence index (Ifarraeguerri and Chang [6])
and 3) moment index (Jones and Sibson [7] and Chiang and Chang [1]).

The concept of using entropy index emerges from the statistical behaviour for very high dimensional data sets in lower dimensions due to Renyi [11] and Huber [4]. Projection is a convolution, which due to CLT 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 [6] is given by

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

or,

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

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 [6] have reduced the dimension from 210 to a meagre 11 for a 256×256 HYDICE sensor scene. The component images resulted from PCA and PP as exhibited in [6] show some potential of the methodology. But the computation involved in such a methodology is highly intensive although there is no mention of it in [6]. It appears that computation time required with such a methodology for a whole hyperspectral scene which is usually of 256×3128 with 242 channels would be unrealistically large.

In the present work we adopt the Ifarraeguerri and Chang's [6] information divergence index for PP to reduce the dimension "almost as good as optimally" with a technique that leads to faster computaional 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 whole data set, we replace "best for sure" with "good enough with high probability"as per ordinal optimization suggested by Ho [3]. This "softening of the goal" eases computaional burden in our problem and it is much easier to find best projection within top 100 than to get best projection so that some meaningful solution can be obtained in a resonable time. To do so we reduce the search space from $|W| = 8 \times 10^5 (= 256 \times 3128)$ to 1000 random samples ($= N$, say) from the whole data set. Then, the probability that at least one sample will yield one of the best projection on the top $100 (= n)$ is approximately 0.1175 which is computed from the expression $1 - (1 - \frac{n}{|W|})^N$ [3]. This implies that we need $\frac{1}{0.1175} \simeq 8$ iterations of the procedure (on the average) to gurantee that our projection is among the 100 best projection. 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 MRF segmentation scheme.

The MRF procedure we adopted here yields a segmented output which subsequently can be clustered into a specified number of classes in the absence of groundtruth. The output so obtained only reveals different unidentified classes. When groundtruth samples are provided it would generate complete classified output by some cluster validation scheme. The originality of the paper lies in underlining how PP can be exploited in a reasonable time frame with MRF based segmentation and ordinal optimisation for analysing hyperspectral images. A schematic representation of the methodology is exhibited in Fig. 1.

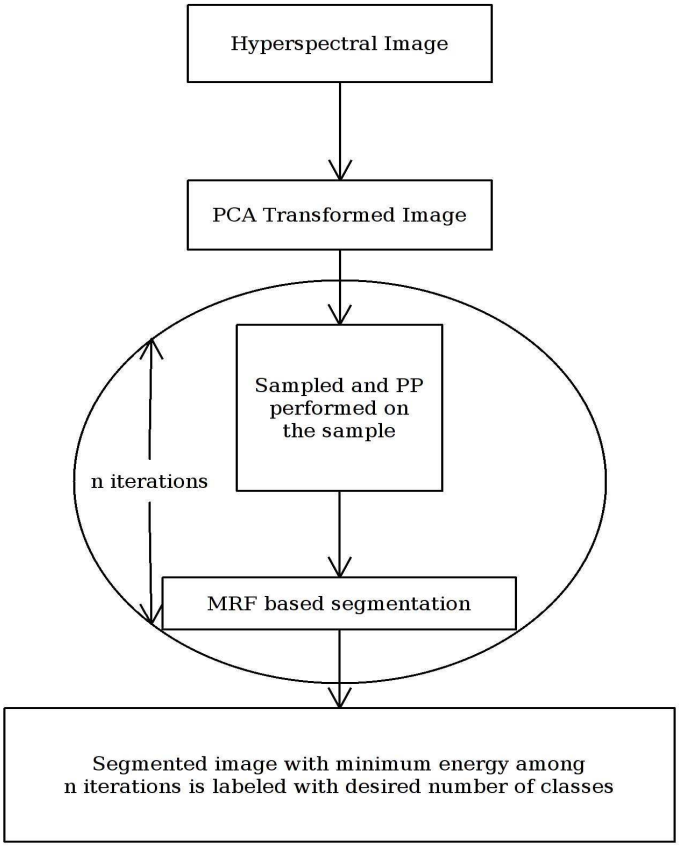

Fig 1: Workflow

2 Projection Pursuit:

Projection pursuit procedure that we shall follow has been proposed by Ifarraguerri and Chang [6] but for the

completeness of the work, the methodology is briefly described here. 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 with the highest value of the projection index 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.

A random sample from the hyperspectral image data consisting of N pixels, over d bands is arranged in a matrix $\mathbf{X}_{d\times N}$. As an initial reduction of dimension, PCA is performed using the well known "covariance free method" [14] and 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

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

a copy of which is saved as $\mathbf{Z}_{orig}$.

As described earlier, this procedure assumes that some pixel spectrum in $\mathbf{Z}$ gives a fair approximation to the optimal projection and moreover, optimality is regarded here as least normality. From the construction of $\mathbf{Z}$, l^{th} column of $\mathbf{Z}$, viz., $\mathbf{z}_l$ is the spectrum of the l^{th} pixel in the image. So we project the whole data set along $\mathbf{z}_l$ and take the calculated projection score vector $\mathbf{p} = \mathbf{Z}^T\mathbf{z}_l$, the index i of p_i 's , (the components of **p**) running from $\frac{-N}{2}$ to $\frac{N}{2}$. For comparison with the normal distribution we quantize the normal distribution as $\mathbf{q} = (q_i)_{N\times 1}$, $q_i = \int f dx$, where f is the normal density and integration is taken from $(i-1)\times h$ to $i\times h$, i running from $\frac{-N}{2}$ to $\frac{N}{2}$ and h is the bin size (class interval), in the range (maximum pixel value-minimum pixel value). 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 colums $\mathbf{w_r}$ 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 same reasoning, is certain to produce one of the best solutions. The best solution (projection) among all these iterations is identified in MRF based segmentation scheme described below, the criteria being minimum segmentation energy.

3 MRF model based segmentation scheme

We follow the scheme of Sarkar et al. [12] in defining the MRF on a region adjacency graph(RAG) of initial oversegmented regions- the details are omitted here. Our discussion here is directed in formulating the energy function in MRF based segmentation approach. Minimizing this energy function will result in a 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 component channels. This implies that better the revealation of the hidden clustering characteristics of the high dimensional point cloud in lower dimension, more uniform the segmented region.

(ii) Two distinct adjacent regions R_i and R_j should be as dissimilar as possible with respect to the measured characteristic as evident from the selected component channels.

The multi-component channel 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 $R = \{R_i; 1 \le i \le Q\}$ is a set of nodes and E is a set of edges connecting them. With appropriate neighborhood system a MRF is defined (see details in [12]). The posterior probability distribution is given by

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

where $Z_{ps} = \sum_{\mathbf{X}} e^{-U_{ps}(\mathbf{x}|\mathbf{y})}$. The events $\{\mathbf{X} = \mathbf{x}\}$ and $\{\mathbf{Y} = \mathbf{y}\}$ represent respectively a specfic labelling configuration and a specific realization. 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 potential to achieve the desired objective.

Incorporating the above two constraints we define the clique potential (energy) function [12] 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 $\mathbf{W}_{ij} = \frac{1}{\nu_{ij}}[\sum_{k=1}^{n_i}(\mathbf{Y}_{ik} - \mathbf{M}_i) * (\mathbf{Y}_{ik} - \mathbf{M}_i)^{'} + \sum_{k=1}^{n_j}(\mathbf{Y}_{jk} - \mathbf{M}_j) * (\mathbf{Y}_{jk} - \mathbf{M}_j)^{'}]$.

$\mathbf{M_i}$ 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} is a binary variable taking values 0 and 1. ν_{ij} takes the value 1 when the regions (in clique potential) are homogeneous with respect to the multi-component channel pixel intensity values. If not then η_{ij} takes the value 0. It may be noted that η_{ij} = 1 indicates that $x_i = x_j$.

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. 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 $\mathbf{s}_{pooled} = \frac{\mathbf{S}_i + \mathbf{S}_j}{\nu_{ij}} = \mathbf{W}_{ij}$, the comparative criterion needed here is based on Hotelling's T^2 statistics.

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 $P[T^2 > F_\alpha] = \alpha$ [as in [12, p. 1106]].

The segmented image so obtained is 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 more the revealtion of the hidden clustering characterstics of the high dimensional data set by PP in lower dimension the better is the segmented output with lesser and lesser energy. Corresponding to every thousand random samples chosen from the hyerpsectral 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 iteration. For n number of such iteraion (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. With eight($= n$) iterations only, each iteration takes a small duration of time, we ensure that the segmented output (with minimum energy) is among the best 100 cases.

In the absence of any supervised knowledge (training samples), 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 and identified as forest in accordance with some supervised knowledge, if available. In the same way, crop areas and fallows etc also comprise fragmented regions as per similar pixel intensities which may be combined. This implies that our MRF model based segmented output can be tuned into a desirable number of classes in the absence of any ground truth knowledge. This is a kind of unsupervised crude cluster validation technique. However, the above tuned output that corresponds to the desired number of classes 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 of an unsupervised classifier and a supervised classifier.

The simple technique that we follow for such a tuning comprises two steps with inputs as the segmented image and one component image and is carried out as follows

1) Determine maximum and minimum value of region means in segmented (region) output. Let those values be M_{max} and M_{min}.

2) For each region mean M_j in the region output we fill the region following the condition if

$$(M_j - M_{min}) \leq \frac{i \times (M_{max} - M_{min})}{N_c} \quad (12)$$

where N_c is the desire number of classes,
then $G = 255 - \frac{(i-1)\times 255}{N_c - 1}$ where G is the new gray value for entire segment (region) and i iterates through 1 to N_c.

The tuned segmented output thus exhibits the different regions in as many gray shades as the desire number of classes. In order to make the regions (classes) of the images more vivid and distinct we subsequently use the technique of Jinxiu et al [9] from gray to pseudocolor.

4 Experimental Results

The data analysed in this paper is an archive image data acquired from NASA, EO-1 satellite (Hyperion sensor) 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 West Bengal, India. The area was imaged on 17 January 2003. The hyperspecral image 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 and corrupted with noise have been removed and finally 143 bands have been used as an input to our methodology for unsupervised labelling. The bands used are ([12-52, 77-120, 143-164 ,189-224])

Since the pass date was 17 January, the crops that are usually cultivated in the aforementioned area during this period are potato (three varieties), mustard (scanty), vegetables (cauliflower, brinjal, cabbage, bittergourd etc). Besides, the landcover of the area comprises meadows, paddy stubbles, isolated areas that are being ploughed for next crop, small pockets of homestead, shrubs, small rivers with shallow water and deep water in parts, sand in the river banks, cluster of trees of different species along the river banks and isolated laterite areas among others. The number of classes would be around thirty odd or so as exact groundtruth was not available.

At 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 2, 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 "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 homogenity (uniformity) measure (value of the energy function associated with the underlying MRF).

The one which yields least energy (implying the best partitioning) in the segmentation stage gives us the desired case. After carrying out PP on all 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 the absence of groundtruth samples, as in the case of our present study, tuning of the segmented output is carried out with desired number of landcover classes and is exhibited with a pseudo colour code [9]. Fig. 2(a) and 2(b) respectively exhibit the original image and the segmented output based on PCA component images labelled with 30 classes. In Fig. 3(a) and 3(b), we show two segmented outputs based on PP projected component images with minimum and maximum energy respectively for the 8 iterations and are labelled with as many classes as of Fig. 2(b). Although the pseudo colors of the regions appears to be similar in Fig. 2(b), 3(a) and 3(b) but there are 30 different shades of color. We note that the course of the river in the top part is very distinct only in Fig. 3(a). Likewise the course of the three other rivers are also distinctly visible in Fig. 3(a) as compared to Fig. 2(b) and Fig. 3(b). This fact entails that the complex variability of the landscape cover surface is expected to segregate appropriately in Fig. 3(a). In order to examine the above, a similar subset of each of the Fig. 2(b), 3(a) and 3(b) are compared respectively in Fig. 4(a), 4(b) and 4(c). In Fig. 4(a) we have eight different regions while in Fig 4(b) and Fig 4(c) there are nine although the study area covered are the same in the three figures. This fact only explains, as said above that the separation of classes is best determined in Fig. 4(b) and Fig. 4(c). Further, the segments displayed are most vivid in Fig 4(b) as compared to that of Fig 4(a) and Fig 4(c). In particular, the different crop regions, the canal and the bridge over the river are best identified in Fig 4(b). This figure (Fig 4(b)) is the segmented output that yields the minimum energy for the number of iterations considered and hence one of the best projection onto the lower dimension and thus brings out the potential of the methodology. When labelled with desired number of classes this labelled image may be compared with any supervised classification technique. After PCA is performed, each iteration comprising PP and MRF segmentation takes 7 minutes 22 seconds on a Intel Pentium D 3.00GHz CPU.

5 Conclusion

PP procedure with information divergence index identifies some collection of projection 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 be a potential methodology 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 suggests that the complex variability of the landscape surface cover over earth has been potentially analysed with the proposed methodology. When detail groundtruth is available a cluster validation technique may be adopted to determine classification accuracy of the complex variablity of landscape surface cover on earth. In cluster validation stage the same groundtruth sample size as is usually gathered for multispectral imagery would work, thus reducing the "curse of dimensionality" in this aspect. The methodology is not computationally intensive.

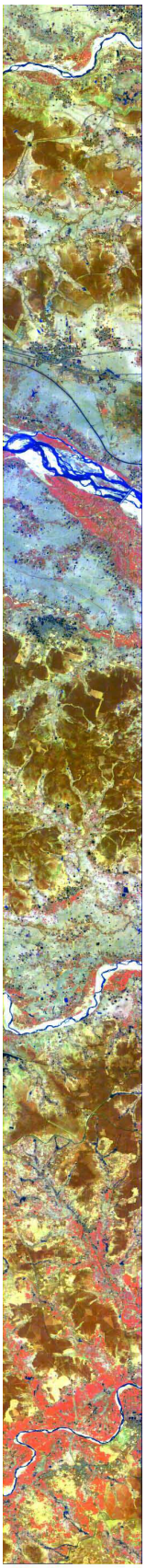

Fig 2(a): Original Image

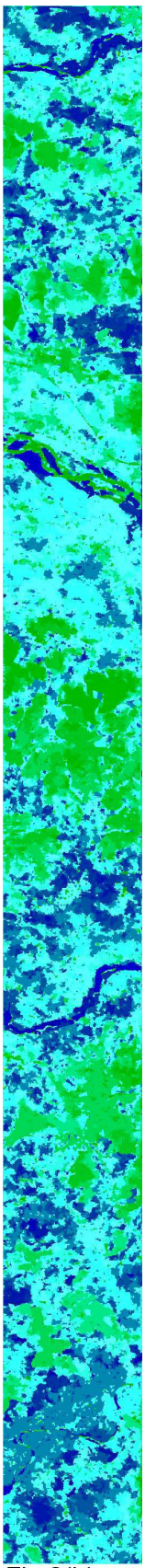

Fig 2(b): PCA Segmented Image

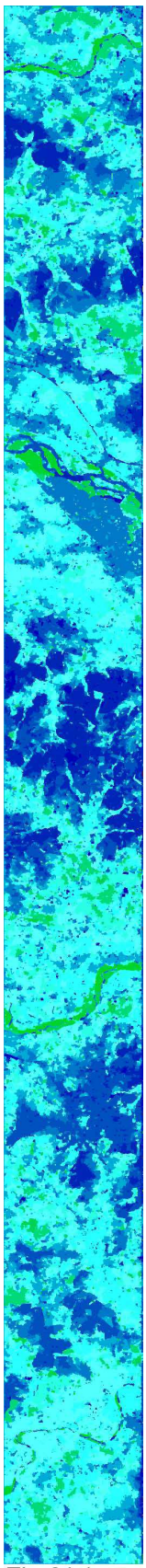

Fig 3(a): Segmented Image - Min Energy

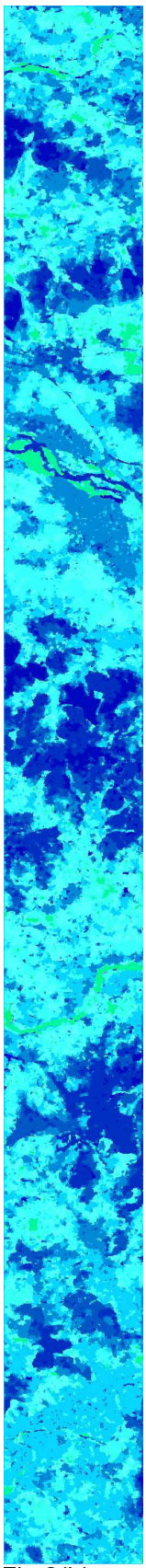

Fig 3(b): Segmented Image - Max Energy

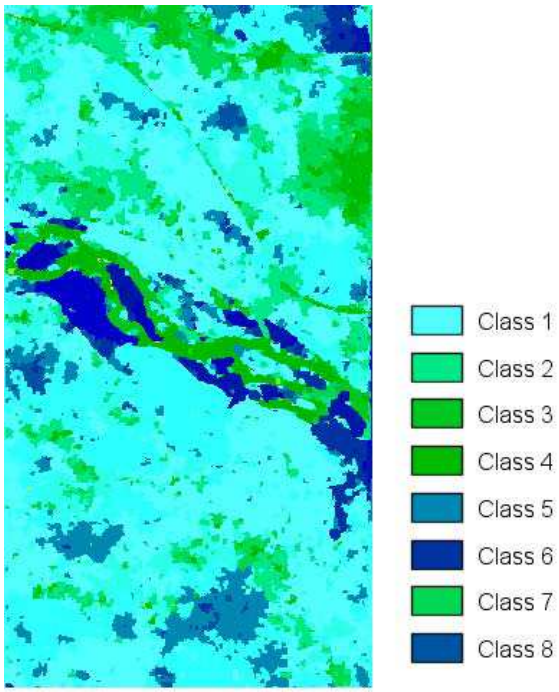

Fig 4(a): Subscene of fig 2(b)

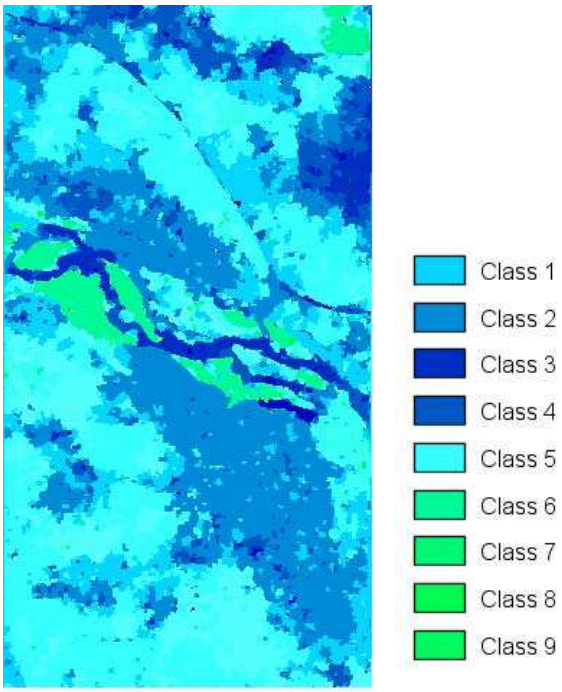

Fig 4(c): Subscene of fig 3(b)

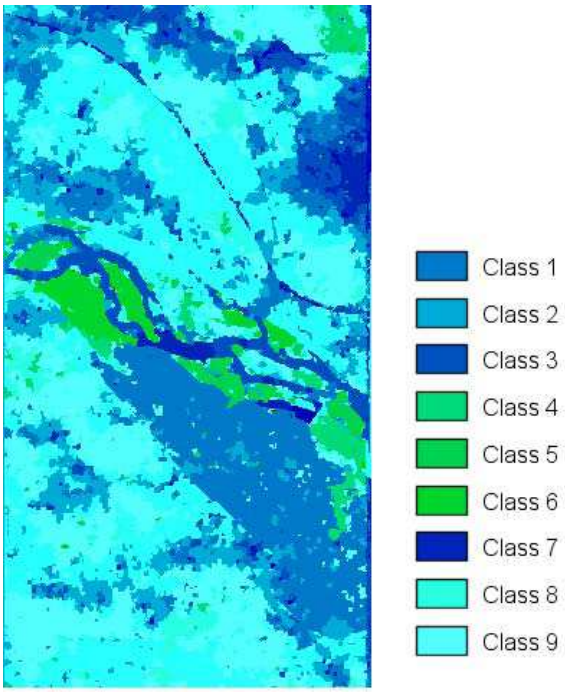

Fig 4(b): Subscene of fig 3(a)

Acknowledgement 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] 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.

[2] 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.

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

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

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

[6] 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.

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

[8] 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.

[9] 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.

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

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

[12] 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.

[13] 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.

[14] 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.