Chapter 5

A subspace regression, two phase label optimization, and efficient edge preservation scheme(EPS), for relaxation labelling

This chapter studies two relaxation labelling based methods, namely:(1) two-phase label optimization and (2) edge preservation-based scheme for improvement in accuracy of predicted labels. The two-phase label optimization method has described in proposed method-1 and Experimental Result Analysis-1. The EPS method has illustrated in the proposed method-2 and Experimental Result Analysis-2. Since the relaxation methods have applied in the prediction probability of pixel labels, we have used subspace regression methods to obtain the probabilistic prediction of pixel labels before applying the relaxation labelling.

5.1 Introduction

This work introduces a two-step algorithm which deals with spectral mixing issue [93] as well as performs the empirical study of continuous labelling in high dimensional(HD)

images [81] for over-segmentation within class labels. Therefore, we have proposed a two-step framework that first computed the classification accuracy and then improved the accuracy by optimizing a clustering function. In step-1, we have applied a subspace regression [137] followed by an alpha expansion method to obtain the classified HD image. For a detailed explanation, the readers may refer to [137]. This method better classifies the HD-HSI(High-Dimensional Hyperspectral image) image by removing the spectral mixing problem, which is a well-known problem in the HD image domain. In previous literature, this method has also denoted as smlrsub and smlrsubmll, which we have represented as subspaceMLR and subspaceMLRGC in our work for more clarity. The clustered image of step-1 is used for further investigation to improve the classified result by optimizing a variational image function using cluster energy minimization. In step-2, we have used a variational image function optimization-based approach to improving the classified image clusters, obtained from step-1, by investigating some interesting over-segmented contours. In this context, we refer the readers to [105], for a state-of-art image objective function, and it's a label optimization technique that we have used in the section-3.2, for a complete understanding of our approach. In various past works such as [81, 87, 104], the variational image function acts as an objective function that contains the description of the image regions and their boundaries. These variational function can be of two types:1- continuous and 2- discrete function. The continuous image function resulted in minimal clusters by using cluster merging [80]. However, they suffer from local minimum by gradient descent, and they are sensitive to initialization of clusters [110, 80]. The discrete image function, such as graph cut segmentation methods, has been used recently due to their fast convergence and obtaining a global minimum in an optimization process [80]. But the graph-based discrete methods also have a limitation of having discrete clusters and a large number of labels [116]. Therefore, we have optimized a combination of continuous and discrete image function in two phases, to obtain the optimal number of clusters. The RBF kernel on HD image has used as a function that provides the functional output as a square Euclidean distance between image values and initially estimated cluster centers. The discrete function is an α expansion graph cut method which is a well known NP-hard method [79]. The objective of this study is to optimize the combination of continuous image function(RBF kernel-based) and discrete image function(graph cut) to obtain the optimal number of clusters [98, 112, 83, 106]. Once we computed the optimal clusters, then we have found the updated cluster centers [84, 109, 81]. These cluster centers have resulted in a new continuous image function, which provides the improved data cost of our optimization

framework. The data cost received by the above optimization has fused with the data cost of the subspaceMLR method, obtained from step-1, by using a linear decision fusion method. The step-2 has performed in two phases. The formulation of step-2 and data cost fusion has discussed in the proposed method section. In the first phase of step-2, we have updated the cluster centers by minimizing cluster energy. This energy minimization has stopped until some stopping criteria have met. The energy minimization has resulted in improved cluster centers. In the second phase of step-2, the RBF kernel-based image function has updated using improved cluster centers obtained from phase-1. Classification probabilistic result from step-1 and updated image function from step-2 has transformed into a spectral data-cost. Subsequently, the data-cost of step-1 and step-2 have fused with the linear decision fusion method. Finally, The graph-cut method has applied to the fused spectral data-cost(Dc) and a spatial smoothness cost(Sc). The fusion of data-costs has resulted in a significant improvement in accuracy.Experimental results on synthetic and real HSI datasets has obtained in the Experimental results section. Section-5.2 and 5.3 show the formulation and experimental analysis of the two-phase label optimization method. Figure-5.1 shows the process flow diagram of our proposed two-step method-1. In Proposed method-2, we have studied another relaxation labelling approach in which a spatial relaxation has exploited the neighbourhood relationship between the pixels of an image to minimize the spatio-spectral distortion. Application of spatial relaxation with spectral data can lead to reduce the noise effect and increase the class characterization. Such methods can also be applied to estimate the posteriors of the probabilistic classifier, to increase the classifier's final accuracy. In proposed method-2, we have introduced an edge-based feature fusion method, which helps characterize the class labels of HD Image in a better sense. It is an iterative method that exploits the spatial information from an image in such a manner that it assumes the feature preservation in vertical and horizontal directions for each pixel. By combining the subspace regression-based probabilistic method, the proposed method gives better accuracy for benchmark HSI datasets. Before this, we have implemented a fast Bayesian subspace regression method to achieve the posterior probabilities, for our edge feature relaxation method. The subspace regression approach in Method-2 is based on HySime projection followed by an EPS based edge preservation scheme. Figure-5.11 has denoted the process flow of our method-2. We have performed the experimental evaluation of the EPS optimization method in Experimental Analysis-2. Finally, we have compared the results with some recently proposed methods, and α expansion graph cut optimization method, which is an efficient technique to fuse the contextual knowledge in posterior

probabilities.

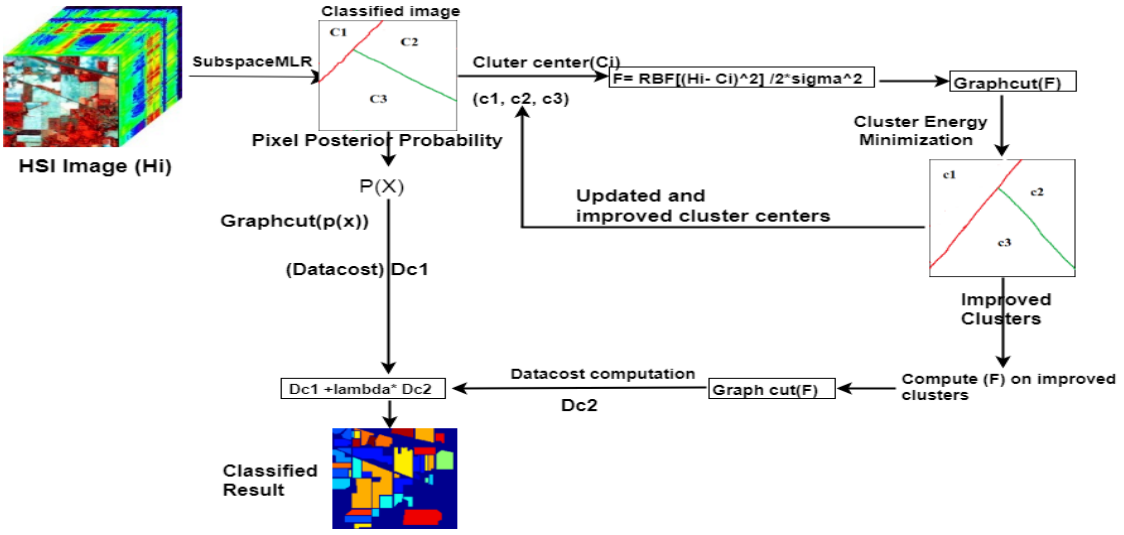


FIGURE 5.1: Process flow of proposed method

5.2 Proposed Method-1

In this section, we have formulated our two-step framework. In section-5.2.1(step-1), we have explained the details of the subspaceMLR method. The cluster centers of the classified image obtained from step-1 has used in step-2 to improve the classification accuracy. In section-5.2.2, we have formulated the step-2, which is a two-phase optimization technique. In phase-1, we have calculated the cluster centers of the classified image by taking an arithmetic mean of the principal component's spectral values in each cluster. Subsequently, we have formulated an RBF kernel-based continuous image function in which the square Euclidean difference between the image and cluster centers has taken as input. In phase-2, the continuous RBF kernel has optimized with graph-cut based discrete function. The optimization summary is as follows:

The outcome of RBF kernel-based continuous function has considered as data cost for graph-cut based discrete function, which gives the cumulative energy and the improved cluster labels. The enhancement in cluster labels resulted in the improved cluster centers and decreased in cluster energy by a fixed fraction. The improved clusters centers have provided improved data cost by using RBF kernel. The improved data cost obtained from

step-1 and step-2 has fused with linear decision fusion, which significantly improves the classification accuracy.

5.2.1 Step-1:Subspace Projection and Logistic Regression method:SubspaceMLR

In this section, we have obtained the procedure for calculating the robust posterior probabilities for the mixed pixel environment of HD images. Let for an image $H_i = \{x_{i1}, x_{i2}..., x_{in}\}$, the $(x_{i1},\ x_{i2}, .., x_{in})$ represents the data pixels where $i\varepsilon N$ and N=[1,2,3,...,n] pixels. The purpose of pixel-wise prediction is to determine the posterior probability of a pixels belonging to set of given labels $y_i\varepsilon(1,2,..,K)$, where K shows the classes. The posterior has calculated by using an estimated probability of pixel that are related to K classes .i.e. $p(y_i = K|x_i)$. Hence, the problem has transformed into MAP (maximum a posterior) estimate. The discriminant equation is: $y_i = k_1$ if $p(y_i = k_1|x_i) > p(y_i = k_2|x_i)$, where $(k1, k2\ \varepsilon\ K)$. Many other methods for instance [106, 83, 112], have suggested to determine the MAP estimates. Numerous algorithms such as [106, 83, 112], are proposed to solve this MAP based estimate. In present work, we have adapted the mlrsub approach that has proposed in the reference [95], therefore, we refer the readers to this reference. This has lead to model the MAP function directly to posterior probabilities of prediction estimates for each class. The discriminative and generative probability, .i.e $p(x_i|y_i)$ and $p(y_i)$, has resulted in class posteriors .i.e $p(y_i|x_i)$. The method has formulated with a regressor w as:

$$\hat{p}(y_i = k|x_i, w) = \frac{exp(w_{(k)}^T \phi_{(k)}(x_i))}{\Sigma_{k=1}^K exp(w_{(k)}^T \phi_{(k)}(x_i))} \tag{5.1}$$

In this method, new subspaces have determined before the predictions because it has found that high dimensional dataset can dwell in a low dimension with preserved information. Therefore, the classes are assumed to be the orthonormal basis vectors. The feature vectors i.e. $\phi_{(k)}(x_i)$ can be shown as:

$$\phi_{(k)}(x_i) = [||x_i^2||, ||x_i^T U_{(k)}||^2]$$

where $U_{(k)}$ is the feature space of length k. The annotation for this method is subspaceMLR, in place of mlrsub, for a better sense. The finally predicted HD image, which improved

in step-1, has achieved by equation-1. The data cost has also evaluated in step-1 by using posteriors obtained from equation-1.

5.2.2 Step-2:Two phase optimization method: paraKERNALGC

In this section, we have formulated and optimized a class label updater function, which updates the centroid of classes(cluster center) by updating the cluster labels. The update in cluster labels has resulted in a new cluster center that gives us an updated data cost. The updated data cost has used in the α expansion based graph-cut function to improve the cluster labels by minimizing the cluster energy. Two-phase label update procedure and it's functional description in the section-5.2.2.1 and section-5.2.2.2(phase-1) has been taken from [105]. Therefore, we refer the readers to [105] for a better understanding of the objective function described in equation-5.2 and label update process related in equation-5.6 to 5.10. In section-5.2.2.1(phase-1), we have formulated a combination of continuous and discrete image function. Section-5.2.2.2 describes the optimization procedure to minimize the function to obtain improved cluster labels. In section-5.2.2.3(phase-2), the enhanced cluster labels have transformed into improved data cost using continuous RBF kernel. In section-5.2.3, we have described the method that has fused the data cost obtained from step-1 and step-2 using linear decision fusion. We have denoted the combined function as paraKERNALGC.

5.2.2.1 Functional description

Let us have HD image I divided into N regions. θ is a function that map our image pixel 'a' to given labeled class L as θ : a ε I $\rightarrow \theta$ (a)ε L. 'a' is the point or spectral values of the image, L is the labels, S is smoothness term, N_t is noise term or data cost, c is cluster center obtained from step-1 and b is the neighbourhood points of a. Mapping function θ allocates each data point of the image I to a labelled region. L is the number of regions with cardinality less than N. The optimization function has defined as:

$$f(\theta) = N_t(\theta) + \mu S(\theta) \tag{5.2}$$

- N_t is noise term or data cost
- μ is constant
- S is smoothness term

Let we have a point 'a' on image which have 'b' points in the neighbourhood then noise term of objective function is

$$N_t(\theta) = \sum_{a \varepsilon I} N_{t(a)}(\theta(a)) = \sum_{l \varepsilon L} \sum_{(a \varepsilon S_l)} (c_l - I_a)^2 \tag{5.3}$$

In this equation c_l is cluster centre or centroid of region l. The smoothness term is:

$$S(\theta) = \sum_{(a,b) \varepsilon N} s_{(a,b)}(\theta(a), \theta(b)) \tag{5.4}$$

where N is a neighbours of pixel point (a,b). $s_{(a,b)}(\theta(a), \theta(b))$ is smoothness term with differences between truncated square values [81, 80]. $s_{(a,b)}(\lambda(a), \lambda(b)) = min(Constant^2, |c_{\theta(a)} - c_{\theta(b)}|^2)$. The optimization equation has transformed as:

$$f(c_l, \theta) = \sum_{l \varepsilon L} \sum_{a \varepsilon R_l} (c_l - I_a)^2 + \alpha \sum_{(a,b) \varepsilon N} s_{(a,b)}(\theta(a), \theta(b)) \tag{5.5}$$

In the next two subsections(phase-1 and 2), we have optimized the equation-5.5 by using gradient descent-based energy minimization and graph cut method. Phase-1 denoted the label update by gradient descent based energy minimization, and phase-2 describes the Graph cut segmentation approach.

5.2.2.2 Phase-1: Class Label update using gradient descent

In this phase, the equation-5.5 is minimized using label update with gradient descent methodology, as followed: Differentiate equation-5.5 w.r.t c_k where $c_k \varepsilon L$

$$\frac{df(c_l, \theta)}{dc_k} = 2 \sum_{a \varepsilon S_k} (c_k - I_a) + 2\alpha \sum_{(a,b) \varepsilon N, \theta(a) = k, |c_{\theta(a)} - c_{\theta(b)}| < constant} |c_{\theta(a)} - c_{\theta(b)}| \tag{5.6}$$

We have considered the four neighbourhood of s_k .i.e. N = 4. If b is the neighbours of a then $\theta(a) \neq \theta(b)$ and $|c_{\theta(a)} - c_{\theta(b)}| < constant$. Therefore, equation-5.6 can be simplified as:

$$\frac{df(c_l,\theta)}{dc_k} = 2\sum_{a\varepsilon s_k}(c_k - I_a) + 2\alpha \sum_{(a\varepsilon B_k)}\sum_{b\varepsilon N}(c_k - c_{\theta(b)}) \tag{5.7}$$

B_k is boundary of each regions S_k. To minimize above equation, put $\frac{df(c_l,\theta)}{dc_k}$=0

$$2\sum_{a\varepsilon s_k}(c_k - I_a) + 2\alpha \sum_{(a\varepsilon B_k)}\sum_{b\varepsilon N}(c_k - c_{\theta(b)}) = 0 \tag{5.8}$$

$$\sum_{a\varepsilon s_k} I_a + \alpha \sum_{(a\varepsilon B_k)}\sum_{b\varepsilon N} c_{\theta(b)} = c_k \times S_k^{\star} + c_k \alpha \sum_{a\varepsilon B_k} N^{\star} \tag{5.9}$$

From above equation, the label optimization function c_k is turn out to be as

$$c_k = \frac{\sum\limits_{a\varepsilon s_k} I_a + \alpha \sum\limits_{(a\varepsilon B_k)}\sum\limits_{b\varepsilon N} c_{\theta(b)}}{S_k^{\star} + \alpha \sum\limits_{a\varepsilon B_k} N^{\star}} \tag{5.10}$$

$\star$ denotes the cardinality of regions or class boundary. By using the updated cluster center c_k, we have calculated the data cost by using the RBF kernel function described in section-3.2.

5.2.2.3 phase-2:Partition Update with Graph Cut

In the phase-2, we have applied a α expansion graph cut method to obtain the cluster energy and updated cluster labels by using data cost obtained from phase-1. Phase-1 has resulted in updated cluster centers c_k, updated cluster labels, and minimized energy. The updated cluster centers will iteratively pass into graph cut based function. This iterative process has continuously decreased the cluster energy and updates the labelled regions according to phase-1. The α expansion graph cut method consists of a process that finds the N-cut in a graph with minimum cost. For a detailed explanation of this method, we refer the readers to [86].

5.2.3 Fusion of Datacosts using Regularizer (λ)

The data costs obtained from section-5.2.1 and 5.2.2 has been fused with a linear decision fusion method using regularizer λ. ?*lambda* refers to the weight of data cost obtained by the optimization process discussed in step-1 and step-2. The value of λ lied between (0-1), and it has observed that classification accuracy is increasing at λ. The data cost from step-1 is:

$$Dc_1 = -log[\hat{p}(y_i = k|x_i, w))] \tag{5.11}$$

Dc_1 is calculated by using log likelihood values obtained from equation-1 of step-1.

$$Dc_2 = -RBF\left(\frac{(H_i - c_i)^2}{2\sigma^2}\right) \tag{5.12}$$

Dc_2 is calculated using RBF kernel based continuous image function described in step-2. Where H_i is the spectral values of HSI image and c_i is the updated cluster centers obtained from phase-1 of step-2.

$$Dc_{Fusion} = Dc_1 + \lambda Dc_2 \tag{5.13}$$

s.t $0 \leq \lambda \leq 1$. The combined data cost (Dc_{fusion}) has used with spatial smoothness cost (Sc) of HD image in graph-cut based α expansion method to obtain the finally improved classification result. In equation-5.13, λ=0 implies that only subspace regression-based probabilities have used in the classification process. The impact of regularizer λ on overall accuracy has discussed in the experimental section.

5.3 Experimental Result Analysis-1

In this section, we have tested the efficiency of our proposed method on both simulated and real Hyperspectral datasets. Our initial interest is to run the experiments with a synthetic image to evaluate the proposed scheme in a controlled situation. Our main interest is to run the method with real HSI images. The scale parameter of the RBF kernel is set to $\sigma = 1$ for synthetic and real image experiments. ρ is an information control parameter, and we have studied the impact of σ on Real images in the parameter analysis section. We have

generated the synthetic images by using two antipodal vectors having opposite signs. Real HD images have taken from "Grupo de Inteligencia Computacional (GIC) Buscar."

5.3.1 Experiment 1: Application on synthetic image

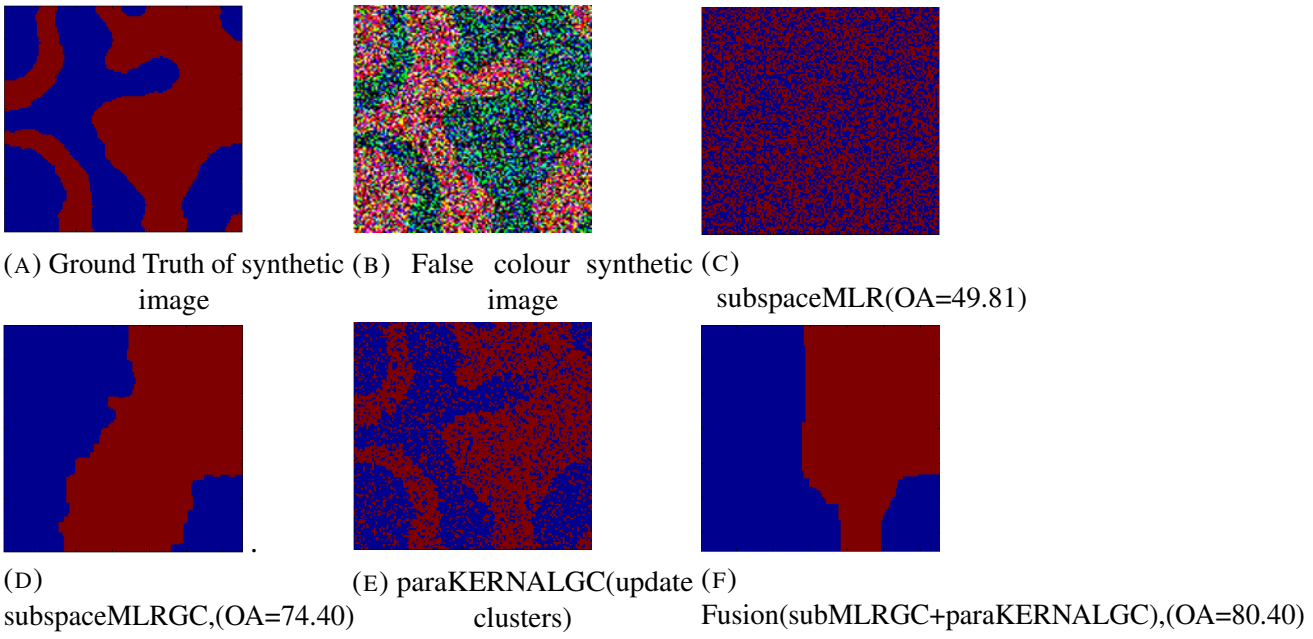
In this experiment, we have set up a synthetic Hyperspectral image to observe the effect of the parameter used in the proposed method on overall accuracy in a synthetic image environment. Our synthetic image is a binary image having two classes which have generated by using two antipodal vector [94]. For synthetic image, we have first created multilevel binary prior to having two labelled classes as $y \varepsilon L$ and L=[1,2] with image size (130×130) and mean=2. The data vector x has simulated as :

$$x_{z_i} = m_{z_i} + n_{z_i} \; \forall \; i \; \varepsilon \; S \; and \; z_i \; \varepsilon \; L^n \tag{5.14}$$

Where L is the number of labels and S is data points or pixels in a synthetic image. Here,x_{z_i} is synthetic image, m_{z_i} is known antipodal spectral values and $[n_{z_i}]$ is zero means Gaussian noise with covariance $\sigma_1^2 I$, that is, $[n_{z_i} \sim N(0, \sigma_1^2 I)]$. We have generated a binary image of antipodal values having 50 bands so number of class Labels are L=2, data points $x_{z_i} \; \varepsilon \; R^{50}$, the antipodal spectral values are $m_{z_i} = +1 or -1$. The synthetic image of labelled class L has shown in figure 5.2(a), where classes are $L_i = 1, 2$ For detailed use of antipodal values in the synthetic image we refer the readers to [94]. The proposed two-step method has performed on the synthetic binary image, and results are shown in figure-5.2(c), (d), (e), (f). The overall accuracy of proposed fusion method(subspaceMLR+paraKERNALGC) is OA=80.09 which is significantly higher than the subspaceMLR(OA=49.81%) and subspaceMLRGC(OA=74.40%) method, as shown in figure-5.2(c), (d), (f). Figure-5.2(e) shows the updated cluster results obtained from two-step optimization. The effect of various parameters on overall accuracy has shown in subsequent sections as

(1) In the first experiment, we have evaluated the effect of the training sample size on the overall accuracy.

(2) In the second inspection, we have stabilized the impact of the σ_1 on the overall accuracy.



(A) Ground Truth of synthetic image (B) False colour synthetic image (C) subspaceMLR(OA=49.81)

(D) subspaceMLRGC,(OA=74.40) (E) paraKERNALGC(update clusters) (F) Fusion(subMLRGC+paraKERNALGC),(OA=80.40)

FIGURE 5.2: Classification and Clustering with spectral and parametric kernel method in projected subspaces

5.3.1.1 Experiment 1:Impact of Training Sample Size on OA

In the first experiment with synthetic images, we have depicted the effect of training samples on overall accuracy. The parameters that has been fixed in this experiment are: the synthetic image noise $sigma_1 = 1.2$, RBF kernel standard deviation $\sigma = 0.2$ (Continuous image function) and value of $\lambda = 1$. We have analyzed the Overall accuracy(OA) on the training samples increasing with a step size of 500, as shown in figure-5.3. It has observed that the OA in subspaceMLRGC is increasing when training samples are increasing, and it is maximum at 1500 training samples. In the proposed fusion method(subspaceMLR+paraKERNALGC), the overall accuracy is always increasing when training samples are increasing. The Fusion method has achieved 92.32% accuracy, which is significantly higher than the subspaceMLR and subspaceMLRGC methods. it has shown that the obtained overall accuracy massively depends upon training samples,which is key factor,affected by the dimensionality.

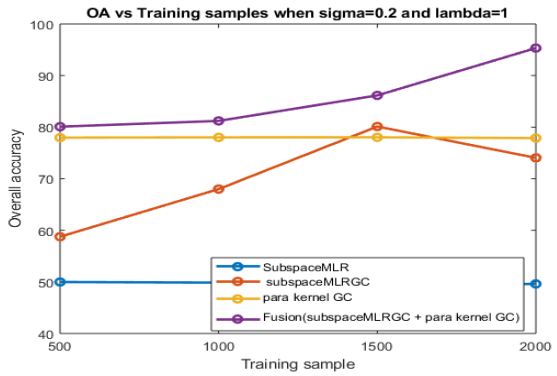


FIGURE 5.3: Overall Accuracy(OA) Vs Training samples

5.3.1.2 Experiment 2:Impact of Parameter σ_1 on OA :

In the second synthetic image experiment, we observed OA's effect on a synthetic image noise σ_1. In this experiment we have fixed some experimental parameters as: the RBF kernel standard deviation $\sigma = 0.2$, training samples =1000, and $\lambda = 1$. However, this parameter set up is inadequate but leads to a robust and better set of results with our proposed method. Figure-5.4, shows the OA vs σ_1 plot. When σ_1 increases, the OA is also decreases in case of subspaceMLRGC and proposed subspaceMLR+paraKERNALGC(fusion) method. The parakernalGC and subspaceMLRGC methods have a minimal effect of σ_1 on overall accuracy. In our Fusion results, it has observed that OA is not significantly decreasing with σ_1. A comprehensive inspection of the proposed methods, that is, the evaluation of the proposed method's performance on the real hyperspectral images, must be performed. It has done in the subsequent section with real-time images.

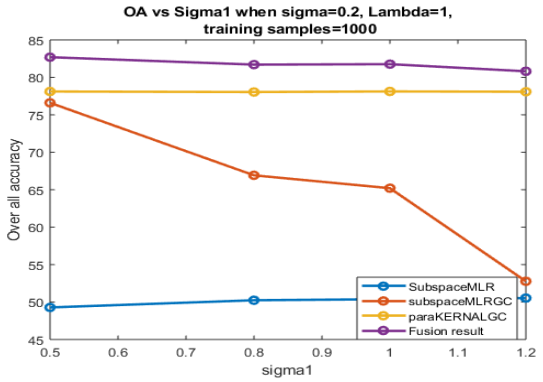


FIGURE 5.4: Overall Accuracy(OA) Vs σ_1

5.3.2 Experiment 2:Experiments with real images data set

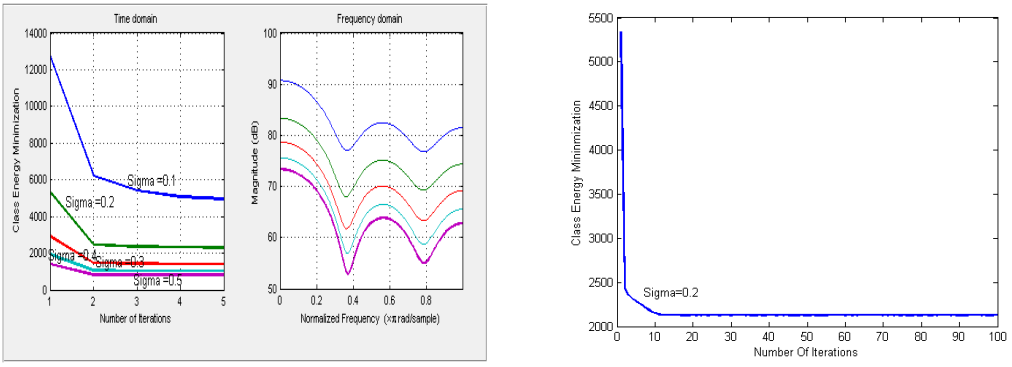
The experiments of proposed methods on a synthetic image has to lead to improve the classification accuracy significantly. Therefore, in this section, we have to perform the experiments on real HD images. To explore the proposed subspaceMLR+paraKERNALGC and paraKERNALGC methods, we have analytically tested the widely used Hyperspectral datasets obtained from "Grupo de Inteligencia Computacional (GIC) Buscar". To analyze the acceptance of proposed method, we have compared the proposed method results with some popular supervised learning schemes(comparative methods) like linear discriminant analysis (LDA), quadratic discriminant analysis (QDA) and Support vector machine(SVM)[139, 100, 108], that are proposed in [138, 102, 53]. For comparison, we have applied the HySime method for subspace projection before applying the LDA and QDA.

5.3.2.1 Observation 1:On Indiana Pines Dataset

In this experiment, we have used the Indiana Pines(IP) image to study the performance of the proposed method. The image consists of 145×145 pixels in 200 channels. The ground truth contains 16 mutually exclusive classes with 10366 validation data samples. It is a standard data set to authenticate the performance in hyperspectral imaging research issues and settled the problem using a critical presence of spectrally mixed pixels in the class data value. For data sampling, we have used low training samples as $9-10\%$ of available data. Table-5.1 shows the classification results achieved by the proposed subspaceMLR+paraKERNEGC method. To implement the two-step subspaceMLR+paraKERNALGC method with the IP dataset, in step-1, we have calculated the classified IP image using subspaceMLR method. In step-2, we have computed the cluster centers obtained from step-1 and performed the two-phase optimization. In the second step, we explored the local cluster information in our image, in which we have assumed that the number of regions is unknown, and we have to fix those labelled regions. The paper [81], is commonly applied for graph cut clustering when the number of regions is unknown. This method uses a continuous image space function, which inherently maps the data image with the help of the RBF kernel and optimizes the image function with a discrete graph cut function. The proposed fusion based 'subspaceMLR+paraKERNALGC'

method is achieving an overall accuracy(OA) of 94.15%, which is significantly higher than the previously proposed subspaceMLR, and subspaceMLRGC method's accuracy, which is 68.17 % and 93.65% respectively. The figure-5.5(a) and (b) shows the decrease in cluster energy after 5 and 100 iterations of optimization function. It has observed that our optimization is convex since cluster energy is minimizing after each iteration. To analyze the effect of σ on cluster label update, we have derived the improved clusters by our proposed scheme using different values of σ= 0.1, 0.3, 0.4, 0.5, and 0.7 1.0. σ is the standard deviation of continuous RBF image function. As σ of RBF kernel increases, cluster size also increases, therefore it is unable to detect small contrast regions. Therefore we have selected σ=0.2 for our current experiment, which is providing the updated clusters very close to ground truth. Figure-5.6(a) to (r) have represented the clusters, contours, and cluster labels mapping on the original image by using the paraKERNALGC method, discussed in step-2. These outcome images show that our paraKERNALGC method can handle a different kind of noise in the classes and small contrast values. Finally, we have fused the data cost(DC) resulting from the paraKERNALGC method with data cost(DC) of subspaceMLR. After fusion, in the subspaceMLR+paraKERNALGC method, our accuracy of classification raised to 94.15% from 93.65%, which is a remarkable aspect. Further, we have implemented some comparative methods like HySime+LDA, HySime+QDA, HySime+LDAGC, HySime+QDAGC, and SVM, to compare the results with the proposed method, shown as comparative methods in Table-5.1. The comparative methods have proposed in [94] and denoted as LDA, QDA, LDAGC, QDAGC, and SVM for short. The suffix GC in subspaceMLRGC, LDAGC, and QDAGC has denoted the graph cut approach, which has used for spatial information fusion in the subspaceMLR method. As a result, the graph cut(GC) method has appropriately increased classifier performance. We have not described the operational detail of SVM, HySime+LDA, HySime+QDA, HySime+LDAGC and HySime+QDAGC methods which have discussed in [77]. LDA and LDAGC methods have concluded with 65.63% and 79.83% accuracy, respectively, which is 19 % and 15 % lower than our proposed method. In QDA and QDAGC implementation, we obtained 80.81% and 91.17% accuracy, which is 14% and 4% lower than our proposed method. The SVM method provides significant accuracy .i.e. 80.81%, although it is lower than the proposed method but higher than the subapceMLR method. We have not implemented the SVM methods in this paper; results has taken from [113]for comparison. Graph cut optimization of SVM results .i.e. SVMGC(.i.e.SVM graph cut) is not investigated because posterior probability of SVM is not reliable, since it

is a hard classifier. Figure-5.7(a) shows the ground truth, and figure-5.7(b), (c), (d), (e), (f) are the outcome of the comparative methods discussed above. Figure- 5.7(g), (h), (i) represents the label contours, label mapping on the image, and Label energy integration with IP image, respectively. Figure-5.7(j) shows the final results obtained from fusion-based proposed method .i.e. subspaceMLRGC+paraKERNALGC.



(A) Energy minimization after 5 iterations in IP Dataset

(B) Energy minimization after 100 iterations in IP Dataset

FIGURE 5.5: Cluster energy minimization

TABLE 5.1: Class wise and Overall performance of proposed method in Indiana Pines Dataset

	samples	subspaceMLR+paraKERNALGC	subspaceMLR	subspaceMLRGC	LDA	QDA	SVM	LDAGC	QDAGC
alfalfa	54	96.15	80.77	96.15	0	0	94.48	81.48	84.62
corn_no till	1434	88.77	38.92	78.43	59.26	80.77	71.34	78.64	86.89
corn_min till	834	88.98	57.74	90.16	73.84	78.02	68.34	57.08	85.03
corn	234	100	65.29	100	51.09	76.03	85.23	91.79	98.26
grass/pasture	497	94.88	73.02	94.19	61.94	88.95	73.19	68.51	93.43
grass/tree	747	97.63	94.21	97.63	70.03	92.25	96.94	98.92	96.14
grass/pasture-mowed	26	100	76.92	100	93.35	94.21	77.77	100	84.62
Hay-windrowed	489	100	96.44	100	92.31	84.62	85.89	100	96.95
oats	20	100	80.00	100	99.49	96.24	73.74	100	0
soyabean_no till	968	80.36	54.10	89.23	100	0	86.31	82.60	92.75
soyabean_min till	2468	99.70	65.04	99.70	57.60	79.16	87.03	73.02	88.43
soyabean_clean till	614	98.89	77.45	98.71	44.43	67.40	92.71	71.60	91.31
wheat	212	100	100	100	64.20	82.26	94.27	99.11	99.33
woods	1294	99.83	98.07	99.83	99.11	98.67	97.42	89.20	96.77
bldg-grass tree	380	73.62	33.88	76.87	82.58	94.46	99.13	84.64	91.35
stone-steel tower	95	100	97.14	100	66.07	76.60	90.29	79.17	100
OA	-	94.15	68.17	93.65	65.63	80.81	80.56	79.41	91.17
AA	-	94.93	74.31	95.06	69.71	74.35	77.81	84.73	86.62

5.3.2.2 Observation 2:On Salinas Valley Image Dataset

In this section, we have used another HD image, an AVIRIS sensor imagery of Salinas Valley(SV), California. The image data has 86×83 pixels with 204 bands in 6 exclusive

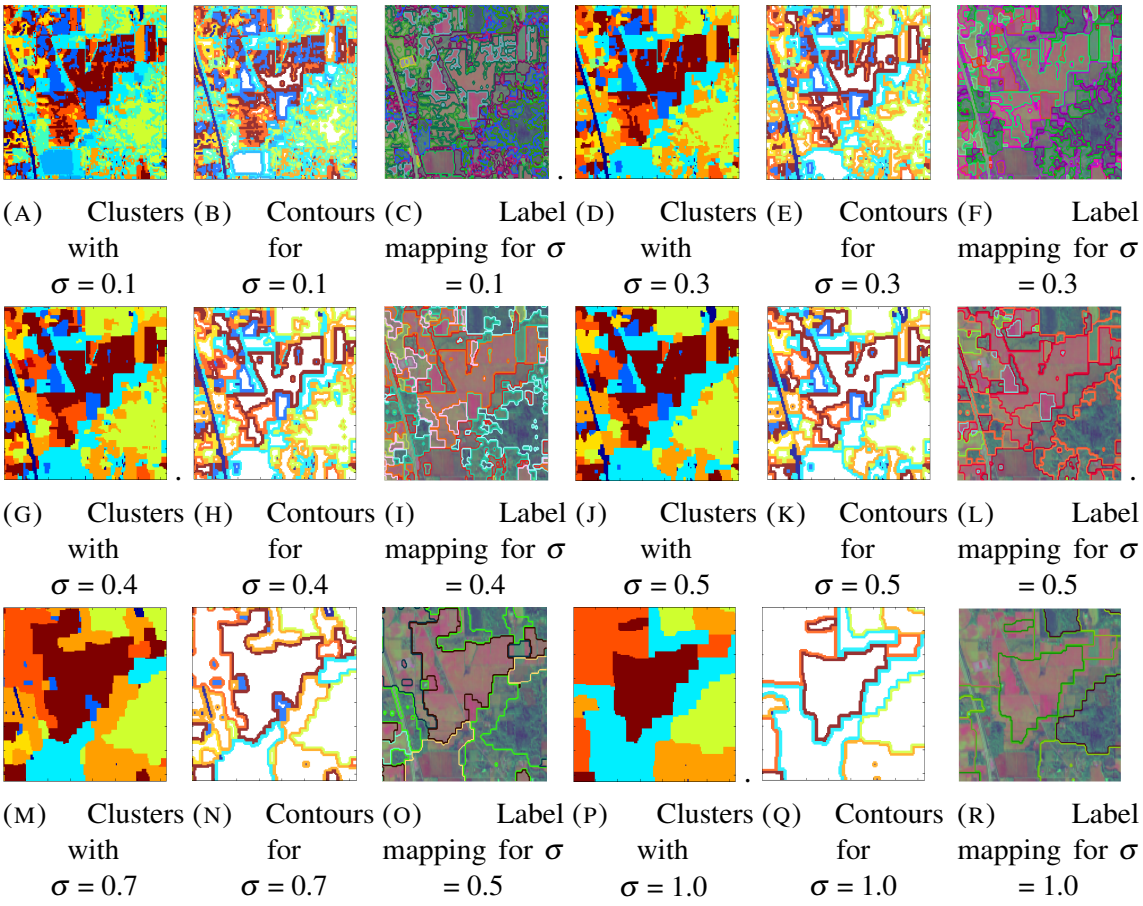
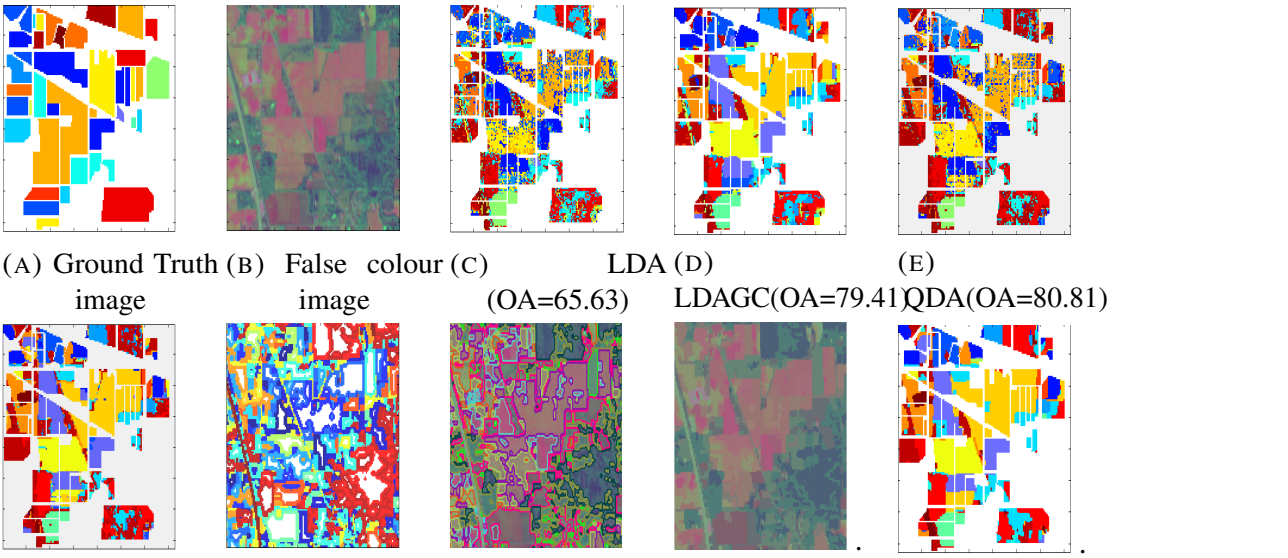


FIGURE 5.6: Cluster,Contours and Label mapping on image by paraKERNELGC method on Indaian pines data

ground truth classes. Some channels have removed since they are non-informative channels. This dataset has 6889 possible labelled samples for training and testing. Figure-5.9(a), shows a ground-truth image having six classes. The percentage of training and testing set is similar to experiment-1. We have applied our proposed two-step method in the SV dataset and found that cluster energy is decreasing with an increasing number of iterations, as shown in figure-5.8(a). The optimization in step-2, i.e., paraKERNALGC, has resulted in updated and improved clusters for the SV dataset, as shown in figure-5.8. Figure-5.8(b) to 5.8(j) have depicted the updated clusters, label contours, and cluster label mapping on image for different standard deviation i.e. σ=0.1, 0.5, 0.7 for RBF image function in paraKERNALGC method. Further, we have applied our two-step fusion-based method to identify the versatility of the proposed method. We have carried out the performance evaluation on the proposed subspaceMLR+paraKERNALGC method for a given validation image set. Table-5.2 is depicting the classification accuracy of proposed

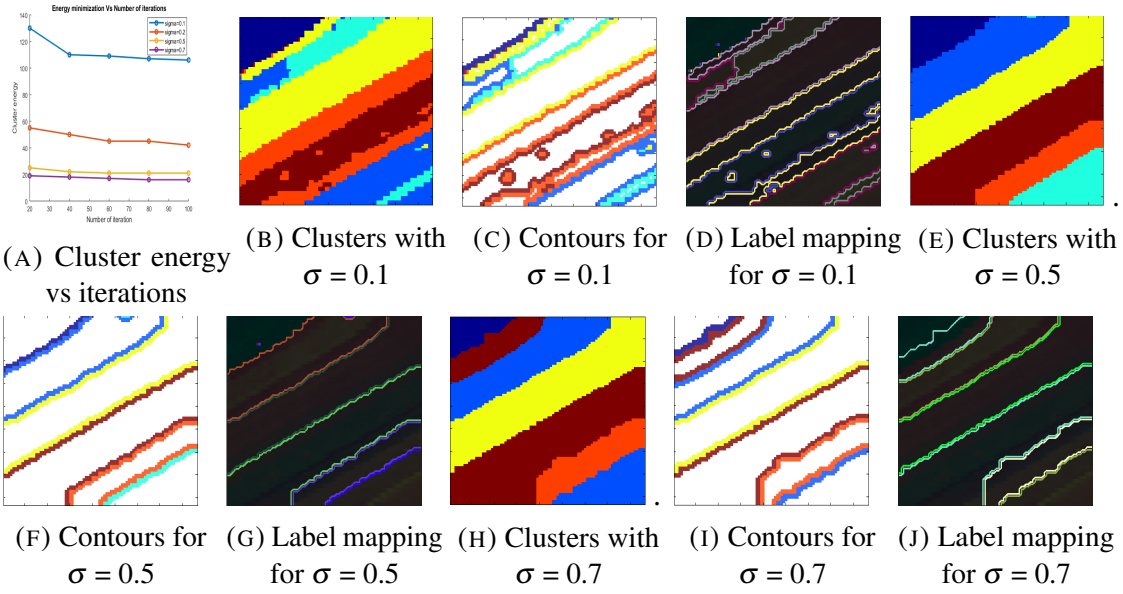


(A) Ground Truth image (B) False colour image (C) LDA (OA=65.63) (D) LDAGC(OA=79.41) (E) QDA(OA=80.81)

(F) QDAGC result(OA=91.17) (G) Label contours at σ=0.2 (H) label overlap on image at σ=0.2 (I) Label energy image at σ=0.2 (J) subspaceMLR+paraKERNALGC

FIGURE 5.7: Classification and Clustering with subspaceMLR and for Indaian pines data

and comparative algorithms, as mentioned in experiment-1. The proposed method.i.e. subspaceMLR+paraKERNALGC, has achieved the overall accuracy(OA) as 99.86% in the above experimental set-up. subspaceMLR and it's graph cut implementation .i.e. subspaceMLRGC, is achieving an accuracy of 97.73% and 98.86% respectively. LDA, QDA, LDAGC, and QDAGC attain accuracy as 98.87%, 99.44%, 98.95%, and 99.44%. In this case, the subspaceMLR+paraKERNALGC scheme is outperforming than other comparative methods. Figure-5.9 is representing the classification and clustering results carried out by proposed and comparative methods. For detailing, figure-5.9 (b),(c), (d) shows the updated cluster, cluster contours, and cluster label mapping in the image for σ=0.2. Figure-5.9(e) represents the proposed subspaceMLR+paraKERNALGC method result for two-step fusion. Figure-5.9(f) and 5.8(g) shows the subspaceMLR and subspaceMLRGC outcomes. Figure-5.9(h), (i), (j), (k) have depicted the comparative method results i.e., LDA, LDAGC QDA, QDAGC, as discussed in Observation-1. Along with this, Table-5.2 represents a comparison of the proposed subspaceMLR+paraKERNALGC method with some previously proposed approaches for the SV dataset.



(A) Cluster energy vs iterations (B) Clusters with $\sigma = 0.1$ (C) Contours for $\sigma = 0.1$ (D) Label mapping for $\sigma = 0.1$ (E) Clusters with $\sigma = 0.5$

(F) Contours for $\sigma = 0.5$ (G) Label mapping for $\sigma = 0.5$ (H) Clusters with $\sigma = 0.7$ (I) Contours for $\sigma = 0.7$ (J) Label mapping for $\sigma = 0.7$

FIGURE 5.8: Cluster energy, Cluster labels,Contours and Label mapping on image by paraKERNELGC method on Salinas valley data

5.3.3 Comparative Analysis

In this section, we have compared the proposed method results with some recently proposed state-of-art methods such as DWSSR(KNN), DWSSR(KNN), SSR(K-means), SSR(SVM), K-Mean, and RPCA+K-Means. In table-5.3, the overall classification(OA) accuracies of previously proposed methods have shared for IP and SV datasets. For IP dataset, our proposed subspaceMLR+paraKERNALGC method is achieving 94.15% accuracy which is 10% higher than FRPCAGL(SVM), 15% higher than DWSSR(SVM), 21% higher than DWSSR(KNN), 23% higher than SSR(SVM), 31% higher than SSR(KNN), 14% higher than K-means, and 2% higher than RPCA+K-means. The subspaceMLR is also achieving an accuracy of 93.65%, which is significantly better than FRPCAGL(SVM), DWSSR(SVM), DWSSR(KNN), SSR(SVM), SSR(KNN), K-means,

TABLE 5.2: Class wise and Overall performance of proposed method in Salinas Valley Dataset

	samples	subspaceMLR+paraKERNALGC	subspaceMLR	subspaceMLRGC	LDA	QDA	LDAGC	QDAGC
Brocoli_green_weeds_1	391	100	100	100	0	0	99.38	99.37
Corn_senesced_green_weeds	1343	100	96.73	100	99.38	99.37	99.12	98.08
Lettuce_romaine_4wk	616	98.88	88.10	98.88	98.88	98.00	94.55	99.81
Lettuce_romaine_5wk	1525	100	100	100	94.55	99.63	100	100
Lettuce_romaine_6wk	674	100	100	100	100	100	99.83	100
Lettuce_romaine_7wk	799	99.86	99.31	99.86	99.83	100	99.02	100
OA	-	99.86	97.73	98.86	98.87	99.44	98.95	99.44
AA	-	99.79	97.36	99.79	82.10	82.83	98.65	99.54

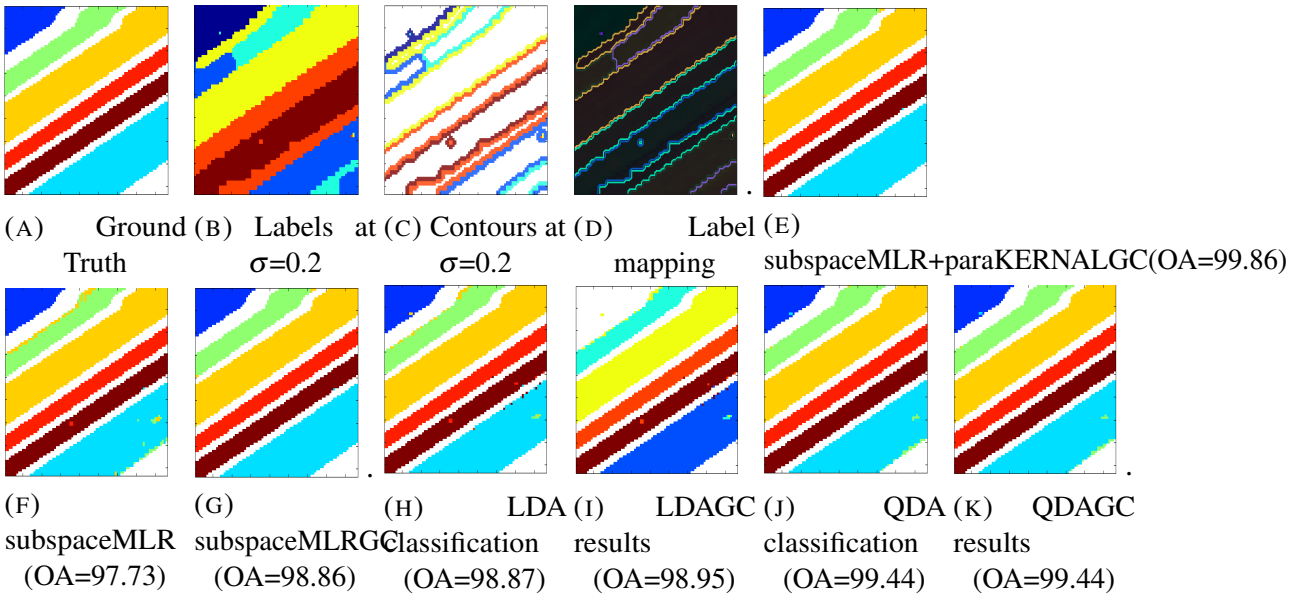


FIGURE 5.9: Classification and Clustering results(in%) with spectral spatial attribute,spectral data,LDA and QDA in projected subspaces for Salinas valley Data Set

and RPCA+K-means. For the SV dataset, the proposed subspaceMLR+paraKERNALGC methods are achieving higher accuracy than FRPCAGL(SVM), K-means, and RPCA+Kmeans method with a significant margin. Proposed subsaceMLRGC+paraKERNALGC is achieving the accuracy of 99.86% which is greater than the FRPCSGL(SVM(99.37%)), K-means(99.21%), RPCA+Kmeans(99.26%) methods. For the SV dataset, overall accuracy is above 99% for both proposed and past methods, but the proposed scheme is achieving greater accuracy than FRPCAGL(SVM), K-Means, and RPCA+K-Means methods.

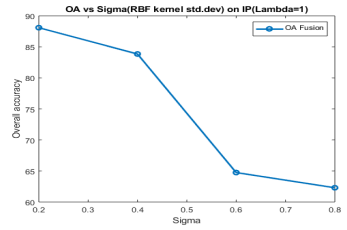
TABLE 5.3: Overall accuracy(in %) for recently proposed methods

	FRPCAGL(SVM)	DWSSR(SVM)	DWSSR(KNN)	SSR(SVM)	SSR(KNN)	K-means	RPCA+K-means
IP	84.58	79.89	73.80	71.86	63.53	80.84	82.41
SV	99.37	-	-	-	-	99.21	99.26

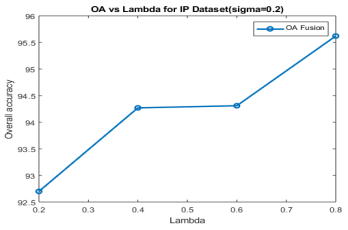
5.3.4 Parameter Analysis

In this section, we have analyzed the parameters used in the two-step proposed method .i.e. σ and λ. σ is the standard deviation of continuous image function(RBF kernel), and

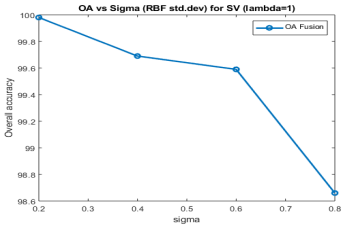
λ is regularizer, which is the weight of data cost in paraKERNALGC(step-2) method when fuses with the data cost of subspaceMLR method(step-1). It has observed that the overall accuracy(OA) is decreasing when the value of σ increases for both IP and SV datasets. The OA is increasing when regularizer weight λ for paraKERNALGC is increasing. The increase in λ is adding optimized information, which increases the accuracy of classification. Figure-5.10(a), (b) shows the effect of σ and λ on overall accuracy for IP dataset and figure-5,10(c), (d) depicted for SV dataset.



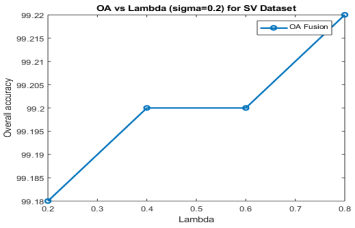
(A) OA vs Sigma for IP



(B) OA vs Lambda for IP



(C) OA vs Sigma for SV



(D) OA vs Lambda for SV

FIGURE 5.10

5.3.5 Computational Time

Our proposed method is computationally efficient for IP, SV and synthetic datasets. Our proposed method exhibits the fusion of two algorithms .i.e. subspaceMLR and paraKERNALGC. In the IP dataset, the execution time for subspaceMLR is 3.48 seconds, and for paraKERNALGCis 4.25 seconds. For the SV dataset, the computational time is 3.43 and 4.01 seconds for subspaceMLR and paraKERNALGC, respectively. The time taken for the synthetic image is 4.76 seconds for the subspaceMLR method. Therefore, we have observed that our proposed scheme is computationally robust and efficient for the pixel-wise classification of high dimensional images.

5.4 Proposed Method-2

In this section, we have initially formulated a pixel-wise classifier for posterior estimation, which is an essential input for our proposed edge preservation scheme (EPS). Further, we have improved the posterior probability by using our relaxation scheme. For example, in 9×9 image, as shown in figure-5.11, (p1, p2...,p9), the posterior probability is obtained from the subspaceMLR method and (g1, g2..., g9) are the edge-based gradient feature for each pixel. Let the final probability of central pixel with posterior p5 has to calculate. The final probability (p_final) has been calculated using weighted sum of **(Neighbours edge gradient value** $\times$ **Neighbour probability)** i.e. (p1*g1+p2*g2+...+p9*g9). This term can also be denoted as smoothness term. Smoothness weight λ is used as a regularizer to compute the weighted sum of neighbourhood probability for this pixel. Finally, the weighted sum has added with the posterior probability of pixel, as described in figure-5.11 to obtain the final posterior probability (p_final). The computation of p_{final} has performed in a serial manner where edge feature based gradient filter and N-neighbour probability-based spatial filter has traversed along the complete image using the Gauss-Seidel method as described in algorithm-2. The error after each iteration is an absolute difference between the initial probability and probability after fusion of image feature. The reduction in error is leading to an early convergence of the proposed method.

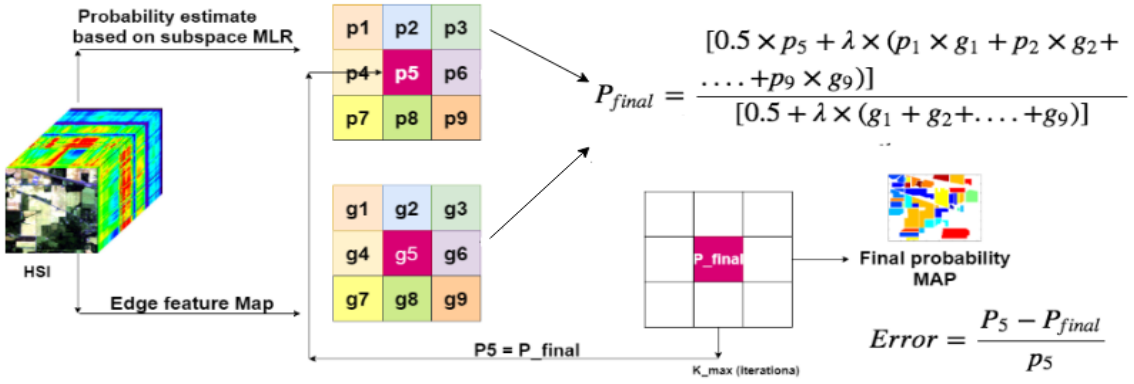


FIGURE 5.11: Process flow diagram for EPS optimization of central pixel

5.4.1 Posterior probability based pixel Learning Method:subspaceMLR

For a HD image, $H_i = \{x_{i1}, x_{i2}..., x_{in}\}$, $(x_{i1}, x_{i2}, .., x_{in})$ denotes the pixels vector where $i \varepsilon C$ and C=[1,2,3,...,n] for n pixels of an image. Our objective in pixel-wise probabilistic classifier is to calculate a decision probability of individual pixel for a set of assigned labels $y_i \varepsilon (1, 2, .., K)$, where K denotes the number of classes. The decision probability is calculated by a posterior estimation of probability for each pixel belonging to K classes .i.e. $p(y_i = K|x_i)$. Therefore, our problem is converted into a MAP(maximum a posteriori) based estimate, as described in problem statement section. Discriminant function is defined as: $y_i = k_1$ if $p(y_i = k_1|x_i) > p(y_i = k_2|x_i)$, where $(k1, k2\ \varepsilon\ K)$. The algorithm such as [128], has proposed to solve this MAP based estimate. In this paper, we have used subspace based Hysime method, proposed in the reference [122], and Multinomial regression to project the subspace and obtain the probability of prediction. By this method, we have modelled our MAP equation directly to posterior estimate of pixels probability for each class label. In this paper, the discriminative and generative probability, .i.e $p(x_i|y_i)$ and $p(y_i)$, is resulting in MAP based class prediction .i.e $p(y_i|x_i)$. This model is formulated as:

$$\hat{p}(y_i = k|x_i) = \frac{exp(\phi_{(k)}(x_i))}{\Sigma_{k=1}^{K} exp(\phi_{(k)}(x_i))} \tag{5.15}$$

The subspace projection idea has based on a fundamental assumption that any high dimensional dataset can reside in lower space without loss of information. Therefore, classes have represented as orthonormal basis vectors. Hence the feature vector function $\phi_{(k)}(x_i)$ is denoted as:

$$\phi_{(k)}(x_i) = [||x_i^2||, ||x_i^T U_{(k)}||^2]$$

where $U_{(k)}$ is the feature vector of length k. This method has annotated as subspaceMLR in this work.

5.4.2 Edge Preservation Scheme(EPS) based relaxation

In this section, we introduced a spatial edge relaxation based scheme that logically smoothing the original image and the posterior probabilities, obtained from previously

proposed subspaceMLR method. Let the posterior probabilities proposed by subspaceMLR methods are $p = [p_1, ..., p_k]$ ε $R^{K \times k}$ and probability vector, $p_s = [p_s(1), ..., p_s(K)]^T$ for $s \varepsilon S$, is a K dimensional multivariate vector in s^{th} class. Let $d_i = [d_i(1), d_i(2), ..., d_i(k)] \varepsilon R^{k \times K}$ be the vector of finally obtained probabilities by proposed Edge Preservation Scheme. We have transformed this method to an optimization problem having objective function as:

$$F(d) = \underset{d}{argmin} \left(\frac{1}{2} \mid\mid d - p \mid\mid^2 + \lambda \sum_i \sum_{j \varepsilon N_i} \phi_j \mid\mid d_j - d_i \mid\mid^2 \right) \tag{5.16}$$

such that:

$$d_i \geq 0 \ , \ 1^T d_i = 1 \ and \ \lambda \simeq 0.5; \tag{5.17}$$

we have supported the constraints by an evidence that vector d_i, that represents the probability 1, is a column vector of K 1's. λ denotes the regularizers that controls the effect of neighbourhood smoothing terms in proposed objective function. N_i is i_{th} neighbour of pixels, which is considered as eight neighbours in our paper. ϕ_j is an edge feature value, where $j \varepsilon S$ is a neighbourhood site in feature image:

$$\phi(x) = exp \left(- \sum_{i=1}^{u} P.F(X^{(i)}) \right) \tag{5.18}$$

P.F denotes the gradient-based Prewitt edge detector filter, which detects the edges at discontinuous points like 0 or 1. $X^{(i)}$ is a particular channel of the image in which Prewitt function has applied in both the horizontal and vertical directions. In the above objective function, the first term quantifies the ill-fitting of data, and the second term advocates the smoothness weighted by edge feature ϕ_j, in classification output. The value of ϕ_j is higher in continuous regions and lower at discontinuities. In the objective function, the first term resulted in noisy classification. The second term causes smoothness while applying the edge image ϕ in such a way that it does not smooth the discontinuities. Now we want to refer some past edge preservation schemes such that total variation (TV) and anisotropic diffusion [120, 135]. We have taken the objective function similar to the cases mentioned above. In past researches, the authors have minimized the objective functions with the total variation or constrained quadratic problem- solving. Whereas, we have minimized our objective function by iterative Gauss-Seidel method, which is converging very fast due to the convex objective. The optimization methodology for gauss-seidel function, as in

Algorithm 2 Gauss Seidel Method for EPS optimization

$Inputs \leftarrow p \; or \; X, \lambda, \; i, E = \| \; p \; \|, \omega$
$Output \leftarrow d$
$t \leftarrow 1$
for $[t = 1 : (E^{t+1} - E^{t}) : \omega; t < i]$ **do**
  for $band \leftarrow 1 \; to \; max(band)$ **do**

$$d_i^{t+1}(band) = \frac{0.5 \times p_i(band) + \lambda \times \sum\limits_{j \varepsilon N_i} \phi_j d_j^t(k)}{0.5 + \lambda \times \sum\limits_{j \varepsilon N_i} \phi_j}$$

  end for

$$d_i^{t+1} = \frac{d_i^{t+1}}{\sum\limits_{band=1}^{max(band)} d_i^{t+1}(band)}$$

$$E^{t+1} = \frac{\|d^{t+1} - d^t\|}{\|d^t\|}$$

end for

algorithm-2, has obtained and modified from [129]. In this work, we have predetermined the edges with the Prewitt function. However, some other edge identification function may also be used. It has observed that the predetermination of channel-wise edge feature leads to a unique solution since the number of channels is extremely large and, the problem is convex. Iterative Gauss-Seidel function minimizes the objective function for d_i^k and projects the posterior probabilities on minimum simplex for each pixel value. We have shown the pseudo-code in algorithm-2. 'i' denotes the number of iterations to converge, and E is an error after each iteration. $E^{t+1} = \frac{\|d^{t+1} - d^t\|}{\|d^t\|}$ shows the error parameters and ω depicts the parameter that controls error convergence. We have used this algorithm to improve the posterior probabilities as well as spectral values directly obtained from the image. subspceMLR+EPS term had denoted when we applied this optimization as post-processing of posterior probabilities. The term, im+EPS, is used when this algorithm has directly applied to image values as preprocessing or post-processing step. In the above algorithm, X denotes the original image, and P denotes the posterior probabilities obtained from subspaceMLR. Lastly, we have analytically shown that the algorithm converges swiftly with both types of data .i.e image value and posterior probability.

5.4.3 Multilevel logistic prior diffusion in Posterior with Graph cut (GC)

In recent times, the integration of the SVM and other methods with MRF based contextual prior has also increased the marginal probability up-to remarkable level[87, 123]. Therefore, in this section, we have formulated this approach, as discussed in [119], to compare with our proposed method analytically. Let, the classification labels obtained from the previously formulated process as pseudo-class labels. Let the classified labels are $\hat{y}$, and we have to find the final labelled outcome as y. We need to solve the following optimization:

$$y = \underset{y \varepsilon L}{arg\ max}\ [logP(y|\hat{y})] \tag{5.19}$$

$$y = \underset{y \varepsilon L}{arg\ max}\ [logP(\hat{y}|y) + logP(y)] \tag{5.20}$$

the

$$logP(\hat{y}|y) = \sum_{l=1}^{L} 1\{y_i = l\} log(\hat{y}_i l) \tag{5.21}$$

Where $l \varepsilon L$ is the available labels, it is an image segmentation task that forces the adjacent pixels to reside in the same class. This process has dramatically increased the classification accuracy. The contextual probability can be written as:

$$p(y) = \frac{1}{Z} e^{\mu \sum\limits_{(i,j) \varepsilon C} (y_i - y_j)} \tag{5.22}$$

where C is the neighbourhood pixels. From the likelihood estimates obtained from equation-(5.15) and equation-(22), we can write the final MAP estimate as:

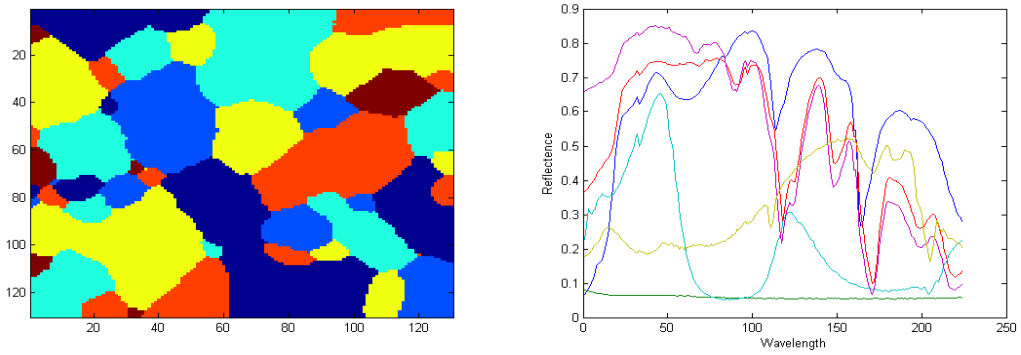
$$y = \underset{y \varepsilon k^n}{argmin} \left[\sum_{i \varepsilon s} -logp(y_i|x_i, \hat{w}) - \mu \sum_{1\ j} (\delta(y_i - y_j)) \right] \tag{5.23}$$

5.5 Experimental Result Analysis-2

Before the characterization of our outcome with a synthetic and real image, it is first mandatory to set up the parameters in our experiments. In this section, a small number of labelled training samples are used to speak for large classes. In subspaceMLR implementation part, the spectral control factors are τ=190,190,100,190 and training samples are L= 1000,1000,1000 and 500 for simulated, real Indiana pines, Pavia and Salinas valley dataset respectively. For class label relaxation on posterior probabilities and image values, we have used Regularizer λ=0.9,0.9,0.5,0.9 for HSI datasets, respectively. The maximum iteration for error detection has considered as 20 in each case.

5.5.1 Setting up the Experiment for Simulated Dataset

The first experiment has a set-up with a simulated image created using the 6-spectral signatures of classes obtained from the USGS spectroscopy lab, as shown in figure-5.12(b). A multilevel logistic prior based image has generated, as shown in figure-5.12(a). The



(A) Ground truth (B) spectral signatures

FIGURE 5.12: simulated image with spectral signatures

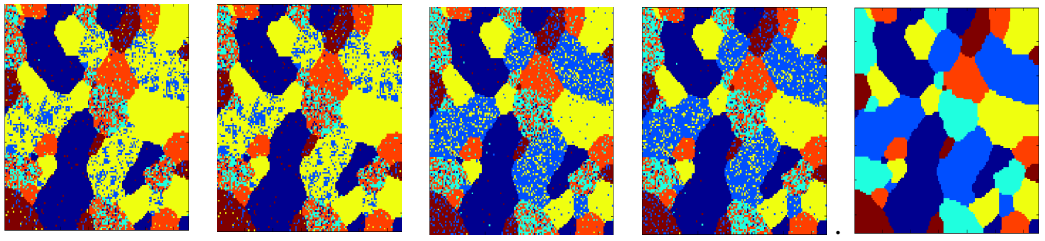
image is generated by spectral responses of 6 mineral classes given for 224 channels of AVIRIS sensor.

5.5.2 Experiment 1:Modelling on synthetic image dataset

In this section, a simulated image is created by generating a 6 class image with multilevel logistic prior using a mixing method described as:

$$x_j = \sum_{j=1}^{J} m_k \gamma_k j + n_j \tag{5.24}$$

Where k is the number of class, assumed as 6. m_k's are the spectral signatures for k^{th} label. Hence, x_i's are obtained by synthetically mixing of spectral signatures for every pixel. γ_k is the abundance constants that assumed to be fixed during the simulation. In various literature, γ_k is shown to be very erratically distributed in class labels, so we have considered it to reside in lower subspaces and linearly distributed [122, 126]. Edge preservation scheme(EPS) has applied to posteriors obtained from the subspaceMLR method and the original image having regularizer λ=0.9. Maximum iterations have been fixed as k_{max}=20. Improved posterior estimates are obtained to predict the final labels. However, we have also applied a α expansion-based graph-cut(GC) optimization method to improve posterior results for real HSI's in subsequent sections. In MRF based graph-cut(GC) optimization, smoothness constant, $\mu = 2$, is applied on synthetic image of size $[130 \times 130 \times 224]$. n_i is normal or Gaussian noise added to the image .i.e $n_i \sim N(0, \sigma^2 I)$. We have assumed x_i having classes $y_i = k_1$; then, we specified the $\gamma_i^{k_1}$



(A) Eps_im method (OA=70.89) (B) Eps_im_Eps method (OA=70.90) (C) subspaceMLR (OA=83.42) (D) subspaceMLR_Eps (OA=86.25) (E) subspaceMLR_Eps_GC (OA=98.93)

FIGURE 5.13: Classification and class relaxation outcome

as abundance factor for a class and $\gamma_i^{k_2}$ (where $k_1 \varepsilon K$ but $k_2 \nsubseteq k_1$) is abundance factor of rest other classes. It has been assumed that the abundance of other classes creates the problem of spectral mixing in a pixel value. Therefore, $\gamma_i^{k_1} = \gamma$ is set up with a uniform distribution for this experiment. We have applied constant abundance values for the

complete dataset. Some analytical experiments have deftly performed so that various useful aspects of our proposed EPS based subspaceMLR method has discovered in a robust supervised setting.
(1) In the first observation, we evaluate the impact of regularizer λ for edge preservation (EPS) on the overall accuracy.
(2) In the second experiment, we figured out the effect of error(E) on the number of iterations performed to obtain improved posterior distribution.
(3) In the third experiment, we have analysed the noise effect(σ) on OA.

5.5.2.1 Exercise 1:Effect of Regularizer λ on OA

In this section, we considered the error variance $\sigma = 0.4$, spectral control for subspaces $\tau = 0.9$, and abundance factor $\gamma = 0.7$ for subspaceMLR method. Subsequently, the posterior probability of estimate has improved by the proposed scheme(.i.e EPS), in which local information has obtained with Prewitt gradient and maximum iterations, .i.e. 20. The ε is eight neighbourhood information for the proposed optimization objective. In Figure-5.14, We have analytically observed that overall accuracy in the subspaceMLR method is increasing with an increase in λ from 0 to 0.8, and later, it decreases. Maximum OA has obtained when class relaxation-based EPS method has performed on the posterior distribution of subspaceMLR, .i.e subspace+EPS. On applying EPS relaxation on the simulated image, OA is increasing with an increase in λ for both the cases:(1) when EPS is used as preprocessing step, and (2) when EPS has used as a post-processing step. It is crucial to notice that only the first part of the optimization equation has considered when the value of λ is 0, and the initially formulated subspaceMLR method remains effective for prediction. λ=1 will lead to making both the terms, .i.e. noise, and smoothness, effective in optimization equation-5.17. Figure-5.14 plots the derived OA outcomes as a function of λ, with maximum iterations. It has fixed as K_{max}= 20. In this figure, the class relaxation process indeed relies upon regularizer λ. Remarkably, for $0.5 \leq \lambda \leq 0.8$, the class relaxation scheme is providing a significantly better result for the given simulated image. Figure-5.13(a), (b), (d), and (e) are showing the image mapping results after applying the EPS method. Figure-5.13(c) shows the subspace-based MLR result when EPS relaxation is not applied.

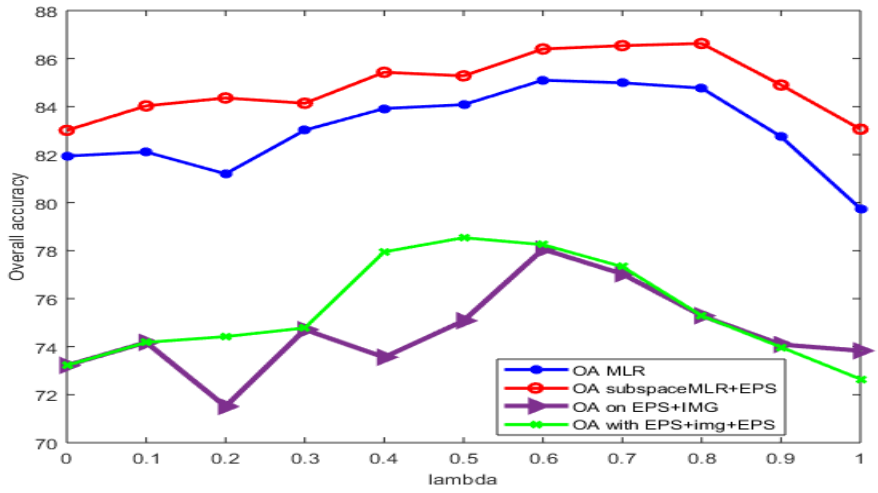


FIGURE 5.14: Overall Accuracy Vs Regularizer(λ)

5.5.2.2 Exercise 2:Effect of error (Err) on iterations(itr):

Figure-5.15 depicts the Error convergence of the subspaceMLR+EPS method, which is a post-processing step, on various values of λ. The proposed method is converging swiftly in 20 iterations to achieve minimum error. The error function is following a similar convergence pattern for every possible value of λ. The convergence error is minimum at a maximum value of λ. Therefore, we have considered λ=0.9 and maximum iterations K_{max}=20 for the rest of the simulated and real HSI experiments.

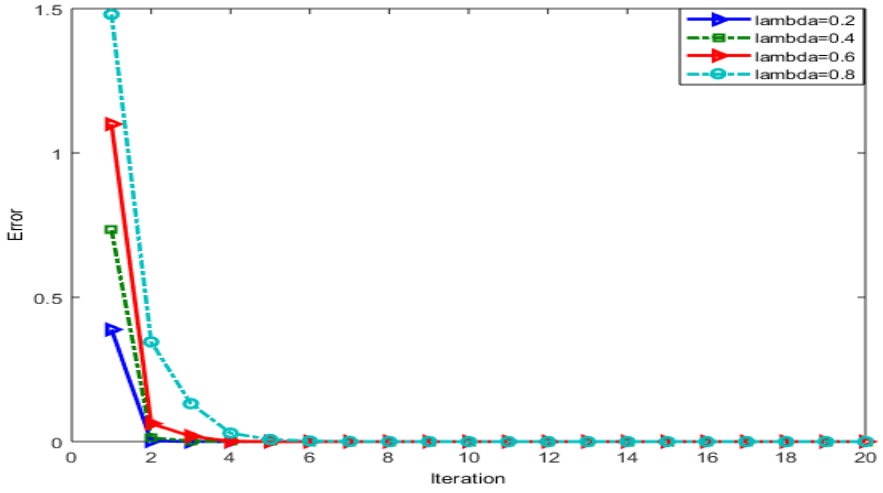


FIGURE 5.15: Error term Vs Number of Iterations

5.5.2.3 Exercise 3:Effect of noise term sigma on OA:

In the third synthetic image experiment, we have derived the noise effect on overall accuracy, as shown in figure-5.16, which depicts the OA results calculated by the proposed schemes using various noise values (σ). From various conclusions that can be drawn out, the foremost is that the careful use of local features of an image for all possible neighbourhood will lead to improving the performance of both subspaceMLR and subspaceMLR+EPS method. It has noticed that, for each value of σ, the overall accuracy of relaxation-based subspaceMLR+EPS methods is superior to subspaceMLR. The efficiency of the proposed scheme decreases as the value of noise σ increases. At low noise values, both approaches are showing better performance than high noise values.

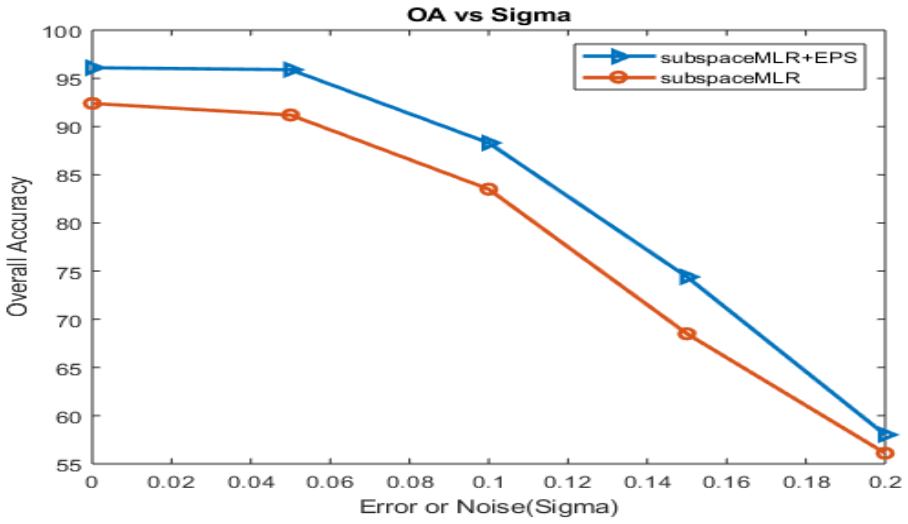


FIGURE 5.16: Overall Accuracy Vs Error term

5.5.3 Experiment 2:Modelling with Real Images Data Set

Three different datasets, collected by AVIRIS and ROSIS sensor, are used to figure out the proposed EPS optimization method for real images. We have first reported the parameters and constants used in this section. Indian Pines(IP) dataset consist of $145 \times 145 \times 224$ pixels with 16 exclusive classes in ground truth. The Pavia University(PU) dataset has a size

of $610 \times 340 \times 103$ pixels with ground truth having nine different classes. In the Salinas Valley (SV) dataset, we have taken a subset of original HSI, which contains the image size as $86 \times 83 \times 204$ pixels with 3.7 meters resolution. The ground truth image contains six exclusive classes. Figure-5.19(g) shows the ground truth, in which we have used 6889 available pixels for training and testing. In Indian Pines(IP) and Salinas Valley(SV) data set, we have chosen random training samples from the available ground truth. The rest of the data has used for validation testing. But, for Pavia University(PU) dataset, we have generated 500 training samples by selecting 60 points per class. Corresponding λ has selected as 0.9, which is very high, to regularize the spatial smoothing. Maximum iteration for convergence K_{max} is 20, in each experiment.

TABLE 5.4: Class-wise and Overall performance of proposed method in Indiana Pines(IP) Dataset

	samples	KCRC	MV	JKCRC	EMARCK	MCK	MCKL	RMCK	EPF	EMP	(subspaceMLR)	(subspaceMLR +EPS)	(EPS+img)	(EPS+img+EPS)	(subspaceMLR+GC)
alfalfa	54	46.67	54.12	45.10	80.20	70.98	86.47	73.92	98.67	92.71	72.22	98.15	98.15	98.15	92.59
corn_no till	1434	70.24	82.25	84.52	88.99	86.14	90.68	88.05	94.04	89.39	50.77	66.39	93.65	92.05	74.69
corn_min till	834	56.49	68.81	74.60	92.78	86.40	92.89	88.85	96.89	83.99	47.12	56.83	93.88	96.04	67.51
corn	234	40.41	71.22	64.64	66.35	68.69	85.32	72.79	95.78	81.79	50.43	76.07	97.86	100	84.62
grass/pasture	497	86.59	90.97	89.70	90.06	90.44	93.35	91.21	97.74	90.78	86.52	95.98	98.19	96.38	90.14
grass/tree	747	95.18	96.50	76.52	97.87	99.13	96.45	99.04	94.32	95.84	94.78	100	98.93	98.26	97.59
grass/pasture-mowed	26	76.25	9.58	0	95.00	93.33	91.25	96.25	100	88.00	92.31	57.69	100	73.08	100
Hay-windrowed	489	98.60	99.78	99.78	99.50	99.70	98.06	99.66	99.07	99.61	98.57	100	100	100	100
oats	20	36.67	10.00	20	68.33	71.11	76.67	74.44	40.00	78.24	95.00	0	100	05.00	100
soyabean_no till	968	65.56	79.82	84.61	88.72	86.38	85.50	95.25	92.88	90.63	62.60	84.50	96.49	97.00	81.82
soyabean_min till	2468	85.79	99.80	99.68	96.66	95.77	92.83	83.88	90.09	92.91	62.32	81.00	92.30	93.64	78.12
soyabean_clean till	614	60.03	94.12	83.02	86.64	80.00	86.74	83.88	97.13	87.83	50.65	58.31	93.81	95.03	58.47
wheat	212	98.31	100	94.38	99.35	99.45	98.41	99.40	99.95	97.45	99.53	100	99.53	99.53	99.53
woods	1294	95.96	99.76	99.77	98.20	99.02	96.27	98.93	95.18	99.43	91.58	99.77	97.84	98.76	98.76
bldg-grass tree	380	36.95	41.94	39.22	90.33	75.76	84.49	79.89	92.55	97.99	47.37	77.37	99.47	100	83.42
stone-steel tower	95	85.11	91.89	91.89	90.33	88.89	88.89	90.67	96.75	98.66	90.53	100	94.74	96.84	100
OA	-	77.57	88.55	88.62	92.99	90.94	91.99	91.91	93.91	92.36	68.09	81.54	95.53	95.79	82.71
AA	-	70.93	74.41	72.96	86.95	86.95	90.27	88.73	92.57	91.58	74.52	78.25	97.18	90.04	87.95

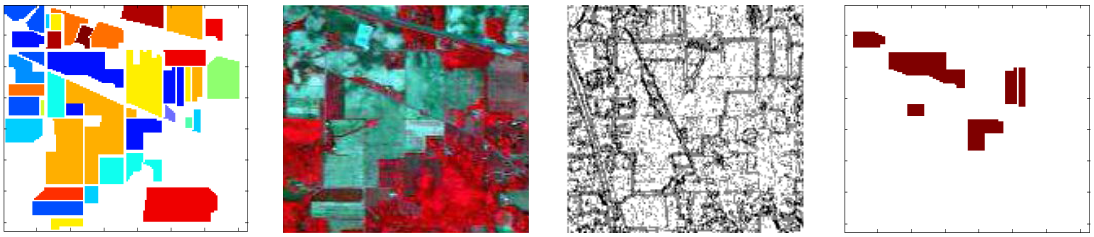
5.5.3.1 Observation 1:On Indiana Pines(IP) Dataset

In this experiment, we have used the Indiana Pines image to validate the efficiency of the proposed method. Table-5.4 shows the classifier accuracies for the image set mentioned above, in which, metrics like OA and AA denotes the overall and average accuracies. This table shows the accuracy of each class in the ground truth, in which we have compared the results with some recently proposed state-of-art methods. The proposed method results have shown on the right side of the vertical line in Table-5.4, and the previous method

results are on the left side. The accuracy, class-wise and overall for comparative methods are borrowed from [131, 144], to compare with our proposed method. From outcomes depicted by Table-5.4, we have concluded that the performance of our proposed method is state-of-art since proposed subspaceMLR+EPS have an accuracy, OA=81.54%, which is outperforming the KCRC(OA=77.57%) method. The proposed relaxation on direct image values, .i.e. EPS+img and EPS+img+EPS, is achieving OA=95.53% and OA=95.79% which are better accuracies than the previously proposed methods [131] for MV(OA=77.57%), JKCRC(OA=88.62%), EMARCK(OA=92.09%), MCK(OA=90.94%), MCKL(OA=91.99%) and RMCK(OA=91.91%). Our proposed method is outperforming another recently proposed work .i.e. [144], in which OA is computed for EPF(OA=93.91%) and EMP(OA=92.36%). It has noticed that for the IP dataset, our proposed method is achieving 2% to 18% higher accuracies than previous methods. In regards to class-wise accuracy shown in Table-5.4, for example, OA obtained by EPS+img+EPS method for Corn, grass/pasture classes are increasing from 95.78%, 97.74% to 98.19%, 96.04% respectively from EPS method, i.e., a previous method. A similar conclusion can be drawn by comparing individual accuracy for other classes; three of them, namely grass/pasture, Hay-windrowed, and oats, achieve 100% accuracy in the EPS+img method. Figure-5.17(a), 5.17(b) are ground truth, and false-colour composite images, respectively. Figure-5.17(c) represents the edge features generated by the gradient-based Prewitt filter. Figure-5.19 represents the images generated by the proposed methods. Figure-5.19(a),(c),(e),(g) represents the labelled map image, which exhibits a good quality of boundary delineation for classes. Figure-5.19(b),(d),(f),(h) depicts the results in a mask of available labelled classes. Figure-5.18 emphasizes the change in the posterior probability of the pixels for *corn_no_till* class, in IP dataset. In the same way, other class probability can also be shown. With the results shown in figure-5.19(c),(e),(g), we can wind-up by the conclusion that our proposed scheme is smoothing the far border regions and conserving the local structure during label prediction.

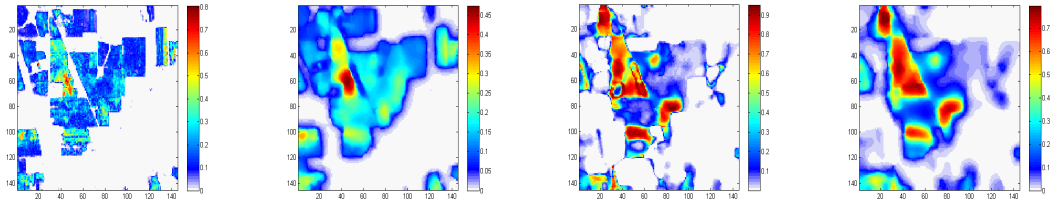
5.5.3.2 Observation 2:On Pavia University(PU) Rosis Sensor dataset

Table-5.5 analyses the proposed and past method results acquired for Pavia University(PU) dataset, from which various conclusions can be drawn. In this method, we have used only 500 samples for training and rest for testing. We have compared the



(A) Ground Truth (B) False colour image (C) Local feature Map (D) Class-Corn-no till

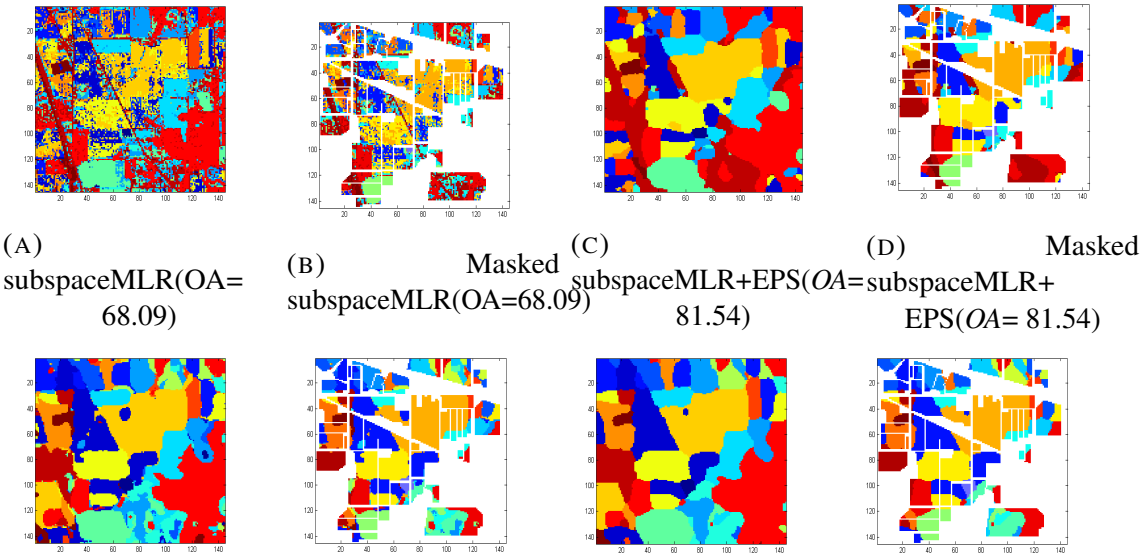
FIGURE 5.17: (a)Ground truth (b)False color image (c)Feature map (d)One class map, for Indian Pines Image



(A) subspaceMLR ($OA_{Corn-notill}$= 50.57) (B) subspaceMLR+Eps ($OA_{Corn-notill}$= 66.39) (C) Eps+im ($OA_{Corn-notill}$= 93.65) (D) subspaceMLR ($OA_{Corn-notill}$= 92.05)

FIGURE 5.18: Class probability image for Corn-no till class in 4 cases for Indian Pines Image

proposed method results from [141, 131, 144], which also shares the classification results for limited samples of PU dataset. It can be observed that our proposed method, subspaceMLR+EPS, is achieving an accuracy of 75.00%, which is better than the accuracy obtained in RASP-ASSSC (OA=40. 90%), KASP-ASSC(OA= 31.32%), LI-ASSC(OA=49.29%), LLE-ASSSC(OA= 51.51%) and LI-ASP(OA=28.78%). The KCRS (OA=79.16 %) method achieves almost similar accuracy to our proposed method .i.e. subspaceMLR+EPS. In this case, only SVM(OA= 80.99%) is performing better than our proposed method. The subspaceMLR+GC(graph-cut) method is providing 93.11% accuracy, which is 13% higher than SVM and 14% higher than the KCRC method. The SVM results are also showing better accuracy .i.e 80.99% than proposed subspaceMLR+EPS, EPS+img, and EPS+img+EPS methods for PU dataset. In this dataset, the edge relaxation on the image is also showing improvement, but probability-based relaxation is achieving better outcomes than direct image relaxation. Therefore, we have concluded that the relaxation approach is improving the classification output significantly. First and foremost, it is remarkable that the proposed relaxation approach exhibited excellent performance using small training samples. Figure-5.20 is



(A) subspaceMLR(OA= 68.09) (B) Masked subspaceMLR(OA=68.09) (C) subspaceMLR+EPS(*OA*= 81.54) (D) Masked subspaceMLR+ EPS(*OA*= 81.54)

(E) EPS+ img (*OA*= 93.53) (F) Masked EPS+img (*OA*= 93.53) (G) EPS+img+EPS (*OA*= 95.79) (H) Masked EPS+img+ EPS(*OA*= 95.79)

FIGURE 5.19: Classification Results and Accuracy(in%), obtained For Indian Pines

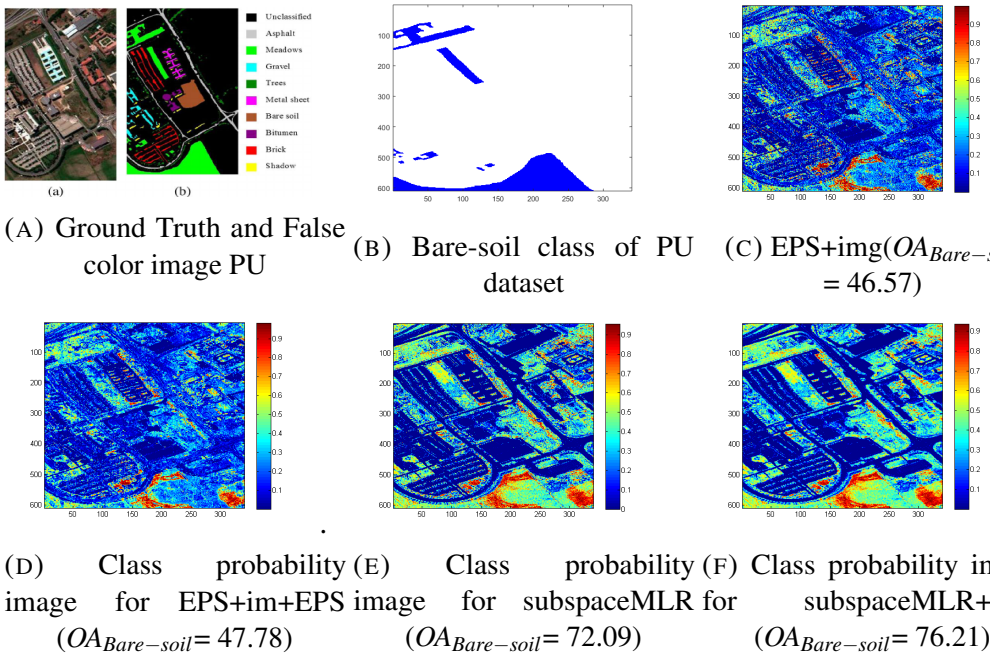
showing the classifier outcome of proposed methods. Figure-5.21 depicts the variation in posterior probability for Bare-soil class.

TABLE 5.5: Class wise and Overall performance of proposed method in Pavia University (PU) Dataset

	samples	RASP-ASSC	KASP-ASSC	LI-ASSC	LLE-ASSC	LI-ASP	KCRC	SVM	(subspaceMLR)	(subspaceMLR+EPS)	(EPS+img)	(EPS+img+EPS)	(subspaceMLR+GC)
asphalt	6631	44.78	44.87	73.74	77.86	29.23	65.80	83.71	61.68	67.98	45.27	49.49	92.05
Bare soil	18649	89.97	89.98	86.41	81.82	54.44	79.53	92.25	72.09	76.21	46.57	47.78	99.44
Bitumen	2099	0	0	0.01	0	2.36	75.53	81.58	44.40	46.45	56.17	60.60	45.55
Bricks	3064	63.29	63.50	96.00	54.24	27.71	91.04	92.59	73.86	75.49	78.56	79.99	82.31
Gravel	1345	95.17	94.31	30.66	98.65	57.60	98.97	70.32	99.03	99.55	95.17	96.13	99.93
Meadows	5029	22.34	22.57	24.78	32.21	15.26	79.39	70.25	68.28	74.35	71.74	76.50	98.85
Metal sheets	1330	0	0	0	0	0	0	93.42	69.10	74.21	72.63	78.87	62.78
Shadows	3682	0.2	19.69	50.73	49.65	7.5	75.35	96.62	77.43	83.24	52.25	57.58	97.96
Trees	947	14.11	14.65	95.91	30.26	100	99.29	92.41	99.37	99.79	53.43	55.86	99.89
OA	-	40.90	31.32	49.29	51.51	28.78	79.16	80.99	70.61	75.00	55.07	57.87	93.11
AA	-	40.90	31.32	49.29	51.51	28.78	84.20	88.28	73.92	77.48	63.53	66.98	86.53

5.5.3.3 Observation 3:On Salinas Valley (SV) Dataset

In this section, we have experimented on the third AVIRIS sensor image data-set .i.e. Salinas Valley(SV) data-set. We have used 15% data for training and a complete data-set



(A) Ground Truth and False color image PU

(B) Bare-soil class of PU dataset

(C) EPS+img($OA_{Bare-soil}$ = 46.57)

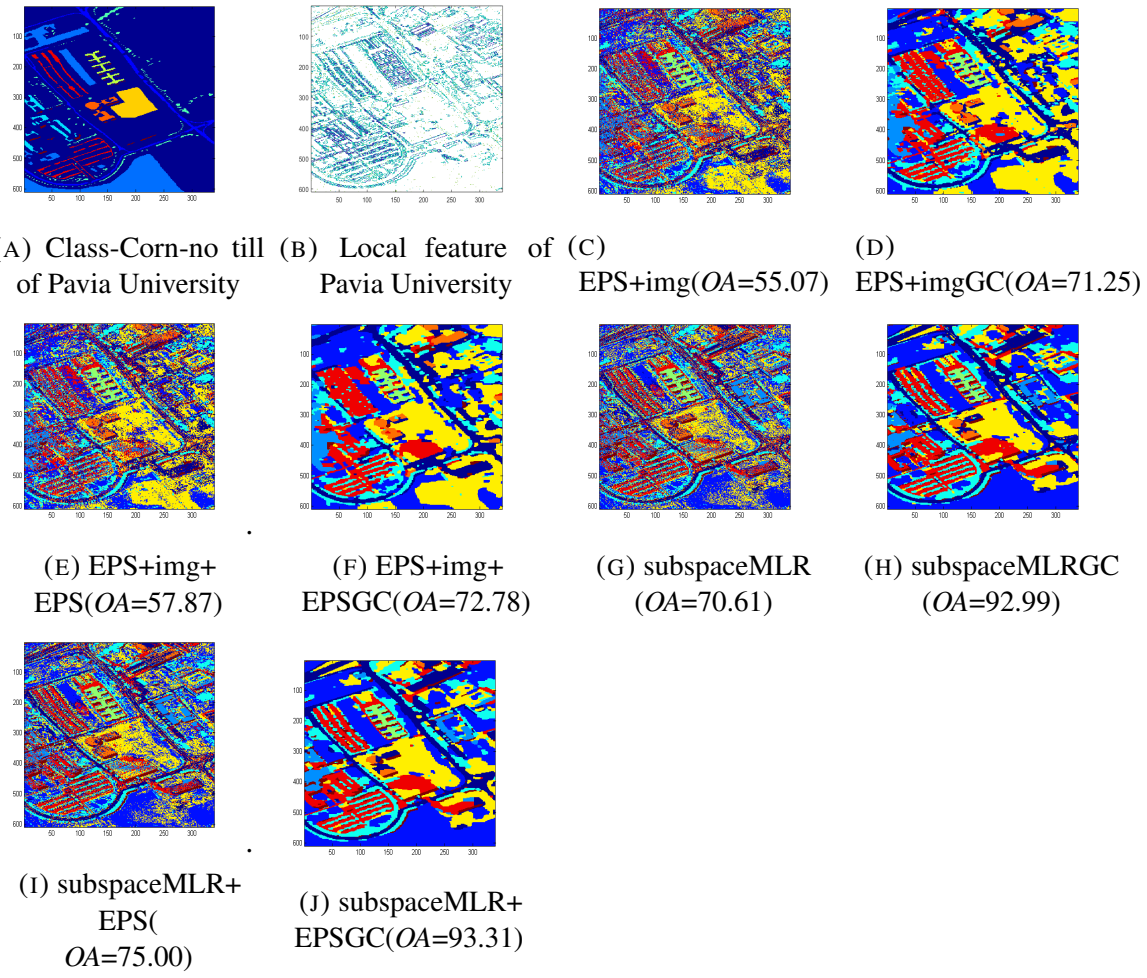
(D) Class probability image for EPS+im+EPS ($OA_{Bare-soil}$= 47.78)

(E) Class probability image for subspaceMLR ($OA_{Bare-soil}$= 72.09)

(F) Class probability image for subspaceMLR+EPS ($OA_{Bare-soil}$= 76.21)

FIGURE 5.20: Class probability image for Bare-soil class of Pavia university dataset

for validation testing. Table-5.6 shows the OA and AA, which is the overall and average accuracy obtained by our proposed method. In this experiment, we have observed that the accuracies achieved by our proposed relaxation-based subspaceMLR+EPS method are remarkably high, .i.e. OA=99.10%. This method is obtaining 6% higher accuracy than the subspaceMLR method, which has OA=92.40%. Direct image-based feature fusion methods .i.e. EPS+img and EPS+img+EPS are also resulting in improvement of OA than subspaceMLR. The proposed EPS+img+EPS and EPS+img method have achieved OA=95.10% and 91.15%, respectively. In class-wise accuracy, the subspaceMLR+EPS method is concluded as a state-of-art method for Salinas Valley data-set, since in each class, it has achieved greater than 99% accuracy. Figure-5.22(a) shows the edge base feature image, and Figure-5.22(b) shows the class2 image for which probability images have been shared. Figure-5.22 (h) to (l), shows the results obtained by our proposed method, whereas, figure-5.22(c) to (f) has depicted the posterior probability distribution of the class2 .i.e. "Corn_senesced_green_weeds". We have compared the proposed method results with recently proposed work [49, 130, 58, 53, 140], for SV datasets. In the proposed work, we have used the subset of the original Salinas Valley image(SV), whereas, in [49], the complete image has been observed. The proposed method subspaceMLR+EPS is achieving 7% greater accuracy than LBP-ELM, 6% greater than SVM-RFS, ten % greater



(A) Class-Corn-no till of Pavia University (B) Local feature of Pavia University (C) EPS+img(*OA*=55.07) (D) EPS+imgGC(*OA*=71.25)

(E) EPS+img+ EPS(*OA*=57.87) (F) EPS+img+ EPSGC(*OA*=72.78) (G) subspaceMLR (*OA*=70.61) (H) subspaceMLRGC (*OA*=92.99)

(I) subspaceMLR+ EPS(*OA*=75.00) (J) subspaceMLR+ EPSGC(*OA*=93.31)

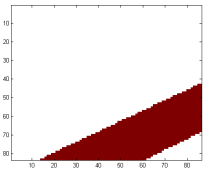
FIGURE 5.21: Classification results for Pavia university dataset

than 1-D CNN, 5% greater than CNN-PPF and almost similar accuracy to spatial-spectral optimization(SSO) method proposed in [49].

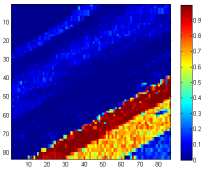
5.5.3.4 Overall performance on IP, PU, and SV datasets

In this section, we have compared the performance of our proposed method in terms of Overall Accuracy(OA) and Average Accuracy(AA). The OA has obtained by computing the fraction of correctly predicted samples for the total number of samples in that class. AA is a simple arithmetic average of correctly predicted samples in each class. Figure-5.23(a) and 5.23(b) has denoted the OA and AA for IP, PU, SV datasets to validate our EPS based proposed method. In figure-5.23(a), the EPS+img+EPS method has achieved maximum

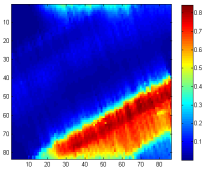
(A) Feature map for Salinas valley dataset



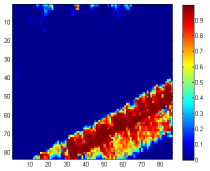
(B) Class2(Corn _senesced_ green_weeds)



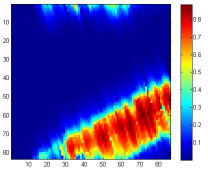
(C) Class probability for subspaceMLR (OA_{class2}= 91.59)



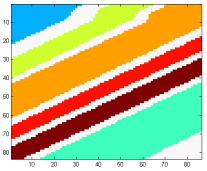
(D) Class probability for pps+im (OA_{class2}= 98.06)



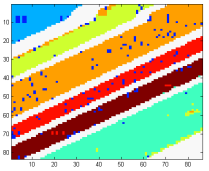
(E) Class probability for subspaceMLR (OA_{class2}= 89.72)



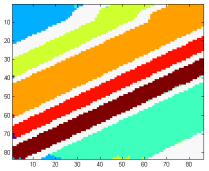
(F) Class probability for subspaceMLR (OA_{class2}= 91.59)



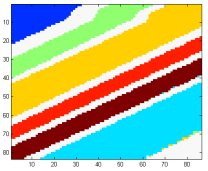
(G) Ground Truth(6 class) Salinas valley dataset



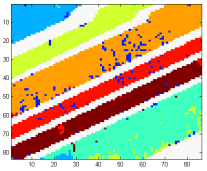
(H) Class probability for subspaceMLR (OA= 93.77)



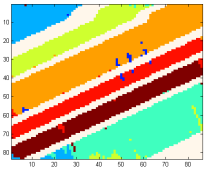
(I) Class probability for subspaceMLR+EPS (OA= 99.25)



(J) Class probability for subspaceMLRGC (OA= 99.73)



(K) Class probability for EPS+img(OA= 94.74)

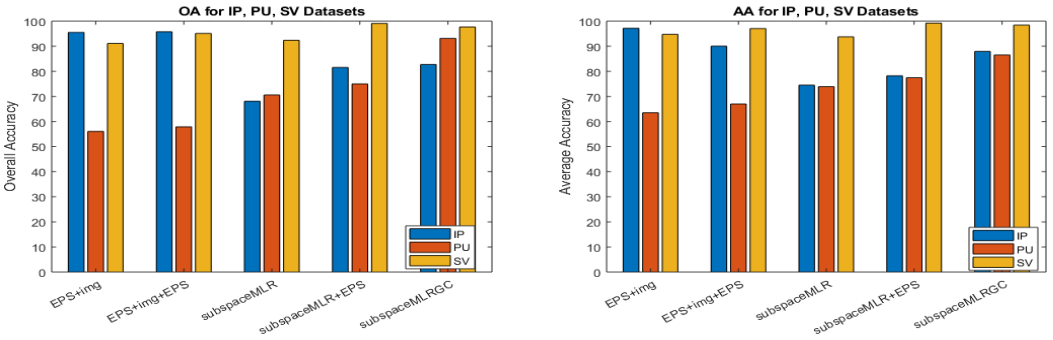


(L) Class probability for EPS+img+EPS (OA= 97.04)

FIGURE 5.22: (a)-(f):Class probability images for class2 of Salinas Data, and (g)-(l):Classification results for Salinas valley Dataset

TABLE 5.6: Class wise and Overall performance of proposed method in Salinas Valley (SV) Dataset

	samples	LBP-ELM	SVM-RFS	1-DCNN	CNN-PPF	SSO	(subspaceMLR)	(subspaceMLR+EPS)	(EPS+img)	(EPS+img+EPS)
Brocoli_green_weeds_1	2009	99.75	99.55	97.34	100	100	94.41	99.49	98.21	99.49
Corn_senesced_green_weeds	3278	95.03	96.49	91.29	98.60	99.35	95.91	98.06	89.72	91.59
Lettuce_romaine_4wk	1068	96.82	98.78	97.58	99.54	100	88.67	99.68	98.86	98.54
Lettuce_romaine_5wk	1927	100	100	100	100	99.94	95.31	99.87	91.15	97.77
Lettuce_romaine_6wk	916	98.25	99.13	99.02	99.44	99.44	93.42	99.41	94.51	98.22
Lettuce_romaine_7wk	1070	97.94	98.97	95.05	98.96	99.08	94.92	99.00	95.99	96.62
Brocoli_green_weeds_2	3726	99.87	99.92	99.29	99.88	100	-	-	-	-
Fallow	1976	99.60	99.44	99.51	99.60	99.72	-	-	-	-
Fallow_rough_plow	1394	99.64	99.86	99.66	99.49	99.67	-	-	-	-
Fallow_smooth	2678	98.81	98.02	96.97	98.34	99.52	-	-	-	-
Stubble	3959	99.67	99.70	99.60	99.97	99.84	-	-	-	-
Celery	3579	99.66	99.69	99.49	100	99.97	-	-	-	-
Grapes_untrained	11271	84.04	84.85	72.25	88.68	98.95	-	-	-	-
Soil_vinyard_develop	6203	99.89	99.58	97.53	98.33	100	-	-	-	-
Vinyard_untrained	7268	72.96	76.38	76.83	83.53	99.49	-	-	-	-
Vinyard_vertcal_trellies	1807	99.06	99.56	98.94	99.31	100	-	-	-	-
OA	-	92.42	93.15	89.28	94.80	99.58	92.40	99.10	91.15	95.10
AA	-	96.31	96.87	95.02	97.10	99.70	93.77	99.24	94.74	97.04



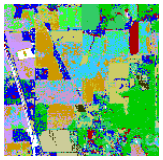
(A) Overall Accuracy for IP, PU, SV Datasets (B) Average Accuracy for IP, PU, SV Datasets

FIGURE 5.23: Overall and Average Accuracy for IP, PU, SV Datasets

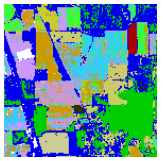
OA, .i.e. 95.79% for IP dataset. subspaceMLRGC and subspaceMLR+EPS methods have achieved the maximum OA=93.11% and OA=99.10% for PU and SV datasets, respectively. In figure-5.23(b), we have observed that EPS+img, subspaceMLRGC, subspaceMLR+EPS methods have obtained greater average accuracies, .i.e. 97.18%, 86.53% and 99.24%, than other methods. It was found that the EPS method had achieved greater accuracy in PU and SV datasets when combined with probability values. In the IP dataset, the EPS method is achieving better accuracies when combined with direct image values.

5.5.4 Contrast Experiments

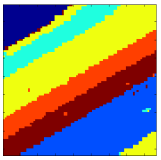
In this section, the contrast experiments have been performed to compare the proposed method. Hysime+LDA and Hysime+QDA are applied to investigate the classification performance. Hysime [122, 126] is a minimum error based subspace projection method that removes the noise and projects the subspace of HD image. We have applied an affine projection on projected subspaces to obtain original size with preserved information. The mean square error, projection error, and noise power plot are shown in figure-5.24-(g), (h), (i). The classified images are shared in figure-(5.24)-(a) to (f). Table-5.7 is showing the Hysime+LDA, Hysime+QDA, Mlr_BREG [127], SMLR_SpTV [81, 137], KMNF, OMNF, OKMNF. In this section, our proposed method has outperformed in terms of overall accuracy than the previously proposed methods. The proposed method achieves 95.79%, 93.11%, and 99.10% accuracy for IP, PU, SV dataset, which is significantly higher than the comparative methods, present in this section. In Table-5.7, we have mentioned the results of previous methods for comparative analysis with our proposed method.



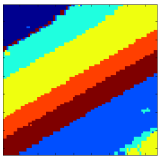
(A) Hysime+LDA for IP (OA= 67.28%)



(B) Hysime+QDA for IP (OA= 74.82%)



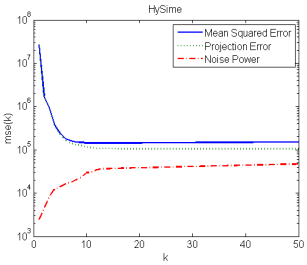
(C) Hysime+LDA for SV (OA= 98.89%)



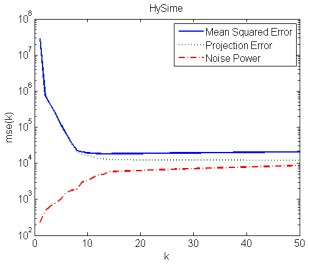
(D) Hysime+QDA for SV (OA= 99.24%)

(E) Hysime+LDA for PU (OA= 64.91%)

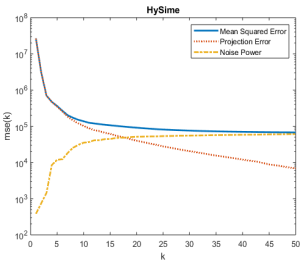
(F) Hysime+QDA for PU (OA= 57.77%)



(G) Mean Square Error vs Dimensions for Hysime on IP



(H) Mean Square Error vs Dimensions for Hysime on SV



(I) Mean Square Error vs Dimensions for Hysime on PU

FIGURE 5.24: Hysime+LDA and Hysime+QDA results on IP, PU, SV Datasets [(a)to(f)] and MSE, Projection error and noise power in Hysime projection [(g)to(i)]

TABLE 5.7: Overall accuracy for Inidain pines and pavia dataset for past methods

Data	Acc.	Hysime _LDA	Hysime _QDA	Mlr _BREG	SMLR _SpTV	KMNF	OMNF	OKMNF
Indiana Pines	OA	67.28%	74.82%	76.35%	76.10%	80.30%	86.82%	87.06%
	AA	69.14%	63.24%	86.22%	86.92%	-	-	-
Pavia Univ.	OA	78.32%	73.08%	76.20%	78.43%	88.25%	92.13%	92.55%
	AA	81.09%	81.09%	83.86%	87.05%	-	-	-
Salinas Valley	OA	98.89%	99.24%	-	-	-	-	
	AA	82.38%	82.81%					

5.6 Conclusion

In this chapter, we have proposed two approaches in the proposed method as well as experimental analysis one and two sections, for the fusion of subspace regression and cluster label optimization-based cluster energy minimization and an EPS based framework for relaxation-based classification in HD data cube images. In experiment-1, The image labelling problem has divided into a two-step framework. In the first step, we have transformed the HD image data in suitable subspaces and applied Multinomial regression,

resulted in a classified image. In the second step, we explored a two-phase approach as:(1) a continuous label update with minimum class energy and (2) clustering by expansion combinatorial graph cut method. Further in experiment-2, Firstly, A fast subspace-based classification scheme for HD images was implemented based on the pixel-wise logistic regression classification. Secondly, gradient-based adaptive local neighbourhood features information has integrated with pixel-wise posterior results of classification. Such a scheme encourages the neighbourhood values to reside in the same labelled class and improves the classification posteriors learned by the subspaceMLR method. The combined exploitation of spatial-spectral knowledge is entitled to increase the classifier result. The proposed frameworks have resulted in better accuracies as compared to previously proposed approaches. Calibration of various parameters has also described in the synthetic image and real image application part, which is also advocating the improvement in the proposed scheme.