A Survey on Trends and Techniques used in Hyperspectral Image Processing

Sonia Sarmah
Assam Don Bosco University
Assistant Professor, Dept. of CSE &IT
School of Technology, Assam

Sanjib Kr. Kalita, PhD
Gauhati University
Assistant Professor, Dept. of Computer Science,
Guwahati, Assam

ABSTRACT

Recent advancement in remote sensing facilitates collection of hyperspectral images (HSIs) in hundreds of bands which provides a potential platform to detect and identify the unique trends in land and atmospheric datasets with high accuracy. But along with the detailed information, HSIs also pose several processing problems such as- 1) increase in computational complexity due to high dimensionality. So dimension reduction without losing information is one of the major concerns in this area and 2) limited availability of labeled training sets causes the ill posed problem which is needed to be addressed by the classification algorithms.

Initially classification techniques of HSIs were based on spectral information only. Gradually researchers started utilizing both spectral and spatial information to increase classification accuracy. Also the classification algorithms have evolved from supervised to semi supervised mode. This paper presents a survey about the techniques available in the field of HSI processing to provide a seminal view of how the field of HSI analysis has evolved over the last few decades and also provides a snapshot of the state of the art techniques used in this area.

General Terms

Classification algorithms, image processing, supervised, semi supervised techniques.

Keywords

Hyperspectral images (HSIs), remote sensing, dimension reduction, ill-posed problem,.

1. INTRODUCTION

Hyperspectral imaging collects information across hundreds of spectral bands (which extend beyond the visible spectrum as well) with relatively narrow bandwidths. With such a detail, HSIs provide ample spectral information to detect and identify unique trends in land and atmospheric datasets with more accuracy than is possible with other types of remotely sensed data. For example the Airborne Visible-infrared Imaging Spectrometer (AVIRIS) technology collects image data in 220-224 bands with 10nm spectral resolution and spectral coverage 0.4-2.5 μm. But this increased number of bands provides challenges to the current techniques for analyzing such data. For example, the high dimensionality of the datasets causes computational complexities. So dimension reduction without losing information is an important issue. Another important issue is the collection of labeled training samples, which requires expensive ground campaign. This results in high number of spectral bands and relatively low number of labeled training samples which poses the well-known Hughes phenomenon [1].

Conventional HSI analysis techniques have focused on spectral properties only, without incorporating information on the spatially adjacent data. Algorithms were proposed both to reduce the dimensionality of the datasets and to take advantage of highly correlated spectrally adjacent bands to exploit the information in the datasets effectively. However, towards the end of last decade, researchers have integrated analysis of both spatial and spectral information simultaneously and have achieved substantially improved accuracy [2]-[5]. Kernel based methods such as Support Vector Machines (SVMs) have also been widely used for HSI classification [6], [7]. Kernel Fisher Discriminant (KFD) analysis is another interesting and effective method in this category [8]. Gabor wavelets have also been introduced to hyperspectral region classification in [9]-[11]. Multinomial logistic regression [12] is one of the state of the art discriminative techniques in this area of research.

The rest of this paper is organized as follows. Section 2 presents some of the classification techniques used in hyperspectral images. A brief conclusion and discussion is given in Section 3.

2. CLASSIFICATION TECHNIQUES USED IN HYPERSPECTRAL IMAGES

2.1 Band selection method

Each HSI scene is represented by an image cube where the third dimension represents the spectral range. This creates enormous data for computer processing and transmission. In literature many criteria such as divergence [13], have been introduced for band selection to find the bands which are crucial and significant in terms of information conservation. Chien-I Chang et al. in 1999, proposed a band selection method that works in two phases - 1) band prioritization phase, where bands were assigned priority according to the contained information based on eigenanalysis method and 2) band decorrelation phase, where divergence was used to decorrelate prioritized bands. Finally the band selection was done by an eigenanlysis-based band prioritization in conjunction with the divergence based band decorrelation method [14].

For the first phase, i.e, band prioritization phase total four criteria were proposed, out of which two were Principal Component Analysis (PCA) based criteria and remaining two were classification based criteria. The two PCA based methods used two different transformations which resulted in two different loading matrices. From these matrices two different criteria for band prioritization were derived. The Maximum Variance PCA (MVPCA) transformed the data coordinates in such a fashion that the first principal component is along the maximum variation. In maximum Signal to Noise Ratio PCA (MSNRPCA), principal

components were evaluated by Signal to Noise Ratio (SNR) rather than variance. The classification based criteria included the minimum misclassification canonical analysis (MMCA), which was derived from Fisher discriminant function and orthogonal subspace based projection criteria (OSP) is another classification based criteria[15],[16]. After prioritization of all the bands, divergence based band decorrelation was further used to remove redundant and insignificant bands. If the divergence between two bands fell below the prescribed threshold, then the band with lower priority was removed. Fig. 1 depicts pictorial representation of the devised method. The experiment was performed on a HYDICE (Hyperspectral Digital Imagery Collection Experiment) image, which is a radiance data collected using 210 bands with 10nm spectral resolution and spectral coverage 0.4-2.5 µ. The experiment results showed that not more than 12 bands were needed for the proposed band selection method. So the proposed band selection method could effectively reduce band dimensionality with very little loss of information in HSI classification. Table 1 shows the result of the experiment in terms of total number of bands required and band power ratio for the four approaches- MVPCA, MSNRPCA, MMCA and OSP. Here band power ratio is the classification power of the selected band subsets divided by that of all bands being used.
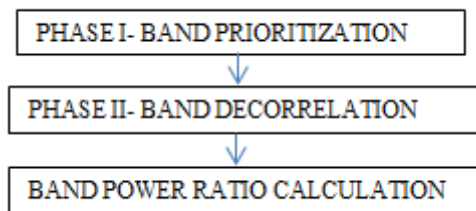

Fig.1: Steps in band selection method

Table 1. Experimental results of band selection method

Method	MVPCA	MSNRPCA	MMCA	OSP
Number of Bands	10	11	10	12
Band Power Ratio	0.0896	0.2260	0.2859	0.1400

Shailesh Kumar et al. in 2001 presented another algorithm which took into account the fact that adjacent bands are generally correlated [17]. They proposed a set of best-bases top down and bottom up algorithms for classification of HIS data. The top down algorithms recursively partitioned the bands into two sets of bands and replaced each final set of bands by its mean value. In bottom up algorithms an agglomerative tree was built by merging highly correlated adjacent bands and projected them onto their Fisher Direction, yielding high discrimination among the classes. Both these algorithms were used in a pair wise classifier framework where the original C class problem was divided into a $\binom{c}{2}$ class problem. Bayesian pair wise classifier framework [18],[19] as shown in Fig. 3 was used to decompose a C class problem into a set of $\binom{c}{2}$ class problem for all pairs (wi,wj), $1<i<j<C$. Each of these two class problems were solved independently and their results were combined to obtain the final result for the original C class problem. The experiment was carried out on AVIRIS Kennedy Space Center (KSC) dataset. The image consisted of 224 spectral bands but the algorithms were applied only to 183 bands excluding the water absorption band. The experiment results showed

improvements of 5% to 10% accuracies using smaller number of features relative to the other techniques like Local Discriminant Bases (LDB) proposed by Coifman and Wickerhauser [20], extended by Saito and Coifman [21] and Segmented Principal Components Transformation (SPCT) proposed by Jia and Richards [22] .
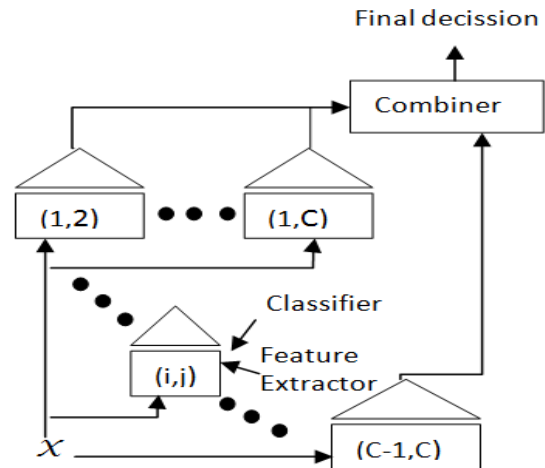

Fig.2: Pairwise classifier architecture: (C, 2) pairwise classifier with respective feature selectors

The comparison of the results is listed in the Table 2. Quantity inside [ ] represents standard deviation over 66 pairwise classifiers and quantity inside ( ) represents standard deviation across ten samples. OA stands for Overall Accuracy.

Table 2. OA (in %) for various best bases algorithms using pairwise classifier framework

	GLD B-BU	GLDB-TD	SPCT	LDB
Test Accuracy	99.4 [1.31]	97.52 [4.20]	93.26 [4.40]	93.53 [4.14]
No. of features	1.2 [0.42]	2.67 [0.73]	10.57 [2.42]	7.11 [1.88]
OA	95.3 (2.42)	86.3 (3.56)	78.7 (3.47)	79.3 (4.12)

2.2 Kernel based methods

Survey reveals that many supervised algorithms for HSI classification have been proposed by various researchers using Artificial Neural Network (ANN), Radial Basis Function Neural Networks (RBFNNs) [23]-[25]. But one of the major issues in HSI dataset is high number of spectral bands and relatively low number of labeled training samples, which poses the well-known problem of the curse of dimensionality referred to as Hughes phenomenon [1] in literature. To tackle this problem researchers switched to the Kernel based methods, such as Support Vector Machines (SVMs) or Kernel Fisher Discriminant [KFD] analysis and were able to get excellent performance in terms of accuracy and robustness. The main reason behind this success was the ability of kernels to handle large input spaces efficiently, to work with low numbered labeled training samples and to deal with noisy samples in a robust way [6],[7]. Apart from SVMs and KFD, a number of other powerful kernel based learning classifiers such as Support Vector Clustering (SVC) [26], regularized AdaBoost (Reg-AB) algorithms [27] were proposed and successfully applied in the field of HSI classification. In 2005 Gustavo Camps-Valls and Lorenzo Bruzzone in their work [28], showed the comparison of

accuracy obtained by various kernel based methods by carrying out experiments on a portion of an AVIRIS Indian Pine dataset. Total nine classes were used to generate 4757 training samples and 4588 test samples. Table 3 summarizes the experimental results after a coarse feature selection (200 input bands) and results in the original (noisy) dataset (220 input bands). The result was expressed in terms of OA and kappa statistic (κ) in the test set for different kernel classifiers - Linear Discriminant Analysis (LDA), Regularized RBF neural network (REG-RBFNN), SVMs with RBF kernel (SVM-RBF) and with polynomial kernel (SVM-POLY), and Kernel Fisher Discriminant (KFD) analysis (with RBF kernel).

They claimed that SVMs revealed excellent results in terms of computational cost, accuracy, robustness to common level of noise (i.e., Gaussian, uniform or impulsive), and ensures sparsity. The only drawback was that they could not provide probabilistic output directly. Reg-AB showed almost comparable result to those offered by SVMs working efficiently with low number of labeled samples. KFD exhibited good accuracies in normal situations, but were in average inferior to those obtained with SVMs and Reg-AB. Regularized Radial Basis Function Neural Network (Reg-RBFNN) [29] offered an acceptable tradeoff between accuracy and computational cost.

In 2006 Gustavo Camps-Valls et al. presented composite kernel machines to enhance classification by taking into account spectral, spatial and local cross information in HSIs [30]. Authors used the properties derived from Mercer's conditions by which a scaled summation of (positive definite) kernel matrices are valid kernels, which provided good results in other domains [31], [32]. The experiments were carried out using different classifiers (some were spectral classifiers only and some were composite classifiers) to compare the results. These classifiers were- 1) Spectral Kernel (Kw) which uses only spectral features, 2) Contextual Kernel (Ks) which uses only spatial features, 3) stacked kernel (K{s,w}) in which feature vectors were built from the concatenation of spectral and spatial features using the stacked feature approach, 4) Direct Summation Kernel (Ks+ Kw), which is a simple composite kernel combining spectral and spatial features, 5) Weighted Summation Kernel (μKs+(1-μ)Kw) ,where μ is a positive real-valued free parameter(0< μ<1), which is tuned in the training process and constitutes a tradeoff between the spatial and spectral information to classify a given pixel which is a composite kernel that balances spatial and spectral content 6) cross information kernel classifiers (Ksw, Kws) which took into account the cross relationship between spatial and spectral information. In addition to that, two standard methods were also included for base line comparison: bLOOC +DAFE+ECHO, which use both contextual and spectral information and Euclidean classifier [33] which only uses spectral information. They used the AVIRIS Indian Pine dataset for evaluating the proposed method. First the experiments were carried out using only a part of the scene, called the subset scene, consisting of 68X86 pixels and containing four labeled classes. Then the whole scene, consisting of the full 145X145 pixels, containing 16 classes, was used for conducting the experiments. In both the datasets, 20% of the labeled samples were used for training and the rest for validation. Table 4 summarizes the experimental results of the classifiers in terms of OA. It was shown that all kernel based methods produced better results and the composite kernel classifiers improved the results obtained by the usual spectral kernel by 11% in the whole scene and by 4% in the subset scene.

Table 3.Experimental results for various kernel based methods

METHOD	220 input bands		200 input bands	
	OA[%]	κ	OA[%]	κ
LDA	82.08	0.79	82.32	0.79
Reg-RBFNN	91.39	0.90	88.36	0.86
SVM-RBF	94.31	0.93	91.34	0.90
SVM-Poly	94.44	0.93	91.74	0.90
KFD	91.54	0.90	90.69	0.89
Reg-AB	93.50	0.92	90.65	0.89

Table 4. OA[%] for various spectral and composite classifiers

CLASSIFIERS	Subset Scene	Whole Scene
Spectral classifiers	OA[%]	OA[%]
Euclidian	67.43	48.23
bLOOC +DAFE+ECHO	93.50	82.91
Kw	95.90	87.30
Kw developed	95.10	88.55
Spatial-spectral classifiers	OA[%]	OA[%]
Mean		
Ks	93.44	84.55
K{s,w}	96.84	94.21
Ks+ Kw	97.12	92.16
μKs+(1-μ)Kw	97.43	95.97
Ks+ Kw+ Ksw+ Kws	97.44	94.80
Mean and standard deviation		
Ks	94.86	88.00
K{s,w}	98.23	94.21
Ks+Kw	98.26	95.45
μKs+(1-μ)Kw	98.86	96.53

2.3 Semi supervised learning methods

In 2006, Lorenzo Bruzzone et al. in their work [34], proposed a semi supervised classification method that exploited both labeled and unlabeled samples to address the ill posed problem (also known as Hughes phenomena), which occur mainly due to small size training sets which can force the over fitting or under fitting of learning algorithms. They used semi supervised SVMs known as transductive SVMs(TSVMs)[35][36], an iterative algorithm which gradually search for a reliable separating hyper plane with transductive processes that incorporates both labeled and unlabeled samples in training phase. The proposed TSVM classifier merged the advantages of semi supervised methods with those of kernel based methods. Experiment was carried out on Landsat 5 Thematic Mapper image. In order to simulate ill posed classification problems, a random sub sampling strategy was applied. From 4549 original training patterns, experiments with 10,20,30,40, 50 and 100 training samples were designed. In all the datasets, seven features and six land-cover classes were considered for analysis. They have claimed the experimental results as summarized in Table 5, which shows the average increase of the OA provided by the proposed TSVM, with respect to standard ISVM for all ten realizations of the training sets made up of 10, 20, 30, 40, 50, and 100 samples. The average increase in the accuracy (avg) and the standard deviation (std dev) for the ten realizations of each different size were also reported.

Table 5. Average increase of OA in TSVM with respect to standard ISVM

Dataset	Size 10	Size 20	Size 30	Size 40	Size 50	Size 100
1	10.68	7.64	2.67	0.89	3.76	3.19
2	4.73	5.18	3.36	4.21	7.60	1.01
3	12.33	9.66	3.68	2.18	4.0	4.73
4	3.64	7.89	-0.49	8.86	5.90	0.08
5	10.23	3.11	3.11	1.86	2.63	0.67
6	10.76	3.03	6.19	1.90	5.54	-0.20
7	11.24	11.16	6.19	9.50	1.05	1.94
8	9.26	6.07	3.80	1.17	2.26	2.14
9	3.84	8.57	15.04	1.94	3.32	0.53
10	7.84	9.10	9.66	3.40	1.33	0.49
Avg	8.46	7.14	5.32	3.59	3.74	1.46
Std. dev	3.26	2.74	4.34	3.11	2.10	1.55

2.4 Multinomial logistic regression based methods

Sparse Multinomial Logistic Regression [37] is another technique which was introduced in the field of HSI processing. The multinomial logistic regression method has the advantage of learning the class distributions by themselves. In 2009, Jun Li et al. presented a semi supervised classification and segmentation approach based on sparse multinomial logistic regression [38]. They considered both the spectral and spatial information to enhance the classification result. The contextual information inherent to the spatial configuration of the image pixels was modeled by Multi Level Logistic (MLL) Markov- Gibbs random field. The algorithm was implemented in two main steps- 1) a learning step to infer the class distributions, and 2) segmentation by inferring the labels of posterior distribution built on the learned class distributions and on MLL prior. The class distributions were modeled with multinomial logistic regression, where the regressors were computed with LORSAL (Logistic Regression via Variable Splitting and Augmented Lagrangian) algorithm [39]. The maximum a posterior (MAP) segmentation was computed via a min-cut based integer optimization algorithm. Experiments were carried out with AVIRIS Indian Pine Scene. The results they claimed are depicted in Table 6, where [S] denotes supervised algorithm, [Semi-S] denotes the proposed semi supervised algorithm, [L] denotes the LORSAL algorithm. It was shown that the proposed semi supervised algorithm yield better result compared to other state of the art techniques.

Table 6. OA [%] for supervised and semi supervised approaches

Methods	Number of labeled samples per class					
	5	10	15	20	25	120
[S]	66.83	77.98	81.31	84.46	89.33	96.71
[Semi-S]	72.62	80.15	83.50	87.14	91.54	96.99
[L]	58.10	66.56	70.24	72.84	78.88	84.79

In 2013, Jun Li et al, extended their work to design another semi supervised learning algorithm by developing a new soft sparse multinomial logistic regression model (S2MLR) which exploits both hard and soft labels [40]. These labels correspond to labeled and unlabeled training samples respectively. Hyperspectral labels are dominated by mixed pixel, so assigning single (hard) label to a pixel may be a potential source of error. The use of soft classification has

been very rarely used in hyperspectral imaging. Some researchers claimed that consideration of soft labels led to good performance [41], [42]. The proposed S2MLR exploits the concept of soft labels to generate unlabeled training samples. This algorithm has been specially designed to address the presence of highly mixed pixels in real HSIs. The experiments were carried out first using AVIRIS KSC dataset and then using Indian Pines Scene. The third experiment was carried out using ROSIS (Reflective Optics System Imaging Spectrometer) Pavia City dataset .The experiments were carried out by randomly selecting a very limited number of labeled samples for training purpose to show the good capacity of the proposed algorithm to tackle the ill posed problem. Table 7 shows the experimental results in terms of OA (in %) and κ (in %) for the AVIRIS KSC dataset using the proposed SSL with 36 labeled (three per class) and 1226 unlabeled training samples. They claimed that SSL with soft labels gave better results than supervised algorithms and SSL with hard labels.

Table 7. Experimental result using ssl with 36 labeled (three per class) and 1226 unlabeled training samples.

Methods		Accuracy Measure	
		OA	κ
Supervised		66.29	62.30
SSL	soft	75.54	72.48
	hard	74.85	71.70

2.5 Gabor filters

2-D Gabor Filters were applied to multiband images band by band or by considering opponent relationship between pair of bands [43]. These filters were also used to represent spectral dimension of HSI dataset [9]. Recently 3-D Gabor filters have been used in this field as they have the ability to extract joint spatial and spectrum information. In 2010, Bau et al. developed a model for spectral/spatial based on 3-D Gabor filters [10]. The dimensions extracted by 3-D Gabor filter were very huge. So only those features were selected that emphasizes most significant spectral/spatial differences between various classes in a scene. Authors used AVIRIS HSI dataset acquired over Pocomoke City, containing 224 bands. They used all the 224 bands for the experiment and claimed that 3-D Gabor filter successfully separated samples with similar constituent material but with different spatial structure. 3-D Gabor filters also outperformed spectral or GLCM features. The experiment results are shown in Table 8. D' represents the number of features which gave the optimized classification rate.

Table 8.Comparision of results using spectral features and Gabor filter

Experiment	Optimized Feature Sets			All Features		
Vector Type	Train	D'	Test	Train	D	Train
Spectral	82%	24	73%	77%	224	71%
Gabor	89%	11	83%	85%	26	81%
GLCM PCA	79%	13	74%	76%	32	76%
GLCM RGB	85%	26	78%	81%	96	77%
Spectral+ Gabor	97%	8	91%	84%	250	78%

In 2011, Linlin Shen et al. proposed a 3-D Gabor wavelet based approach for pixel based hyperspectral imagery classification [11]. Initially, they designed a set of Gabor wavelets with different frequencies to extract signal variances in space, spectrum and joint spectral/spatial domains. Then a

feature selection and fusion process was proposed to eliminate redundant features and to make the fused features more discriminative. The proposed method was applied on Indian Pines Scene and KSC dataset. The performance of the proposed fused Gabor features was compared with spectral and individual Gabor filter using K-NN [44] and SVMs as classifiers. They claimed that the fused Gabor features achieved 99.89% accuracy when 75% of the total samples were used as gallery. The performance of the fused Gabor feature for Indian Pine and KSC images was further improved to 96.04% and 95.36% respectively considering only 5% of the total samples as gallery. Table 9 shows the experimental result in terms of OA (in %).

Table 9. Experimental results using individual and fused 3D Gabor wavelets

Method	Indian Pine		KSC	
	SVM	KNN	SVM	KNN
Individual Gabor	93.41	86.49	80.08	95.02
Fused Gabor	96.04	93.38	85.78	95.36

In their work in 2013, authors proposed a symmetrical uncertainty based and Markov-blanket-based approach to select informative and non-redundant Gabor features for HSI classification [45]. The extracted Gabor features were ranked by their information contained and started by an empty set, features were added one by another by investigating the redundancy with already selected features. The performance of the of the proposed method was tested on Indian Pine Scene using different gallery size, i.e., for each class, 5%,10%,25% and 50% of the available labeled samples were used as gallery. The remaining samples were used as test sets. The experiments were performed using K-NN and SVMs separately as classifiers and the results were compared with those obtained by using spectrum features and fused Gabor features. Table 10 summarizes the result obtained using 5% of the samples as gallery. It was shown that the proposed selected Gabor feature method significantly reduced feature dimensions and thus improves efficiency of classification.

Table 10. Experimental results using fused gabor and selected Gabor

Features	Dim	KNN	SVM
		OA [%]	OA [%]
Spectrum	220	71.08	67.15
Fused Gabor	149.6	92.44	95.31
Selected Gabor	136.2	92.89	94.66

2.6 Tensor algebra based methods

Construction of some tensor embedded frameworks for dimension reduction has been proposed in [46]-[48]. In 2013, Liangpei Zhang et al. in their work proposed a supervised manifold learning algorithm referred to as Tensor Discriminative Locality Alignment (TDLA) for spectral-spatial feature representation and dimension reduction for HSIs under the umbrella of multi linear algebra [49]. The proposed algorithm was tested on three datasets. Dataset1 - a HYDICE dataset, dataset 2- a ROSIS dataset and dataset 3- an AVIRIS dataset. In these datasets total 7, 6 and 10 major classes were analyzed respectively using SVM and Neural Network (NN) separately. The experiment results are listed in

Table 11 in terms of OA (in %) and к (%). The proposed method significantly improved the classification accuracies but still there is a scope of improvement for input feature representations, addressing more discriminative features. Authors concluded that the proposed tensor representation could preserve the original spatial constraints of a certain pixel to a great extent and it could process a more comprehensive high-order feature directly.

Table 11. Performance of TDLA over different datasets

Classi-fier	Dataset 1		Dataset 2		Dataset 3	
	OA	к	OA	к	OA	к
SVM	97.66	97.26	96.42	95.69	89.54	87.06
NN	89.90	88.19	84.91	81.81	71.22	67.09

Recently Silvia Varelo et al., in 2013 have proposed a region based hierarchical approach for HSI classification based on Binary Partition Tree (BPT). This can be interpreted as a set of hierarchical regions stored in a tree structure [50]. In BPT the nodes represents image regions and the branches represents inclusion relationship among the nodes. The leaf nodes represent the regions of an initial partition, the root node represents the entire image scene and the intermediate nodes represent the regions formed by merging of two child nodes. The BPT is constructed using iterative bottom up merging algorithm. The creation BPT relies on two important notions – 1) the region model which specifies how regions are represented and how to model the union of two regions, and 2) the merging criteria which define the similarity of neighboring regions and hence determines the order in which regions are going to be merged. Authors have discussed some of the region models and merging criteria and a pruning strategy has also been discussed to remove redundant sub trees from the original tree. Once the BPT is constructed, the fixed tree structure allows implementation of advanced application dependent techniques on it. Urban HYDICE dataset and ROSIS Pavia dataset were used to evaluate the performance of the proposed method. Table 12 summarizes the experimental result.

Table 12. OA: using simple SVM and Pruned BPT

Experimental Data sets	OA [%] Simple SVM	OA [%] Pruned BPT
Pavia University Dataset	88.58	92.96
HYDICE dataset	87.74	94.69

3. CONLUSION

Over the time many techniques have been proposed by researchers to meet the challenges in the field of HSI analysis. Table 13 summarizes the finding s of different methods discussed in this paper. Although hyperspectral imaging is a source of enormous information as discussed it has some inherent issues which are needed to be addressed. Handling of high dimensional dataset and limited availability of training samples are the main issues in this field. Researchers are still working in the area of dimension reduction. Although the usage of semi supervised algorithms and SVMs has solved the second problem up to some extent, this area is still open for research

Table 13. Findings of different methods in HSI analysis

Method	Findings
Band Selection	Up to 94% of the bands were reduced to produce almost similar classification result.
Kernel Based Methods	These methods revealed excellent result in terms of accuracy, computational cost and robustness. Composite kernels produced up to 98.86% classification accuracy.
Semi Supervised Learning Algorithms	These methods showed up to 8.86 improvement than the supervised methods by working with very low number of training samples
Multinomial Logistic Regression	Semi supervised algorithms using MLL showed significant improvement in classification in classification accuracy compared to supervised methods while working with very limited number of training samples.
3-D Gabor Filter	These methods considered both spectral and spatial features and produced up to 94.66% accuracy using only 5% of the samples available.
Tensors	Up to 97.66% classification accuracy were achieved using these methods.
Pruned BPT	These methods showed 4-7% improvement in classification accuracy than normal SVM.

4. REFERENCES

[1] G. F. Hughes, "On the mean accuracy of statistical pattern recognizers," IEEE Trans. Inf. Theory, vol. IE-14, no. 1, pp. 55–63, Jan. 1968.

[2] M. Fauvel, J.A. Benediktsson, J. Chanussot, and J.R. Sveinsson, "Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles," IEEE Transactions on Geoscience and Remote Sensing, vol. 46, no. 11, pp. 3804–3814, 2008.

[3] A. Plaza, J. A. Benediktsson, J. W. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, A. Gualtieri, M. Marconcini, J. C. Tilton, and G. Trianni, "Recent advances in techniques for hyperspectral image processing," Remote Sensing of Environment, in press, 2009.

[4] J. Borges, J. Bioucas-Dias, and A. Marc¸al, "Evaluation of Bayesian hyperspectral imaging segmentation with a discriminative class learning," in Proc. IEEE International Geoscience and Remote sensing Symposium, Barcelona, Spain, 2007.

[5] R. S. Rand and D. M. Keenan, "Spatially smooth partitioning of hyperspectral imagery using spectral/spatial measures of disparity," IEEE Trans. Geosci. Remote Sens., vol. 41, no. 6, pp. 1479–1490, Jun. 2003.

[6] F. Melgani and L. Bruzzone, "Classification of hyperspectral remotesensing images with support vector machines," IEEE Trans. Geosci. Remote Sens., vol. 42, no. 8, pp. 1778–1790, Aug. 2004.

[7] G. Camps-Valls, L. Gómez-Chova, J. Calpe, E. Soria, J. D. Martín, L. Alonso, and J. Moreno, "Robust support vector method for hyperspectral data classification and

knowledge discovery," IEEE Trans. Geosci. Remote Sens., vol. 42, no. 7, pp. 1530–1542, Jul. 2004

[8] M. Murat Dundar and A. Landgrebe, "A cost-effective semisupervised classifier approach with kernels," IEEE Trans. Geosci. Remote Sens., vol. 42, no. 1, pp. 264–270, Jan. 2004.

[9] M. Shi and G. Healey, "Hyperspectral texture recognition using a multiscale opponent representation," IEEE Trans. Geosci. Remote Sens., vol. 41, no. 5, pp. 1090–1095, May 2003.

[10] T. C. Bau, S. Sarkar, and G. Healey, "Hyperspectral region classification using a three-dimensional Gabor filterbank," IEEE Trans. Geosci. Remote Sens., vol. 48, no. 9, pp. 3457–3464, Sep. 2010.

[11] L. Shen and S. Jia, "Three-dimensional Gabor wavelets for pixel-based hyperspectral imagery classification," IEEE Trans. Geosci. Remote Sens., vol. 49, no. 12, pp. 5039–5046, Dec. 2011.

[12] D. B¨ohning, "Multinomial logistic regression algorithm," Annals of the Institute of Statistical Mathematics, vol. 44, pp. 197–200, 1992.

[13] S. D. Stearns, B. E. Wilson, and J. R. Peterson, "Dimensionality reduction by optimal band selection for pixel classification of hyperspectral imagery," Applicat. Digital Image Processing XVI, vol. 2028, pp. 118–127, 1993.

[14] Chein-I Chang, Qian Du, Tzu-Lung Sun and Mark L. G. Althouse, "A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification", IEEE Trans. Geosci. Remote Sens., vol. 37, no. 6, Nov. 1999

[15] C.-I Chang, X. Zhao, M. L. G. Althouse, and J.-J. Pan, "Least squares subspace projection approach to mixed pixel classification in hyperspectral images," IEEE Trans. Geosci. Remote Sensing, vol. 36, pp. 898–912, May 1998.

[16] C.-I. Chang, T.-L. E. Sun, and M. L. G. Althouse, "An unsupervised interference rejection approach to target detection and classification for hyperspectral imagery," Opt. Eng., vol. 37, pp. 735–743, Mar. 1998.

[17] Shailesh Kumar, Joydeep Ghosh, and Melba M. Crawford "Best-bases feature extraction algorithms for classification of hyperspectral data", IEEE Trans. Geosci.Remote Sensing, vol. 39, no. 7, July 2001.

[18] S. Kumar, J. Ghosh, and M. M. Crawford, "A versatile framework for labeling imagery with large number of classes," in Proc. Int. Joint Conf. Neural Networks, Washington, DC, 1999.

[19] M. M. Crawford, S.Kumar,M. R. Ricard, J. C. Gibeaut, and A. Neuenshwander, "Fusion of airborne polarimetric and interferometric SAR for classification of coastal

environments," IEEE Trans. Geosci. Remote Sensing, vol. 37, pp. 1306–1315, May 1999.

[20] R. R. Coifman and M. V. Wickerhauser, "Entropy-based algorithms for best basis selection," IEEE Trans. Inform. Theory, vol. 38, pp. 713–719, Mar. 1992.

[21] N. Saito and R. R. Coifman, "Local discriminant bases, in Mathematical Imaging: Wavelet Applications in Signal and Image Processing II," in Proc. SPIE, vol. 2303, 1994, pp. 2–14.

[22] X. Jia and J. A. Richards, "Segmented principal components transformation for efficient hyperspectral remote-sensing image display and classification," IEEE Trans. Geosci. Remote Sensing, vol. 37, pp. 538–542, Jan. 1999.

[23] L. Bruzzone and D. Fernández-Prieto, "A technique for the selection of kernel-function parameters in RBF neural networks for classification of remote-sensing images," IEEE Trans. Geosci. Remote Sens., vol. 37, no. 2, pp. 1179–1184, Mar. 1999.

[24] G. Giacinto and L. Bruzzone, "Combination of neural and statistical algorithms for supervised classification of remote-sensing images," Pattern Recognit.Lett., vol. 21, no. 5, pp. 399–405, 2000.

[25] L. Bruzzone and R. Cossu, "A multiple-cascade-classifier system for a robust and partially unsupervised updating of land-cover maps," IEEE Trans. Geosci. Remote Sens., vol. 40, no. 9, pp. 1984–1996, Sep. 2002.

[26] A. Ben-Hur, D. Horn, H. Siegelmann, and V. Vapnik, "Support vector clustering," J. Mach. Learn. Res., vol. 2, pp. 125–137, 2001.

[27] G. Rätsch, B. Schökopf, A. Smola, S. Mika, T. Onoda, and K.-R. Müller, "Robust ensemble learning," in Advances in Large Margin Classifiers, A. Smola, P. Bartlett, B. Schölkopf, and D. Schuurmans, Eds. Cambridge, MA: MIT Press, 1999, pp. 207–219.

[28] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," IEEE Trans. Geosci. Remote Sens., vol. 43, no. 6, pp. 1351–1362, Jun. 2005.

[29] B. Schölkopf, K.-K. Sung, C. Burges, F. Girosi, P. Niyogi, T. Poggio,and V. N. Vapnik, "Comparing support vector machines with Gaussian kernels to radial basis function classifiers," IEEE Trans. Signal Process., vol. 45, no. 11, pp. 2758–2765, Nov. 1997.

[30] Gustavo Camps-Valls, Luis Gomez-Chova, Jordi Muñoz-Marí, Joan Vila-Francés, and Javier Calpe-Maravilla, "Composite kernels for hyperspectral image classification", IEEE Geosci. Remote Sens. Letters, vol. 3, no. 1, Jan. 2006.

[31] B. Mak, J. Kwok, and S. Ho, "A study of various composite kernels for kernel eigenvoice speaker adaptation," in Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing, vol. 1, May 2004, pp. 325–328.

[32] J.-T. Sun, B.-Y.Zhang, Z. Chen, Y.-C.Lu, C.-Y. Shi, and W. Ma, "GECKO: A method to optimize composite kernels for web page classification," in Proc. IEEE/WIC/ACM Int. Conf. Web Intelligence, vol. 1, Sep. 2004, pp. 299–305

[33] S. Tadjudin and D. Landgrebe, "Classification of high dimensional data with limited training samples," Ph.D. dissertation, School of Elect. Eng. Comput.Sci., Purdue Univ., West Lafayette, IN, May 1998. Tech. Rep. TR-ECE-98-9.

[34] Lorenzo Bruzzone, Mingmin Chi and MattiaMarconcini, "A novel transductive SVM for semisupervised classification of remote-sensing images", IEEE Trans. Geosci. Remote Sens., vol. 44, no. 11, Nov. 2006.

[35] V. N. Vapnik, "The nature of statistical learning theory", 2nd ed. Berlin, Germany: Springer-Verlag, 1999, pp. 237–240, 263–265, and 291–299.

[36] A. Gammerman, V. Vapnik, and V. Vowk, "Learning by transduction," in Proc. Uncertainty Artif. Intell., Madison, WI, Jul. 1998, pp. 148–156.

[37] B. Krishnapuram, L. Carin, M. Figueiredo, and A. Hartemink, "Sparse multinomial logistic regression: Fast algorithms and generalization bounds," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 6, pp. 957–968, 2005.

[38] J. Li, J. Bioucas-Dias, and A. Plaza, "Semi-supervised hyperspectral image classification based on aMarkov random field and sparse multinomial logistic regression," in Proc. IEEE Int. Geosci. Remote Sens. Symp., 2009,vol. 3, pp. 817–820.

[39] J. Bioucas-Dias and M. Figueeredo, "Logistic regression via variable splitting and augmented lagrangian tools," Tech. Rep., Instituto Superior T´ecnico, TULisbon, 2009.

[40] Jun Li, José M. Bioucas-Dias and Antonio Plaza,"Semisupervised hyperspectral image classification using soft sparse multinomial logistic regression", IEEE Geosci. Remote Sens. Letters, vol. 10, no. 2, Mar. 2013.

[41] X. Zhu, Z. Ghahramani, and J. Lafferty, "Semi-supervised learning using Gaussian fields and harmonic functions," in Proc. 20th ICML Conf., 2003, pp. 912–919.

[42] J. Li, J. Bioucas-Dias, and A. Plaza, "Exploiting spatial information in semi-supervised hyperspectral image segmentation," in Proc. IEEE WHISPERS, 2010, pp. 1–4.

[43] A. Jain and G. Healey, "A multiscale representation including opponent color features for texture recognition," IEEE Trans. Image Process., vol. 7, no. 1, pp. 124–128, Jan. 1998.

[44] T. Cover and P. Hart, "Nearest neighbor pattern classification," IEEE Trans. Inf. Theory, vol. 13, no. 1, pp. 21–27, Jan. 1967.

[45] Linlin Shen, Zexuan Zhu, Sen Jia, Jiasong Zhu, and Yiwen Sun, "Discriminative Gabor feature selection for hyperspectral image classification", IEEE Trans. Geosci. Remote Sensing Letters, vol. 13, no. 1, Jan. 2013.

[46] S. Yan, D. Xu, Q. Yang, L. Zhang, X. Tang, and H.-J.Zhang, "Multilinear discriminant analysis for face recognition," IEEE Trans. Image Process., vol. 16, no. 1, pp. 212–220, Jan. 2007.

[47] D. Xu, S. Yan, L. Zhang, S. Lin, H.-J. Zhang, and T. S. Huang, "Reconstruction and recognition of tensor-based objects with concurrent subspaces analysis," IEEE Trans.

Circuits Syst. Video Technol., vol. 18, no. 1, pp. 36–47, Jan. 2008.

[48] N. Renard and S. Bourennane, "Dimensionality reduction based on tensor modeling for classification methods," IEEE Trans. Geosci. Remote Sens., vol. 47, no. 4, pp. 1123–1131, Apr. 2009.

[49] Liangpei Zhang, Lefei Zhang, Dacheng Tao and Xin Huang, "Tensor discriminative locality alignment for hyperspectral image spectral–spatial feature extraction", IEEE Trans. Geosci. Remote Sensing, vol. 51, no. 1, Jan. 2013.

[50] Silvia Valero,PhilippeSalembier and Jocelyn Chanussot, "Hyperspectral image representation and processing with binary partition trees", IEEE Trans. Imsage Process., vol. 22, no. 4, Apr. 2013.