International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol.8, No.10 (2015), pp.11-18
http://dx.doi.org/10.14257/ijsip.2015.8.10.02

Marker Selection using Support Vector Machine Over-fitting for Very Low Training Sample Analysis of Hyperspectral Image Classification

Farid Muhammad Imran

Shaanxi Key Laboratory of Information Acquisition and Processing, Center for Earth Observation, School of Electronics and Information, Northwestern Polytechnical University, Xi'an, 710129, China
imran@mail.au.edu.pk

Abstract

In this paper we have proposed a new marker selection technique using Support Vector Machine over-fitting. Markers are the most reliable pixels in a class. We used our proposed technique to do classification of hyperspectral image with very low training samples, as low as one pixel per class. We have used both spectral and spatial information to improve the classification results. The spatial information is extracted using Extended Morphological Profiles with duality. Nonparametric supervised feature extraction methods are used to eliminate the redundant and irrelevant information in both spatial and spectral domains. In the end we have done experimentation to verify our proposed approach. The experimentation results show that when non-parametric weighted feature extraction method is used we get better classification results. The classification maps shows that even with just one training sample per class we still can get a reliably reasonable classification map.

Keywords: *Classification, feature extraction, hyperspectral images, support vector machine*

1. Introduction

From the last two decades a lot of work has been done in hyperspectral remote sensing technology [1]. Detailed physical analysis of object structures is possible by using the advanced hyperspectral imaging sensors that are able to capture hundreds of narrow spectral channels [2].

It is now commonly accepted that using the spatial and spectral information simultaneously provides significant advantages in terms of improving the performance of classification techniques [3]. In addition, small structures are now better identified due to recent advances in spatial resolution of sensors [2]. It is well known that contextual information, i.e., the inter pixel dependency is also useful for the classification of hyperspectral images (HSIs) [4].

Apart from many other problems, one of the problems that many scientists faced in remote sensing is the limited number of training samples. A lot of work and methods have been presented to improve the classification accuracy with limited number of training samples.

In this paper very low training sample analysis (VLTSA) is performed using spatial-spectral classification scheme shown in Figure 1. Apart from the spectral information, the spatial information is extracted using Extended Morphological Profiles (EMP) with duality property (EMPD), which improves the classification accuracy better than the conventional EMP because it reduces the shape noise [5]. The original HSI is first

ISSN: 2005-4254 IJSIP

normalized and then it is used for principal component analysis (PCA) and feature extraction (FE) analysis, this also

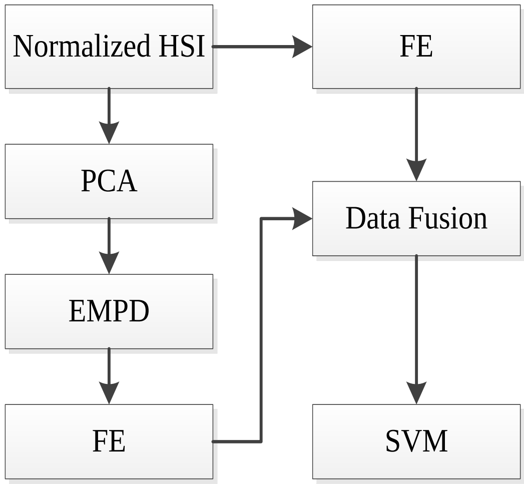


Figure 1. The Main Flow Diagram on which we have Implemented Our Approach

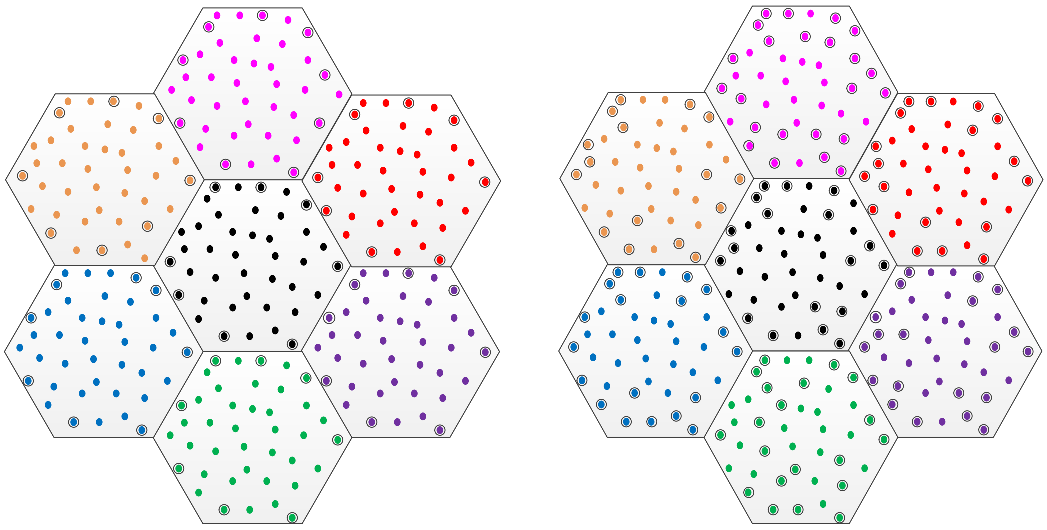
helps to improve the classification accuracy. For FE, two nonparametric supervised FE techniques are used, named Decision Boundary Feature Extraction (DBFE) and Nonparametric Weighted Feature Extraction (NWFE). Support Vector Machine (SVM) for classification is used because it can handle both spatial and spectral information very efficiently. Figure 1, summaries the main flow of our work.

In supervised classification, data set is labelled based upon the available ground truth data (GTD). The training and testing samples are picked randomly. If the training samples to be picked are very limited for example only one or two, then it is very important that the picked training samples should be reliable and must represent the class for which they have been labelled for. Training samples are picked randomly, so there is a chance that they will be picked at the boundary of the class, where there is likelihood that they may not belong to the class for which they are labelled of. They may represent the neighbouring class as the classes are usually very close to each other in HSIs. So, the probability of reliability of limited training samples that have been chosen for the analysis must be increased.

One idea that comes to the mind is that the pixels at the center are more accurate to represent the class than the pixels at the boundary. Many algorithms have been proposed to choose more reliable pixels as region marker. In [6], Tarabalka *et al*. choose markers by analyzing probabilistic SVM classification results. In [7] Multiple Classification technique has been presented for marker selection in such a way that the complementary benefits of each classifier are exploited, while their weaknesses are avoided. In [8] Multiple Spectral-Spatial Classifier is presented for marker selection, which is further used for marker-controlled region growth, based on a minimum spanning forest algorithm [6]. All the above mentioned techniques for marker selections are quite complex and time consuming as well. We propose a new marker selection technique called *marker selection using SVM over-fitting*, which is described as follows:

2. Marker Selection using SVM Over-fitting

Over-fitting is usually avoided in SVM but it can be used for selecting markers. SVM uses the support vectors, which usually lie around the boundary of the different classes to separate them, rather than using the whole class vectors. In SVM, over-fitting can be done by simply



(a) **(b)**

Figure 2. (a) Seven classes example with support vectors in circle after SVM and (b) when number of support vectors are increased using SVM over-fitting

changing the parameters C and γ. If over-fitting is performed then there will be more number of support vectors and the vectors in different classes that are not the support vectors can be chosen for markers. This technique is very simple and quick.

The basic idea of the marker selection using SVM over-fitting technique is presented in Figure 2. Figure 2(a) shows the seven class example with support vector in circle. The support vectors lies at the boundary of the different classes. If we increase the number of support vectors by modifying the parameters C and γ or we over-fit our data the number of support vectors increase around the boundary, shown in Figure 2(b) and then we will pick the non-support vectors from each class as they are at the centre of the classes compare to the support vectors. The picked non-support vectors are randomized for training samples, which are then used for VLTSA.

Figure 3 shows how the marker selection using SVM over-fitting technique is implemented on the Figure 1. It can be seen from the Figure 3 that we are feeding in the output of the proposed approach (marker selection using SVM over-fitting technique, with block named VLTSA_nonSV) into the FE because we are using the nonparametric supervise feature extraction techniques. Each block of VLTSA_nonSV is explained in Figure 4. Figure 4 explains our proposed approach i.e. the ground truth data is used to separate the training and testing samples from each class with 50 percent ratio. From the training samples, SVM model is formed and then n non-support vectors are selected from each class. This information is feed into the nonparametric supervise feature extraction techniques. For the FE, VLTSA_nonSV is performed only once. Referring to the Figure 3 it can be seen that the spectral and spatial information after FE is fused using the concatenate vector, which will then be used for final classification using SVM. After data fusion the VLTSA_nonSV techniques is performed on spectral and spatial information, which then be used for classification using SVM. At this stage the VLTSA_nonSV technique with the final SVM is repeated 100 times for the reliability of the classification results, as VLTSA is performed, where the numbers of training samples are very low.

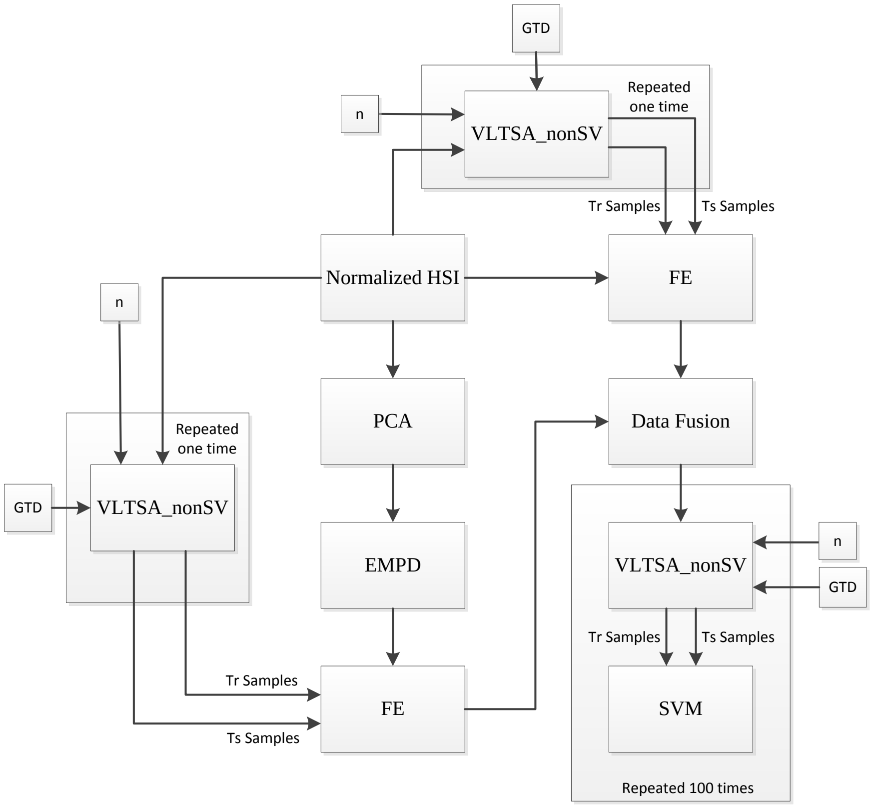


Figure 3. Proposed SVM Over-fitting Approach; Implemented on the Figure 1

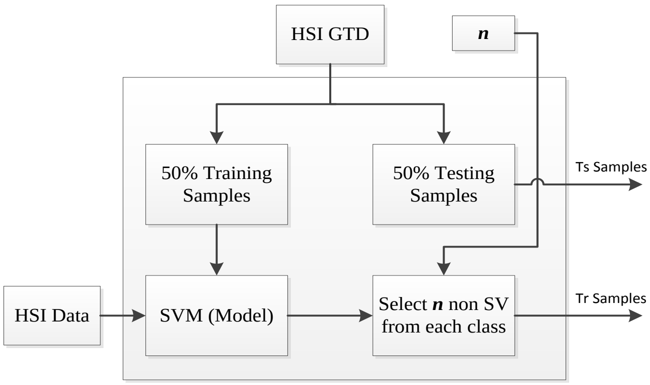


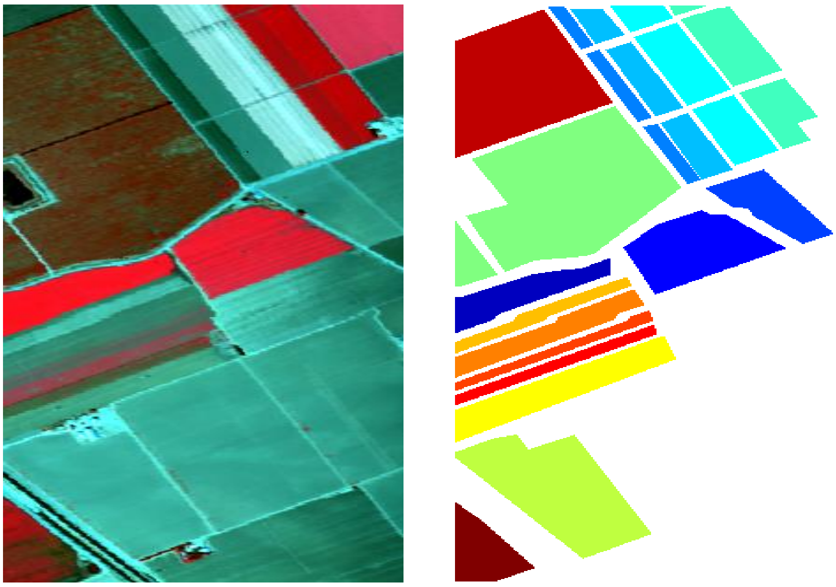
Figure 4. The VLTSA_nonSV Block in Figure 3

3. Non-Parametric Supervised Feature Extraction techniques

Two supervised nonparametric FE techniques named Decision Boundary Feature Extraction (DBFE) and Nonparametric Weighted Feature Extraction (NWFE) are used to extract the spatial features in our work. Nonparametric FE is based on a nonparametric extension of the scatter matrices. There are at least two advantages of using the nonparametric scatter matrices. First, they are generally of full rank. This provides the ability to specify the number of extracted features desired and to reduce the effect of the singularity problem. It is in contrast to parametric discriminant analysis, which usually can only extract $L-1$ features [9], where L is the number of classes. Second, the nonparametric nature of scatter matrices reduces the effects of outliers and works well even for non-normal data sets and most of the hyperspectral data sets are non-normal [9].

In DBFE all features useful for discriminating the classes can be extracted from the decision boundary [10]. The decision boundary feature matrix (DBFM) [11] is formed by using the vector norm at the decision boundary. The vector norm is the normal vector to the line connecting the two pair of training samples belonging to different classes. DBFE is very much dependent of the number of training samples and can be computationally intensive.

Kuo and Landgrebe [12] proposed NWFE using the advantage of Discriminant Analysis Feature Extraction (DAFE) and DBFE and eliminating their disadvantages. The main ideas of NWFE are putting different weights on every sample to compute the weighted means and defining new nonparametric between class and within-class scatter matrices in order to obtain more than $L-1$ features [9].



(a) **(b)**

Figure 5. (a) Three Channel Colour Composite of Salinas Data Set and (b) its Ground Truth Map

4. Experimental Results

AVIRIS Salinas data set is used for our experimentation. Salinas data set has 512 by 217 pixels with 204 bands in spectral dimension. Three channel colour (RGB) composite of its data set is shown in Figure 5(a) and it's GTD in (b) with 16 mutually exclusive classes. This data set is chosen because it is a challenging classification problem as most of the classes have the similar spectral.

The criteria used to compare classification results involve Overall Accuracy (OA), Average Accuracy (AA) and the kappa coefficient (*k*). Time analysis is not done as it is obvious that lesser the feature, faster will be the processing time. *MATLAB* is used for morphological operations while *MultiSpec* software is used for feature extraction. The SVM classification is done using *LIBSVM* [13]. In our study, one-against-one strategy is used for SVM using radial basis kernel. Throughout the experiments, the normalized HSI data set is used, which is feed into *MultiSpec* software for FE, as shown in Figure. 3. The concatenate vector is used for data fusion; to combine spatial and spectral information. For PCA, DBFE and NWFE the number of features are selected based on cumulative percentage of 99%. For VLTSA, every step is performed once up-till data fusion. See Figure 3. After that, training samples are randomly selected 100 times and then averaged for the reliability of the results. Every 100 times SVM parameters C (4, 8, 16, 32, 64) and γ (1, 2, 4) are determined using five-fold cross-validation.

Table 1 summarizes the VLTSA of Salinas data set using proposed marker selection SVM over-fitting technique. The experiment is performed when only one to ten training samples per class are chosen. The experiments are divided into five parts i.e. analyzing spectral information only; analyzing spatial information only; analyzing spatial and spectral information together; analyzing spectral and spectral information together with feature extraction using DBFE and finally analyzing spectral and spectral information together with feature extraction using NWFE. It can be seen from Table 1 that for Salinas data set the spatial information gives better classification results as compare to when only spectral information is used. It can also be observed from Table 1 that the classification results improve when non-parametric feature extraction methods are used. FE removes the redundant and irrelevant information from the spectral and spatial information and hence gives us better classification accuracy compared to when no FE technique is used. When NWFE is used for FE we get the best results; even for one training sample per class we get an OA of 78% and when only three training samples per class is used we get an OA of 83%, which is very good compare to the low number of training samples.

Table 2 shows the number of features used for classification results when DBFE and NWFE techniques are used for FE. The first digit in the bracket represents the spectral feature and the second digit represents the spatial feature and the digit outside the bracket represents the sum of spectral and spatial features. It can be seen in Table 2 that with only 11 features, when NWFE is used as FE, we get an accuracy of 78% when only one training sample per class is used.

Figure 6 shows the classification map when only one and ten training samples per class are used for classification using our proposed techniques of marker selection using SVM over-fitting. The classification maps shown in Figure 6 are when both spatial and spectral information are used and is reduced using NWFE technique as FE. It can be seen that with just one training pixel per class still a reasonable classification map can be obtained. Using our proposed technique of marker selection using SVM over-fitting, the reliability of pixel belongs to the class for which it has been labelled for is increased. This helps us to choose

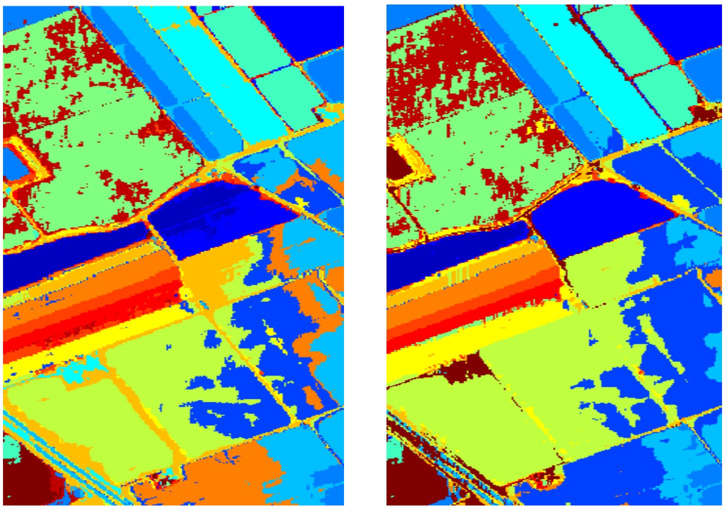
Table 1. Overall (OA) and Average (AA) classification accuracy in percentage. TS stands for Training sample/s and MPs stands for measuring parameters

TS	MPs	Spectral	EMPD	Spectral + EMPD	DBFE 99%	NWFE 99%
1	**OA**	76.110	77.304	75.847	77.314	**77.882**
	AA	79.300	81.969	81.160	82.186	**82.957**
	kappa	0.7338	0.7483	0.7326	0.7488	**0.7542**
2	**OA**	78.505	79.383	78.667	**80.224**	80.180
	AA	81.636	83.853	83.115	83.874	**85.348**
	kappa	0.7603	0.7715	0.7636	**0.7805**	0.7795
3	**OA**	79.884	82.502	82.329	82.687	**83.111**
	AA	83.349	87.415	87.532	87.337	**87.635**
	kappa	0.7757	0.8059	0.8042	0.8079	**0.8125**
4	**OA**	80.319	84.338	84.065	84.098	**85.173**
	AA	83.697	88.860	88.739	88.799	**89.837**
	kappa	0.7804	0.8262	0.8232	0.8235	**0.8352**
5	**OA**	80.858	85.107	84.382	84.725	**85.896**
	AA	84.323	89.773	89.312	89.559	**90.610**
	kappa	0.7863	0.8347	0.8268	0.8306	**0.8432**
6	**OA**	81.055	85.843	85.453	85.503	**86.079**
	AA	84.442	90.342	90.041	90.340	**90.891**
	kappa	0.7885	0.8428	0.8386	0.8391	**0.8452**
7	**OA**	81.369	86.600	86.322	86.277	**87.337**

	AA	84.824	91.094	90.920	90.887	**91.626**
	kappa	0.7920	0.8512	0.8482	0.8477	**0.8592**
8	**OA**	81.590	86.905	86.610	86.736	**87.670**
	AA	85.118	91.354	91.260	91.126	**91.851**
	kappa	0.7944	0.8546	0.8514	0.8528	**0.8630**
9	**OA**	81.619	87.228	87.180	86.699	**87.880**
	AA	85.062	91.627	91.547	91.283	**92.184**
	kappa	0.7947	0.8582	0.8577	0.8523	**0.8653**
10	**OA**	81.858	87.563	87.399	87.160	**88.443**
	AA	85.437	91.819	91.789	91.762	**92.334**
	kappa	0.7974	0.8619	0.8601	0.8574	**0.8715**

Table 2. Total Number of Features Selected during VLTSA for Salinas

No. of Training Samples	Salinas	
	DBFE 99%	NWFE 99%
1	11(5, 6)	11(7, 4)
2	8(5, 3)	11(7, 4)
3	31(13, 18)	12(7, 5)
4	37(17, 20)	18(12, 6)
5	39(17, 22)	18(12, 6)
6	52(29, 23)	19(13, 6)
7	52(28, 24)	19(13, 6)
8	48(24, 24)	20(14, 6)
9	54(29, 25)	21(15, 6)
10	55(33, 22)	23(17, 6)



(a) **(b)**

Figure 6. Classification maps of Salinas obtained from VLTSA (a) using NWFE as FE when only one training sample per class is used and (b) when only ten training samples per class are used.

more reliable pixels for VLTSA and hence results in better classification maps even with very low training sample.

5. Conclusion

In this paper, a new marker selection technique is proposed using SVM over-fitting. The technique is implemented when both spatial and spectral information are extracted for hyperspectral image classification. Spatial information is extracted using Extended Morphological Profile with duality. Nonparametric feature extraction techniques are used to reduce the redundant and irrelevant information from the spatial and spectral information. VLTSA is performed on the hyperspectral image. One to ten training samples per class are examined and it is concluded that better classification accuracy is obtained when NWFE is used as FE. It is also investigated that even using very low training samples still a reasonably fine classifications can be obtained by using the marker selection technique of SVM over-fitting.

References

[1] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," IEEE Geosci. Remote Sens. Mag., **(2013)**, vol. 1, no. 2, pp. 6–36.
[2] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Advances in spectral–spatial classification of hyperspectral images," Proceedings of the IEEE, **(2013)**, vol. 101, no. 3, pp. 652–675.
[3] B. Song, J. Li, M. D. Mura, L. Peijun, A. Plaza, J. M. Bioucas-Dias, J. A. Benediktsson, and J. Chanussot, "Remotely Sensed Image Classification Using Sparse Representations of Morphological Attribute Profiles," Geoscience and Remote Sensing, IEEE Transactions on, **(2014)**, vol. 52, no. 8, pp. 5122–5136.
[4] D. A. Landgrebe, "Signal Theory Methods in Multispectral Remote Sensing", Wiley, Hoboken, NJ, **(2003)**.
[5] F. M. Imran and M. He, "Extended Morphological Profiles with duality for hyperspectral image classification," to be published in Proc. of IEEE WHISPERS, **(2015)** June 2-5; Tokyo, Japan.
[6] Y. Tarabalka, J. Chanussot and J. A. Benediktsson, "Segmentation and classification of hyperspectral images using minimum spanning forest grown from automatically selected markers," Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Trans. on, vol. 40, no. 5, **(2010)**, pp. 1267–1279.
[7] Y. Tarabalka, J. A. Benediktsson, J. Chanussot and J. C. Tilton, "Multiple spectral-spatial classification approach for hyperspectral data," IEEE Trans. Geosci. Remote Sens., vol. 48, no. 11, **(2010)**, pp. 4122–4132.
[8] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot and J. C. Tilton, "Advances in Spectral-Spatial Classification of Hyperspectral Images," Proceedings of the IEEE, vol. 101, no. 3, **(2013)**, pp. 652–675.
[9] K. Bor-Chen, and D. A. Landgrebe, "Nonparametric weighted feature extraction for classification," IEEE Trans. Geosci. Remote Sens., vol. 42, no. 5, **(2004)**, pp. 1096–1105.
[10] L. Chulhee, and D. A. Landgrebe, "Feature Selection Based on Decision Boundaries," in Proc. IEEE IGARSS, **(1991)**, vol. 3, pp. 1471–1474.
[11] L. Chulhee, and D. A. Landgrebe, "Decision boundary feature extraction for nonparametric classification," IEEE Trans. Systems, Man and Cybernetics, vol. 23, no. 2, **(1993)**, pp. 433–444.
[12] P. O. Gislason, J. A. Benediktsson, and J. R. Sveinsson, "Random forests for land cover classification," Pattern Recognition Letters, vol. 27, no. 4, **(2006)**, pp. 294–300.
[13] C. -C. Chang and C. -J. Lin, LIBSVM: a library for support vector machines, **(2011)**, http://www.csie.ntu.edu.tw/~cjlin/libsvm/

Author

Muhammad Imran Farid, he received his BSc degree in both mathematics and physics from University of the Punjab, Pakistan, in 1998 and the M.Sc. degree in electronics from Quaid-i-Azam University, Islamabad, Pakistan, in 2000. He also received his MSc degree in radio frequency communication systems from University of Southampton, U.K, in 2005. He is currently pursuing the Ph.D. degree in image processing at Northwestern Polytechnical University, Xian, China. He is currently working as a Lecturer in Air University, Islamabad, Pakistan and is on study leave for PhD. His research interests include remote sensing, pattern recognition and astronomy.

