International Journal of Multimedia and Ubiquitous Engineering
Vol.10, No.12 (2015), pp.207-216
http://dx.doi.org/10.14257/ijmue.2015.10.12.21

Noise Effects on Feature Mining Non-Parametric Supervised Feature Extraction Techniques

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 two famous and commonly used feature mining non-parametric supervised feature extraction techniques (NSFETs) called Non-parametric Weighted Feature Extraction (NWFE) and Decision Boundary Feature Extraction (DBFE) are analyzed to see their efficiency in the presence of noise. In particularly these feature extraction techniques are used in classification as they give better classification accuracy. This study reveals that NSFETs are very sensitive to noise because of which the number of features increases and we get low classification accuracy. In order to see the behavior of NSFETs, spatial and spectral information from hyperspectral image classification is used. The experimental results show that in the presence of noise, spectral information is much more effected than the spatial information when features are extracted using the NSFETs. It is also examined that NWFE is more affected by noise than DBFE. The linear filtering technique is used just before the classifier in order to mitigate the noise effects in NSFETs. Using linear filtering just before the classifier does improve the final classification accuracy but with high number of spatial and spectral features. This does not satisfy the one of the main purpose of feature extraction and that is feature reduction.

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

1. Introduction

Hyperspectral Imaging (HSI) has a huge data set and sometimes huge data set can reduce the effectiveness of the data mining. Some attributes in data may not contribute for a meaningful model. The irrelevant attributes add noises and processing time, which affects the model accuracy and real-time performance. Some attributes in data may represent the same feature, which adds skewness in the logic of algorithm and results in affecting the model accuracy as well. Apart from this, higher the dimensionality of processing space, higher the computational cost. Sometimes in order to minimize the effects of noise, correlation and high dimensionality, some kind of dimension reduction methods including linear and nonlinear ones need to be done as a preprocessing step for data mining [1-3].

ISSN: 1975-0080 IJMUE
Copyright ⓒ 2015 SERSC

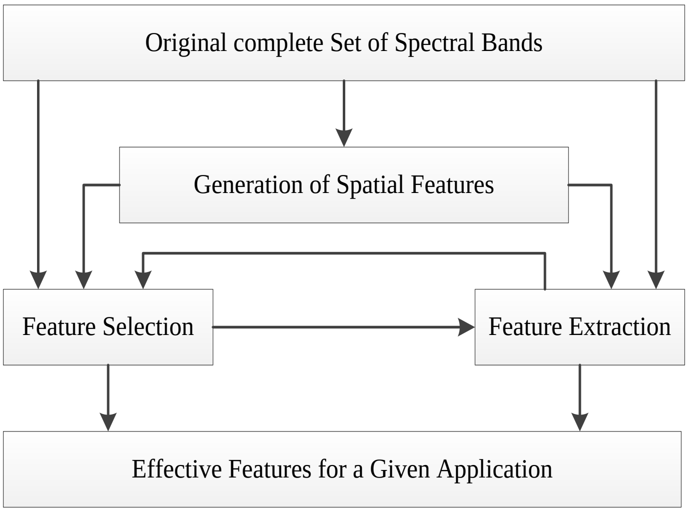


Figure 1. Possible Operations for Feature Mining (Only One Path is Used for Any Specific Approach) [1]

Theoretical and practical problems may arise when the dimensionality of the data in the spectral domain increases. The geometrical concepts that applied to the first three dimensional problems do not necessarily applied to the higher dimensionality space [4-5]. For example, when the dimensionality is very high, than the data have the tendency to lies in the tails, which contradicts with the bell-shaped functions. In classification such problems are referred as curse of dimensionality. Hughes [6] has shown that if the dimensionality of a problem is higher than a certain limit (number of training samples) then it will affect the classification accuracy. On the other hand the higher spectral resolution is helpful to discriminate between different classes but the complexity leads to poorer classification accuracy. To mitigate these phenomena usually feature selection (FS) and feature extraction (FE) or both are performed before the hyperspectral image classification [5]. FS methods choose features from the original feature set. In order to filter out the unimportant and redundant features the criteria like Information Gain, Correlation and Mutual Information *etc.* can be used. FE generates a small number of new features based on some criteria via transformation matrix to get the optimum subspace. In recent years spatial features are also extracted from the HSI along with the spectral features and then the features are selected using FS and FE. The whole of this process is called feature mining [7-8]. Figure 1 shows the possible operations for spatial and spectral feature mining [1]. In this paper, nonparametric supervised feature extraction techniques (NSFETs) are used, since they have been proved previously that for HSIs, nonparametric FE techniques give better classification results.

In this paper the spatial and spectral classification scheme used to measure the noise analysis in NSFETs are shown in Figure 2. Apart from the spectral information, the spatial information is extracted using Extended Morphological Profiles (EMP) [9] with duality property (EMPD), which improves the classification accuracy better than the conventional EMP because it reduces the shape noise [10]. The original HSI is first normalized and then it is used for principal component analysis (PCA) and feature extraction (FE) analysis. For FE, two NSFETs are used. Support Vector Machine (SVM) for classification is used because it can handle both spatial and spectral information very efficiently. Figure 2, summaries the flow of our work.

 Copyright ⓒ 2015 SERSC

2. Non-Parametric Supervised Feature Extraction

Nonparametric FE is based on a nonparametric extension of the scatter matrices. There are at least two advantages of using the nonparametric scatter matrices. Firstly, they are generally full rank. This provides the ability to specify the number of extracted features desired and to reduce the effect of the singularity problem. This is in contrast to parametric discriminant analysis, which usually can only extract $L-1$ (number of classes minus one) features [11]. In a real situation, this may not be enough. Secondly, 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 [11].

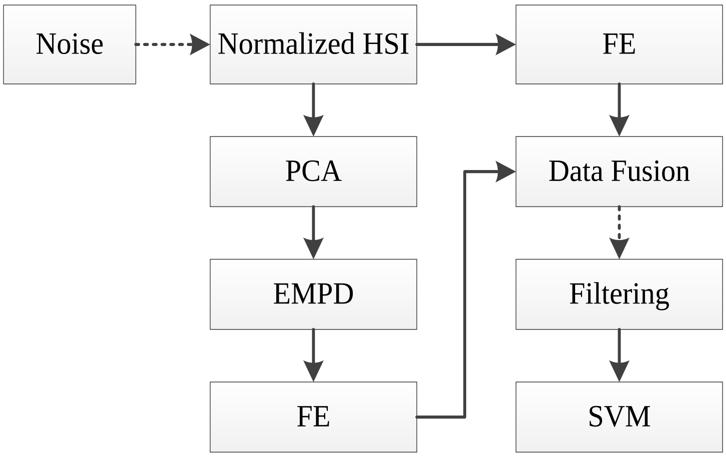


Figure 2. Spatial and Spectral Classification Scheme Based on EMPD and Linear Filtering

For our experiments, two feature mining NSFETs are used, named Decision Boundary Feature Extraction (DBFE) and Nonparametric Weighted Feature Extraction (NWFE).

2.1. Decision Boundary Feature Extraction (DBFE)

This method is based directly on the decision boundary. All features that can be useful for discriminating the classes can be extracted from the decision boundary [12-13]. The concepts of extracting discriminately informative and discriminately redundant features, using decision boundary feature matrix (DBFM) from the decision boundary are presented in [14]. The DBFM is formed 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. If N_i is the unit vector norm at point X on the decision boundary, then the DBFM is defined as

$$DBFM = \frac{1}{L}\sum N_i N_i^T \tag{1}$$

where L is the number of training samples. The eigenvector of non-zero eigenvalues of DBFM is the necessary feature vector to achieve the same classification accuracy as the original space. DBFE is very much dependent on the number of training samples and can be computationally intensive.

Copyright ⓒ 2015 SERSC

2.2. Nonparametric Weighted Feature Extraction (NWFE)

Kuo and Landgrebe [11] proposed NWFE by using the advantage of Discriminant Analysis Feature Extraction (DAFE) and DBFE, and by eliminating their disadvantages. The main ideas of NWFE are to put different weights on every sample to compute the weighted means and to define new nonparametric scatter matrices between and within class, in order to obtain more than $L-1$ features [11]. The nonparametric scatter matrices between and within class are defined as

$$S_b = \sum_{i=1}^{L} \frac{P_i}{L-1} \sum_{\substack{j=1 \\ j \neq 1}}^{L} \sum_{k=1}^{N_i} \lambda_k^{(i,j)} \left(x_k^{(i)} - M_j\left(x_k^{(i)}\right)\right)\left(x_k^{(i)} - M_j\left(x_k^{(i)}\right)\right)^T \tag{2}$$

$$S_w = \sum_{i=1}^{L} P_i \sum_{k=1}^{N_i} \lambda_k^{(i,j)} \left(x_k^{(i)} - M_i\left(x_k^{(i)}\right)\right)\left(x_k^{(i)} - M_i\left(x_k^{(i)}\right)\right)^T \tag{3}$$

Where $x_k^{(i)}$ refers to *k*th sample from class i, N_i the training sample size of each class, P_i the prior probability of each class and L the total number of classes.

Furthermore

$$\lambda_k^{(i,j)} = \frac{dist\left(x_k^{(i)}, M_j\left(x_k^{(i)}\right)\right)^{-1}}{\sum_{l=1}^{N_j} dist\left(x_l^{(i)}, M_j\left(x_l^{(i)}\right)\right)^{-1}} \tag{4}$$

$$M_j\left(x_k^{(i)}\right) = \sum_{l=1}^{N_j} w_l^{(i,j)} x_l^{(i)} x_l^{(j)} \tag{5}$$

Where

$$w_l^{(i,j)} = \frac{dist\left(x_k^{(i)}, x_l^{(j)}\right)^{-1}}{\sum_{l=1}^{N_j} dist\left(x_k^{(i)}, x_l^{(j)}\right)^{-1}} \tag{6}$$

And *dist.* (*a*, *b*) denotes the Euclidean distance between *a* and *b*. The resulting transformation is obtained by selecting *n* eigenvectors of $s_w^{-1} s_b$ that represent *n* largest eigenvalues.

3. Linear Filtering

Linear Filtering is one of the old and efficient techniques of signal processing. Briefly, a linear function is evaluated on each pixel $x_{m,n}$ on an image with respect to its neighborhood (usually a small rectangular shape) to compute a new pixel $y_{m,n}$ as shown in the Figure 2.

A linear filter in two dimensions has the general form

$$y_{m,n} = \Sigma_s \Sigma_t h_{s,t} x_{m-s,n-t} \tag{7}$$

 Copyright ⓒ 2015 SERSC

Where x is the input, y is the output, and h is the filter impulse response, which can be used for smoothing, sharpening, edge detection *etc.* of an image. The right hand side of the equation above is the convolution between h and x.

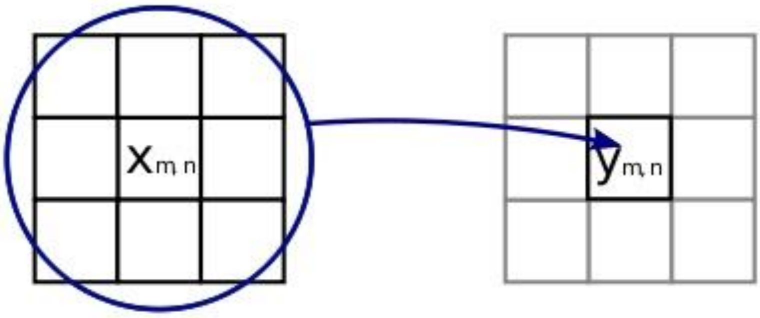


Figure 3. An Example of a 3x3 Window for Linear Filtering

For our experiment, h is used as a truncated Gaussian Filter with padding. A 2D isotropic Gaussian has the form:

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/(2\sigma^2)} \tag{8}$$

Where x and y are the distances from the origin to the X and Y axes respectively, and σ is the standard deviation.

The effect of Gaussian Filter is to smooth an image. The degree of smoothness is determined by the standard deviation of Gaussian Filter. The Gaussian Filter outputs a weighted average of each pixel's neighborhood, with the weighted average more towards the central pixels. Because of this, it provides gentler smoothing and preserved edges better than a similar sized mean filter [15].

4. Experimental Results

4.1. Data Set

One HSI urban data set called Pavia University is used for our experimentation. This image had been taken from an airborne ROSIS-03 (Reflective Optics System Imaging Spectrometer) optical sensor. ROSIS-03 sensor uses 115 bands with spectral coverage from 0.43 to 0.86μm. The spatial resolution is 1.3m per pixel. Apart from this, its data is atmospherically corrected but not geometrically.

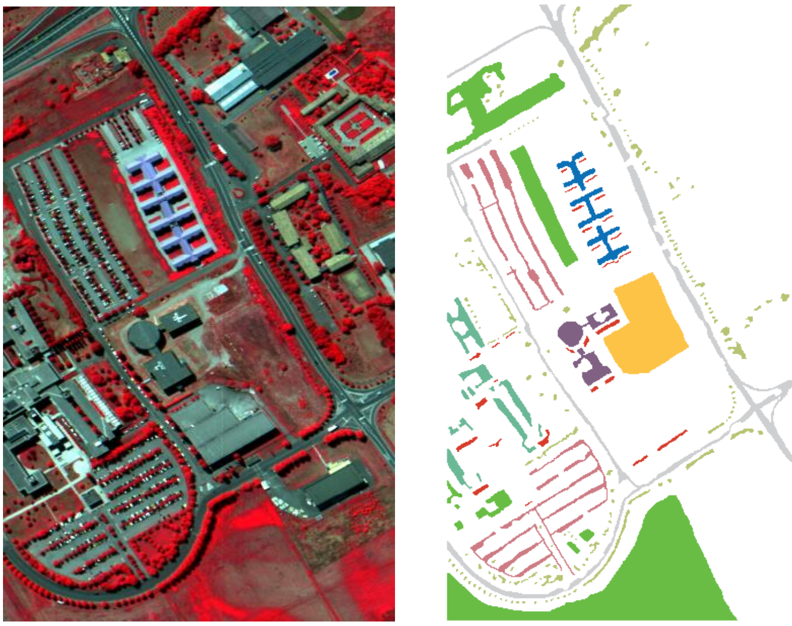
Pavia University data set has 610 by 340 pixels with 103 bands in spectral dimension. Three channel color composite of its data set is shown in Figure 3(a) and its Ground Truth Data (GTD) in (b). Nine selected classes with number of training and testing split are mentioned in Table 1. The color-map used in Pavia University GTD is also mentioned in Table 1.

4.2. Experimental Setup

The number of training and testing samples split of each class of Pavia University data set is mentioned in Table 1. The criteria used to compare classification results involve Overall Accuracy (OA), Average Accuracy (AA) and the kappa coefficient. *MATLAB* is used for morphological operations while *MultiSpec* [5] software is used for feature extraction. The SVM classification is done using *LIBSVM* [16]. In our study, one-against-one strategy is used for SVM with Gaussian kernel. The parameters C (4, 8, 16, 32, 64) and γ (1, 2, 4) are selected using five-fold cross-validation. To enhance the reliability of the experiments, all experiments were performed five times using randomly selected training samples and then the accuracy measuring parameters (OA, AA, and kappa) are averaged. Every five times SVM parameters C and γ are determined using five-fold cross-validation.

Table 1. Training and Testing Samples with Class Color Information

Pavia University			
Class	**Color**	**Train**	**Test**
Asphalt		548	6083
Meadows		540	18109
Gravel		392	1707
Trees		524	2540
Metal Sheets		256	1089
Soil		532	4497
Bitumen		375	955
Bricks		514	3168
Shadows		231	716
	Total	3912	38864



(a) (b)

Figure 4. (a) Pavia University Data and (b) Its Ground Truth Map

Throughout the experiments, the normalized HSI data set is used, which is feed into *MultiSpec* software for FE, as shown in Figure 2. The concatenate vector is used for data fusion; to combine spatial and spectral information. Gaussian Filter with padding technique is used for our experimentation and is truncated at 4 times of the standard deviation. For Gaussian filtering, sigma (variance) value is 15.5 when Pavia University data set is used. Sigma is found using hit and trial method, based on getting better classification accuracy. For PCA, DBFE and NWFE the number of features are selected based on cumulative Eigen value percentage of 99%. For noise analysis first five PCA principal components are selected, irrespective of cumulative Eigen value percentage.

 Copyright ⓒ 2015 SERSC

4.3. Noise Analysis

Random Gaussian noise is artificially introduced directly after the normalize data sets and then the data sets are feed into the *MultiSpec* software for FE. Only signal to noise ratios (SNRs) of 5dB and 1dB are used for noise analysis, which can be calculated as

$$SNR = 10\log_{10}\left(\operatorname{var}(image)\big/\operatorname{var}(noise)\right) \tag{9}$$

When noises of SNR 5dB and 1dB are introduced, the first 5 principal components of PCA are selected irrespective of the cumulative Eigen value percentage to perform MPs, which results in a total of 45 EMPD bands. If we restrict ourselves to cumulative percentage of 99% then first 100 principal components need to be selected and the total numbers of spectral features are 103. With this large amount of principal components it will not be feasible to do EMPD. Therefore we restrict to first 5 principal components when noise is introduced. It is to be noted that when no noise is introduced cumulative percentage of 99% is achieve with only 3 principal components. This shows how badly the spectral features are affected by noise.

Table 2. WITHOUT FILTERING: Noise and Feature Analysis with SNR of 5dB and 1dB when NWFE Technique is used for FE

FE	NWFE 99%		
Noise (SNR)	No	5dB	1dB
Features (spectral, spatial)	33(12, 21)	133(101, 32)	135(101, 34)
OA	95.875	85.285	78.757
AA	96.571	88.218	84.013
kappa	0.9440	0.8043	0.7235

Table 3. WITHOUT FILTERING: Noise and Feature Analysis with SNR of 5dB and 1dB when DBFE Technique is used for FE

FE	DBFE 99%		
Noise (SNR)	No	5dB	1dB
Features (spectral, spatial)	68(42, 26)	101(73, 28)	105(76, 29)
OA	95.601	83.193	79.089
AA	96.010	87.966	84.788
kappa	0.9402	0.7782	0.7276

Table 2 shows noise effects on the spatial and spectral features of Pavia University dataset. When NWFE is used for FE with no noise, the total number of features is 33; 12 spectral and 21 spatial. This is quite low compare to the total number of spectral and spatial features of 148; 103 spectral and 45 spatial. When noise of 5dB and 1dB is introduced, the total number of features increases to 133 and 135 respectively. With 5dB noise the spectral features increases to 101 and spatial feature increases to 32 and with 1dB noise the spectral features increases to 101 and spatial features increases to 34. With cumulative percentage of 99% almost all the spectral features are included. Compare to spectral feature, spatial feature are affected less by noise. The OA classification accuracy also reduces to 85.285% and 78.757% when noises of 5dB and 1dB are introduced in the HSI respectively.

Table 3 shows the noise and feature analysis when DBFE technique is used for FE. When no noise is used the total number of features extracted using DBFE is 68; 42 from spectral and 26 from spatial. When noise of 5dB and 1dB is introduced in HSI the number of features increases to 101 and 105 respectively. With 5dB noise spectral features increases to 73 and spatial features increases to 28. Similarly with 1dB noise spectral features increases to 76 and spatial features increases to 29. It is to be noted that compare to NWFE the DBFE is less affected by noise. The spatial features are almost the same when noises of 5dB and 1dB are introduced compare to no noise case. The spectral information is less affected by noise as compare to NWFE.

Table 4. WITH FILTERING: Noise and Feature Analysis with SNR of 5dB and 1dB when NWFE Technique is used for FE

FE	NWFE 99%			
Feature (spectral, spatial)	133(101, 32)		135(101, 34)	
SNR	5dB		1dB	
Filtering	No	Yes	No	Yes
OA	85.285	99.961	78.757	99.930
AA	88.218	99.933	84.013	99.897
kappa	0.8043	0.9995	0.7235	0.9990

Table 5. WITH FILTERING: Noise and Feature Analysis with SNR of 5dB and 1dB when DBFE Technique is used for FE

FE	DBFE 99%			
Feature (spectral, spatial)	101(73, 28)		105(76, 29)	
SNR	5dB		1dB	
Filtering	No	Yes	No	Yes
OA	83.193	99.945	79.089	99.956
AA	87.966	99.899	84.788	99.917
kappa	0.7782	0.9993	0.7276	0.9994

Table 4 summarizes the noise and feature analysis for Pavia University data set when NWFE technique is used for FE. The OA increases to 99% when Gaussian filtering is performed just before the classifier SVM. Similarly Table 5 summarizes the noise and feature analysis for Pavia University data set when DBFE technique is used for FE. The OA increases to 99% when filtering is introduced just before classifier irrespective of the noise. Although we get the high accuracy but the total number of features are very high for both NWFE and DBFE techniques.

One of the main purposes of doing the FE is to reduce dimension. This we cannot achieve when feature mining NSFETs are introduced by noise. NSFETs are very sensitive to noise. Without noise they perform very well and give us very good classification accuracy with less number of features but when noise is introduced their performance goes down rapidly. The number of features to be extracted using NSFETs are very high. Therefore NSFETs may not be the good choice for FE in the harsh environment; where there are more chances of the presence of noise.

5. Conclusion

This paper explains the noise effects on two state of the art feature mining NSFETs named NWFE and DBFE. NSFETs are tested on spatial and spectral information extraction technique for HSI. It is analyzed that both FE techniques are very sensitive to noise. Comparing both, NWFE is more affected by noise than DBFE. When NWFE is used the spectral information is too much effected by noise resulting in high spectral features. In both techniques spatial features are less

 Copyright ⓒ 2015 SERSC

affected by noise than spectral feature. By using Gaussian Filtering just before SVM mitigates the effects of noise and we get the high classification accuracy results but this does not satisfy the one of the main purposes of feature extraction and that is feature reduction. Therefore it is concluded that NSFETs may not be the good choice for FE in the harsh environment.

References

[1] X. Jia, K. Bor-Chen and M. M. Crawford, "Feature Mining for Hyperspectral Image Classification", Proceedings of the IEEE, vol. 101, no. 3, **(2013)**, pp. 676–697.
[2] X. Jia and M. He, "Feature Mining for Hyperspectral Data", Invited Tutorial at IEEE GRSS 4th Workshop on Hyperspectral Image and Signal Processing, Shanghai, **(2012)**.
[3] H. Mingyi, "Signal and Information Processing in Hyperspectral Imaging Remote Sensing", Invited Distinguished Lecture IEEE ICIEA, **(2014)**.
[4] M. G. Kendall, "A course in the geometry of n-dimensions", Dover Publication, New York, **(1961)**.
[5] D. A. Landgrebe, "Signal Theory Methods in Multispectral Remote Sensing", John Wiley and Sons, New Jersey, **(2003)**.
[6] G. F. Hughes, "On the mean accuracy of statistical pattern recognizers", IEEE Transactions on Information Theory, vol. 14, no. 1, **(1968)**, pp. 55–63.
[7] P. Dollar, Z. Tu, H. Tao and S. Belongie, ''Feature mining for image classification", Proc. IEEE Conf. Computer Vis. Pattern Recognition, **(2007)**, pp. 1–8.
[8] Q. Liu, A. H. Sung, Z. Chen and J. Xu, "Feature mining and pattern classification for steganalysis of LSB matching steganography in grayscale images", Pattern Recognition, vol. 41, **(2008)**, pp. 56–66.
[9] M. Fauvel, J. A. Benediktsson, J. Chanussot and J. R. Sveinsson, "Spectral and Spatial Classification of Hyperspectral Data Using SVMs and Morphological Profiles", IEEE Trans. Geosci. Remote Sens., vol. 46, no. 11, **(2008)**, pp. 3804–3814.
[10] F. M. Imran and Mingyi He, "Extended Morphological Profiles with duality for hyperspectral image classification", to be published in Proc. IEEE WHISPERS, Tokyo, Japan, June 2-5, **(2015)**.
[11] 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.
[12] L. Chulhee and D. A. Landgrebe, "Feature Selection Based on Decision Boundaries", in Proc. IEEE IGARSS, vol. 3, **(1991)**, pp. 1471–1474.
[13] L. Chulhee and D. A. Landgrebe, "Feature extraction based on decision boundaries", Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 15, no. 4, **(1993)**, pp. 388–400.
[14] 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.
[15] R. C. Gonzalez and R. E. Woods, "Digital Image Processing", 3rd Edition, Prentice-Hall, Inc., New Jersey, **(2006)**.
[16] C. C. Chang and C. J. Lin, "LIBSVM: a library for support vector machines", http://www.csie.ntu.edu.tw/~cjlin/libsvm/, **(2011)**.

Author

Muhammad Imran Farid, 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.

Copyright ⓒ 2015 SERSC