

Variogram Fractal Dimension Based Features for Hyperspectral Data Dimensionality Reduction

Kriti Mukherjee · Jayanta K Ghosh ·
Ramesh C. Mittal

Received: 19 November 2011 / Accepted: 19 July 2012 / Published online: 4 August 2012
© Indian Society of Remote Sensing 2012

Abstract In this paper a new approach for fractal based dimensionality reduction of hyperspectral data has been proposed. The features have been generated by multiplying variogram fractal dimension value with spectral energy. Fractal dimension bears the information related to the shape or characteristic of the spectral response curves and the spectral energy bears the information related to class separation. It has been observed that, the features provide accuracy better than 90 % in distinguishing different land cover classes in an urban area, different vegetation types belonging to an agricultural area as well as various types of minerals belonging to the same parent class. Statistical comparison with some conventional dimensionality reduction methods validates the fact that the proposed method, having less computational burden than the conventional methods, is able to produce classification statistically equivalent to those of the conventional methods.

Keywords Hyperspectral · Fractal · Variogram · Dimensionality reduction · Computational complexity

K. Mukherjee (✉) · J. K. Ghosh
Civil Engineering Department,
Indian Institute of Technology,
Roorkee, Uttarakhand 247667, India
e-mail: mukherjee.kriti@gmail.com

R. C. Mittal
Mathematics Department, Indian Institute of Technology,
Roorkee, Uttarakhand 247667, India

Introduction

Spectral imaging of the earth using hyperspectral sensors designed on the principle of charged coupled devices (CCD) is a recent advancement of remote sensing. These CCDs are able to detect the energy reflected from moving earth objects in very narrow wavelength bands because of their high signal to noise ratio and very small discharge time. One image is obtained at each band of wavelengths and the total number of images in each scene is equal to the number of bands the sensors are designed with (generally within 10 and 1000). Thus, for each plot on the ground surface that is represented by the spatial resolution of the sensor, intensity values equal to the number of bands, determined by the spectral resolution of the sensor, can be obtained. The detailed spectral response of a pixel assists in providing accurate and precise information of the ground cover under measurement. This makes hyperspectral data useful to study subtly different classes and deal with applications like target recognition, anomaly detection and background characterization. Due to the huge data volume associated with each scene of hyperspectral data, this data type requires more specific attention to the complexity of data receiving, storing, transforming and processing. In particular, due to the high dimensionality of this data the analysis of the images becomes a complex problem. Some researchers studied the characteristics of the high dimensional space and their implications for hyperspectral-data analysis. Kendall (1961) proved that the volume in a hypercube has a tendency to concentrate

in the corners and in a hyperellipsoid in an outside shell. Consequently, the high-dimensional space is mostly empty. Furthermore, Hughes (1968) showed that with limited number of training samples, classifier performance improves with dimensionality initially and then declines. Moreover, Fukunaga (1989) proved that in given circumstances, the required number of training samples is linearly related to the dimensionality for a linear classifier and to the square of the dimensionality for a quadratic classifier; also this situation for nonparametric classifier like neural network gets worse. Under these circumstances and difficulties, a large number of classes of interest and a large number of available spectral bands need a large number of training samples, which unfortunately are expensive or tedious to acquire. As a result, either the class statistics must be estimated from the limited training sample set or otherwise different feature selection/ extraction based dimensionality reduction methods are required in order to get good classification accuracy. Some novel methods of classification of hyperspectral data have also been developed by various authors. Support vector machine (SVM) is one such tool to classify hyperspectral data much more efficiently even with a limited number of training samples and thus overcoming Hughes phenomenon (Hughes 1968). This is basically a linear learning machine based on the principle of optimal separation of classes. However, the high algorithmic complexity and extensive memory requirements are some serious limitations associated with SVMs (Horvath, 2003).

Feature selection based approaches reduce the dimensionality of the data by selecting a representative subset of the original features. Since only few of the all available bands are chosen, this approach always suffers from loss of original information. Feature extraction methods, on the other hand, preserve most of the desired original information (Pu and Gong 2004). However, many of the feature extraction algorithms project the data to a new coordinate system and the physical meaning of the original hyperspectral data, after transforming to the new coordinate system gets lost. Spectral space based feature extraction methods are another alternative, which operate on the spectral response curves of each pixel of hyperspectral data. The physical meaning of the data, i.e., the pattern of the spectral response curve (SRC) is taken into consideration in such dimensionality reduction algorithms and at the same time, no information from the original data is ignored or unattended since the features are generated using the whole of the SRC. Because of

these advantages of spectral space based dimensionality reduction methods, this research work has also been concentrated on developing a new methodology for dimensionality reduction using fractal mathematics and applied on each SRC of the hyperspectral data.

The fractal dimension computation method used in this dimensionality reduction approach is known as variogram method. Fractal dimension was first used to reduce the dimensionality of hyperspectral data by Ghosh and Somvanshi (2008) and there after applied by Ghosh et al. (2008), Ghosh et al. (2009) and Ghosh and Mukherjee (2009). They reduced the high dimensional hyperspectral data to a single dimension. This paper is a modification of the fractal based method and based on generating more than one fractal based features from the hyper spectral data, which are able to distinguish subtly different land cover classes.

Theory of Fractal Dimension

An object on the earth surface has a characteristic dimension associated with it. According to Mandelbrot (1977), if the object is having an irregular shape like a curved line, or like the coast of Britain, its dimension is not an integer, but a fraction. Formally, a fractal is defined as a set for which the Hausdorff-Besicovitch (or fractal) dimension strictly exceeds the topological dimension (Mandelbrot 1977). In short, fractal dimension measures the geometric complexity of an object by determining how completely the fractals embed themselves in normal Euclidean space (Mandelbrot, 1982). Hence fractal is used as a tool for spatial and contextual information extraction from remote sensing images. While a straight line has a dimension of exactly one, a fractal curve will have a dimension between one and two, depending on how much space it takes up as it curves and twists. However, a straight line and a crooked coast line both have same topological dimension 1 and a smooth surface and a rugged topographic surface both have a dimension 2. In other words, part of the information about the irregular objects is lost necessarily in top as follows.

$$Y_{MXN} = \begin{bmatrix} Y_{1,1} & Y_{2,1} & \dots & Y_{M,1} \\ Y_{1,2} & Y_{2,2} & \dots & Y_{M,2} \\ \dots & \dots & Y_{i,k} & \dots \\ Y_{1,N} & Y_{2,N} & \dots & Y_{M,N} \end{bmatrix}$$

For any pixel, y_i , the spectral response curve (SRC) can be expressed as the vector, $y_i = \{y_{i,1}, y_{i,2}, \dots, y_{i,k}, \dots, y_{i,N}\}^T$ and the bands can be expressed as another vector, $x = \{1, 2, \dots, k, \dots, N\}^T$, where T represents vector transpose operation. When the vector y_i is plotted across vector x by joining each successive points $(k, y_{i,k})$ by straight lines, the curve obtained is called the SRC of the pixel at the IFOV of the hyper spectral sensor. SRC thus obtained very closely represents the actual ground response pattern. The SRCs of hyperspectral data, being irregular, also have unique fractal dimension associated to them characterising their shape or degree of irregularity.

Variogram Fractal Dimension

The fractal dimension using variogram method is computed by taking a large number of pairs of points of different band gaps along the SRC and computing the differences in their radiance/relectance values. Then the fractal dimension is obtained from the log-log plot of (expected differences in radiance)² vs. distance in terms of number of bands between the point pairs (Klinkenberg 1994).

The variogram is actually a graph of the semi-variance (γ_h) (Davis 1973) of values given for points separated by a certain distance, given by Eq. 1 below.

$$\gamma_h = \frac{1}{2} [E(Z(x) - Z(x+h))^2] \tag{1}$$

Where, $Z(x)$ and $Z(x+h)$ are the spectral response values along the SRC at band number ‘x’ and ‘x+h’ respectively, where ‘h’ is the difference between the two bands. If all pairs of points along the curve lying at distance $\pm h$ are picked, the semivariance for all such ‘N’ number of points is given by

$$\gamma_h = \frac{1}{2N} \sum_{i=1}^N [Z(x_i) - Z(x_{i+h})]^2 \tag{2}$$

The plot of γ_h vs h is known as the semivariogram of the given curve. Thus, for computing fractal dimension using variogram approach, first step is to compute the semivariance γ_h for different band gaps ‘h’. Then $\log(\gamma_h)$ vs. $\log(h)$ is plotted and regression method is used to calculate the slope of the line, which is given

by $4-2D$, where D is the fractal dimension of the curve.

Preprocessing of the Hyperspectral Data

This step comprises of smoothing and interpolation of the hyperspectral data. Smoothing step is important to get a uniform fractal dimension for SRCs of the same class, but corresponding to different pixels. Since the hyper-spectral data is in general noise prone, often there are sharp transitions of spectral response values from one band to the next. These sharp transitions are generally due to noise present in the data. As a result, the SRCs of different pixels of the same class may be affected by noise differently and give wrong information about the class characteristics. The difference in SRCs of different pixels belonging to the same class would produce different fractal dimension values and the results would be highly erroneous. Therefore, it is very important to reduce the noise from the SRCs, in order to maintain uniformity of class response from different pixels and smooth out the response. To achieve this, a low pass filtering has been carried out on the data. In the first step, the mean spectral response over all the pixel response values of the hyperspectral data is computed and then the resultant SRC is transformed to its frequency domain by performing N point fast Fourier transform (FFT) on the SRC (Oppenheim et al. 2005). The frequency at which the cumulative power contributed just exceeds a user defined threshold percentage is selected as a range R for application of inverse Fourier transform. In the next step, FFT is again applied followed by inverse FFT over range R for all the SRCs of the hyperspectral data. Within range R the power contributed in the transformed domain are passed with no attenuation and all frequencies outside the range R are completely attenuated. Thus the smoothed version of the original SRC has been obtained.

Now, from the fractal dimension computation algorithm discussed in section 2.1, it is obvious that more number of data points help in defining the regression line more accurately. Keeping this in mind, the smoothed SRC has been interpolated by using cubic spline interpolation technique to increase the number of data points available in the SRC y to say, L (Ahlgberg et al. 1967). The new signal so obtained is expressed as z in the following section.

Proposed Methodology for Dimensionality Reduction

Computation of Fractal Dimension

A moving window of length say 'a' (where a is integer part of L/n, and n is the number of reduced dimension) has been defined. The mth window W_m is given by

$$W_m = u[p - (m - 1)a] - u[p - ma], \text{ for } m = 1, 2, \dots, n \quad (3)$$

Where, $u[p]$ is the conventional discrete time unit step function. The window has been moved along the interpolated SRC. Then, the mth portion (z_m) of the spline interpolated SRC on which D computation is performed is given by

$$z_m = z^* W_m, \text{ for } m = 1, 2, \dots, n. \quad (4)$$

Thus each z_m is a vector having 'a' data points. Variogram method for fractal dimension computation has then been applied on each z_m for $m=1,2,\dots,n$ and n fractal dimensions have been computed, each of them represented by D_m (for $m=1,2,\dots,n$), corresponding to n parts of the smoothed SRC.

Feature Generation

Each D_m obtained in section 4.1 has been multiplied by the energy associated with corresponding portion of the SRC. Since each SRC and therefore each part of the SRC is equivalent to a discrete time uniformly sampled signal, therefore the energy associated (E) with the signal in each window is given by (Oppenheim et al. 2005)

$$E_m = \sum_{j=1}^a z_{m,j}^2 \text{ for } m = 1, 2, \dots, n \quad (5)$$

where $z_{m,j}$ is the j'th data point value in m'th window of the spline interpolated SRC. Thus, m'th feature F_m in an SRC is given by

$$F_m = D_m \times E_m, \text{ for } m = 1, 2, \dots, n \quad (6)$$

Thus, n new features have been generated corresponding to n sections of the SRC reducing N dimensions of hyper spectral data to n dimensions.

The flow diagram of the methodology starting from smoothing is as shown in Fig. 1.

Optimum Reduced Dimensional Set Selection

The n fractal energy features generated by the proposed methodology have then been used for classification using Maximum likelihood classifier using the training pixels and the accuracy of classification has been obtained using the testing pixels. This procedure is followed by varying the number of features from 5 to 20 and the number of dimensions giving maximum overall accuracy has been defined as optimum reduced dimensional set.

Study Area and Data

Three study areas having increasing number of classes to be distinguished have been investigated.

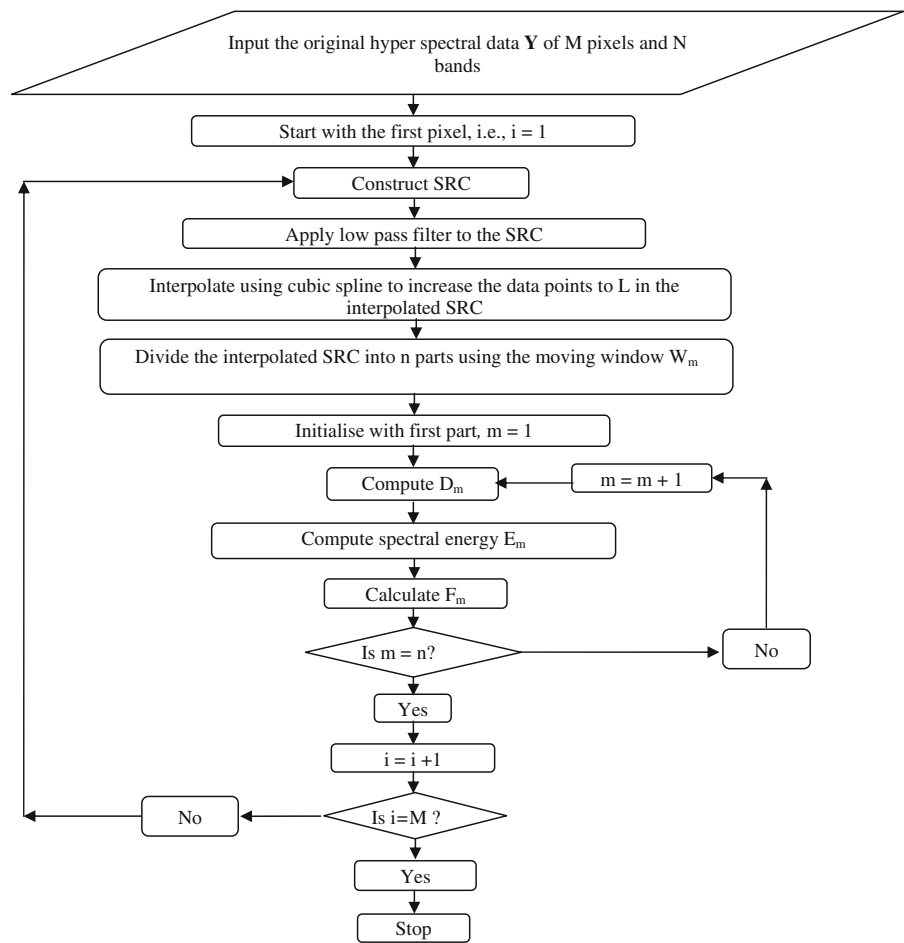
Study Area I

The data for first study area is of an urban area available at website <http://www.agc.army.mil/Hypercube/>. This data has been captured by HYDICE (Hyperspectral Digital Imagery Collection Experiment) sensor and is a 16 bit data having 307 pixels, 307 lines and 210 bands. HYDICE collects data of 210 bands over the range 0.413–2.504 microns. From the 210 bands, the water absorption bands and noisy bands have been removed resulting in 188 final bands for analysis. There are five land use/land cover classes in this image which are two types of road (represented as R1 (208 training and 213 testing samples used) and R2 (344 training and 369 testing samples used)), concrete (C) (449 training and 457 testing samples used), barren land (B) (330 training and 388 testing samples used) and trees (T) (203 training and 223 testing samples used). The first three classes, road 1, road 2 and concrete belong to the third level of Anderson's classification scheme (Anderson 1976), trees belong to the second level and barren land belongs to the first level.

Study Area II

The second study area of medium complexity is located in America and known as Indian Pines. The data is actually a portion of an AVIRIS (Airborne Visible InfraRed Imaging Spectrometer) image readily available online from the website <http://dynamo.ecn.purdue.edu/~biehl/Multispec/documentation.html>. The image had

Fig. 1 Methodology to generate fractal based features



been collected over NW Indiana’s Indian Pines test site in June 1992 and has 16 classes. AVIRIS hyperspectral data consists of 224 bands over the range 0.38–2.5 microns. The spectral resolution of the data is about 10 nm and the spatial resolution is 20 m. 4 among all the available bands contain no data and therefore have been removed, keeping the remaining 220 bands for analysis. The water absorption bands and noisy bands (104–108, 150–163, 220) have also been removed from the original data resulting in a total of 200 bands finally. The image is actually a subset of the original AVIRIS image and has 145 lines and 145 samples in each line. Among all the 16 classes, the number of labelled samples available for some of the classes is very less and has not been used in this study. 9 among the 16 classes have been chosen for classification and accuracy assessment. There are basically two types of classes of level I, i.e., agriculture and rangeland available in this image (Anderson 1976). All five crop types, two of corn type

(Corn notill (CN) and Corn Min (CM)) and three of soybean type (Soybeans-notill (SN), Soybeans min (SM) and Soybean clean (SC)) fall in level IV under row crops of cropland that belong to agricultural area. The number of training/testing samples used for the classes CN, CM, SN, SM and SC are respectively 231, 213, 266, 477 and 200 respectively. The crops in this image are in the early growth stage and therefore have only about 5 % crop cover. Remaining field area is soil covered with residue from previous year’s crop. The no-till, min and clean labels indicate the amount of crop residue remaining from previous year. No-till corresponds to a large amount of residue, min-till has a moderate amount and clean-till have the least amount of residue from the previous year. The other four classes, Grass/Pastures (GP), Grass/Trees (GT), Hay Windrowed (HW) and Woods (W) are the level III classes that fall under grassland which belong to the range land. The numbers of training/testing samples used for these

classes are 118, 210, 189 and 217 for GP, GT, HW and W respectively.

Study Area III

The third study area of highest complexity is a mining area which is also a portion of an AVIRIS image freely available online at <http://aviris.jpl.nasa.gov/html/aviris.freedata.html>. This is also a 16 bit data having 533 pixels, 477 lines and 224 bands. The image had been collected over Cuprite, Nevada in June 1997 and is a mining area in southern Nevada with mineral and little vegetation. The site, located approximately 200 km northwest of Las Vegas is a relatively undisturbed acid-sulphate hydrothermal system exhibiting well exposed alteration mineralogy consisting principally of kaolinite, alunite, and hydrothermal silica. The noisy and water absorption bands are removed to give 197 final bands for analysis. 14 different classes have been chosen from this study area for further analysis which are K-Alunite (KA) (202 training and 216 testing samples identified), Na-Alunite (25 training and 32 testing samples identified), Kaolinite-wxl (KW) (205 training and 208 testing samples identified), Kaolinite-pxl (KP) (200 training and 199 testing samples identified), Kaolinite+Smectite or Muscovite (KSM) (205 training and 202 testing samples identified), Hallyosite (H) (182 training and 181 testing samples identified), Alunite+Kaolinite and/or Muscovite (AKM) (214 training and 213 testing samples identified), Calcite (C) (237 training and 225 testing samples identified), Calcite+Montmorillonite+Kaolinite (CMK) (201 training and 204 testing samples identified), Na-Montmorillonite (Na-M) (227 training and 237 testing samples identified), Low/med-Al Muscovite (LM) (210 training and 209 testing samples identified), High Al Muscovite (HM) (228 training and 213 testing samples identified), Chalcodony (Ch) (207 training and 219 testing samples identified) and Nontronite (N) (200 training and 200 testing samples identified). All the classes, being minerals of various types belong to level IV of Anderson's classification scheme.

The data used in this study are the radiance data, in their original format, since any type of preprocessing may actually distort the shape of the class distribution in the data (Landgrebe 2003). If the training samples

are collected from the data itself, many of the observation variables are accounted for and there is a much reduced need for complex preprocessing. The number of training samples was chosen in such a way so that a minimum of 5 samples for each class are available after dimensionality reduction.

Results

The smoothed SRCs obtained by applying the preprocessing step mentioned in section 3 have been interpolated to generate SRCs having $M=600$ data points so that a minimum of 30 data points are available for each segment of the SRC when the number of reduced dimension chosen by the user is maximum, e.g., 20. Now, starting with reduced dimension as $n=5$, the number of data points 'a', for computing D within the moving window W_m , have been found to be 120 [$a = \lfloor 600/5 \rfloor$]. F_m is then computed for $m=1, 2, \dots, 5$, i.e., at each position of W_m . Thus, five features from each SRC have been computed resulting original N dimensions reduced to $n=5$ dimensions. A maximum likelihood classifier (MLC) is then used to classify land cover classes using these five features. The same steps mentioned in section 5 are again followed for $n=6, 7, \dots, 20$ and classification accuracies have been computed for each data set. The optimum number of reduced dimension was found as 10 for study area I giving a classification accuracy of 94.24 %, 18 for study area II giving classification accuracy of 94.20 % and 14 for study area III giving classification accuracy of 96.23 %.

The number of features of the optimum feature sets were again varied from 2 (since MLC classifier requires at least 2 features for classification) to the maximum number of features available in the optimum reduced dimensional set with an increase of 1 feature in each step for each data set. The plots of overall accuracy vs. number of features for each study area using the optimum reduced dimensional set are shown in Fig. 2. It is observed from Fig. 2 that maximum accuracy of 94.42 % has been achieved using the first 7 features of the optimum reduced dimensional set for study area I, using all the 18 features for study area II and all the 14 features for study area III. Thus it can be concluded that for simple data, where the number of classes to be distinguished is low and the class spectra are also widely different from each other, such as study area I, the

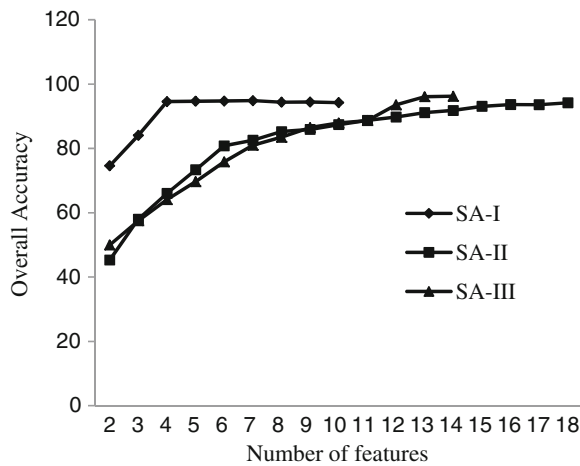


Fig. 2 Variation of overall accuracy with number of features of the optimum reduced dimensional set for Study Area I, II and III

number of required features is less. However, when the data is more complex, i.e., the number of classes is more and they are spectrally very similar to each other, then more number of informative features is necessary for producing accurate classification.

Next, to validate the proposed method based dimensionality reduction, the results of classification using the proposed method have been compared with those of some other conventional methods of dimensionality reduction, such as principal component analysis or PCA (Davis 1973), minimum noise fraction or MNF (Green et al. 1988) and Independent Component Analysis or ICA (Robila 2004). In each case the number of features chosen for comparison was the same as the number of features producing maximum overall accuracy using the proposed method. The overall accuracy results obtained by using the proposed and conventional dimensionality reduction methods on all three study areas are given in Table 1.

Further, a statistical test to check the equivalence between the proposed method and the conventional methods mentioned above has been performed. The

test is based on the number of correctly and incorrectly classified pixels from both classification results under comparison (Foody, 2004). A two by two matrix is formed for comparing each two classification results as shown in Table 2.

In the next step, the statistical significance of the difference between two classification results is computed using the following equation.

$$\text{Confidence interval} = p_1 - p_0 \pm z_{\alpha/2}(\text{SE}) \tag{7}$$

where, p_1 is the proportion of correctly classified samples of proposed method based classification and p_0 is the same for the classification using the conventional method. If $\alpha=0.05$ be chosen as the zone of indifference, $z_{\alpha/2}=1.96$ from statistical table. SE is the standard error of the estimate and is given by (refer Table 2)

$$SE = \frac{\sqrt{f_{12} + f_{21} - (f_{12} - f_{21})^2/n}}{n} \tag{8}$$

If the confidence interval lies within the zone of indifference then the two classification results under comparison are not significantly different from each other at 95 % significance level and vice versa (Foody 2009).

The results using the above method are shown in Table 3 for study area I, II and III. It can be observed from Table 3 that for study area I, the match in the number of correctly classified pixels is maximum with PCA (1552) followed by ICA (1534) and MNF (1506). The number of correctly classified pixels common for both classifications under comparison for study area II shows that here MNF and proposed method produced maximum number of common correctly classified pixels (1948) followed by PCA (1929) and ICA (1890). Similarly for study area III, the number of common correctly classified pixels of the proposed method based result is maximum with

Table 1 Overall accuracy values using the proposed and conventional methods of dimensionality reduction

Method	Overall Accuracy (%)		
	Study area I (9 features)	Study area II (18 features)	Study area III (14 features)
Proposed	94.42	94.20	96.23
PCA	95.21	94.20	96.45
MNF	94.67	95.71	97.61
ICA	96.30	92.46	96.51

Table 2 Assessment of the significance of difference between two classifications using McNemar test

		Proposed method (C1)		
Conventional method (C2)		Correct	Incorrect	Total
Correct		f_{11}	f_{12}	$f_{11} + f_{12} = \sum_{i=1}^k n_{ii}$ of C2
Incorrect		f_{21}	f_{22}	$f_{21} + f_{22} = n - \sum_{i=1}^k n_{ii}$ of C2
Total		$f_{11} + f_{21} = \sum_{i=1}^k n_{ii}$ of C1	$f_{12} + f_{22} = n - \sum_{i=1}^k n_{ii}$ of C1	n

MNF (2634), followed by ICA (2611) and PCA (2607).

Next, based on the values of Table 3, Fig. 3 has been produced. It has been found that all the confidence intervals lie within the zone of indifference in Fig. 3. Thus, differences between the proposed method and the conventional methods are actually insignificant statistically. It is also established that classification based on proposed method and those from conventional methods are equivalent.

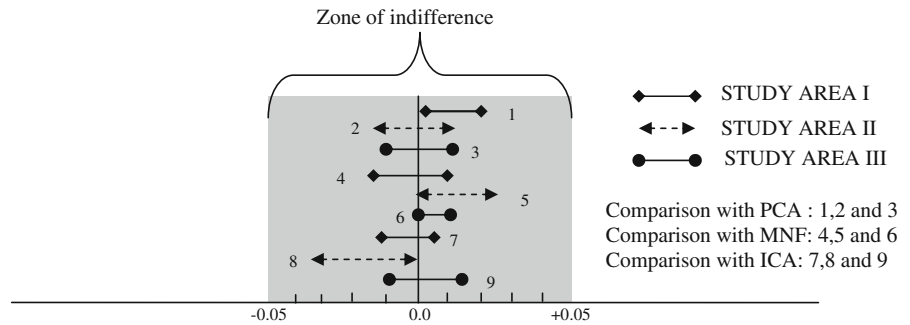
Another important fact is that the conventional methods of dimensionality reduction mentioned above are based on matrix operations. The aim of PCA is to compute the eigen values and eigen vectors of the covariance matrix of the N dimensional hyperspectral data to extract the transformed components. The complexity involved in the whole process is given by $O(MN^2 + N^3)$ (Moigne et al 2002). MNF is noise adjusted PCA where the first step is to determine the noise covariance matrix (Σ_N) and then the noise to signal

covariance matrix is obtained ($\Sigma_N \Sigma^{-1}$). Finally eigen values and eigen vectors of this matrix are computed to find the new components. Thus clearly the computation involved in MNF is more than PCA. One matrix inverse operation involves complexity of the order of $O(N^3)$ (Cormen et al 2001) leading to total complexity of $O(2 N^3 + MN^2)$ i.e., twice as complex as PCA. In ICA, the aim is find the independent source signals (S) from observed spectral response (X) by the use of an unmixing matrix (W). It starts from an initial unmixing matrix and iteratively goes on computing S till it fulfils certain criteria. The complexity of ICA varies from $O(N^4)$ to $O(N^5)$ where N is the dimension (Comon 1994). Whereas, the proposed fractal dimension based method of dimensionality reduction are applied on each pixel response and based on three simple steps. In the first step the whole SRC is segmented into Q sections. On each section the fractal dimension computation algorithm is applied. Finally the fractal dimension of each section is multiplied by

Table 3 Comparison of VF based proposed and other conventional methods of dimensionality reduction

		Study area I			Study area II			Study area III		
Conventional methods		VFD								
		C	IC	T	C	IC	T	C	IC	T
PCA	C	1552	22	1574	1929	69	1998	2607	53	2660
	IC	13	63	76	69	54	123	47	51	98
MNF	C	1506	49	1555	1948	82	2030	2634	58	2692
	IC	59	36	95	50	41	91	20	46	66
ICA	C	1534	19	1553	1890	71	1961	2611	51	2662
	IC	31	66	97	108	52	160	43	53	96
	T	1565	85	1650	1998	123	2121	2654	104	2758

Fig. 3 Confidence interval of proposed and conventional dimensionality reduction methods for study area I, II and III



corresponding spectral energy associated with the section. Variogram method is based on computation of variance followed by least square linear regression, both of which are functions of the number of data points P and involve complexity of $O(P^2)$. The same computation is performed for each of the Q segments and the complexity of the proposed method for each pixel is $O(QP^2)$, and for M pixels the complexity is given by $O(MQP^2)$. It is therefore evident that the complexity associated with the proposed method is much less than those of the conventional methods of dimensionality reduction mentioned above, since both Q and P are less than N. Table 4 shows the complexity of the conventional and proposed methods.

Conclusion

In this paper, a new approach for dimensionality reduction method for hyperspectral data has been proposed based on variogram fractal dimension based features. The approach is unsupervised in the sense that the user need not have any prior knowledge regarding the training samples or number of classes present in the data for application of the steps for dimensionality reduction. It is only required to provide the number of features in the reduced dimension as input. However, to choose the optimum feature set, knowledge of reference ground truth data is required since the choice is based on classification accuracy achieved with each set of features. It

Table 4 Computational Complexity of Conventional and Proposed Method of Dimensionality Reduction

Method	Algorithm used	Complexity
PCA	$\sum_{PC} = W_{PC} \sum W_{PC}^T$ Where Σ is the covariance matrix of the N dimensional hyperspectral data. By using equation $(\sum -\lambda_k I) \vec{e}_k = 0$, Where, $\lambda_k =$ k'th eigen value, $\vec{e}_k =$ k'th eigen value	$O(MN^2 + N^3)$ (Moigne et al 2002) Where, M is the number of pixels and N is the number of bands in the original hyperspectral data.
MNF	$\sum_{MNF} = W_{MNF} \sum_N \sum^{-1} W_{MNF}^T$ Where, Σ_N is the noise covariance matrix.	$O(2 N^3 + MN^2)$
ICA	$X = AS + n$, Where, X is the pixel vector (known), A is the mixing matrix (unknown), S is source (Unknown). Finds Unmixing matrix W iteratively such that $S = WX$ and each S is independent	$O(N^4)$ to $O(N^5)$ (Comon, 1992)
VFD	Computed the semivariance γ_h for different distances 'h'. Plotted $\log(\gamma_h)$ vs. $\log(h)$ to obtain variogram fractal dimension	$O(MQP^2)$

has been observed that the features obtained by multiplying the Variogram fractal dimension values with the energy of the corresponding signal are able to provide very good classification accuracy at a reduced cost of computational complexity. Therefore, it may be concluded that proposed variogram fractal energy feature based method is a viable alternative for reduction of dimensionality of hyperspectral data.

References

- Ahlberg, J.H., Nilson, E.N., & Walsh, J.L. (1967). The theory of splines and their applications. Academic Press: Newyork and London, ch 2.
- Anderson, J.R. (1976). A land use and land cover classification system for use with remote sensor data. URL:<http://landcover.usgs.gov/pdf/anderson.pdf>
- Comon, P. (1994). Independent component analysis, a new concept? *Signal Processing*, 36, 287–314.
- Comen, T.H., Leiserson, C.L., Rivest, R.L., Stein, C. (2001) Introduction to algorithms, 2nd Edition, MIT Press, McGraw Hill Book Co.
- Davis, J.C. (1973) Statistics and data analysis in geology, 2nd edition. John Wiley & Sons, pp 239–246
- Foody, G. M. (2004). Thematic map comparison: evaluating the statistical significance of difference in classification accuracy, *Photogrammetric Engineering and Remote Sensing* 70(5), 627–633.
- Foody, G. M. (2009). Classification accuracy comparison: hypothesis tests and the use of confidence intervals in evaluations of difference, equivalence and non-inferiority. *Remote Sensing of Environment*, 113, 1658–1663.
- Fukunaga, K. (1989). Effect of sample size in classifier design. *IEEE Pattern Analysis and Machine Intelligence*, 11(8), 873–885.
- Ghosh, J.K. & Mukherjee, K. (2009) Fractal based dimensionality reduction of hyper-spectral images using power spectrum method, advances in computational vision and medical image processing : methods and applications. In J. M. R. S. Tavares, R. M. Natal Jorge (Eds), Springer Publication, pp 89–94. (ISBN: 978-0-415-57041-1)
- Ghosh, J. K., & Somvanshi, A. (2008). Fractal based dimensionality reduction of hyperspectral images. *Journal of Indian Society of Remote Sensing*, 36, 235–241.
- Ghosh, J. K., Somvanshi, A., & Mittal, R. C. (2008). Fractal feature for classification of hyperspectral images of Moffit field, U.S.A. *Current Science*, 94, 356–358.
- Ghosh, J. K., Singh, A., & Mukherjee, K. (2009). Detection of anomaly in water body using hyperspectral images. *Journal of Geomatics*, 3(2), 53–56.
- Green, A. A., Berman, M., Switzer, P., & Craig, M. D. (1988). A transformation for ordering multispectral data in terms of image quality with implications for noise removal. *IEEE Transactions on Geoscience and Remote Sensing*, 26(1), 65–74.
- Horvath, G. (2003). *Neural networks in measurement systems, advances in learning theory: methods, models and applications* (p. 392). In J. A. K. Suykens, G. Horvath, S. Basu, C. Micchelli, J. Vandewalle (Eds.), Amsterdam, The Netherlands: IOC Press.
- Hughes, G.F. (1968). On the mean accuracy of statistical pattern recognizers. *IEEE Transactions on Information Theory*, 14 (1)
- Kendall, M.G. (1961). A course in the geometry of n-dimensions. Hafner Publishing
- Klinkenberg, B. (1994). A review of methods used to determine the fractal dimension of linear features. *Mathematical Geology*, 26(1), 23–46.
- Landgrebe, D.A. (2003) Signal theory methods in multispectral remote sensing. John Wiley & Sons
- Mandelbrot, B. B. (1977). *The fractal geometry of nature*. New York: W.H. Freeman and Company.
- Mandelbrot, B. B. (1982). *The fractal geometry of nature*. San Francisco: W.H. Freeman and Company.
- Moigne, J. L., Cole-Rhodes, A., Eastman, R., El-Ghazawi, T., Johnson, K., Kaewpijit, S., Laporte, N., Morissette, J., Nethanyahu, N. S., Stone, H. S., & Zavorin, I. (2002). Multiple sensor image registration, image fusion and dimension reduction of earth science imagery. *IEEE Proceedings of the fifth international conference on information fusion*, 2, 999–1006.
- Oppenheim, A.V., Schafer, R.W., & Buck, J.R. (2005). Discrete time signal processing, 2nd Edn, Pearson Education.
- Pu, R., & Gong, P. (2004). Wavelet transform applied to EO-1 hyperspectral data for forest LAI and crown closure mapping. *Remote Sensing of Environment*, 91, 212–224.
- Robila, S.A., ed (2004). Advanced image processing techniques for remotely sensed images. Springer-Verlag.