A New Hyperspectral Image Classification Approach Using Fractal Dimension of Spectral Response Curve

S. Abolfazl Hosseini, Hassan Ghassemian, Senior Member IEEE
Faculty of Electrical and Computer Engineering
Tarbiat Modares University
Tehran, Iran
abolfazl.hosseini@modares.ac.ir, ghassemi@modares.ac.ir

Abstract—In classification of hyperspectral images, methods like PCA, LDA, and ICA have simple structure and relatively good results, but an important deficiency is that these methods are not sensitive to the order of primary features. For every pixel of a hyperspectral image we have a vector of measured quantities corresponding to reflection coefficients of consecutive wavelengths which is called as spectral reflectance curve (SRC). So the ordaince of measured data might have some information that could be useful in classification. SRCs due to the high degree of snuggle they have, could be considered as fractals. In this paper we suggest a new approach of fractal dimension feature extraction based on FD of adjacent overlapping intervals and using their principal components as feature vector elements. These new features applied as inputs of a statistical per pixel classifier of segmented image, produced by K-means clustering method, whit Mahalanobis distance. A majority voting step ended classification process. We achieved meaningful improvement of correct classification rate comparing to classic PCA method.

Keywords: Classification, Feature Extraction, Fractal Dimension, Hyperspectral Image, Spectral reflectance curve.

I. INTRODUCTION

Airborne and space-borne imaging of earth surface started at the middle decades of last century in the form of panchromatic and multispectral pictures. Later, hyperspectral imagery technology was introduced as a combination of imaging and spectrum technique. This kind of imaging is composed of hundreds of spatially coregistered images corresponding to different spectral channels and construct a fine spectral signature or spectral response curve (SRC) by obtaining narrow and almost continuous spectrum of electromagnetic wave from visible light to short wave infrared (SWIR). Fig. 1 illustrates a hyperspectral image cube and a typical SRC belong to a pixel of image [1]. SRCs show distribution of the energy in terms of wavelengths reflection coefficients. These spectral responses are -in many cases- unique to each type of LC and could show how much the LCs are different from each other, so different LCs can be discriminated based upon difference in their SRCs. For its high spectral resolution, each pixel of hyperspectral image can represent spectrum curve of ground-objects in a good way. The procedure of recognition ground-objects characteristics by analysing SRCs, and then classifying them is named as spectral feature analysis. If one hyperspectral image consists of N bands, any pixel corresponds to one point in N-dimensional feature space, and each component of the point coordinate is proportional to the reflectivity coefficient of pixel at a distinct wave band. The large number of spectral bands potentially can improve ability of classification, but it also creates some problems. Increasing spectral band number could increase classification accuracy when those bands are useful in discriminating the classes. In hyperspectral images, spectral bands are associated with narrow and sometimes, overlapping bandwidths. For this reason adjacent bands have correlation, and the information of this kind of images has some degree of redundancy. So the dimensionality could be reduced without losing significant information and separability among classes [2]. Finding the optimum dimension of subspace (Optimum number of features) for classification depends on many factors like the number of training samples, feature extraction approach (supervised or unsupervised), classification system (supervised / unsupervised, pixel based / region based), and type of classifier (statistical / non-statistical, parametric / nonparametric). Some of the famous feature extraction algorithms are: Principal Component Analysis (PCA), Independent Component Analysis (ICA), feature extraction using projection pursuit, feature extraction by decision boundary, linear discriminant analysis (LDA) and one dimensional wavelet [3,4,5].
PCA is a linear transformation which transforms data from primitive space to a feature space with linear independent components. This new space has orthogonal axis so that their directions are defined by covariance matrix feature vector of training samples. Since eigenvalue of covariance matrix is proportional to the power of data with corresponding eigenvector direction, we can obtain an effective dimension of data by specifying nonzero eigenvalues [6]. PCA guarantees that in remained features the suitable data for detecting objects are kept. The difference between ICA and PCA is that the obtained features of PCA are linear independent while the ones for ICA are statistically independent. ICA such as PCA has linear and nonlinear methods [7, 8]. Projection pursuit algorithm can be defined as a search to get the most optimum space for linear projection of high dimensional data to the low dimensional space by optimizing a criterion. This criterion can be parametric or nonparametric. Hall criterion is an example of nonparametric criterions [9]. These indicators generally need a large number of training data to compute. Decision boundary feature extraction algorithm acts based on Bayesian decision theory and neglecting redundant features. But its computation load is so complex [10]. LDA tries dispersion of each class in obtained features become less and the distance between classes become more. Fisher ratio is used for this purpose. The important problem of this algorithm reveals when the mean between classes is low which cause having invalid feature extractions. Moreover, if mean of one class has a much difference with the other classes, it affects to compute covariance matrices of classes. Consequently, it will not tend to an effective feature extraction. The other problem of this algorithm is that if there are m classes, the maximum feature number which can be extracted is m-1 [11]. Therefore, the final dimension is limited by applying this algorithm.

Although methods like PCA, LDA, and ICA have simple structure and relatively good results, but there are some deficiencies when they are used in hyperspectral Image classification. In addition to what was mentioned in previous paragraph, an important deficiency is that these methods are not sensitive to the order of primary features. For every pixel of a hyperspectral image we have a vector of measured quantities corresponding to reflection coefficients of consecutive wavelengths. So the ordinance of measured data might have some information that could be useful in classification. It is clear that two sequences which have equal values in different places don’t have the same FD. In addition, FD is a measure of snuggle. SRCs belong to different LCs have non equal degree of crumpleness and FD could represent this specification well. Unfortunately this type of information isn’t utilised by many feature extraction methods. Hosseini and Ghassemian [12] introduced a method that utilises the information of data ordinance, based on fractal dimension of each pixel SRC. They showed that SRC of groundpixels in hyperspectral images, due to the high degree of snuggle they have, could be considered as fractals and explained by their fractal properties, especially by their FD.

They showed that dividing every SRC to several nonoverlapping adjacent intervals and constructing a vector containing FD of these intervals yields to considerable classification results. Specially, better performance was achieved when these fractal dimension features merged with some of principal component of original data points to use in classification.

In this paper we suggest a new approach of fractal dimension feature extraction based on FD of adjacent overlapping intervals and using their principal components as feature vector elements. Block diagram of classification scheme is depicted in Fig. 2. As a pre-classification step, a segmentation process has been performed, using K-MEANS clustering technique in original data space. Experiments results showed improvement in correct classification rate and some other assessment parameters.

In section II the concept of FD and a simple way for calculating FD is presented. Spectral data sets used in this research are introduced in section III. Also, this section contains a brief explanation of proposed method and the results of our experiments. Then it will be terminated with discussion about our experiments results. At the end of paper you find conclusion part.

II. FRACTAL DIMENSION

For many of seemingly complex forms found in nature (like coastlines, mountains and clouds) there are both a description and a mathematical model provided by Mandelbrot’s geometry [13]. In the manner of Mandelbrot, who used the term fractal for the first time, a fractal is a set for which the Hausdorff-Besicovich dimension exceeds the topological dimension strictly. In another word a fractal is a shape made of parts which are similar to the whole in some way.

The basic quality of natural fractals is statistical self-similarity which is quantified by fractal dimension (FD), but many phenomena can be considered as fractals because they have a crumple nature. Fig.3 illustrates three typical SRCs.
corresponding to three pixels belonging to different LCs of a hyperspectral image and Crumple nature of them is depicted in this figure. Fortunately fractal geometry as a new branch of mathematics is appropriate for describing irregular crumple shapes of real world.

One of the central concepts of fractal geometry is self-similarity or scaling and it is closely connected with our intuitive notion of dimension. A one dimensional object is an object which can be divided into N identical parts such that each part is a scaled down version of the whole by the ratio of r=1/N. continuing this manner a D dimensional object is an object which can be divided into N self-similar parts, each of which is scaled down by ratio of 

\[ r = \frac{1}{\sqrt[N]{D}} \]  

or \[ N = \frac{1}{r^D} \]  

and (by generalisation) the fractal (or self-similarity) dimension can be defined as: 

\[ D = \frac{\log N}{\log \frac{1}{r}} \]

Some of methods for calculating FD are: Box counting, walking-Driver, Prism counting, Epsilon-Blanket, Perimeter-area Relationship, Fractional Brownian motion, Power spectrum, and Hybrid methods.

In practice for computing FD of a curve we must map it into unit square and then use one method to calculate its FD. In this paper we use Hausdorff dimension (D_h) of a set in a metric space which is given by [14]:

\[ D_h = \lim_{\epsilon \to 0} \frac{-\ln[ N(\epsilon) ]}{\ln(\epsilon)} \]  

Where N(\epsilon) is total number of radius \epsilon open balls we need to cover the set. In case of a line of length L consisting of segments of length 2\epsilon each, there will be N(\epsilon) = L/(2\epsilon) segments in the line. Thus, (1) will be changed to:

\[ D_h = \lim_{\epsilon \to 0} \frac{-\ln(L) + \ln(2\epsilon)}{\ln(\epsilon)} \]  

\[ = \lim_{\epsilon \to 0} \left[ 1 - \frac{\ln(L) - \ln(2\epsilon)}{\ln(\epsilon)} \right] \]  

\[ = \lim_{\epsilon \to 0} \left[ 1 - \frac{\ln(L)}{\ln(\epsilon)} \right] \]  

In this study, the SRC of a LC, a planar curve in a space, was considered as the metric space. Then it was scaled linearly into a normalized space, to map the original curve into an equivalent metric space. The first scaling normalizes band number of data point in the abscissa as:

\[ x_i^* = \frac{x_i}{x_{max}} \]  

The second scaling normalizes the ordinate as:

\[ y_i^* = \frac{y_i - y_{min}}{y_{max} - y_{min}} \]  

Where y_i is reflectance coefficient value of the i-th band of curve and y_{min} and y_{max} are the minimum and maximum values of the sample values respectively, of the considered curve over the entire band numbers. These two linear scaling map the N values of the SRC into another space that belongs to a unit square. This unit square may be visualized as covered by a net of N(sample number) x N (sample value) cells. Each of them contains one point of the scaled SRC. Calculating L of the scaled SRC and taking \epsilon = 1/(2 x N') [where N' (the number of segments) = N - 1], (2) becomes:

\[ x_i^* = \frac{x_i}{x_{max}} \]  

\[ y_i^* = \frac{y_i - y_{min}}{y_{max} - y_{min}} \]  

In this study, the SRC of a LC, a planar curve in a space, was considered as the metric space. Then it was scaled linearly into a normalized space, to map the original curve into an equivalent metric space. The first scaling normalizes band number of data point in the abscissa as:

\[ x_i^* = \frac{x_i}{x_{max}} \]  

The second scaling normalizes the ordinate as:

\[ y_i^* = \frac{y_i - y_{min}}{y_{max} - y_{min}} \]  

Where y_i is reflectance coefficient value of the i-th band of curve and y_{min} and y_{max} are the minimum and maximum values of the sample values respectively, of the considered curve over the entire band numbers. These two linear scaling map the N values of the SRC into another space that belongs to a unit square. This unit square may be visualized as covered by a net of N(sample number) x N (sample value) cells. Each of them contains one point of the scaled SRC. Calculating L of the scaled SRC and taking \epsilon = 1/(2 x N') [where N' (the number of segments) = N - 1], (2) becomes:

\[ x_i^* = \frac{x_i}{x_{max}} \]  

\[ y_i^* = \frac{y_i - y_{min}}{y_{max} - y_{min}} \]  

In this study, the SRC of a LC, a planar curve in a space, was considered as the metric space. Then it was scaled linearly into a normalized space, to map the original curve into an equivalent metric space. The first scaling normalizes band number of data point in the abscissa as:

\[ x_i^* = \frac{x_i}{x_{max}} \]  

The second scaling normalizes the ordinate as:

\[ y_i^* = \frac{y_i - y_{min}}{y_{max} - y_{min}} \]  

Where y_i is reflectance coefficient value of the i-th band of curve and y_{min} and y_{max} are the minimum and maximum values of the sample values respectively, of the considered curve over the entire band numbers. These two linear scaling map the N values of the SRC into another space that belongs to a unit square. This unit square may be visualized as covered by a net of N(sample number) x N (sample value) cells. Each of them contains one point of the scaled SRC. Calculating L of the scaled SRC and taking \epsilon = 1/(2 x N') [where N' (the number of segments) = N - 1], (2) becomes:

\[ x_i^* = \frac{x_i}{x_{max}} \]  

\[ y_i^* = \frac{y_i - y_{min}}{y_{max} - y_{min}} \]  

In this study, the SRC of a LC, a planar curve in a space, was considered as the metric space. Then it was scaled linearly into a normalized space, to map the original curve into an equivalent metric space. The first scaling normalizes band number of data point in the abscissa as:

\[ x_i^* = \frac{x_i}{x_{max}} \]  

The second scaling normalizes the ordinate as:

\[ y_i^* = \frac{y_i - y_{min}}{y_{max} - y_{min}} \]  

Where y_i is reflectance coefficient value of the i-th band of curve and y_{min} and y_{max} are the minimum and maximum values of the sample values respectively, of the considered curve over the entire band numbers. These two linear scaling map the N values of the SRC into another space that belongs to a unit square. This unit square may be visualized as covered by a net of N(sample number) x N (sample value) cells. Each of them contains one point of the scaled SRC. Calculating L of the scaled SRC and taking \epsilon = 1/(2 x N') [where N' (the number of segments) = N - 1], (2) becomes:
\[ D_h = \lim_{N \to \infty} \left[ 1 - \frac{\ln(L)}{\ln(1/2N)} \right] \]
\[ = \lim_{N \to \infty} \left[ 1 - \frac{\ln(L)}{\ln(1) - \ln(2N)} \right] \] (5)
\[ = \lim_{N \to \infty} \left[ 1 + \frac{\ln(L)}{\ln(2N)} \right] \]

Therefore, the fractal feature:
\[ D_h = 1 + \frac{\ln(L)}{\ln(2N)} \] (6)

Where:
\[ L = \sum_{i=1}^{N} \text{dist}(i, i+1) \] (7)

The approximation to \( D \) expressed in (5), improves as \( N \to \infty \).

III. EXPERIMENTS AND DISCUSSION

A. Study Regions

Three hyperspectral data sets, whose Specification is mentioned in this section, have been considered for our experiments.

**Salinas scene**

A high geometric resolution (3.7-meter pixels) picture over Salinas Valley, California which is collected by AVIRIS sensor in 224 bands. 20 water absorption bands, in this case bands: [108-112], [154-167], and 224 are discarded. Picture dimensions are 512*217 and radiometric resolution is 12 bits. Salinas’s groundtruth differentiates 16 classes. It includes vegetables, bare soils, and vineyard fields.

**DC Mall scene**

This hyperspectral image was photographed by HYDICE sensor with 210 bands and is related to a region of Washington State. 19 noisy bands are omitted. Image dimensions are 276*307 pixels and radiometric resolution is 8 bits. This image contains 6 classes of roof, road, lawn, tree, water and shade.

**Pavia University scene**

This 115 spectral bands picture (from 0.43 to 0.86 \( \mu \)m spectral range) is acquired by the ROSIS sensor over university of Pavia, Italy. Picture dimensions are 610*610 pixels, but some of the samples contain no information and have to be discarded before the analysis. Also 12 most noisy channels have been removed. The spatial resolution of picture is 1.3 meters. Image groundtruth contains 9 classes.

As an example, colour images of Pavia university scene is illustrated in Fig. 4. The figure is produced by computing mean intensity of visible light bands, and assigning them to blue, green and red colours.

B. Method, Experiments and Classification Results

Before applying our feature extraction method, we performed a K-MEANS clustering algorithm for the pictures as a pre-classification step. After that, each pixel SRC is divided to some overlapping intervals with length \( L \) (3 \( \leq \) \( L \) \( \leq \) N, N is hyperspectral image band number) and then we computed FD of each interval. By this way, an \( m \) dimensional vector of FDs (\( m= N-L+1 \)) was constructed. For example if \( N=204 \), as Salinas data is, and \( L=34 \), then band intervals are :{(1-34),(2-35)…(171-204)} and \( m=171 \). After that we used first \( D \) (1 \( \leq \) \( D \) \( \leq \) 12) principal components of these new features for per pixel classification step, employing a statistical classifier whit Mahalanobis distance. After that, a majority voting over each segment has been done to assign all pixels of each pixel to one class. Correct Classification Rate (CCR) was compared to the state that principal components of original data have been used and have been shown in Tables I-III. Also Fig.5 illustrates effect of \( L \) on CCR when we use 12 PC of FD features for Pavia scene. As it could be seen, the best values of \( L \) are in range (10-20) and it is almost true for other dimension, but for other images, this range varies. Fig.6 depicts diagram of best values of \( L \) in terms of feature number. Also in Fig. 7 the class map of algorithm for 12 feature numbers and \( L=75 \) for Pavia image is shown.

C. Discussion

As it could be found from results, FD features of pixel SRC in hyperspectral image have capability of extracting the information which exists in ordinance of data point.
Employing this information causes a considerable improvement in CCR. Also as it could be seen from TABLE II, using these new features will delete the ripple of CCR that produces by PCA for some kinds of data. The main reason for this effect is that SCR of pixel in hyperspectral images are very crumple and feature extraction methods like PCA, do not pay attention to this fact. By using fractal dimension of overlapping segments on SRC, we formed a new arrangement of data sets and then by performing principal component analysis over this new arrangement, the redundancy of data has been omitted. Therefore we exploit the advantages of feature reduction besides of advantages of sequence discipline. Results of experiment are considerably acceptable both in low and high radiometric resolution data set. Although the computational load of feature extraction step increases due to FD calculation, however, the PCA part of algorithm takes less time because size of new feature vector is smaller than original data vector. For example when N=204 and L=34, then m=171 which means 15% size reduction for input vector of PCA algorithm. This yields time consumption in calculating of Inverse matrix for PCA technique.

IV. CONCLUSION

A new method for feature extraction in hyperspectral images was introduced. The method tries to keep two kinds of information which don’t be kept by conventional feature extraction methods.

**TABLE I: results for Salinas Data**

<table>
<thead>
<tr>
<th>Number of features</th>
<th>CCR for Principal components of primary data</th>
<th>Best CCR for principal components of FD features</th>
<th>L for best result</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>48.76</td>
<td>71.25</td>
<td>17</td>
</tr>
<tr>
<td>2</td>
<td>81.68</td>
<td>87.02</td>
<td>34</td>
</tr>
<tr>
<td>3</td>
<td>87.98</td>
<td>92.01</td>
<td>101</td>
</tr>
<tr>
<td>4</td>
<td>91.16</td>
<td>92.75</td>
<td>102</td>
</tr>
<tr>
<td>5</td>
<td>91.76</td>
<td>92.14</td>
<td>125</td>
</tr>
<tr>
<td>6</td>
<td>91.61</td>
<td>92.98</td>
<td>122</td>
</tr>
<tr>
<td>7</td>
<td>91.49</td>
<td>93.38</td>
<td>106</td>
</tr>
<tr>
<td>8</td>
<td>91.53</td>
<td>93.69</td>
<td>106</td>
</tr>
<tr>
<td>9</td>
<td>91.69</td>
<td>93.68</td>
<td>101</td>
</tr>
<tr>
<td>10</td>
<td>93.78</td>
<td>93.88</td>
<td>70</td>
</tr>
<tr>
<td>11</td>
<td>93.84</td>
<td>93.98</td>
<td>71</td>
</tr>
<tr>
<td>12</td>
<td>93.86</td>
<td>93.79</td>
<td>75</td>
</tr>
</tbody>
</table>

**TABLE II: results for DC Mall Data**

<table>
<thead>
<tr>
<th>Number of features</th>
<th>CCR for Principal components of primary data</th>
<th>Best CCR for principal components of FD features</th>
<th>L for best result</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>72.48</td>
<td>76.20</td>
<td>14</td>
</tr>
<tr>
<td>2</td>
<td>98.46</td>
<td>95.45</td>
<td>23</td>
</tr>
<tr>
<td>3</td>
<td>98.44</td>
<td>96.34</td>
<td>93</td>
</tr>
<tr>
<td>4</td>
<td>96.53</td>
<td>98.90</td>
<td>73</td>
</tr>
<tr>
<td>5</td>
<td>73.91</td>
<td>97.95</td>
<td>23</td>
</tr>
<tr>
<td>6</td>
<td>91.33</td>
<td>99.30</td>
<td>53</td>
</tr>
<tr>
<td>7</td>
<td>95.91</td>
<td>99.17</td>
<td>53</td>
</tr>
<tr>
<td>8</td>
<td>78.70</td>
<td>99.57</td>
<td>53</td>
</tr>
<tr>
<td>9</td>
<td>93.24</td>
<td>99.49</td>
<td>53</td>
</tr>
<tr>
<td>10</td>
<td>93.35</td>
<td>98.65</td>
<td>53</td>
</tr>
<tr>
<td>11</td>
<td>93.43</td>
<td>99.49</td>
<td>53</td>
</tr>
<tr>
<td>12</td>
<td>85.49</td>
<td>94.05</td>
<td>53</td>
</tr>
</tbody>
</table>

**TABLE III: results for Pavia University Data**

<table>
<thead>
<tr>
<th>Number of features</th>
<th>CCR for Principal components of primary data</th>
<th>Best CCR for principal components of FD features</th>
<th>L for best result</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>45.55</td>
<td>47.24</td>
<td>57</td>
</tr>
<tr>
<td>2</td>
<td>68.60</td>
<td>62.08</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>68.60</td>
<td>66.04</td>
<td>57</td>
</tr>
<tr>
<td>4</td>
<td>70.30</td>
<td>71.99</td>
<td>20</td>
</tr>
<tr>
<td>5</td>
<td>74.28</td>
<td>72.72</td>
<td>21</td>
</tr>
<tr>
<td>6</td>
<td>76.59</td>
<td>80.87</td>
<td>13</td>
</tr>
<tr>
<td>7</td>
<td>75.67</td>
<td>83.13</td>
<td>13</td>
</tr>
<tr>
<td>8</td>
<td>77.00</td>
<td>84.72</td>
<td>11</td>
</tr>
<tr>
<td>9</td>
<td>78.01</td>
<td>84.72</td>
<td>13</td>
</tr>
<tr>
<td>10</td>
<td>78.29</td>
<td>84.63</td>
<td>12</td>
</tr>
<tr>
<td>11</td>
<td>77.59</td>
<td>85.22</td>
<td>15</td>
</tr>
<tr>
<td>12</td>
<td>80.18</td>
<td>84.77</td>
<td>20</td>
</tr>
</tbody>
</table>

First, information of sequence discipline (ordinance of reflectance coefficients in spectral response curve) and the second, information existing in crumpleness of SRC.
Experiments were performed by using of segmented images produced through a K-MEANS clustering step and combination of spatial and spectral information by majority voting over segments.

Figure 7: class map of Pavia scene produced by proposed method

ACKNOWLEDGMENT

This research is done by support of Iran communication research centre under contract number 18133/500 T by Identification code: 90-01-03. The authors gratefully acknowledge that organization for its support.

REFERENCES