Comparative Analysis of Scattering and Random Features in Hyperspectral Image Classification

Nikhila Haridas*, V Sowmya, Soman K P

Centre for Excellence in Computational Engineering and Networking, Amrita Vishwa Vidyapeetham, Amritanagar, Coimbatore-641112, India

Abstract

Hyperspectral images (HSI) contain extremely rich spectral and spatial information that offers great potential to discriminate between various land cover classes. The inherent high dimensionality and insufficient training samples in such images introduce Hughes phenomenon. In order to deal with this issue, several preprocessing techniques have been integrated in processing chain of HSI prior to classification. Supervised feature extraction is one such method which mitigates the curse of dimensionality induced by Hughes effect. In recent years, new strategies for feature extraction based on scattering transform and Random Kitchen Sink have been introduced, which can be used in context of hyperspectral image classification. This paper presents a comparative analysis of scattering and random features in hyperspectral image classification. The classification is performed using simple linear classifier such as Regularized Least Square (RLS) accessed through Grand Unified Regularized Least Squares (GURLS) library. The proposed approach is tested on two standard hyperspectral datasets namely, Salinas-A and Indian Pines subset scene captured by NASAs AVIRIS sensor (Airborne Visible Infrared Imaging Spectrometer). In order to show the effectiveness of proposed method, a comparative analysis is performed based on feature dimension, classification accuracy measures and computational time. From the comparative assessment, it is evident that classification using random features achieve excellent classification results with less computation time when compared with raw pixels(without feature extraction) and scattering features for both the datasets.

Keywords: Classification; Feature Extraction; Hyperspectral; Random Kitchen Sink; Scattering transform

* Corresponding author. Tel.: +0-999-500-8969
E-mail address: nikhila.haridas92@gmail.com
1. Introduction

The recent advance in sensor technology in the field of remote sensing has made it possible to capture images in a broad range of the electromagnetic spectrum, covering the visible and infrared region (0.4 to 2.5 μm). Hyperspectral sensors acquire data of a scene in hundreds to thousands of contiguous narrow spectral bands with high spectral and spatial resolution. Among the various hyperspectral image processing tasks, HSI classification has drawn much attention in recent years. The abundant spatial and spectral information present in hyperspectral images allows identification and discrimination of materials in the scene with great precision. However, the well-known difficulty in classification of hyperspectral images is Hughes phenomenon (curse of dimensionality) where, only limited training information is available for higher number of spectral channels. Moreover, the presence of redundant information and the increase in computational time due to high dimensionality may have a bad effect on HSI classification. Thus, the use of preprocessing techniques such as denoising, feature extraction, dimensionality reduction etc. has become an integral part of hyperspectral image classification systems.

In literature, a number of different approaches exist for preprocessing and feature extraction in HSI. Jiangfeng Bao et al. in\(^1\) used spectral derivative features for classification of hyperspectral data. Other well-known techniques for preprocessing in hyperspectral imagery include Principal Component Analysis (PCA)\(^2\), Singular Value Decomposition (SVD) and Independent Component Analysis (ICA). Yuan Yan Tang et al. in\(^3\) proposed three dimensional scattering wavelet transform which captures the spectral spatial information for HSI classification. Ali Rahimi et.al\(^4\) introduced Random Kitchen Sink algorithm to significantly reduce the computation time needed for training kernel machines and for improving the classification accuracy. Inmaculada Dpido et al. in\(^5\) performed a comparative study of different unmixing based feature extraction techniques for HSI classification.

This paper presents the performance comparison of different feature extraction techniques in HSI classification. Three different features of hyperspectral images including raw pixels, scattering features and random features were considered for the assessment. The parameters used in the analysis of proposed work are feature dimension, classification accuracy and computational time. The study demonstrates that Random Kitchen Sink is a fast and efficient feature extraction mechanism which improves the performance of large scale HSI classification. Rest of the paper is outlined as follows. An overview of feature extraction techniques is given in section 2. A brief introduction to GURLS library is presented in section 3. Section 4 describes the methodology of the proposed work. Section 5 reports the experimental results and analysis and finally, section 6 concludes the paper.

2. Feature Extraction Techniques

This section gives a brief description about the conceptual basics of the feature extraction algorithms namely, Scattering Transform and Random Kitchen Sink.

2.1. Scattering Transform

An important characteristic required for features extracted for classification is that, they need to be translation invariant which do not affect the human ability to recognize. Apart from translation invariance, signal representation also need to be continuous with respect to transformations such as scaling or rotations and should contain enough information in order to discriminate between various signal classes. Scattering features\(^6\) build informative signal representations through a nonlinear unitary transform which are stable and invariant with respect to small deformations. Scattering transform is therefore an efficient method to extract features based on the translational, rotational and convolutional invariance of image pixels. These informative invariant features are well adapted for image classification. The rich information present in hyperspectral image pixels is scattered along multiple paths implemented in a convolutional network architecture. A scattering vector is computed by cascading wavelet transform convolutions, modulus operators and local averaging. An open source software library called ScatNet is available for the computation of scattering vector. The detailed mathematical description of scattering operator is given by Bruna and Mallat\(^6\). Scattering Network output is given by,
where ’x’ is the input pixel vector, \( \psi \) is the wavelet, \( \phi \) is the scaling function and \( \lambda \) is the scale factor.

2.2. Random Kitchen Sink

Random Kitchen Sink is an efficient algorithm used to generate randomized features whose inner products approximate popular kernels\(^4\). The key idea behind the use of Random Kitchen Sink algorithm is that, it captures nonlinear properties of the data by mapping it to a high dimensional feature space probably an infinite dimensional space by using a mapping function, \( \phi \). Consider RBF kernel in defining Random Kitchen Sink algorithm. The expression for RBF kernel is given as

\[
k(x, y) = \langle \phi(x)\phi(y) \rangle = e^{-\|x-y\|^2} = e^{-\|\Sigma^{-1/2}(x-y)\|^2} = e^{-z^2} = k(z)
\]

where \( z = x - y \) i.e. the kernel function is completely dependent on the difference between two data points and \( \Sigma = \sigma^{-1}I \) is the covariance matrix and \( I \) is an \( N \times N \) identity matrix with mean zero. This can be equivalently written in the form of

\[
k(x, y) = \frac{1}{\sqrt{m}} \begin{bmatrix}
\cos(x^T \Omega_1) \\
\cos(x^T \Omega_2) \\
\vdots \\
\cos(x^T \Omega_m)
\end{bmatrix}
\cdot \frac{1}{\sqrt{m}} \begin{bmatrix}
\cos(y^T \Omega_1) \\
\cos(y^T \Omega_2) \\
\vdots \\
\cos(y^T \Omega_m)
\end{bmatrix} = \langle \phi(x), \phi(y) \rangle
\]

That is, each data point is mapped to cosines and sines to form a 2\( m \)-tuple vector. Further, the inner product of the mapped data points is used to approximate the Gaussian kernel. The detailed mathematical description of RKS is given by Ali Rahimi et.al\(^4\).

3. GURLS- A Least Squares Library for Supervised Learning

Grand Unified Regularized Least Squares commonly known as GURLS is a software library used for supervised learning. The library is suited for multi-category/multi-label problems for machine learning tasks such as classification, regression etc. The fact that it reduces a learning problem into a simple linear algebraic problem makes it computationally efficient for learning tasks. GURLS library exploits the use of latest tools in linear algebra for generating solution\(^7\). The library is based on the Regularized Least Squares (RLS) which can easily handle large scale/high dimensional datasets which are difficult to manage for other linear learning methods. The favourable properties of RLS makes it an obvious choice as compared to many existing learning algorithms which go complicated when dealing with multi-classification problems.

Regularized Least Squares: RLS takes the input hyperspectral data and calculates the weight matrix \( W \) required for the classification. The objective function can be equivalently expressed in terms of the Frobenius norm of a matrix as,

\[
\min_W \|WX - Y\|_F^2 + \lambda \|W\|_F^2
\]
This expression is equivalent to
\[
\min_{W} \text{tr}((WX - Y)(WX - Y)^T) + \text{tr}(\lambda W^TW)
\]  
That is weight matrix \(W\) can be obtained as,
\[
W_{p \times n_b} = Y_{p \times N}^T X_{N \times n_b}^T (X_{N \times n_b} X_{N \times n_b}^T + \lambda I_{n_b \times n_b})^{-1}
\]
where \(X\) is the input hyperspectral data, \(Y\) is the label vector, \(p\) is the number of class and \(n_b\) represents the number of bands.

4. Methodology

**Preprocessing**: In this stage, the entire 3D hyperspectral cube, \(X \in \mathbb{R}^{n_1 \times n_2 \times n_b}\) is loaded for processing. Each image in HSI cube is of size \(n_1 \times n_2\) and contains total \(n_b\) bands. Highly noisy bands in the HSI data are removed manually in advance as they lack detailed information for land cover classification. For the ease of processing, the data is further represented in 2D form as \(\tilde{X} \in \mathbb{R}^{n_b \times N}\) where, \(N\) is the total number of pixels in the image. Each pixel vector \(x_i\) belongs to \(\mathbb{R}^{n_b}\).

**Feature Extraction**: Scattering Transform and Random Kitchen Sink are the two feature extraction techniques used.

- **Scattering transform**: Scattering features are extracted from the 2D hyperspectral image, \(X\) using ScatNet software package implemented in matlab. The scat function takes each pixel, \(x_i\) for \(i = 1, 2, 3, \ldots, N\) as its input along with set of linear operators corresponding to wavelet transforms. The informative signal representations of input data can be computed by varying the linear operators such as wavelet operators. Furthermore, scattering features can be generated by varying the filter types and its parameters. Morlet filter 2D is used in this experiment. The scattering coefficients are computed with a cascade of wavelet modulus operators and is implemented in a convolutional network architecture where the wavelet operator gives filter coefficients.

- **Random Kitchen Sink**: Random features are generated using Random Kitchen Sink algorithm where the input hyperspectral data is mapped to a higher dimensional feature space i.e., each pixel vector \(x_i\) is mapped to \(k\) dimensional space where \(k >> n_b\) by taking the sines and cosines of each pixel. When the data is projected to a higher dimensional feature space, the data separability increases which helps in better classification.

**Classification**: The data points and labels for training and testing are generated from the extracted features and ground truth information. Let the training data, training labels, testing data and predicted labels be represented by \(X_tr, Y_tr, X_te\) and \(Y_pr\) respectively. The weight matrix, \(W\) is calculated from training data \(X_tr\), and training labels, \(Y_tr\) according to the Eq. 6. With the calculated weight matrix, \(W\) the output class labels are predicted using the equation \(Y_pr = WX_te\). The predicted labels are used for quantitative and qualitative assessment.

5. Experimental Results

This section describes about the dataset, accuracy assessment measures and experimental results obtained by the proposed technique.

5.1. Dataset Description

The experiment is performed on the on two standard hyperspectral datasets namely, Salinas A and Indian Pines subset. Salinas A contains 86 x 83 pixels located within the Salinas scene at [samples, lines] = [591-676, 158-240]. The ground truth contains 6 classes which are mutually exclusive. A subset portion of the Indian Pines image, called the subset scene was also used to evaluate the performance of the proposed method. The subset is a part of 145 x 145 scene and contains pixels located within Indian pines scene at [27-94] [31-116]. The size of subset portion is 68 x 86 with 224 spectral channels. The subset scene contains 4 different classes with similar spectral signatures.
5.2. Accuracy Assessment Measures

In hyperspectral image classification, accuracy assessment is done both qualitatively and quantitatively. In qualitative assessment, a thematic map generated with the classification is considered to be accurate if it gives an unbiased representation of the ground truth image. Classification error in this case is the visual discrepancy between the generated thematic map and the reality. In order to quantify the accuracy assessment, a confusion matrix or error matrix is generated for the entire testing data. The classification accuracy measures such as Class wise Accuracy, Overall Accuracy, Average Accuracy and Kappa Coefficient may be derived from the confusion matrix.9.

5.3. Results and Discussion

This section presents comparative and quantitative assessment of different feature extraction methods in hyperspectral image classification. The study investigates the role of scattering (SF) and random features (RF) for improving the classification accuracy in HSI. In order to show the effectiveness of the feature extraction, raw pixels of hyperspectral image are also adopted for comparison. The analysis is based on the feature dimension, quantitative accuracy assessment measures and computational time.

5.3.1. Analysis of feature dimension

The role of feature dimension in increasing the classification accuracy is evaluated by empirically varying the number of input features given for classification. The training sets for the experiment are generated by randomly selecting samples from each class in equal ratio. The experiment is carried out on both the datasets with 10%, 20%, 30% and 40% training set. Fig. 1 and 2 shows the classification maps of the same. All the pixels other than the background pixels were used for validation. In the experiment with Salinas-A, a total of 5348 pixels were used to generate the training set excluding the background pixels. From the obtained results, it is clear that Salinas A produce excellent classification results for all the three input features even with 10% training set. The experiments with Indian pines subset scene demonstrates that better classification results are attained only with the larger training set (40%). This is because of the similarity in spectral signatures of various classes which leads to misclassification. The analysis of feature dimension shows that classification accuracy increases with the increase in feature dimensionality for both the datasets. Also, higher classification accuracies are obtained for Salinas A with limited number of training samples.

5.3.2. Analysis of accuracy assessment results

Based on the aforementioned training sets, quantitative assessment measures such as Overall Accuracy (OA), Classwise Accuracy (CA), Average Accuracy (AA) and Kappa Coefficient (K) were computed for both the data sets. Classification maps were also generated in each case for qualitative assessment. In order to assure statistical stability, the average classification results of three runs are given. Table 1 and 2 shows the comparison of land cover classification results of Salinas-A and Indian pines subset with 10%, 20%, 30% and 40% training set. It is worth noting that, CA, OA, AA and K of random features dominates for both datasets compared to other feature extraction approaches. Even though Salinas A produce satisfactory results without feature extraction (OA=81.80%, AA=81.95% and K = 0.7753), use of scattering and random features surpass the results produced by raw pixels. In the case of Indian pines subset scene, significant improvement in classification accuracy is obtained after incorporating feature extraction i.e for 40% training set OA improved from 61.78% to 74.92% and 93.34% by using scattering and random features. Despite the fact that scattering features improves the classification accuracy when compared to pixel features, they are computationally complex and time consuming. In comparison with raw pixels, the increments in OA for Salinas A dataset with limited training set (10%) of scattering and random features are 4.69% and 16.51% respectively and in the case Indian pines subset scene, OA shows significant increment of 19.6% and 41.22% for scattering and random features respectively.
Fig. 1: Training set, Classification map (without feature extraction (Pixels), Scattering Features (SF), Random Features (RF)) of Salinas A scene. C1, C2, C3, C4, C5, C6 represents the corresponding class numbers as given in Table 1.

<table>
<thead>
<tr>
<th>Class name</th>
<th>Percentage of training set</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10%</td>
</tr>
<tr>
<td>1 Broccoli green weeds</td>
<td>81.58</td>
</tr>
<tr>
<td>2 Corn senesced green weeds</td>
<td>76.99</td>
</tr>
<tr>
<td>3 Lettuce_romaine_4wk</td>
<td>88.14</td>
</tr>
<tr>
<td>4 Lettuce_romaine_5wk</td>
<td>85.31</td>
</tr>
<tr>
<td>5 Lettuce_romaine_6wk</td>
<td>78.33</td>
</tr>
<tr>
<td>6 Lettuce_romaine_7wk</td>
<td>81.35</td>
</tr>
</tbody>
</table>

Overall Accuracy: 81.80%
Average Accuracy: 81.95%
Kappa Coefficient: 0.775

Table 1. Land cover classification results for the Salinas A scene for different feature dimension.
Fig. 2: Training set, Classification map (without feature extraction (Pixels), Scattering Features(SF), Random Features(RF)) of Indian Pines subset scene. C1, C2, C3, C4 represents the corresponding class numbers as given in Table 2.

Table 2. Land cover classification results for the Indian pines subset scene for different feature dimension

<table>
<thead>
<tr>
<th>No</th>
<th>Class name</th>
<th>Percentage of training set</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>10%</td>
</tr>
<tr>
<td></td>
<td>Pixels</td>
<td>SF</td>
</tr>
<tr>
<td>1</td>
<td>Corn notill</td>
<td>36.81</td>
</tr>
<tr>
<td>2</td>
<td>Grass trees</td>
<td>57.12</td>
</tr>
<tr>
<td>3</td>
<td>Soybean notill</td>
<td>52.73</td>
</tr>
<tr>
<td>4</td>
<td>Soybean mintill</td>
<td>49.44</td>
</tr>
<tr>
<td></td>
<td>Overall Accuracy</td>
<td>48.37</td>
</tr>
<tr>
<td></td>
<td>Average Accuracy</td>
<td>32.68</td>
</tr>
<tr>
<td></td>
<td>Kappa Coefficient</td>
<td>0.328</td>
</tr>
</tbody>
</table>
5.3.3. Analysis of computation time

This section describes about the computation time required by each method for both the data sets. The experiment is performed on Intel(R) Core(TM) i7-4790S CPU, 8.00 GB memory, and 64 bit OS using MATLAB 2013. The running time includes the time required for feature extraction and classification. Table 3 shows the time analysis of raw pixels, scattering features and random features and Fig 3 shows the time analysis plot for classification of Salinas A. Among the three methods, random features (RF) have faster prediction time. The analysis clearly shows that scattering features (SF) are highly time consuming compared to raw pixels and random features.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Feature</th>
<th>Time/sec for different % of training data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>Salinas A</td>
<td>Pixels</td>
<td>20.59</td>
</tr>
<tr>
<td></td>
<td>SF</td>
<td>1024.59</td>
</tr>
<tr>
<td></td>
<td>RF</td>
<td>14.60</td>
</tr>
<tr>
<td>Indian Pines</td>
<td>Pixels</td>
<td>18.78</td>
</tr>
<tr>
<td></td>
<td>SF</td>
<td>996.40</td>
</tr>
<tr>
<td></td>
<td>RF</td>
<td>12.82</td>
</tr>
</tbody>
</table>

Table 3. Time analysis of feature extraction techniques

6. Conclusion

This paper presents comparative analysis of feature extraction techniques such as scattering transform and Random Kitchen Sink for supervised land cover classification in hyperspectral images. In the experiment, we propose the benefit of feature extraction for improving the classification accuracy of remotely sensed hyperspectral data. The study demonstrates that scattering and random features accelerates the performance of simple linear classifiers such as Regularized Least Squares. Among the three methods (raw pixels, scattering transform and Random Kitchen Sink) used for comparison, random features proved to be efficient and powerful feature extraction technique with less computation time and excellent classification results. The scattering transform is observed to be computationally complex and time consuming method for feature extraction for high dimensional classification. As future work, the feature extraction techniques such as scattering transform can be implemented in GPU to save the processing time.

References