A Graph-Based Band Selection Method for Hyperspectral Images Using Correlation Matrix

Jintu Kumar Das, Christopher D. Tholou, Alok Anand Minz, and Sonia Sarmah

Abstract Images are classified by analyzing the numerical properties of image features and then are organized into categories. Classification algorithms are typically performed in two phases i.e., training and testing. The number of training samples needed to design a classifier increases with the dimension of feature vector and it is challenging to determine if all features are necessary for the classifier. Therefore, we need a way to reduce the dimension of the feature vector without losing any important information. Remotely sensed hyperspectral images contain hundreds of spectral bands which provide detailed information about the objects. But this increased dimension at the same time also increases the computational complexity of classification. In this paper, we present a graph-based approach for band selection using correlation matrix and mutual information for dimension reduction of hyperspectral images so as to decrease the computational complexity during classification.

Keywords Band selection · Hyperspectral dataset · Correlation matrix · Mutual information · Feature reduction

1 Introduction

Hyperspectral remote sensing is an emerging and multidisciplinary field with many applications such as geology, ecology, atmospheric science, and forensic science. It provides spatial and spectral information simultaneously. The hyperspectral images are represented in a three-dimensional data cube (x, y, λ) for processing and analysis, where x and y represent two spatial dimensions of the scene, and λ represents the spectral dimension [2]. The hyperspectral imaging covers an extensive spectral range providing high potential for discrimination of subtle differences in ground covers. However, due to this high dimensionality, the classification performance for hyperspectral images decreases and may suffer from the curse of dimensionality [9]. As a result, we need to reduce the dimensionality of the such images without losing the

J. K. Das (✉) · C. D. Tholou · A. A. Minz · S. Sarmah
School of Technology, Assam Don Bosco University Azara, Guwahati 781017, Assam, India
e-mail: jintukumardas@gmail.com

original information. Feature reduction is the transformation that maps the data from a high order dimension to a low order dimension [4]. Feature reduction can be implemented with feature selection or feature extraction. Different techniques are already introduced in the past for band selection [3, 4, 6, 7, 11] to find crucial and significant bands present in a hyperspectral image. One of the techniques introduced a supervised feature extraction method based on the discriminant analysis (DA) [4] which uses the first principal component (PC1) to weight the scatter matrices. A graph-based feature reduction method was proposed in [11] which uses super-pixels as input to the proposed method and Simple Linear Iterative Clustering (SLIC) is performed followed by Laplacian Eigenmaps (LE). In [6] authors proposed a feature extraction method where the input hyperspectral images were segregated into multiple subsets containing adjacent bands. Later the bands were merged together by averaging. In subsequent steps, these merged bands were further processed with recursive filtering giving the resulting feature for classification. Local binary pattern (LBF) [6] of the extracted features are formed thereby increasing classification accuracy.

In this paper, we have discussed a graph-based feature reduction method based on the concept of mutual information and a correlation matrix. In the proposed method, band correlation is calculated considering each band as a vertex. Edges are created between bands having equal or greater correlation values than a predefined threshold value. In the next phase, connected components of the graph have been extracted and from each component, the band having the highest mutual information with respect to the ground truth is selected. The process eventually results in a reduced dataset comprising of only significant bands.

2 Background

2.1 Correlation

Correlation is the measure of similarity between two signals [8]. Correlation between two variables X and Y can be found using the formula,

$$r_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{X})^2 \sum (y_i - \bar{Y})^2}} \tag{1}$$

where, r_{xy} is the correlation coefficient of the variables X and Y. x_i and y_i represent the ith value of X and Y respectively in corresponding samples. $\bar{X}$ and $\bar{Y}$ are the mean of the values of X and Y respectively.

2.2 Mutual Information

Mutual Information is the measure of how much a random variable is related to another [10]. The formal definition of mutual information of two random variables X and Y is given by,

$$I(X; Y) = \sum_{y \int Y} \sum_{x \int X} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \tag{2}$$

where, p(x,y) is the joint distribution of X and Y.

3 Graph and Connected Components

Graph—A graph can be defined as $G = (V, E)$ where V represents the set of vertices $v_1, v_2, \ldots$ etc and E represents the set of edges e1, e2, … etc. Each edge is a pair between two vertices (v_i, v_j).

Connected Components—In graph theory, connected components of an undirected graph is a subgraph where any two pair of vertices are connected by at least one path [5] (Fig. 1).
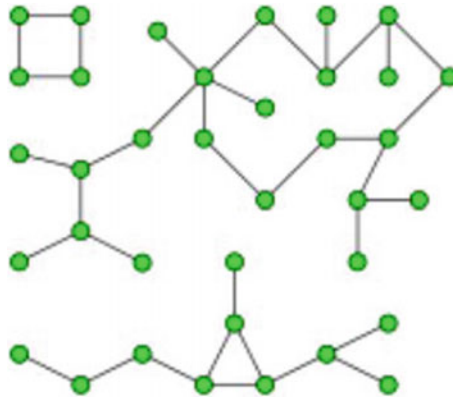
Fig. 1 A graph with 3 connected components

4   Proposed Methodology

In this proposed method, we have introduced a simple graph-based feature selection method using mutual information and a correlation matrix to select the significant bands from a hyperspectral image. The algorithm consists of two phases-graph construction and band selection. The overall time complexity is $O(X * Y * Z) + O(V + E)$ where x, y, z are the height, width, the number of bands present in the image cube and V, E represents the vertices (the bands) and E represents the edges present in the graph. The two phases are discussed in detail in the following sub sections.

4.1   Construction of the Graph

Initially for the construction of the graph the hyperspectral dataset of dimension (x, y, λ), corresponding ground truth and a predefined threshold value of correlation coefficient are taken as input. Then a graph G is constructed considering each band of the input hyperspectral dataset as a vertex. Thus the number of vertices in G is equal to the number of bands in the hyperspectral i.e. λ. An edge is added between a pair of vertices(bands) in G if the correlation coefficient between those two vertices is greater than the input threshold.

4.2   Finding the Connected Components and Band Selection

In this phase, from graph G (constructed in Sect. 1), the connected components are extracted. From each connected component, we select the vertex(band) having the highest mutual information score with ground truth. So, if there are $k(k \leq \lambda)$ connected components in the graph then the total number of selected bands is also 0k0. Finally, a reduced dataset is constructed considering only the selected bands (Fig. 2).

5   Experimental Setup

5.1   Dataset Description

For carrying out the experiment, we have taken 3 datasets acquired by various sensors namely- Indian pines(corrected), Pavia University and Salinas-A. The Indian Pines(corrected) scene was gathered by AVIRIS (Airborne Visible/Infrared Imaging Spectrometer) sensor and consists of 145×145 pixels, 200 spectral bands and 16 identified classes. Salinas-A dataset consists of 86*83 pixels and 204 spectral bands
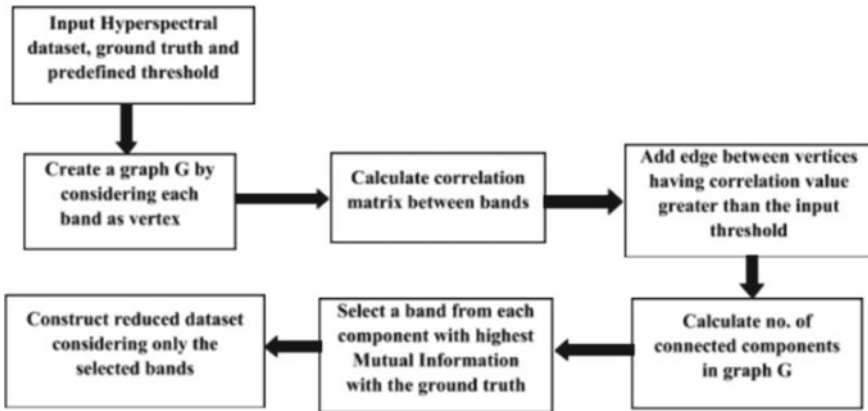
Fig. 2 Proposed method block diagram

and includes 6 classes. Pavia University Scene contains 103 spectral bands, 610*340 pixels and 9 classes [1]. For each dataset a series of experiments with three different threshold values of correlation coefficient—0.95, 0.97, 0.98.

6 Classifier Used

For classification purpose, two classifiers namely Support Vector Machine and Convolutional Neural Network were used separately.

In case of SVM, multiclass SVM with one against all strategy and linear kernel was used for training and testing purpose. The C parameter was set to 1.0 and gamma was set to auto.

A hybrid CNN classifier was also used for the classification. There were 10 hidden layers, 1 input and 1 output layer. The kernel size was taken as (3,3) and number of epoch chosen was 10. For the fully connected layers, the number of neurons were 256 and 128.

7 Evaluation Metrics

For training and testing the SVM classifier tenfold cross-validation method was used. Cross-validation is used to estimate the skill of machine learning model4.2. For CNN the training and testing data was divided into 70% and 30%, respectively. The following evaluation metrics were used to evaluate the performance of the classifiers:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

with TP, FP, TN, FN being number of true positives, false positives, true negatives and false negatives, respectively.

$$\kappa = \frac{p_0 - p_e}{1 - p_e} \tag{4}$$

where κ represents the Cohen Kappa Score, p_0 is the empirical probability of agreement on the label assigned to any sample (the observed agreement ratio), and p_e is the expected agreement when both annotators assign labels randomly. p_e is estimated using a per-annotator empirical prior over the class labels.

8 Results and Discussion

From the series of experiments on the aforementioned datasets, we may observe that using only a small number of bands selected by the proposed methodology, adequate accuracy could be achieved using the SVM and CNN classifer. Table 1 presents the accuracy achieved by using all the bands of the datasets while Tables 2, 3 and 4 present the detailed classification result achieved by using only the selected bands for Indian Pines, Salinas-A and Pavia University dataset. The classification results which are improvement over using all the bands are shown in bold in the respective tables.

For the Indian Pines dataset the number of bands selected for the thresholds were 53, 67 and 96, respectively. From Table 1 it can be seen that using all the 200 bands of Indian Pines the obtained accuracies were 84.39% and 99.40% for SVM and CNN respectively. Table 2 shows that using 96 selected bands SVM gave classification accuracy of 74.92%. But with CNN classifier accuracy increased significantly to 95.32% with the same set of selected bands.

Table 1 Classification results considering all the bands

Dataset	SVM	CNN
Indian Pines	84.39	99.40
Salinas-A	99.92	98.33
Pavia University	91.64	98.81

Table 2 Classification results of Indian Pines on reduced dataset

τ	Bands Retained	SVM		CNN	
		Accuracy	Kappa	Accuracy	Kappa
0.95	53	71.23	73.24	89.48	90.23
0.97	67	73.33	75.13	91.10	91.85
0.98	96	74.92	76.97	95.32	96.32

Table 3 Classification results of Salinas-A on reduced dataset

τ	Bands retained	SVM		CNN	
		Accuracy	Kappa	Accuracy	Kappa
0.95	16	98.95	98.88	98.65	98.10
0.97	20	99.23	98.96	98.83	98.95
0.98	32	99.29	99.15	98.85	98.99

Table 4 Classification results of Pavia University on reduced dataset

τ	Bands retained	SVM		CNN	
		Accuracy	Kappa	Accuracy	Kappa
0.95	3	68.58	56.09	98.78	98.53
0.97	4	69.20	58.04	98.36	98.10
0.98	8	76.58	70.43	98.92	99.23

For Salinas-A dataset both SVM and CNN gave outstanding results with only limited number of selected bands. The obtained accuracy with all the 204 bands of Salinas-A were 99.92 and 98.33%, respectively for SVM and CNN. From Table 3, it may be observed that with only 16 selected bands(which is only .08% of the original number of bands) SVM gave an accuracy of 98.95% and CNN gave an accuracy of 98.65%. Similarly for the Pavia University with all the bands accuracies of 91.64% and 98.81% were achieved by using SVM and CNN respectively. However, for Pavia University dataset, as we can observe from Table 4, SVM performed moderately. But with CNN and using relatively very small number of bands, 8 in our case, high accuracy of 98.92% could be achieved, which was an improvement compared to the same using all the bands.

From the experimental results, it may be observed that using only a small number of bands selected by the proposed methodology, adequate accuracy could be achieved using SVM and CNN classifier. For Indian pines dataset using all the bands and SVM, the obtained accuracy is 84.39%. Using 96 selected bands accuracy of upto 74.92% could be achieved. But using CNN classifier the accuracy of upto 95.32% could be achieved. For Pavia University dataset, as we can observe from Table 4, SVM gives moderate results but with CNN and using relatively very small number of bands, 8 in our case, high accuracy could be achieved. For Salinas-A dataset both SVM and CNN gives outstanding results with only 16 number of bands.

9 Conclusion

In this work, We have proposed an algorithm for graph-based feature reduction, which tackles the challenges posed due to the high computational complexity involvement

while processing the hyperspectral dataset having hundreds or even thousands of bands. We have experimented the proposed method over three different hyperspectral datasets using two classifiers and found that using hybrid CNN classifier the selected bands give close or higher accuracy than using all bands. Our future work will concentrate on the tuning of the hyper-parameters and testing the proposed method on various other large datasets.

References

1. Hyperspectral remote sensing scenes (2019) http://www.ehu.eus/ccw intco/index.php/Hyperspectral Remote Sensing Scenes
2. Eismann MT (2012) Hyperspectral remote sensing. SPIE Bellingham
3. Hinton GE (2006) Salakhutdinov, R.R.: Reducing the dimensionality of data with neural networks. Science 313(5786):504–507
4. Imani M, Ghassemian H (2015) Feature reduction of hyperspectral images: discriminant analysis and the first principal component. J AI Data Mining 3(1):1–9
5. Jain R, Kasturi R, Schunck BG (1995) Machine vision, vol 5. McGraw-Hill, New York
6. MS, D'souza D (2016) Feature extraction of hyperspectral images based on lbp and rf feature extraction techniques. Int J Sci Res (IJSR) 5(5):1977–1979. https://doi.org/10.21275/v5i5.nov 163861
7. Reshma R, Sowmya V, Soman K (2016) Dimensionality reduction using band selection technique for kernel based hyperspectral image classification. Proc Comput Sci 93:396–402
8. Sarmah S, Kalita SK (2016) A correlation based band selection approach for hyperspectral image classification. In: 2016 IEEE 6th international conference on advanced computing (IACC). IEEE, pp 271–274
9. Steinbach M, Erto¨z L, Kumar V (2004) The challenges of clustering high dimensional data. In: New directions in statistical physics. Springer, pp 273–309
10. Wang B, Wang X, Chen Z (2012) Spatial entropy based mutual information in hyperspectral band selection for supervised classification. Int J Numer Anal Model 9(2)
11. Zhang X, Chew SE, Xu Z, Cahill ND (2015) Slic superpixels for efficient graph-based dimensionality reduction of hyperspectral imagery. In: Algorithms and technologies for multispectral, hyperspectral, and ultraspectral imagery XXI. vol 9472. International Society for Optics and Photonics, p 947209