Griffith Research Online

https://research-repository.griffith.edu.au

A hypergraph based semi-supervised band selection method for hyperspectral image classification

Author
Guo, Zhouxiao, Bai, Xiao, Zhang, Zhihong, Zhou, Jun

Published
2013

Conference Title
Proceedings of The IEEE International Conference on Image Processing (ICIP 2013)

DOI
https://doi.org/10.1109/ICIP.2013.6738646

Copyright Statement
Copyright 2013 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Downloaded from
http://hdl.handle.net/10072/57162
A HYPERGRAPH BASED SEMI-SUPERVISED BAND SELECTION METHOD FOR HYPERSPECTRAL IMAGE CLASSIFICATION

Zhouxiao Guo$^1$, Xiao Bai$^{1,*}$, Zhihong Zhang$^2$, Jun Zhou$^3$

$^1$ School of Science and Engineering, Beihang University
Haidian District, Beijing, China
$^2$ Xiamen University, Siming District, Xiamen, China
$^3$ School of Information and Communication Technology
Griffith University, Nathan, QLD 4111, Australia

ABSTRACT
Band selection is a fundamental problem in hyperspectral data processing. In this paper, we present a semi-supervised learning approach and a hypergraph model to select useful bands based on few labeled object information. The contributions of this paper are two-fold. Firstly, the hypergraph model captures multiple relationships between hyperspectral image samples. Secondly, the semi-supervised learning method not only utilizes unlabeled samples in the learning process to improve model performance, but also requires little labeled samples which can significantly reduce large amount of human labor and costs. The proposed approach is evaluated on AVIRIS and APHI datasets, which demonstrates its advantages over several other band selection methods.

Index Terms— Hyperspectral imaging, Band selection, Hypergraph, Image region classification.

1. INTRODUCTION
Nowadays, several imaging sensors mounted on satellites or airplanes can capture hyperspectral imagery with hundreds of contiguous bands across the electromagnetic spectrum. When using hyperspectral imagery for target recognition, many bands having redundant or correlated information for the target materials may influence the effectiveness and efficiency of the performance. Solution to this problem can be divided into two categories. The first category is dimensionality reduction that transforms the high-dimensional data into low-dimension representations via mathematic model, such as principal component analysis (PCA) [1] and independent component analysis (ICA) [2]. The second category is band selection, in which an optimal subset of bands will be selected directly from the original data based on defined criteria [3].

Band selection methods can be further divided into supervised and unsupervised [4]. Supervised methods [5, 6, 7] use training data to establish a predictive model for band selection. Such training data are normally obtained through registration of hyperspectral imagery with ground measurements, or direct manual labelling of target on the images. Unsupervised methods, on the other hand, do not require labelled data to facilitate the judgment on the importance of bands. Due to its convenience, unsupervised methods have been studied extensively in recent years. For example, band index (BI) and mutual information (MI) have been used to measure the statistical dependence between two bands and to evaluate the contribution of each band to classification [8]. Yao et al [9] proposed a Bayesian kernel-based nonlinear band selection method which can reduce the uncertainly by computation of posterior label probabilities. Pabitra et al [10] describe an unsupervised feature selection algorithm based on measuring similarity between features. Sebastiano et al [11] propose a new suboptimal search strategy suitable for feature selection in very high-dimensional remote sensing images. Being convenient to implement, unsupervised methods still can not match the supervised methods in image classification accuracy [12]. Therefore, semi-supervised band selection methods, which combine the advantages of both supervised and unsupervised methods, becomes very interesting.

In this paper, we propose a semi-supervised method [13] based on hypergraph model to select useful bands for object classification. This method commences from the construction of a hypergraph containing all feature points in a hyperspectral image. Then a projection matrix can be learned in a semi-supervised manner. Finally, the optimal bands are selected via solving a minimization problem using the least angle regression algorithm. The effectiveness of this method has been validated via experiments on two hyperspectral datasets. The contributions of this paper are two-fold. Firstly, the constructed hypergraph can represent multiple relationships between hyperspectral imagery samples. Secondly, the semi-supervised learning method not only utilizes unlabeled samples in the learning process to improve model performance, but also requires little labeled samples which can significantly
reduce the large amount of human labor and costs.

The rest of the paper is organized as follows. Section 2 introduces the construction of hypergraph via multivariate mutual information (MII). Section 3 briefly outlines the method of band selection. Section 4 describes the datasets used in the experiments and the image classification results.

2. HYPERGRAPH CONSTRUCTION

A hypergraph is a generalization of a graph in which an edge can connect more than two vertices. In this paper, a hypergraph $G$ is defined as $G = G(X, E, W)$ in which $X$ is a set of elements called vertices, $E$ is a set of non-empty subsets of $X$ called hyperedges. $W$ is a weight function which correspond to each hyperedge. Different from graph edges which connect pairs of nodes, hyperedges can connect an arbitrary number of nodes. In this paper, we only consider $k$-uniform hypergraph in which every hyperedge have the same $K$ vertices. Given a hyperspectral dataset $X = [x_1, ..., x_n] \in \mathbb{R}^{n \times m}$ where $x_i$ presents a feature point in hyperspectral image. A hyperedge weight containing $K$ vertices is computed as follows

$$W_{i_1, i_2, ..., i_K} = \frac{I_{x_{i_1}, x_{i_2}, ..., x_{i_K}}}{H(x_{i_1}) + H(x_{i_2}) + ... + H(x_{i_K})}$$

(1)

where $W_{i_1, i_2, ..., i_K}$ is the weight of hyperedge $i$ which contains $i_1, ..., i_K$ vertices, and $I_{x_{i_1}, x_{i_2}, ..., x_{i_K}}$ is the multivariate mutual information among features $i_1, ..., i_K$ which can be computed as

$$I_{(x_1; x_2; ..., x_N)} = \sum_{k=1}^{N} (-1)^{k-1} \sum_{X \subset \{x_1, x_2, ..., x_N\}, |X| = k} H(X)$$

(2)

where $I_{(x_1; x_2; ..., x_N)}$ [14] is the information content of $[x_1, ..., x_n]$ and equation 2 was introduced by Han [15]. $H(x_1, ..., x_n)$ represents the Shannon entropy of a discrete random variable with possible values and probability mass function. The entropy can be explicitly written as

$$H(X) = \sum_{i=1}^{n} P(x_i) I(x_i) = -\sum_{i=1}^{n} P(x_i) \log_2(P(x_i))$$

(3)

For more than two variables $x_1, ..., x_n$, this expands to

$$H(x_1, ..., x_n) = -\sum_{x_1} \sum_{x_n} \log_2[P(x_1, ..., x_n)]$$

(4)

where $P(x_1, ..., x_n)$ is the probability of $X = [x_1, ..., x_n]$. $P(x_1, ..., x_n) \log_2[P(x_1, ..., x_n)]$ is defined to be 0 if $P(x_1, ..., x_n) = 0$.

It is clear that the greater the value of $I_{(x_1; x_2; ..., x_N)}$ is, the more relevant the $K$ vertices are. In others words, if $I_{(x_1; x_2; ..., x_N)} = 0$, the $K$ vectors vertices are unrelated. Although the hypergraph can represent multiple relationships of features, computation problem arises when building the hyperedge because its number expands quickly with the increase of numbers of total vertices and connections of each hyperedge. For example, for a hypergraph with 100 vertices and each hyperedge connects 4 vertices, $C_4^{100} = 3921225$ hyperedge should be generated. There is a geometric series increase of total hyperedge as well as burgeoning vertices.

In order to reduce the cost of computation, the total number of hyperedge shall be reduced. A straightforward way is to build a hypergraph based on similarity between vertices, and limit the vertices to be connected by a hyperedge. Therefore, we build a hypergraph containing $n$ hyperedges and $n$ is the number of samples in a hyperspectral dataset. For each hyperedge, top $K$ vertices are selected from the dataset based on their Gaussian similarities. To characterize the relationship between two hyperspectral feature points via the weight of hyperedge, we transform the hypergraph into an adjacency matrix $M$. We sum all weights $W_{i_1, i_2, ..., i_K}$ containing $a$ and $b$ vertices as the weight of $M_{a,b}$, then $M_{a,b} = \sum_{i_1, i_2, ..., i_K} W_{i_1, i_2, ..., i_K - 2, a, b}$

(5)

$M$ is an $n \times n$ matrix which should be normalized. To do so, we define a matrix $P = D^{-\frac{1}{2}}MD^{\frac{1}{2}}$, where $D$ is a diagonal matrix with its $D_{ii} = \sum_j M_{i,j}$, so that $P$ is the normalized version of $M$.

3. BAND SELECTION AND IMAGE CLASSIFICATION

After the hypergraph has been transformed into an adjacent matrix, semi-supervised learning method can be used to select hyperspectral bands and perform image classification. Here we aim to get a projection matrix $Q$ which is learned from semi-supervised methods and can be used to classify $X$ by $Y = XQ$. Then we sort all bands based on sparse matrix $S$. This stage consists of four steps, which include label propagation, projection matrix $Q$ learning, band sort based on sparse matrix $S$, and image classification.

3.1. Label Propagation

In this step, we use a recently introduced semi-supervised label propagation method [13]. Given a dataset of pixels $X = [x_1, ..., x_1, ..., x_n] \in \mathbb{R}^{n \times m}$, there are $l$ labeled points and $u = n - l$ unlabeled points. Here we set a class $C + 1$ to record the outlier data.

Defining an initial label matrix

$$L = [(l_1)^T, (l_2)^T, ..., (l_N)^T] \in \mathbb{R}^{N \times C+1}$$

(6)
\[ L_{i,j} = \begin{cases} 1 & \text{if } x_i \text{ is labeled as } l_i = j \text{ or } j = C + 1 \\ 0 & \text{otherwise} \end{cases} \quad (7) \]

Given a following spreading function:
\[ F(t + 1) = \lambda PF(t) + (1 - \lambda)L. \quad (8) \]
where \( \lambda \) is a parameter in \((0, 1)\). The above step should be iteratively repeated until convergence. If \( F^* \) denotes the limit of the sequence \( F(t) \), the final estimation function \( F^* \) can be computed directly without iterations via mathematic methods:
\[ F^* = \lim_{t \to \infty} F(t) = (1 - \lambda)(I - \lambda P)^{-1}L \quad (9) \]

### 3.2. Projection Matrix \( Q \)

Once the label propagation matrix \( F \) is ready, the next step is to learn a projection matrix \( Q \). In other words, we aim to get an affine matrix \( Q \) from \( Y = XQ \). Here we build a linear classifier \( y = Q^T x + b \) where \( x \) is a sample in the hyperspectral imagery and \( b \) is a bias term. If \( y \) is close to \( t_j \) where \( t_j = [0, ..., 0, 1, 0, ..., 0]^T \), \( x \) will be classified into class \( j \). Assuming that the result of classification from the linear classifier is equal to the above semi-supervised learning matrix \( F \), we can define a regression function as follows
\[ \arg \min_{Q,b} \alpha \|Q\|^2 + \sum_{i=1}^{N} \sum_{j=1}^{C} F_{i,j} \|Q^T x_i + b - t_j\|^2 \quad (10) \]
where \( F_{i,j} \) is the label propagation matrix from equation 9. Equation 10 can be solved easily via iterative methods.

### 3.3. Band Sort based on Sparse Matrix \( S \)

Through the above steps, the hyperspectral dataset can be expressed as \( Y = XQ \) where \( Q \) is a projection matrix from equation 10. This is followed by band selection that aims to find an optimal subset of bands. Our idea is to constrain the projection matrix \( Q = [Q_1, ..., Q_c] \) via a sparse selection matrix \( S = [S_1, ..., S_c] \). This can be represented by the following minimization problem
\[ \arg \min_{S} \sum_{k=1}^{c} \|Y_k - XS_k\|^2 \quad (11) \]
where \( S_k \) is an \( m \)-dimensional vector. The optimization problem can be solved by the least angle regression (LAR) algorithm[16], which allows the control of the sparseness of \( S_k \) via tuning a zero-total parameter.

Suppose we want to select \( t \) bands from a total of \( m \) band candidates. Here, we use band score to evaluate every band from the selection matrix \( S \), which is defined as
\[ \text{bandscore}(j) = \max_k |S_{j,k}| \quad (12) \]
where \( j \) is the band index, \( S_{j,k} \) is the \( j \)th line of \( S_k \). After getting scores of all the bands, we then sort the bands in descending order and select the top \( t \) hyperspectral bands.

### 3.4. Image Classification

Finally, we perform image classification based on the selected bands. In the experiment, we have used the SVM classifier because of it has shown excellent classification performance and demonstrated convenience for implementation [17]. The SVM classifier is applied to each feature points with selected bands from the hyperspectral remote sensing images. The classifier adopted Gaussian radial basis function (RBF) as the kernel function and one against all scheme for multi-class classification.

### 4. EXPERIMENTS

![Fig. 1](image1.png)

Fig. 1. (a) Band 60 of the APHI hyperspectral image. (b) The ground truth of (a). (c) Band 113 of the AVIRIS image. (d) The ground truth of (c).

In this paper, experiments were performed on both AVIRIS (Airborne Visible/Infrared Imaging Spectrometer) and APHI (Airborne Push hyperspectral Imager) datasets. The size of the APHI data is 210×150×64. It covers wavelength from 455nm to 805nm. Sample images and their ground truth are shown in Figure 1. In the APHI data, land cover types are paddy, bamboo, tea, pachyrhizus, caraway and water. The AVIRIS data is acquired from an altitude of approximately 20km and has a spatial resolution of 18m. After removing water absorption and other noisy bands, 200
bands were used for testing. In the AVIRIS data, seven land-cover classes are discarded because it is hard to marked the label of hyperspectral samples. The remaining land cover types are corn-notill, corn-min, grass/pasture, grass/tree, hay-windrowed, soybeans-min, soybeans-clean and woods. The numbers of training and testing samples used for classification are shown in Table 1 and Table 2.

In the experiment, the proposed band selection method was compared with two other methods, MVPCA [18] and Affinity Propagation (AP) [19]. MVPCA ranks the bands based on the importance of individual band and its correlation with other bands. AP is a band selection method using affinity propagation. We have also compared the accuracies when all bands were used for classification. The classification results are shown in Figure 2. It is clear that our method has outperformed the alternatives, especially when more than 10 bands have been selected.

Table 1. Number of training and testing samples in the AVIRIS dataset.

<table>
<thead>
<tr>
<th>Class</th>
<th>Training samples</th>
<th>Test samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corn-notill</td>
<td>100</td>
<td>900</td>
</tr>
<tr>
<td>Corn-min</td>
<td>111</td>
<td>556</td>
</tr>
<tr>
<td>Grass/Pasture</td>
<td>140</td>
<td>280</td>
</tr>
<tr>
<td>Grass/Tree</td>
<td>100</td>
<td>250</td>
</tr>
<tr>
<td>Hay-windrowed</td>
<td>80</td>
<td>280</td>
</tr>
<tr>
<td>Soybeans-notill</td>
<td>100</td>
<td>800</td>
</tr>
<tr>
<td>Soybeans-min</td>
<td>300</td>
<td>1500</td>
</tr>
<tr>
<td>Soybeans-clean</td>
<td>50</td>
<td>410</td>
</tr>
<tr>
<td>Woods</td>
<td>100</td>
<td>800</td>
</tr>
<tr>
<td>Total</td>
<td>1081</td>
<td>5776</td>
</tr>
</tbody>
</table>

Table 2. Number of labeled and unlabeled samples in the APHI data.

<table>
<thead>
<tr>
<th>Class</th>
<th>Training samples</th>
<th>Test samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paddy</td>
<td>100</td>
<td>800</td>
</tr>
<tr>
<td>Bamboo</td>
<td>50</td>
<td>280</td>
</tr>
<tr>
<td>Tea</td>
<td>50</td>
<td>280</td>
</tr>
<tr>
<td>Pachyrhizus</td>
<td>30</td>
<td>250</td>
</tr>
<tr>
<td>Caraway</td>
<td>50</td>
<td>200</td>
</tr>
<tr>
<td>Water</td>
<td>60</td>
<td>800</td>
</tr>
<tr>
<td>Total</td>
<td>340</td>
<td>2610</td>
</tr>
</tbody>
</table>

5. DISCUSSION AND CONCLUSION

To conclude, we have presented a novel semi-supervised learning method for band selection and classification on hyperspectral remote sensing images. This method is based on the building of a hypergraph framework that characterizes multiple relationships of a set of feature points. It requires very few labeled samples in the training stage. Experimental results show that the proposed method has outperformed two alternative methods in the literature.

6. REFERENCES


