# HYPERSPECTRAL IMAGE BAND SELECTION VIA GLOBAL OPTIMAL CLUSTERING

*Fahong Zhang*[1], *Qi Wang*[1*], *Xuelong Li*[2]

[1]School of Computer Science and Center for OPTical IMagery Analysis and Learning, Northwestern Polytechnical University, Xi'an 710072, Shaanxi, P. R. China
[2]Center for OPTical IMagery Analysis and Learning, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, Shaanxi, P. R. China

## ABSTRACT

Band selection, by choosing a set of representative bands in hyperspectral images (HSI), is concerned to be an effective method to eliminate the "Hughes phenomenon". In this paper, we present a global optimal clustering-based band selection (GOC) algorithm based on the hypothesis that all the bands in a cluster are continuous at their wavelengths. After the clustering result is obtained, we propose a greedy-based method to select representative bands in each cluster, trying to minimize the linear reconstruction error. Experiment on a real HSI dataset shows that the proposed method outperforms the state-of-the-art competitors.

***Index Terms***— Band selection, hyperspectral image, dynamic programming, global optimal.

## 1. INTRODUCTION

The rich band information in hyperspectral image (HSI) provides an opportunity in a wide range of applications. However, due to the large volume of data cube in hyperspectral images, redundant bands lead to a huge computational complexity as well as "Hughes phenomenon". Band selection method chooses some representative bands of HSIs to reduce the dimension of datasets and effectively solves the aforementioned dilemma. In recent years, a large number of band selection methods have been proposed. They can roughly be divided into supervised and unsupervised method based on whether training samples are used or not. Since the acquisition of labeled samples is a difficult task, sometimes supervised method is not very practical. In this paper, we mainly focus on unsupervised method. According to [1], based on the search strategy employed, unsupervised band selection can be divided into ranking-based methods, clustering-based methods, greedy-based methods and evolutionary-based methods. Ranking-based methods simply assign each band a rank value, and band with higher rank value is preferred to be selected. Constrained band selection (CBS) [2] linearly constrains each band and minimizes the interfering effects of remaining bands to get a rank value. Considered to be a clustering-based method, clustering-based band selection (CBBS) [3] constructs a dissimilarity matrix via mutual-information or Kullback-Leibler divergence, then Wards linkage method is conducted to achieve a clustering result. In volume gradient band selection (VGBS) [4], bands are treated to be points lying in high dimension space. VGBS attempts to choose a set of points which maximize the volume of parallelotope they constitute. It repeatedly removes the band that has the maximum gradient until the desired number of bands have been acquired. Multi-task sparsity pursuit (MTSP) [5], viewed as an evolutionary-based method, constructs an efficient searching strategy for band combinations based on immune clonal strategy (ICS).

In band selection problem, there are $C_L^K$ ways to select a subset with $K$ bands from a HSI with $L$ bands. Thus finding the global optimal solution is a difficult task. In our paper, we propose a clustering-based method which can achieve a global optimal clustering result. Based on the hypothesis that all the bands in a cluster are continuous at their wavelengths, the clustering problem can be transferred to a band separation problem. Specifically, it is to determine $c-1$ separators which can separate the whole bands into $c$ parts. In this way, a global optimal clustering result can be obtained by dynamic programming (DP). After the clustering result is acquired, some bands are selected to minimize the reconstruct error when utilize selected bands to construct the left bands in this cluster. Two main contributions are claimed in this paper. 1) Base on the hypothesis above, we propose a method to obtain a global optimal clustering result by dynamic programming. 2) We present a noise insensitive method to do band selection among the clustering result and experiment on a HSI image shows the superiority of our method.

## 2. GLOBAL OPTIMAL CLUSTERING VIA DYNAMIC PROGRAMMING

### 2.1. INTRODUCTION TO DYNAMIC PROGRAMMING

Dynamic programming is an effective optimization method. In dynamic programming, a complex problem is broken down to a series of subproblems, and each subproblem is broken

down to simpler subproblems. This process is repeatedly conducted until the subproblem is simple enough to solve directly. Then we constantly combine the subproblems to solve the more complex problems and finally solve the original problem. A problem must have two attributes to be solved with dynamic programming, which is optimal substructure and overlapping subproblems attribute. Optimal substructure means that the solution of a problem can be obtained by the combination of its subproblems. Overlapping subproblems is saying that the solution of subproblems could be used many times, so we can store the solution of each subproblem in memory to avoid duplicated calculations.

### 2.2. PROBLEM DEFINITION

We use $\mathbf{x}_l \in R^N$ to denote the $l$-th band, and $\{\mathbf{x}_l\}_{l=1}^{L}$ to denote the whole bands sorted by wavelength. $\mathbf{s} = (s_0, s_1, ..., s_{C-1}, s_C)^T$, in which $0 = s_0 < s_1 < ... < s_{C-1} < s_C = L$, denotes the separators. Due to $\mathbf{s}$, the whole band set is separated into $C$ subsets: $\{\mathbf{x}_l\}_{l=s_0+1}^{s_1}$, $\{\mathbf{x}_l\}_{l=s_1+1}^{s_2}$, ..., $\{\mathbf{x}_l\}_{l=s_{C-1}+1}^{s_C}$. Then we define $D_{inter}$ and $D_{intra}$ as the function of the first $c+1$ separators $\mathbf{s} = (s_0, s_1, ..., s_{c-1}, s_c)^T$:

$$D_{inter}(\mathbf{s} = (s_0, s_1, ..., s_c)^T) = \sum_{i=0}^{c} \sum_{j=s_i+1}^{s_{i+1}} \|M(s_i+1, s_{i+1}) - M(1, L)\|^2, \quad (1)$$

$$D_{intra}(\mathbf{s} = (s_0, s_1, ..., s_c)^T) = \sum_{i=0}^{c} \sum_{j=s_i+1}^{s_{i+1}} \|\mathbf{x}_j - M(s_i+1, s_{i+1})\|^2, \quad (2)$$

in which,

$$c \le min(C, s_c), s_c \le L, \quad (3)$$

$$M(i,j) = \frac{1}{j-i+1} \sum_{k=i}^{j} \mathbf{x}_k. \quad (4)$$

$M(i,j)$ represent mean vectors of $\{\mathbf{x}_l\}_{l=i}^{j}$. $D_{inter}(\mathbf{s} = (s_0, s_1, ..., s_c)^T)$ and $D_{intra}(\mathbf{s} = (s_0, s_1, ..., s_c)^T)$ can be interpreted as inter-cluster distance and intra-cluster distance in the first $s_c$ bands while clustering result is determined by $(s_0, s_1, ..., s_c)^T$. Our target is to solve an optimization problem

$$\max_{\mathbf{s}=(s_0,s_1,...,s_C)^T} \frac{D_{inter}(\mathbf{s})}{D_{intra}(\mathbf{s})}, \quad (5)$$

and find the corresponding $\mathbf{s}$.

### 2.3. PROBLEM CONVERSION

[6] presents a similar optimization problem but they utilize clone selection [7] to obtain a suboptimal solution. Here we propose an improved method to achieve an optimal solution via dynamic programming. In order to meet the suboptimal attribute in dynamic programming, we first do some transformation to $(5)$. We construct a function of $\lambda$:

$$g(\lambda) = \max_{\mathbf{s}=(s_0,s_1,...,s_C)^T} D_{inter}(\mathbf{s}) - \lambda \cdot D_{intra}(\mathbf{s}). \quad (6)$$

Suppose $\lambda^*$ is the solution of $(5)$. It is obvious that $g(\lambda^*) = 0$ and $g(\lambda)$ is a monotone decreasing function. This means $\lambda^*$ is the unique zero of $g(\lambda)$, so dichotomy method can be utilized to obtain $\lambda^*$. We first estimate a search interval of $\lambda^*$: $(left, right)$. By the calculation of $g(\frac{left+right}{2})$, we can constantly half the search interval of $\lambda^*$ and finally obtain $\lambda^*$. As a result, the problem to solve $(5)$ converts to the calculation of $g(\lambda)$.

### 2.4. OPTIMAL SUBSTRUCTURE

For the purpose to find the optimal substructure in $(6)$, we first define subproblems of calculating $g(\lambda)$ as:

$$P(i,j) = \max_{\mathbf{s}=(s_0,s_1,...,s_j)^T, s_j=i} D_{inter}(\mathbf{s}) - \lambda \cdot D_{intra}(\mathbf{s}) \quad (7)$$

for all $i \le L, j \le min(i, C)$. Then the optimal substructure of our problem is revealed as follow.
When $j = 1$:

$$P(i,1) = D_b(1,i) - \lambda \cdot D_w(1,i), \quad (8)$$

and when $j > 1$:

$$P(i,j) = \max_{j-1 \le k \le i-1} P(k, j-1) + D_b(k+1, i) - \lambda \cdot D_w(k+1, i), \quad (9)$$

in which,

$$D_w(i,j) = \sum_{k=i}^{j} \|\mathbf{x}_j - M(i,j)\|^2, \quad (10)$$

$$D_b(i,j) = \sum_{k=i}^{j} \|M(i,j) - M(1,L)\|^2. \quad (11)$$

To explain with it, $D_b(i,j) - \lambda \cdot D_w(i,j)$ actually means the contribution of $\{\mathbf{x}_l\}_{l=i}^{j}$. In $(9)$, suppose the last cluster in the first $i$ bands is $\{\mathbf{x}_l\}_{l=k+1}^{i}$, $P(i,j)$ is actually the optimal sum of two parts for each possible $k$, one indicates the contribution of the last cluster $\{\mathbf{x}_l\}_{l=k+1}^{i}$ and another indicates the solution of subproblem $P(k, j-1)$. Specially, $P(L, C)$ is the solution of the original problem $\lambda^*$. Moreover, in order to find the corresponding $\mathbf{s}$, we use $F(i,j)$ to record the optimal $k$ in $(9)$ for each $P(i,j)$.

### 2.5. IMPLEMENTATION OF GOC

One thing to be noted is that, noises band do great contribution to $D_w$ so they are easier to form a single-band cluster. In order to deal with it, we limit the minimal size of each cluster to 3. This only influences the limitation of $k$ in equation $(10)$ from $j-1 \leq k \leq i-1$ to $3 \cdot (j-1) \leq k \leq i-3$. To sum up the GOC algorithm, a pseudo code is given as follow.

**Algorithm 1** The Global Optimal Clustering Algorithm

**Input:** All bands $X = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_L]$, cluster number $C$.

Yield $A = [\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_L]$, where $\mathbf{a}_l = \sum_{l=1}^{i} \mathbf{x}_l$, so $M(i,j)$ can be represented by $\mathbf{a}_j - \mathbf{a}_{i-1}$ while $\mathbf{a}_0 = \mathbf{0}$.
Get $D_w(i,j), D_b(i,j)$ for each $1 \leq i \leq j \leq L$.
Set $\epsilon \leftarrow e^{-8}$, $left \leftarrow 0$, $right \leftarrow e^{8}$.
**while** $|left - right| > \epsilon$ **do**
    $\lambda \leftarrow \frac{left+right}{2}$.
    Get $P(i,1)$ for each $i \leq L$ according to $(9)$.
    Set each $F(i,1)$ to 0.
    **for** $j \leftarrow 2$ **to** $L$ **do**
        **for** $i \leftarrow j$ **to** $L$ **do**
            Calculate $P(i,j)$ through equation $(10)$ and find the optimal $k$.
            $F(i,j) \leftarrow k$.
        **end for**
    **end for**
    **if** $P(L,C) > \epsilon$ **then**
        $left \leftarrow \lambda$.
    **else**
        $right \leftarrow \lambda$.
    **end if**
**end while**
Get $\mathbf{s} = (s_0, s_1, ..., s_C)$ through $F$.

**Output:** $C+1$ separators $\mathbf{s} = (s_0, s_1, ..., s_C)$.

## 3. BAND SELECTION IN EACH CLUSTER BY A GREEDY BASED METHOD

In this section, a greedy based band selection method is proposed with the obtained $C$ clusters in the previous. Suppose $\Phi_j$ is set of current selected bands in cluster $j$ with size $|\Phi_j|$, and $B_j = [\mathbf{b}_{j,1}, \mathbf{b}_{j,2}, ..., \mathbf{b}_{j,|\Phi_j|}] \in R^{N \times |\Phi_j|}$ is the corresponding matrix. We try to linearly construct the mean of the unselected bands using the selected bands, and our greedy strategy is to select the band which can reduce the linear reconstruction error in the maximum extent. In detail, we define before-selected reconstruction error ($E_1$) and after-selected reconstruction error ($E_2$) as follow:

$$E_1(i) = (I - (B_j^T B_j)^{-1} B_j)\mathbf{y} \tag{12}$$

$$E_2(i) = (I - (\widetilde{B_j}^T \widetilde{B_j})^{-1} \widetilde{B_j})\mathbf{y}. \tag{13}$$

Here $i$ is the indicator of the examined band, and $j$ denotes the cluster to which $\mathbf{x}_i$ belong. $\mathbf{y}$ is the mean of $\{\mathbf{x}_k\}_{k=s_{j-1}+1}^{s_j} - \Phi_j - \{\mathbf{x}_i\}$. $\widetilde{B_j} = [\mathbf{b}_{j,1}, \mathbf{b}_{j,2}, ..., \mathbf{b}_{j,C}, \mathbf{x}_i]$ consists of $B_j$ and $\mathbf{x}_i$. In the proposed method, the band $i$ which can maximize the value of $\frac{E_1(i)}{E_2(i)}$ is selected in each iteration until the desired band number is obtained. To make sure that there is at least one selected band in each cluster, we first select one band in each cluster which has the shortest distance to the mean of the remaining bands.

## 4. EXPERIMENT

### 4.1. EXPERIMENT SETUP AND PARAMETER SETTING

To evaluate the performance of GOC algorithm, we conduct our experiments on Indian Pines. Indian Pines image, as a widely used HSIs, was captured by AVIRIS. It consists of $145 \times 145$ pixels and 224 spectral reflectance bands. There are 16 classes of objects included in the image. In general, the water absorbtion bands are removed and a total 200 bands are utilized in our experiment. To do comparisons, several representative band selection methods are involved as benchmarks: clustering-based band selection with mutual information(CBBS-MI) and Kullback-Leibler (CBBS-KL) [3], VGBS [4], uniform band selection (UBS) [8], and MT-SP [5]. In order to evaluate the classification accuracy of different band selection method, four popular classifiers are used to test the classification accuracy: support vector machine (SVM), k-nearest neighborhood (KNN), linear discriminant analysis (LDA) and classification and regression trees (CART). In our experiment, $5\%$ of samples for each classes are chosen randomly to be training samples, and the rest are used in the testing. To reduce the randomness caused by training samples, the final result is calculated by the mean of 10 individual runs.

In the proposed algorithm, the only parameter to be determined is the number of clusters $C$ and the selected bands number $K$. $K$ controls the number of bands to be selected. For each classifier, we set $K$ every five intervals from 5 to 80. We attempt to set $C = \lfloor \alpha \cdot K \rfloor$, but $C$ is forced to be smaller than $\lfloor \frac{L}{3} \rfloor$ because of the minimal limitation of cluster size. Considering these two factors, $C$ is determined by:

$$C = min(\lfloor \alpha \cdot K \rfloor, \lfloor \frac{\beta \cdot L}{3} \rfloor). \tag{14}$$

Here $\alpha$ and $\beta$ are simply set to 0.8.

### 4.2. EXPERIMENT RESULT

Fig. 1 shows the classification accuracy by SVM verses the number of selected bands in Indian Pines image. The classification accuracy with total bands are plotted as base line. As can be clearly seen in Fig. 1, the proposed GOC method

outperforms the other methods in most of the cases. UBS take the second place owe to the small correlation between its selected bands. MTSP method follows the UBS method but we can see the unstability of MTSP in the cause of the uncertainty of evolution strategy. VGBS and CBBS method do not perform well in Indian Pines image. This is because they are too sensitive to the noisy bands. In VGBS, the band that has the maximum gradient is the most redundant, but the noisy bands usually have smaller gradients and are easier to be selected. In CBBS, noisy bands have larger dissimilarity with other bands and they tend to form single-band clusters. So these two algorithms are more sensitive to the noisy bands. Table 1 displays the accuracy results through four different classifiers. The highest accuracy among four classifiers proved the superiority of our method. In summary, our GOC method obtains a global optimal clustering result and reduces the risk to select noisy bands at the same time, which has significance both in theory and practice.

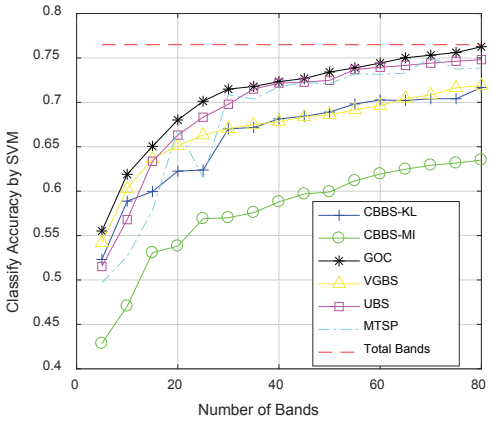


**Fig. 1**. The SVM classification accuracy of different methods versus the number of selected bands in Indian Pines image.

**Table 1**. Classification average accuracy results of Indian Pines through four different classifiers

| Classifiers | SVM | KNN | LDA | CART | Average |
|---|---|---|---|---|---|
| CBBS-KL | 0.6614 | 0.6192 | 0.6338 | 0.5648 | 0.6198 |
| CBBS-MI | 0.5761 | 0.6367 | 0.5795 | 0.5432 | 0.5839 |
| VGBS | 0.6703 | 0.6206 | 0.5635 | 0.5581 | 0.6031 |
| UBS | 0.6938 | 0.6211 | 0.6688 | 0.5762 | 0.6400 |
| MTSP | 0.6802 | 0.6342 | 0.6636 | 0.5735 | 0.6379 |
| GOC | **0.7080** | **0.6531** | **0.6746** | **0.5799** | **0.6539** |

## 5. CONCLUSION

In this paper, we present an algorithm utilizing dynamic programming to achieve a global optimal clustering result, and a noise-insensitive greedy-based band selection method is then designed to get a selected band set. Experiment on Indian Pines in classification accuracy shows that our method outperforms the other methods representing state-of-the-art among four different classifiers. In future work, we will try to find more effective criterions that can be optimized globally via dynamic programming and do comparisons with more band selection algorithms.

## 6. ACKNOWLEDGEMENT

This work was supported by the National Natural Science Foundation of China under Grant 61379094.

# References

[1] Yuan Yuan, Xiangtao Zheng, and Xiaoqiang Lu, "Discovering diverse subset for unsupervised hyperspectral band selection," *IEEE Trans. Image Processing*, vol. 26, no. 1, pp. 51–64, 2017.

[2] Chein-I Chang and Su Wang, "Constrained band selection for hyperspectral imagery," *IEEE Trans. Geoscience and Remote Sensing*, vol. 44, no. 6, pp. 1575–1585, 2006.

[3] Adolfo Martínez Usó, Filiberto Pla, José Martínez Sotoca, and Pedro García-Sevilla, "Clustering-based hyperspectral band selection using information measures," *IEEE Trans. Geoscience and Remote Sensing*, vol. 45, no. 12-2, pp. 4158–4171, 2007.

[4] Xiurui Geng, Kang Sun, Luyan Ji, and Yongchao Zhao, "A fast volume-gradient-based band selection method for hyperspectral image," *IEEE Trans. Geoscience and Remote Sensing*, vol. 52, no. 11, pp. 7111–7119, 2014.

[5] Yuan Yuan, Guokang Zhu, and Qi Wang, "Hyperspectral band selection by multitask sparsity pursuit," *IEEE Trans. Geoscience and Remote Sensing*, vol. 53, no. 2, pp. 631–644, 2015.

[6] Qi Wang, Jianzhe Lin, and Yuan Yuan, "Salient band selection for hyperspectral image classification via manifold ranking," *IEEE Trans. Neural Netw. Learning Syst.*, vol. 27, no. 6, pp. 1279–1289, 2016.

[7] Leandro Nunes de Castro and Fernando J. Von Zuben, "Learning and optimization using the clonal selection principle," *IEEE Trans. Evolutionary Computation*, vol. 6, no. 3, pp. 239–251, 2002.

[8] Chein-I Chang, Qian Du, Tzu-Lung Sun, and Mark L. G. Althouse, "A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification," *IEEE Trans. Geoscience and Remote Sensing*, vol. 37, no. 6, pp. 2631–2641, 1999.