OPTIMAL NEIGHBORING RECONSTRUCTION FOR HYPERSPECTRAL BAND SELECTION

Fahong Zhang[1], *Qi Wang*[1,2*], *Xuelong Li*[3]

[1] School of Computer Science and Center for OPTical IMagery Analysis and Learning,
Northwestern Polytechnical University, Xi'an 710072, Shaanxi, P. R. China.
[2] Unmanned System Research Institute, Northwestern Polytechnical University, Xi'an 710072, Shaanxi, P. R. China.
[3] Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, Shaanxi, P. R. China,
and University of Chinese Academy of Sciences, Beijing 100049, P. R. China.

ABSTRACT

Band selection, as an effective and popular dimensional reduction methods for hyperspectral image (HSI), has raised wide attention in recent years. In this paper, we propose a novel band selection method called optimal neighboring reconstruction (ONR). Compared to conventional methods, ONR mainly has following advantages. 1) It is globally optimal, which means the best combination of bands towards the designed objective function can be achieved. 2) It sufficiently exploits the neighboring structure among bands, so can effectively reduce the redundancy while maintaining the discrimination among bands. Experiments on three real data sets show that the proposed method has excellent performance.

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

1. INTRODUCTION

Hyperspectral image (HSI) records the reflectance of a specific scene to electromagnetic waves in different wavelengths. Owe to the abundant spectral information contained in HSI, a more accurate and specific description of the ground objects can be obtained and applied to various applications [1, 2]. Nevertheless, the high dimension in spectral domain also raises challenges in different aspects such as data storage and processing. To tackle this problem, band selection is proposed as a dimensional reduction technique which aims to select a few of discriminative and low-corrected bands to represent the original data set. According to the use of the labeled samples, band selection can be briefly categorized into supervised and unsupervised methods. Since the acquisition of the labeled samples is usually hard for HSI, the unsupervised methods are more practical in real applications.

According to the employed searching strategy [3], unsupervised band selection can be further divided into ranking-based, clustering-based, greedy-based and evolutionary-based methods [3]. Specifically, ranking-based methods [4] first assign each band a rank value and select the top-rank bands individually. Clustering-based methods [5] consist of two stages. They first sperate the bands into clusters, and then select the most representative bands in each of them. Greedy-based methods [6] are multi-stage procedure. In each stage the current optimal band is selected. Evolutionary-based methods [7] are one-stage procedure. They first generate a random band combination with desired number of bands, and repeatedly update it through some evolutionary algorithms.

In this paper, a novel method called *optimal neighboring reconstruction* (ONR) is proposed. Compared to the above mentioned four kinds of methods, ONR is much more different. First, it is global optimal towards the defined objective function, which means it is not based on some approximate algorithms like what greedy-based or evolutionary-based methods do. Second, it is one-stage and also without pre-processing like clustering-based methods. Hence it is expected to be more robust against different data sets.

The main contributions of this paper are claimed as follow.

1) A effective objective function is proposed in a novel point of view. With consideration of the characteristic that bands with closer wavelengths have stronger correlation [8], a neighboring reconstruction based criterion is presented to evaluate the ability that a certain band subset can linearly reconstruct the original data set.

2) An efficient searching strategy is applied to obtain the optimal band subset towards the proposed criterion. Unlike the previous works which are based on approximate algorithms, dynamic programming [9] is utilized to achieve the global optimal solution.

2. OPTIMAL NEIGHBORING RECONSTRUCTION

This section details the proposed ONR method. First, the proposed objective function is introduced. Second, the optimization method to achieve the global optimum is given.

2.1. OBJECTIVE FUNCTION

Before the introduction of the objective function, we give some notations that will be used throughout the paper. For an arbitrary matrix M, $M_{i,j}$ is its (i,j)th entry, M^i is its ith row and M_j is its jth column. Denote all the band vectors in a HSI as a matrix $X = [X_1, X_2, ..., X_d] \in \mathbb{R}^{n\times d}$, where X_j is the jth band with l_2 norm normalized to 1, n is the number of pixels in each band and d is the number of bands. The indexes of the selected bands are specified by a vector $\mathbf{b} = (b_1, b_2, ..., b_m)^T$, in which m represents the number of the selected bands.

See from the view of linear algebra, band selection can be regarded as a problem to select a serious of bands which can linearly reconstruct the original data set with minimal loss:

$$\min_{\mathbf{b},W} \quad \|E\| \quad s.t. \quad X = [X_{b_1}, X_{b_2}, ...X_{b_m}]W + E. \tag{1}$$

Here $\|\cdot\|$ is a matrix norm, $W \in \mathbb{R}^{m\times d}$ is the coefficient matrix, and E is the reconstruction error matrix. The proposed objective function is raised based on Eq. (1) with two concerns, i.e., noise reduction and neighboring reconstruction.

1) Noise reduction. A common choice of $\|\cdot\|$ is the $l_{2,1}$ norm, which is defined as $\|E\|_{2,1} = \sum_{j=1}^{d}\sqrt{\sum_{i=1}^{n} E_{ij}}$. By the use of $l_{2,1}$ norm, Eq. (1) can be rewritten as:

$$\begin{aligned} \min_{\mathbf{b},W} \quad & \sum_{j=1}^{d} \|E_j\|_2, \\ s.t. \quad & X = [X_{b_1}, X_{b_2}, ...X_{b_m}]W + E. \end{aligned} \tag{2}$$

Nevertheless, noisy bands are always hard to be reconstructed and with large error $\|E_j\|_2$. In order to reduced the effect of noisy bands, we expect that there is a upper bound τ to $\|E_j\|_2$, i.e., if the value of $\|E_j\|_2$ is larger than τ, than it will be set to τ. Motivated by this, Eq. (2) is changed to:

$$\begin{aligned} \min_{\mathbf{b},W} \quad & \sum_{j=1}^{d} f_\tau(\|E_j\|_2), \\ s.t. \quad & X = [X_{b_1}, X_{b_2}, ...X_{b_m}]W + E, \end{aligned} \tag{3}$$

where f_τ is defined as:

$$f_\tau(x) = \begin{cases} x, & x \le \tau, \\ \tau, & x > \tau. \end{cases} \tag{4}$$

2) Neighboring reconstruction. As stated in our previous work [8], strong correlation mainly exists between bands with close indexes. Hence the above linear reconstruction problem is expected to be with two attributes. One is that the coefficient matrix W should be sparse in column, since only few bands have strong correlation with a specific band. Another is that one band should be reconstructed by its neighbors, since they have the largest correlation with it.

A extreme case which can satisfy the above two attributes is that each band X_j is reconstructed by two of its nearest neighbors X_{b_k} and $X_{b_{k+1}}$, where $b_k \le j < b_{k+1}$. More specifically, when $\mathbf{b}$ is fixed, the reconstruction of X_j can be formulated as:

$$\begin{aligned} \min_{Z_j} \quad & f_\tau(\|E_j\|_2), \\ s.t. \quad & X_j = [X_{b_k}, X_{b_{k+1}}]Z_j + E_j, \\ & b_k \le j < b_{k+1}. \end{aligned} \tag{5}$$

where $Z \in \mathbb{R}^{2\times d}$ is a coefficient matrix that records the nonzero elements in W. In special cases, if $j < b_1$ or $j > b_m$, X_j will be reconstructed only by X_{b_1} or X_{b_m}, since there is no available k subject to $b_k \le j < b_{k+1}$. To achieve a united form of objective function, we set $b_0 = 0, b_{m+1} = d+1$ and $X_0 = X_{d+1} = \mathbf{0}$.

Summing up the above considerations, the final objective function is given as:

$$\begin{aligned} \min_{\mathbf{b},Z} \quad & \sum_{j=1}^{d} f_\tau(\|E_j\|_2), \\ s.t. \quad & \forall j, \quad X_j = [X_{b_k}, X_{b_{k+1}}]Z_j + E_j, \\ & b_k \le j < b_{k+1}. \end{aligned} \tag{6}$$

2.2. OPTIMIZATION METHOD

Eq. (6) is not easy to be solved via traditional machine learning methods since it is a combinatorial optimization problem with respect to $\mathbf{b}$. Here we present a special solution to Eq. (6) as follows.

We first define a notation $L_{l,r}$ as the loss when reconstructing $[X_{l+1}, X_{l+2}..., X_{r-1}]$ using $[X_l, X_r]$:

$$\begin{aligned} L_{l,r} = \sum_{j=l+1}^{r-1} \quad & \min_{Z_j} \quad f_\tau(\|E_j\|_2), \\ & s.t. \quad \forall j, \quad X_j = [X_l, X_r]Z_j + E_j. \end{aligned} \tag{7}$$

Then Eq. (6) can be reformulated as:

$$\min_{\mathbf{b}} \quad \sum_{k=0}^{m} \quad L_{b_k, b_{k+1}}. \tag{8}$$

Since now, the original problem is converted to a linearly reconstruction problem as in Eq. (7) and a combinatorial optimization problem as in Eq. (8).

1) Solution to Eq. (7). If f_τ is not employed, Eq. (7) is a linear least square problem, whose minimum norm solution is given as:

$$\begin{cases} Z_j^* = ([X_l, X_r]^T[X_l, X_r])^{-1}[X_l, X_r]^T X_j \\ L_{l,r} = \sum_{j=l+1}^{r-1} \|X_j - [X_l, X_r]Z_j^*\|_2 \end{cases} \tag{9}$$

In fact, Eq. (9) is also one of the solution to Eq. (7) since f_τ is monotonically increasing.

2) Solution to Eq. (8). Eq. (8) can be solved via dynamic programming [8]. We first define an auxiliary variable D as:

$$D_{i,j} = \min_{b_1,b_2,...,b_{j-1}} \sum_{k=0}^{j-1} L_{b_k,b_{k+1}}, \quad s.t. \quad b_j = i, \quad (10)$$

in which $1 \leq j \leq i \leq d+1$. It is easy to see the solution of Eq. (8) is equal to $D_{d+1,m+1}$. Moreover when $j > 1$, there is:

$$\begin{aligned} &\min_{b_1,b_2,...,b_{j-1}} \sum_{k=0}^{j-1} L_{b_k,b_{k+1}} \\ &= \min_{b_{j-1}} \min_{b_1,b_2...b_{j-2}} \sum_{k=0}^{j-2} L_{b_k,b_{k+1}} + L_{b_{j-1},b_j} \\ &= \min_{b_{j-1}} D_{b_{j-1},j-1} + L_{b_{j-1},b_j} \\ &= \min_{j-1 \leq q < b_j} D_{q,j-1} + L_{q,b_j}, \end{aligned} \quad (11)$$

which shows that there is a recursive relation among $D_{i,j}$:

$$D_{i,j} = \min_{j-1 \leq q < i} D_{q,j-1} + L_{q,b_j}. \quad (12)$$

Specifically when $j = 1$, we have $D_{i,1} = L_{0,i}$. Hence all the $D_{i,j}$ can be obtained recursively according to Eq. (12). Once $D_{d+1,m+1}$ is obtained, we need to get the optimal indexes of bands $\mathbf{b}^* = (b_1^*, b_2^*, ..., b_m^*)^T$ corresponding to it. From Eq. (11), it can be inferred that b_j^* is just the argument which maximize $D_{b_{j+1}^*,j+1}$:

$$b_j^* = \arg\min_{j \leq q < b_{j+1}^*} D_{q,j} + L_{q,b_{j+1}^*}, \quad (13)$$

where $b_{m+1}^* = d+1$. Therefore $\mathbf{b}^*$ can be obtained based on Eq. (13).

In summary, the algorithm to optimize the objective function is shown in Algorithm 1.

Algorithm 1 Optimal Neighboring Reconstruction

Input: All bands $X = [X_1, X_2, ..., X_d]$, number of bands m.
1: Get $L_{l,r}$ for each $0 \leq l < r \leq d+1$ according to Eq. (9).
2: Set $D_{i,1} \leftarrow L_{0,i}$ for each $1 \leq i \leq m+1$.
3: **for** $j \leftarrow 2$ **to** $m+1$ **do**
4:     **for** $i \leftarrow j$ **to** $d+1$ **do**
5:         Get $D_{i,j}$ according to Eq. (12).
6:     **end for**
7: **end for**
8: Get $\mathbf{b}^* = (b_1^*, b_2^*, ..., b_m^*)^T$ according to Eq. (13).
Output: The indexes of m selected bands $b_1, b_2, ..., b_m$.

3. EXPERIMENT

In this section, the proposed ONR method is compared with some state-of-the-art methods via classification experiments. First, the experimental setup is introduced. Then the experimental results are shown and some deep some analyses are given.

3.1. EXPERIMENT SETUP

The setup of the experiment includes data sets, comparison methods, classification settings and parameter settings.

1) Data sets. Two data sets named Indian Pines and Salinas are employed in this paper. One can find their description in [10].

2) Comparison methods. The comparison methods include WaLuDi [5], VGBS [6], UBS [4], E-FDPC [11] and MTSP [7].

3) Classification settings. Four popular classifiers are employed in the experiments. They are k-nearest neighborhood (KNN), linear discriminant analysis (LDA), support vector machine (SVM) and classification and regression trees (CART). 10% of the samples for each classes are chosen randomly to train the classifiers, while the rest are used in testing.

4) Parameter settings. About the parameters of the comparison methods, the only parameter for WaLuDi, VGBS, UBS and E-FDPC is the number of selected bands m. For MTSP, its parameters are tuned on Indian Pines and fixed when moving to Salinas. The parameter τ of ONR is set according to the following steps: 1) Sort all the elements in $L_{l,r}$ for each $1 \leq l < r \leq d$ in descending order. Denote the result as $\widetilde{L}$. 2) Estimate the ratio of noisy bands to the whole bands γ. 3) Then τ is set to the rth element in $\widetilde{L}$, where $r = C_d^3 - C_{d'}^3$ and $d' = \lfloor d \cdot \gamma \rfloor$. Here C is the combinatorial number. Finally γ is set to 0.5 for both data sets empirically.

3.2. EXPERIMENT RESULT

The overall accuracies of the comparison methods and ONR are shown in Fig. 1 and Fig. 2 for Indian Pines and Salinas respectively. From Fig. 1, one can observe that the proposed ONR method achieves significant superiority compared to others when SVM, LDA and CART are employed. When KNN is utilized, both ONR and MTSP have promising performance. When referring to Fig. 2, ONR also has satisfactory performance. When LDA is employed, ONR dominates the others and rank the first. When using KNN or CART, ONR and E-FDPC both acquire good performance. While on SVM, ONR and VGBS outperform the others. The above results prove that the proposed method has robust performance against different classifiers and data sets.

Some interesting conclusions can be made based on the promising experiment results. 1) The neighboring reconstruction strategy can better exploit the intrinsic data structure of

HSI. 2) The utilized function f_τ can effectively help to reduce the influence of noisy bands. 3) The employed dynamic programming based optimization method is more powerful compared to traditional methods.

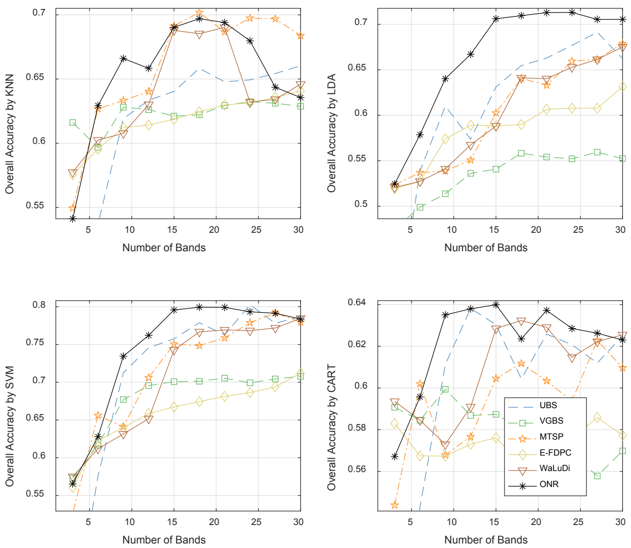

Fig. 1. OA curves on Indian Pines Scene for different band selection methods, in which m is set from 3 to 30 each 3 intervals.

4. CONCLUSION

In this paper, we present an optimal neighboring reconstruction (ONR) method to select the discriminative bands in HSI data set. We first develop a objective function which can fully exploit the neighboring structure of HSI, while minimizing the influence of noisy bands. Then a dynamic programming based method is employed to optimize the above objective function, through which the global optimal solution can be obtained. The proposed method is demonstrated to be effective according to the classification experiments on two real HSI data sets.

5. ACKNOWLEDGEMENT

This work was supported by the National Key R&D Program of China under Grant 2017YFB1002202, National Natural Science Foundation of China under Grant 61773316 and 61379094, and the Open Research Fund of Key Laboratory of Spectral Imaging Technology, Chinese Academy of Sciences.

References

[1] Qi Wang, Zhaotie Meng, and Xuelong Li, "Locality adaptive discriminant analysis for spectralcspatial classification of hyperspectral images," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 11, pp. 2077–2081, 2017.

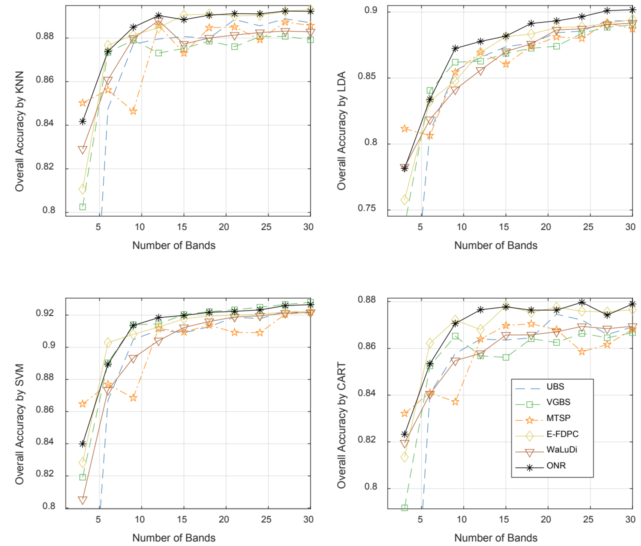

Fig. 2. OA curves on Salinas Scene for different band selection methods, in which m is set from 3 to 30 each 3 intervals.

[2] Qi Wang, Jia Wan, and Yuan Yuan, "Locality constraint distance metric learning for traffic congestion detection," *Pattern Recognition*, 2017.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] F. Zhang, Q. Wang, and X. Li, "Hyperspectral image band selection via global optimal clustering," in *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2017, pp. 1–4.

[9] Raymond P. Jefferis and Kenneth A. Fegley, "Application of dynamic programming to routing problems," *IEEE Trans. Systems Science and Cybernetics*, vol. 1, no. 1, pp. 21–26, 1965.

[10] 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.

[11] S. Jia, G. Tang, J. Zhu, and Q. Li, "A novel ranking-based clustering approach for hyperspectral band selection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 1, pp. 88–102, 2016.