SELECTING BAND SUBSETS FROM HYPERSPECTRAL IMAGE THROUGH A NOVEL EVOLUTIONARY-BASED STRATEGY

Xuefeng Jiang[1], *Lin Zhang*[1], *Fahong Zhang*[12], *Junrui Liu*[1], *Qi Wang*[123*]

[1]School of Computer Science, Northwestern Polytechnical University, Xi'an 710072, Shaanxi, P. R. China
[2]Center for OPTical IMagery Analysis and Learning, Northwestern Polytechnical University, Xi'an 710072, Shaanxi, P. R. China
[3]Unmanned System Research Institute, Northwestern Polytechnical University, Xi'an 710072, Shaanxi, P. R. China

ABSTRACT

Hyperspectral dimensionality reduction by optimal band selection attracts wide attention recently because a few pivotal and physically meaningful bands can not only represent the whole image cube without losing effectiveness but also mitigate the computational burden. In this paper, we construct an efficient searching strategy based on the clonal selection principle to optimize a geometry-based criterion named maximum ellipsoid volume (MEV). The main contributions are two-fold: 1) a subtle relationship that can accelerate the calculation of the criterion and 2) an evolutionary strategy to relieve the heavy computational burden of obtaining the desired bands from numerous quality candidates. The experimental result on a real hyperspectral data demonstrates that the proposed method is effective.

Index Terms— Band selection, hyperspectral image, maximum ellipsoid volume, clonal selection principle.

1. INTRODUCTION

In the past two decades, hyperspectral remote sensing has become the forefront technology in the field of remote sensing. The spectral resolution of hyperspectral image (HSI) can reach 0.01 μm, providing the potential of accurate object identification [1] and detection [2]. Nevertheless, the vast volume of data leads to high storage and transmission costs and a heavy computational burden. Band selection (BS), selecting only a few bands with discriminative information, is an appropriate approach to cope with the problem.

These years, considerable research efforts have been devoted to BS. As a result, a large number of BS methods have been proposed. According to whether the methods make use of the class label of training samples or not, they can be roughly categorized into supervised [3, 4] and unsupervised methods [5]. Owing to the difficulty of obtaining labeled samples, we focus on unsupervised methods, further categorized into four types: ranking-based methods, clustering-based methods, greedy-based methods and evolutionary-based methods in [6]. Some typical works are listed as follows. Based on the eigenvalues and eigenvectors of the variance-covariance matrix, the maximum-variance principal components analysis (MVPCA) [7] method constructs the loading-factors matrix for band prioritization. After prioritizing all bands, it employs a divergence-based band decorrelation method to remove redundant bands. Clustering-based band selection (CBBS) [8] utilizes information measures, such as mutual information and Kullback-Leibler divergence, to construct a dissimilarity matrix, after which Ward's linkage method is applied to the matrix to get the required number of clusters. Volume-gradient-based band selection (VGBS) [9] greedily removes redundant bands in the light of the gradient of volume with regard to HSI and eventually gets the bands with large volume of the ellipsoid [10]. In [11], band selection is transformed into a multitask sparse learning problem. Candidate band subsets evaluated by the introduced criterion are searched based on immune clonal strategy [12].

Briefly speaking, band selection aims to choose k out of L bands in a HSI, which surpasses all the other combinations in term of some specific criterion. Given a proper criterion, there are $C_L^k = \frac{L!}{k!(L-k)!}$ candidate combinations. Assuming $L = 200$, $k = 15$, C_L^k is approximately 1.46×10^{22}. Hence, an exhaustive search is negated in respect of computational complexity, as a consequence of which it is almost impossible to achieve the optimal band subset corresponding to the adopted objection function. In this situation, researchers have to turn to the suboptimal solution.

In this paper, we propose an unsupervised hyperspectral band selection called MEV-IC, which employs a novel evolutionary-based stategy to find the band subset with maximum ellipsoid volume (MEV), a geometry-based criterion. The two contributions of this paper are claimed in the following. First, we find a subtle relationship between the $k \times k$ variance-covariance matrix associated with a specific k-band subset and the $L \times L$ variance-covariance matrix associated with all bands, which speeds up evaluations of possible subsets. Second, we design an evolutionary searching strategy which can reliably and quickly search for the desired band subset from gazillions of possible solutions on the basis of the clonal selection principle.

2. BACKGROUND

In this section, we will review the MEV method [10] initially designed for multispectral images.

From geometric perspective, it's universally known that the variance-covariance matrix associated with a specific k-band subset defines an ellipsoid within the k-D subspace spanned by the k bands. The ellipsoid volume of a k-band subset is $\frac{\pi^{\frac{k}{2}}}{\Gamma(\frac{k}{2}+1)}\prod_{i=1}^{k}\sqrt{\lambda_i}$, where $\sqrt{\lambda_i}$, for each $1 \leq i \leq k$, is the corresponding principal axis of the ellipsoid and $\Gamma(\cdot)$ is the gamma function. Furthermore, the product of the principal axes of the ellipsoid is equal to the square root of the determinant of the corresponding $k \times k$ matrix. Therefore, the volume of the ellipsoid associated with a specific k-band subset and the square root of the determinant of the $k \times k$ variance-covariance matrix of that subset differ by a constant multiple. In [10], Sheffield proved that the joint entropy for data with multivariate Gaussian distribution equals to half the logarithm of the variance-covariance matrix determinant except for an additive constant. Accordingly, selecting the k-band subset with MEV is exactly equivalent to selecting the k-band subset with maximum joint entropy.

However, the MEV method can't be directly applied in HSIs owing to an extreme number of possible combinations.

3. METHOD

This section details the proposed MEV-IC method for unsupervised band selection. First, the relationship between the $k \times k$ variance-covariance matrix associated with a specific k-band subset and the $L \times L$ variance-covariance matrix associated with all bands is further analyzed. Then, a computationally feasible evolutionary search strategy is constructed in order to efficiently obtain the desired band subset from an excess of trillions of candidates.

3.1. Subtle relationship

For convenience, a centralized and normalized HSI cube with a total of L bands is denoted as $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_L]^T \in \mathbb{R}^{L\times N}$, where $\mathbf{x}_i \in \mathbb{R}^N$ describes the i-th spectral band with N pixels. Then an examined k-band subset can be denoted by $\mathbf{M} = [\mathbf{x}_{i_1}, \mathbf{x}_{i_2}, \ldots, \mathbf{x}_{i_k}]^T \in \mathbb{R}^{k\times N}, 1 \leq i_1 < i_2 < \cdots < i_k \leq L$, where k is the number of selected bands. Further, the variance-covariance matrices associated with all bands and the k-band subset can be expressed as follows:

$$\mathbf{D} = \mathbf{M}\mathbf{M}^T = [\mathbf{x}_{i_1}, \mathbf{x}_{i_2}, \ldots, \mathbf{x}_{i_k}]^T([\mathbf{x}_{i_1}, \mathbf{x}_{i_2}, \ldots, \mathbf{x}_{i_k}]^T)^T, \quad (1)$$

and

$$\mathbf{B} = \mathbf{X}\mathbf{X}^T = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_L]^T([\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_L]^T)^T, \quad (2)$$

respectively. For each pair of m and n s.t. $1 \leq m \leq n \leq k$,

$$\mathbf{D}(m, n) = \mathbf{x}_{i_m}^T \mathbf{x}_{i_n} = \mathbf{B}(i_m, i_n). \quad (3)$$

According to (1), (2) and (3), $\mathbf{D}$ can be formed by the i_1-th, i_2-th, ..., i_k-th rows and the i_1-th, i_2-th, ..., i_k-th columns of $\mathbf{B}$, namely that we obtain the variance-covariance matrices associated with all the possible band combinations as long as the variance-covariance matrix $\mathbf{B}$ is calculated, which saves computation time for calculating the determinant of the variance-covariance matrices associated with the quality band combinations.

3.2. Computationally feasible evolutionary strategy

As analyzed in [13], some greedy search strategies such as forward selection and backward elimination can quickly achieve a solution at the cost of the accuracy. However, the solution is poor globally. Fortunately, for such a specific combinatorial optimization problem, the clonal selection algorithm [12] can provide us with a high-quality solution owing to its superb global searching ability. Therefore, we reformulate the clonal selection algorithm to search for the band subset with MEV.

The clonal selection algorithm is motivated by the clonal selection theory that explains how antibody-forming cells react to specific invading antigens and destruct them. In detail, when a body is invaded by an antigen, the original antibody-forming cells of the body produce sufficient antibodies of different affinities to interact with the antigen. Then those antibody-forming cells producing antibodies with high affinities survive, proliferate in proportion to their affinities and mutate stochastically. The process will be repeated until the antibody with desired affinity can be produced. Therefore, the body can form an effective defense against the antigen. Similarly, the combinatorial optimization problem for selecting the band subset with MEV is regarded as the antigen while the chosen band subset is treated as the antibody. It is noteworthy that the antibody-forming cell as the intermediary is removed for simplification. Further, the affinity of an antibody can be defined by the opposite of the function below:

$$F(\mathbf{M}) = \frac{1}{\sqrt{\mathrm{Det}(\mathbf{D})}} = \frac{1}{\sqrt{\mathrm{Det}(\mathbf{M}\mathbf{M}^T)}}. \quad (4)$$

This function can be interpreted as the reciprocal of the square root of the variance-covariance matrix associated with the examined k-band subset. Noting that each $\mathbf{M}$ is a k-subset of $\mathbf{X}$, we can specify it merely by a k-D integer vector in implementation, which indicates which columns of $\mathbf{X}$ are selected. Last but not least, a smaller value of $F(\cdot)$ leads to a better combination.

The clonal selection principle can be summarized as population diversity, genetic mutation and natural selection. First, we set the population size N_0 of the initial antibodies $\mathcal{S}$ heuristically as 13 and select N_0 different k-band subsets at random. Second, each $\mathbf{M}$ in $\mathcal{S}$ produces $N_C(\mathbf{M})$ copies and all the copies should mutate stochastically to new antibodies. It should further be noted that a hash map is supplemented to avoid the repeated antibodies. Third, among the new antibodies produced in the last step, the only one with minmum $F(\cdot)$ value is compared to $\mathbf{M}$, and then if its $F(\cdot)$ value is smaller than $\mathbf{M}$'s, it will replace the original $\mathbf{M}$ in the next iteration; otherwise, it will be discarded. In this paper, $N_C(\mathbf{M})$ is determined by

$$N_C(\mathbf{M}) = \text{Fac} \times \mathit{Ceil}(\min(2, \frac{F(\mathbf{M})}{\text{Q}})), \tag{5}$$

where Fac is a starting variation coefficient fixed to 10, $\mathit{Ceil}(\cdot)$ the round-up function and Q the minimal $F(\cdot)$ value of all the antibodies in $\mathcal{S}$. Moreover, the maximum number of iterations is set as 650. Only if the relative change rate of the smallest $F(\cdot)$ value in the last 100 steps drops to below a predefined threshold 10^{-6} will the updating procedure stop ahead of time.

3.3. Flowchart of MEV-IC

The proposed MEV-IC algorithm is sketched in pseudocode form in Algorithm 1.

Algorithm 1 The MEV-IC algorithm

Input: The observed HSI $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_L]^T$ and the number of selected bands K.

Preprocessing: Normalize $\mathbf{X}$ between 0 and 1 and remove the mean of each hyperspectral band $\mathbf{x}_i$. For convenience, the mean-shifted data is still denoted as $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_L]^T$.

Randomly choose the initial antibodies $\mathcal{S}$ of size N_0 and calculate their $F(\cdot)$ values.

repeat

- **Clone:** Each antibody in $\mathcal{S}$ is cloned proportionally to its $F(\cdot)$ value.
- **Mutation:** All the antibodies produced in the above step mutate independently and stochastically to generate new antibodies.
- **Selection:** Compare each antibody $\mathbf{M}$ in $\mathcal{S}$ to the antibody with minmum $F(\cdot)$ value among the antibodies generated form $\mathbf{M}$ and reserve the antibody with smaller $F(\cdot)$ value.

until The maximum number of iterations is reached or the relative change rate of the smallest $F(\cdot)$ value in the last 100 steps drops to below 10^{-6}.

Output: The antibody (desired band subset) with the smallest $F(\cdot)$ value.

4. EXPERIMENT

4.1. Experimental Settings

In order to verify the effectiveness of the MEV-IC algorithm, we conduct experiments on Pavia University, a widely used HSI captured by ROSIS system in 2002, which consists of 103 hyperspectral bands with 12 non-informative bands removed, each containing 610×340 pixels. In addition, there are 9 classes of available objects in the image. First of all, MEV-IC is compared with VGBS [9] in term of determinant Det because it adopts the same criterion as MEV-IC. Furthermore, in HSI classification, several representative competitors are taken as benchmarks: clustering-based band selection using mutual information (WaLuMI) or Kullback-Leibler divergence (WaLuDi), uniform band selection (UBS) [7], VGBS [9] and MTSP [11]. What's more, four commonly used classifiers are employed to perform the classification: support vector machine (SVM), linear discriminant analysis (LDA), k-nearest neighborhood (KNN), and classification and regression trees (CART). Besides, we take 10% samples for each class at random as the training dataset and the rest as the test dataset.

Table 1. The ratios of the determinant Det of the variance-covariance associated with the bands selected by VGBS to that by MEV-IC

Number of selected bands (k)	3	6	9	12	15	18	21	24	27	30
Ratio (%)	60.4	79.1	67.5	58.1	53.2	49.6	41.4	38.0	41.5	27.8

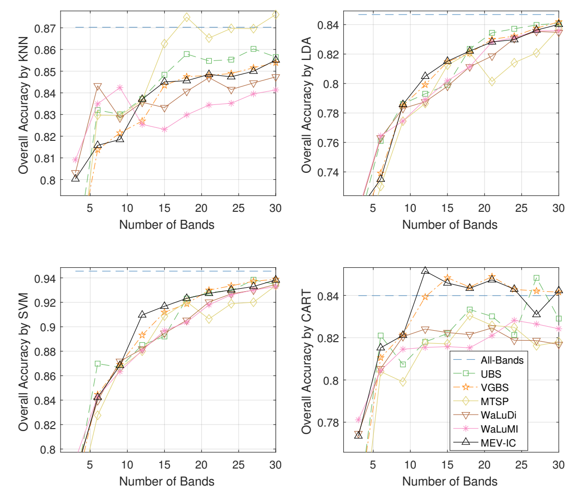

Fig. 1. OA curves on Pavia University for different band selection methods.

4.2. Experimental Result and Discussion

According to Table 1, the determinant Det of the variance-covariance associated with the bands selected by MEV-IC is much greater than that by VGBS, which implies that when we take MEV as the objective function, MEV-IC can find better solutions than VGBS.

Table 2. Average classification accuracies on Pavia University Scene by four different classifiers.

	KNN	LDA	SVM	CART	Mean
UBS	<u>0.8395</u>	**0.8019**	0.8938	0.8169	0.8380
VGBS	0.8315	0.7997	<u>0.8944</u>	<u>0.8279</u>	<u>0.8384</u>
MTSP	**0.8457**	0.7877	0.8831	0.8066	0.8308
WaLuDi	0.8365	0.7975	0.8886	0.8147	0.8343
WaLuMI	0.8315	0.7986	0.8886	0.8146	0.8333
MEV-IC	0.8363	<u>0.8011</u>	**0.8967**	**0.8314**	**0.8414**

Fig. 1 illustrates the overall accuracy (OA) values of four different classifiers. Except for the KNN classifier, the proposed MEV-IC method performs the best in most cases. In addition, Table 2 shows the accuracy results of four different classifiers. (The best and second-best results in each row are highlighted in bold and underlined, respectively.) As for the mean of the accuracy results of four different classifiers (namely the last row in the Table 2), MEV-IC takes over the lead, followed by VGBS, clarifying that MEV-IC can select bands with low correlation. UBS takes the third place due to the low correlation of the uniformly selected bands. When the KNN classifier is employed, MEV-IC do not perform well while MTSP outperform other competitors. That is because the KNN classifier is too sensitive to noise and MTSP succeed to avoid selecting the bands carrying some noise. Moreover, MEV-IC achieves the best or second-best classification results except for the KNN classifier, which further verifies the effectiveness of MEV-IC.

In conclusion, our MEV-IC method is more effective than the compared methods.

5. CONCLUSION

In this paper, an evolutionary-based unsupervised hyperspectral band selection is proposed. We first discover the variance-covariance matrix associated with a specific band subset is the submatrix of the variance-covariance matrix associated with all bands, which accelerates the calculation of the criterion. Then we construct a computationally feasible evolutionary search strategy which relieves the heavy computational burden of obtaining the desired bands based on the clonal selection principle. The result of a real HSI data experiment clearly demonstrates the proposed MEV-IC method is effective and accurate.

6. 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.

7. REFERENCES

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

[2] Qi Wang, Jia Wan, and Yuan Yuan, "Locality constraint distance metric learning for traffic congestion detection," *Pattern Recognition (PR)*, vol. 75, pp. 272–281, 2018.

[3] Xianghai Cao, Tao Xiong, and Licheng Jiao, "Supervised band selection using local spatial information for hyperspectral image," *IEEE Geoscience and Remote Sensing Letters (GRSL)*, vol. 13, no. 3, pp. 329–333, 2016.

[4] He Yang, Qian Du, Hongjun Su, and Yehua Sheng, "An efficient method for supervised hyperspectral band selection," *IEEE Geoscience and Remote Sensing Letters (GRSL)*, vol. 8, no. 1, pp. 138–142, 2011.

[5] Qi Wang, Jianzhe Lin, and Yuan Yuan, "Salient band selection for hyperspectral image classification via manifold ranking," *IEEE Transactions on Neural Networks (TNN)*, vol. 27, no. 6, pp. 1279–1289, 2016.

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

[7] Cheini Chang, Qian Du, Tzulung Sun, and Mark L Althouse, "A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing (TGRS)*, vol. 37, no. 6, pp. 2631–2641, 1999.

[8] Adolfo Martinezuso, Filiberto Pla, Jose Martinez Sotoca, and Pedro Garciasevilla, "Clustering-based hyperspectral band selection using information measures," *IEEE Transactions on Geoscience and Remote Sensing (TGARS)*, vol. 45, no. 12, pp. 4158–4171, 2007.

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

[10] C Sheffield, "Selecting band combinations from multi spectral data.," *Photogrammetric Engineering & Remote Sensing (PERS)*, vol. 51, no. 6, pp. 681–687, 1985.

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

[12] Leandro N De Castro and Fernando J Von Zuben, "Learning and optimization using the clonal selection principle," *IEEE Transactions on Evolutionary Computation (TEC)*, vol. 6, no. 3, pp. 239–251, 2002.

[13] Isabelle Guyon and Andre Elisseeff, "An introduction to variable and feature selection," *Journal of Machine Learning Research (JMLR)*, vol. 3, no. 0, pp. 1157–1182, 2003.