A New Unsupervised Hyperspectral Band Selection Method Based on Multi-objective Optimization

Xia Xu, Zhenwei Shi and Bin Pan

Abstract

Unsupervised band selection methods usually assume specific optimization objectives which may include band or spatial relationship. However, since one objective could only represent parts of hyperspectral characteristics, it is difficult to determine which objective is the most appropriate. In this letter, we propose a new multi-objective optimization based band selection method, which is able to simultaneously optimize several objectives. The hyperspectral band selection is transformed to a combinational optimization problem, where each band is represented by a binary code. More importantly, to overcome the problem of unique solution selection in traditional multi-objective methods, we develop a new incorporated rank based solution set concentration approach in the process of Tchebycheff decomposition. The performance of our method is evaluated under the application of hyperspectral imagery classification. Three recently proposed band selection methods are compared in the experimental part.

Index Terms

Band selection, hyperspectral image, multi-objective optimization

I. Introduction

Hyperspectral images (HSIs) contain abundant materials information with hundreds of narrow bands. However, due to the strong correlation among bands, there is some redundant information in HSI data, which has little influence on land-cover classification [1]. Band selection refers to find the most representative bands that are able to achieve comparable performance with the original data in some applications. Different from some feature extraction methods such as principal component analysis (PCA) [2] and independent component analysis (ICA) [3], band selection can reduce the storage, management and computation cost of HSIs while preserving their physical interpretability. Band selection can be implemented in a supervised or unsupervised manner. Supervised methods usually require class

The work was supported by the National Natural Science Foundation of China under the Grants 61671037, and the Beijing Natural Science Foundation under the Grant 4152031, and the Excellence Foundation of BUAA for PhD Students under the Grants 2017057. *(Corresponding author: Zhenwei Shi.)*

Xia Xu, Zhenwei Shi (Corresponding Author) and Bin Pan are with Image Processing Center, School of Astronautics, Beihang University, Beijing 100191, China, and with Beijing Key Laboratory of Digital Media, Beihang University, Beijing 100191, China, and also with State Key Laboratory of Virtual Reality Technology and Systems, School of Astronautics, Beihang University, Beijing 100191, China, (e-mail: xuxia@buaa.edu.cn; shizhenwei@buaa.edu.cn; panbin@buaa.edu.cn).

knowledge as a *priori*, but it is often unavailable in many real applications. Therefore in this letter we focus on an unsupervised approach.

Unsupervised band selection methods tend to find a representative subset from the whole bands. Traditional band selection methods usually rely on certain statistical properties of HSIs. Geng *et al.* [4] proposed a simplex volume gradient based method to remove redundant bands. Then in [5] a clustering based method was developed which considered each channel map as a data point. Sun *et al.* used sparse representation methods to select the band subset [6], [7]. Zhu *et al.* [8] exploited the structure-aware measures for band informativeness and independence. In [9] a geometrical measures based method was proposed. In [10] overall accuracy and the redundancy were used as the objective function to conduct unsupervised band selection. Recently, some evolutionary algorithms were used for hyperspectral band selection, such as firefly algorithms [11], [12] and particle swarm optimization [13].

However, Among many optimization objective functions for unsupervised band selection, it is difficult to determine which one is the best. Optimizing two or more objective functions simultaneously may lead to better performance.

Inspired by this idea, we attempt to simultaneously minimize several conflicting objective functions via a multi-objective framework. Multi-objective algorithms (MOAs) are suitable to tackle the combinational optimization, which is actually non-convex. For subset selection problems, MOAs have presented superior performance under certain circumstances, when compared with greedy and convex relaxation methods [14]. MOAs began to be studied in hyperspectral image processing since 2016 [15], [16]. In [15], Gong *et al.* introduce multi-objective evolutionary algorithm to hyperspectral band selection for the first time. The motivation of [15] was to provide a series of subsets with different numbers of bands so as to offer more options for decision. Thus it was not suitable in the case that the number of selected bands should be fixed. Generally, MOAs try to find an optimal subset from the solution space. However, in some specific such as hyperspectral band selection, only a single solution is needed. Therefore it is necessary to develop a new approach which could shrink the solution space.

In this letter, we propose an incorporated rank based multi-objective band selection algorithm (IRMoBS) method which is able to simultaneously optimize several objective functions with conflicting relationship. A tri-objective optimization problem is solved in IRMoBS, considering the band number, variance and information entropy of the selected bands. The three objectives are just an available choice. They could be replaced by other preferred objective functions. IRMoBS is developed under the basis of multi-objective evolutionary algorithm based on decomposition (MOEA/D) because of its fast evolutionary speed and low computational complexity [17]. However, MOEA/D would report several non-dominated solutions, which leads to difficulty in making the final decision in the problem of hyperspectral band selection. Therefore, in this letter, we improve MOEA/D by taking the correlation among different individuals into account in the Tchebycheff decomposition approach in MOEA/D. The relationships between individuals are considered as a regularizer. Individuals that have bigger difference to the ideal individual are more likely to be replaced by other closer ones, and this operation is beneficial to the concentration of the population.

The major contributions of this letter contains two aspects:

- An incorporated rank based multi-objective band selection method for hyperspectral image is proposed, which

is able to optimize multiple objective functions simultaneously.

- An improved Tchebycheff decomposition approach is developed based on MOEA/D, where correlation among hyperspectral bands is considered as a regularizer. This improvement is beneficial to the concentration of all individuals.

II. Proposed method

Information entropy, distance metrics such as statistical features, the Bhattacharyya distance, or the Jeffreys-Matusita distance are commonly used objective functions for band selection. Each criteria has its own advantages, but it is hard to say which one is the best. Thus we consider combining several objectives for band selection. Inspired by the fact that MOAs could handle several goals simultaneously, in this letter, a tri-objective optimization based band selection method is proposed under the framework of MOEA/D.

A. Multi-objective Model for IRMoBS

In IRMoBS, the band selection problem is transformed to an ℓ_0 norm based combinational optimization problem, where each band of HSI data is represented by a binary code. Suppose $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_L]^T \in \mathbb{R}^{L\times M}$ is the observed HSI data, which contains L bands and M pixels. The binary vector representation for $\mathbf{X}$ is $\mathbf{b} = [b_1, b_2, \cdots, b_L]$, where $b_i = 1$ if the i-th band is selected and $b_i = 0$ otherwise. Three objectives are designed for IRMoBS, as

$$\begin{aligned} &\min_{\mathbf{b}} \mathcal{F}(\mathbf{b}) = [f_1(\mathbf{b}), f_2(\mathbf{b}), f_3(\mathbf{b})] \\ &\begin{cases} f_1(\mathbf{b}) = \dfrac{1}{|\mathbf{b}|}\displaystyle\sum_{i=1}^{|\mathbf{b}|}\frac{1}{S(\mathbf{x}_{\mathbf{b}}^i)} \\ f_2(\mathbf{b}) = \dfrac{1}{|\mathbf{b}|}\displaystyle\sum_{i=1}^{|\mathbf{b}|}\frac{1}{H(\mathbf{x}_{\mathbf{b}}^i)} \\ f_3(\mathbf{b}) = \big||\mathbf{b}| - k\big| \end{cases} \end{aligned} \tag{1}$$

where $\mathbf{X}_{\mathbf{b}}$ denotes the selected bands data, $S(\mathbf{x}_{\mathbf{b}}^i)$ and $H(\mathbf{x}_{\mathbf{b}}^i)$ are the variance and information entropy of the i-th band respectively, $\big||\mathbf{b}| - k\big|$ is the band number error, $|\mathbf{b}|$ is the cardinality of $\mathbf{b}$, k is the preset band number. Among the three objectives used in Eq. (1), $f_3(\cdot)$ is essential, because it is an important constraint for the number of selected bands. Moreover, $f_3(\cdot)$ is ℓ_0 norm based, and we represent it via binary coding. $f_1(\cdot)$ and $f_2(\cdot)$ are used here because they are very popular and widely used representations about the amount of image information. It is safe to replace them with other preferred objectives.

The Tchebycheff decomposition approach is used in IRMoBS, which could decompose Eq. (1) to a number of scalar single-objective optimization subproblems. Then the i-th subproblem of (1) is

$$\begin{aligned} &\min g_i^{te}(\mathbf{b}|\mathbf{w}_i, \mathbf{z}^*) = \max_{1\le j\le 3}\{w_i^j |f_j(\mathbf{b}) - z_j^*|\} \\ &s.t.\ \mathbf{b} \in \Omega \end{aligned} \tag{2}$$

where $\mathbf{w}_i = [w_i^1, w_i^2, w_i^3]^T$ is the weight vector of the i-th subproblem for the three objectives with $\mathbf{w}_i \geq 0$ and $\sum_{j=1}^{3} w_i^j = 1$, $\mathbf{z}^* = [z_1^*, z_2^*, z_3^*]^T$ is the ideal point of the current population and $z_j^* = \min\{f_j(\mathbf{x})|\mathbf{x} \in \Omega\}$, Ω is the decision space.

B. IRMoBS

Due to the confliction of objectives in MOAs, a set of non-dominated solutions are obtained. They provide a balance of the objectives, so that it is hard to give a rank and make a decision among them. Thus in this letter, to solve the problem, the correlation among different individuals are taken into account in the evolution process. We use an incorporated rank as the correlation criterion. For example, there are two individuals $\mathbf{b}^1 = [b_1^1, b_2^1, \cdots, b_{L_1}^1]$ and $\mathbf{b}^2 = [b_1^2, b_2^2, \cdots, b_{L_2}^2]$. Their corresponding subsets of $\mathbf{X}$ are $\mathbf{X}_{\mathbf{b}^1} = [\mathbf{X}_{b_1^1}, \mathbf{X}_{b_2^1}, \cdots, \mathbf{X}_{b_{L_1}^1}]$ and $\mathbf{X}_{\mathbf{b}^2} = [\mathbf{X}_{b_1^2}, \mathbf{X}_{b_2^2}, \cdots, \mathbf{X}_{b_{L_2}^2}]$. Accordingly, the incorporated subset of $\mathbf{X}$ is $\mathbf{X}_{\mathbf{b}} = \mathbf{X}_{\mathbf{b}^1} \cup \mathbf{X}_{\mathbf{b}^2} = [\mathbf{X}_{b_1}, \mathbf{X}_{b_2}, \cdots, \mathbf{X}_{b_L}]$. Then the incorporated rank of individual $\mathbf{b}^1$ and $\mathbf{b}^2$ is

$$\mathcal{C}(\mathbf{b}^1, \mathbf{b}^2) = \frac{rank(\mathbf{X}_{\mathbf{b}})}{L} \tag{3}$$

where $rank(\mathbf{X}_{\mathbf{b}})$ is the rank of $\mathbf{X}_{\mathbf{b}}$, L is the band number of the incorporated matrix $\mathbf{X}_{\mathbf{b}}$. If there exists a bigger difference between the two individuals $\mathbf{b}^1$ and $\mathbf{b}^2$, the incorporated rank of them is larger. Then we improve the Tchebycheff decomposition approach as follows

$$\begin{aligned} &\min g_i^{ite}(\mathbf{b}|\mathbf{w}_i, \mathbf{z}^*, \mathbf{b}^*) = \max_{1 \leq j \leq 3} \{w_i^j |f_j(\mathbf{b}) - z_j^*|\} + \lambda \cdot \mathcal{C}(\mathbf{b}, \mathbf{b}^*) \\ &s.t.\ \mathbf{b} \in \Omega \end{aligned} \tag{4}$$

where $g_i^{ite}(\mathbf{b}|\mathbf{w}_i, \mathbf{z}^*, \mathbf{b}^*)$ is the objective function of each decomposed subproblem, $\mathbf{b}^*$ is the ideal individual corresponding to the ideal point $\mathbf{z}^*$, λ is a regularizer that used to control the correlation between the individual $\mathbf{b}$ and ideal individual $\mathbf{b}^*$. In view of the physical meaning of $\mathbf{z}^*$, it is defined as $\mathbf{z}^* = \min\{\|\mathcal{F}(\mathbf{b}_i)\|_2 |\mathbf{b}_i \in \Omega, i = 1, ..., p\}$. In the subproblem evolution process, the incorporated rank between all neighbourhood individuals and the ideal individual are calculated respectively to judge which individuals could be updated. Individuals that have larger g^{ite} values, are more likely to be replaced by the others that are closer to the ideal individual. In this case, individuals are more likely to be concentrated to positive areas in the feasible searching space. Then the decision making problem could be avoided. The detail flowpath of IRMoBS is presented in **Algorithm 1**.

In **Algorithm 1**, the population is initialized based on a bit-wise flipping strategy, where each bit of an individual $\mathbf{b}_i (i = 1, 2, \cdots, p)$ is set as 1 with a probability $1/M$. This population is archived in EP. Then the ideal individual $\mathbf{b}^*$ and ideal point $\mathbf{z}^*$ are determined. For each subproblem, the individuals whose Euclidean distances of aggregate coefficient to the subproblem's are closer, are selected as the neighbourhood of it. In the iteration process, a new solution $\mathbf{b}_i'$ is generated based on the bit-wise flipping strategy for each individual $\mathbf{b}_i (i = 1, 2, \cdots, p)$ (line 9). Detailedly, each bit of the individual is flipped to its opposite with a probability $1/M$ and remains unchanged otherwise. The ideal point is updated after comparing the objectives of individual $\mathbf{b}^*$ (line 10, 11). Then the objective values are computed based on (4) for the neighbourhood individuals. It could be seen as a distance metric

Algorithm 1: IRMoBS

Input: Hyperspectral image data $\mathbf{X}$.
Preprocessing:
Set the number of bands as k.
Initialization:
population size p, neighborhood size ns, maximum iteration number T, regularization parameter μ, a population $\mathbf{B} = \{\mathbf{b}^1, ..., \mathbf{b}^p\}$, a set of weight vector $\mathbf{W} = \{\mathbf{w}_1, ..., \mathbf{w}_p\}$, indexes of each subproblem's neighbors $\{\mathcal{N}_1, ..., \mathcal{N}_p\}$, the ideal point $\mathbf{z}^* = [z_1^*, z_2^*, z_3^*]^T$, an population archive EP.
Band Selection:
while $t < T$ **do**
  $t = t + 1$;
  for $i = 1, ..., p$ **do**
    Generate a new individual $\mathbf{b}_i^{'}$ from $\mathbf{b}_i$ based on bit-wise flipping strategy, where each bit is flipped with a probability $1/M$;
    if $\|\mathcal{F}(\mathbf{b}^*)\|_2 > \|\mathcal{F}(\mathbf{b}_j^{'})\|_2$ **then**
      Set $\mathbf{z}^* = \mathcal{F}(\mathbf{b}_i^{'})$ and $\mathbf{b}^* = \mathbf{b}_i^{'}$
    for $j \in \mathcal{N}_i$ **do**
      if $g_i^{ist}(\mathbf{b}_i^{'}|\mathbf{w}_i, \mathbf{z}^*, \mathbf{b}^*) \leq g_j^{ist}(\mathbf{b}_j|\mathbf{w}_j, \mathbf{z}^*, \mathbf{b}^*)$ **then**
        Set $\mathbf{b}_j = \mathbf{b}_i^{'}$ and $\mathcal{F}(\mathbf{b}_j) = \mathcal{F}(\mathbf{b}_i^{'})$
  Update the population archive EP.
Return the solution in EP as $\mathbf{b}^\star$ and record the corresponding band indices.

of the neighbourhood individuals and the ideal individual. Among the neighbourhood individuals, the ones whose objective values are larger than $g_i^{ist}(\mathbf{b}_i')$, are replaced by individual $\mathbf{b}_i'$ (line 12-14). After iteration by iteration, the population is unchanged in the archive EP. Finally, the selected bands are determined according to the solution in EP.

III. Experiments and Discussion

In this section, we evaluate the performance of IRMoBS and some recently proposed methods on two public hyperspectral data sets, Indian Pines and Pavia University[1]. Three recently proposed unsupervised band selection methods, improved sparse subspace clustering (ISSC) based [6], dissimilarity-weighted sparse self-representation (DWSSR) based [7] and dominant set extraction (DSEBS) based [8], are chosen for comparison. The selection quality is evaluated by supervised classification, where overall accuracy (OA) [18], [19] is used as the metric. The results are obtained by three popular classifiers, support vector machine (SVM), extreme learning machine (ELM) and k-nearest neighborhood (KNN). Radial Basis Function (RBF) kernels are adopted in ELM and SVM. The parameters in classifiers are determined by cross validation. We use 10% of the labeled pixels for training and the rest for testing. All the results are conducted 20 times and the average accuracy is reported.

A. Experimental Results on Indian Pines Data Set

Indian Pines data set was collected by AVIRIS sensor, which is a 145×145 pixels size, 20m spatial resolution scene including many natural perennial vegetation. There are 224 spectral bands available with wavelength range

[1]Available online: http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes

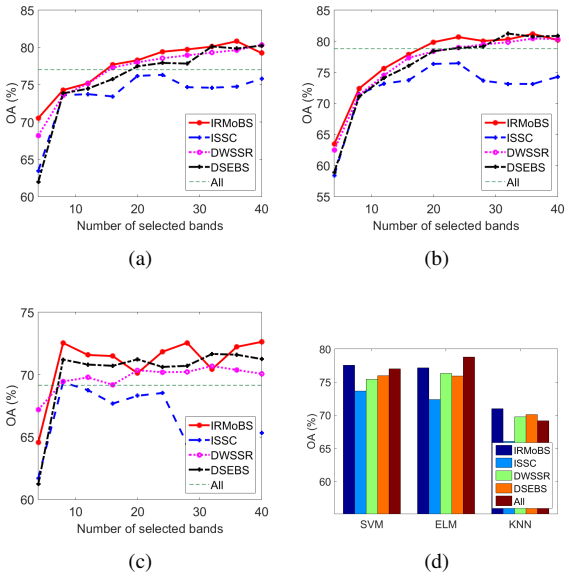


Fig. 1: Classification results on Indian Pines data set. (a)-(c) OA curves by SVM, ELM and KNN classifiers. (d) Average OA bars.

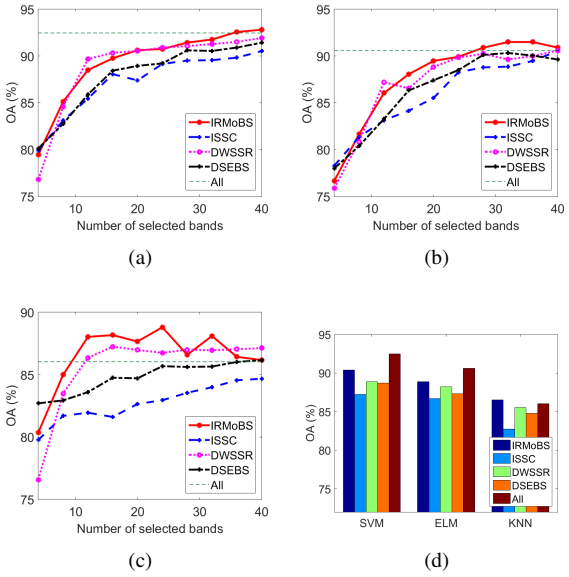


Fig. 2: Classification results on Pavia University data set. (a)-(c) OA curves by SVM, ELM and KNN classifiers. (d) Average OA bars.

400-2500nm. The number of bands is reduced to 200 by removing water absorption bands. Totally 10249 pixels are labeled which is divided into 16 classes.

Fig. 1 displays the comparison results on this data set. From Fig. 1(a)-(c) we find that IRMoBS has presented better performance than others in most cases. When the number of selected bands is small, the classification accuracies improve quickly. However, the results vary little after about 16 selected bands. All the compared methods tend to converge by further increasing the bands number. Specially, classification accuracies without band selection are

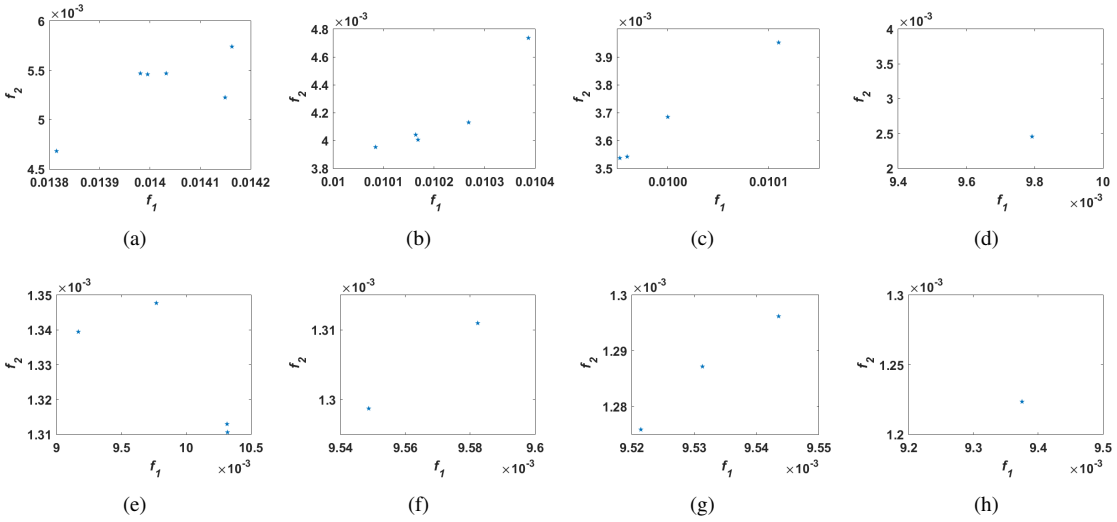


Fig. 3: Experiments on iterations. From left to right are the results by iteration 1, 5, 10 and 30 for (a)-(d) Indian Pines and (e)-(h) Pavia University data sets. The number of selected bands is set as 16.

also illustrated. All the compared methods have presented similar or slight better results. This phenomenon may indicate that band selection really works. In Fig. 1(d), the average accuracies by setting different band number are illustrated. About 1-3% advantages are observed by IRMoBS.

B. Experimental Results on Pavia University Data Set

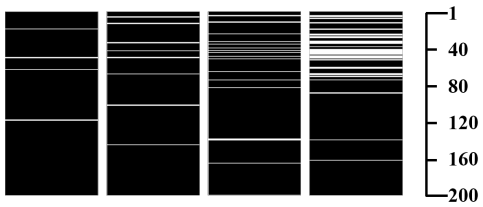
Pavia University data set was acquired by ROSIS sensor which covered a region of 610×340 pixels with 1.3m spatial resolution. 103 spectral bands are available, which ranges from 450 to 850nm wavelength. There are totally 42776 pixels labeled which contains 9 different land-cover materials.

Results on Pavia University data set are shown in Fig. 2. Because the number of training samples used here is larger than that in Indian Pines data set, Fig. 2 have presented high accuracies. IRMoBS still slightly outperforms others. Similar to Fig. 1, there are also significant ascents when the number of bands is small (*e.g.*, 4 or 8), followed by the flat tendency. We may conclude that few bands cannot provide an ideal representation for the original HSI data, and the accuracies comparison is meaningless under this circumstance. Therefore, the latter parts on the curves in Fig. 1 and Fig. 2 are more convincing. Different classifiers also have some influence on the results, but it is not obvious. We can see that IRMoBS reports higher accuracies in all the three classifiers used here after bands number is above 12. Thus we could conclude that IRMoBS is not a classifier-dependent method.

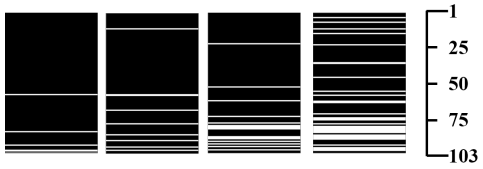
C. Analysis and Discussion

The convergence process of IRMoBS is illustrated in Fig. 3. To make the figure clearer, only residuals in f_1 and f_2 are displayed. As expected, the residuals decline with the increase of iterations. Moreover, since the major motivation of IRMoBS is to obtain a single solution from the solution set, the concentration analysis is more important. In the first iterations, many solutions are observed, just as traditional MOAs. However, with the increase

of iterations, the solution space shrinks, and the residuals continue declining. After 30 iteration, only one or two solutions could be observed. In this case, the decision is relatively easy to make.



(a)



(b)

Fig. 4: Band selection results for (a) Indian Pines and (b) Pavia University data sets. From left to right are the results by band number 4, 8, 16 and 32, respectively. The white lines corresponds to the selected bands.

In Fig. 4 we show the selected bands by white lines for the two data sets. The number of selected bands is respectively set as 4, 8, 16 and 32. It is observed that for AVIRIS data the bands selected by IRMoBS mainly concentrate on the front, especially between 750-950nm wavelength coverage. This result indicates that spectra nearby the boundary of visible and near infrared have more drastic changes, which may contribute to land cover identification. Similar result is also observed in Pavia University data set. Because ROSIS sensor has less spectral bands and narrower wavelength range, the selected bands actually correspond to 730-820nm, which is nearly the same with that in AVIRIS data. Although collected by different hyperspectral sensors, the selected bands by IRMoBS are similar. Therefore, Fig. 4 could demonstrate that IRMoBS is a robust algorithm which is not sensor-dependent.

IV. Conclusion

The performance of unsupervised hyperspectral band selection methods is affected by the optimization objective functions. Designing different objective functions could generate a series of band selection methods. However, it is hard to determine what objective is the most appropriate. In this letter, we propose a new multi-objective based band selection method, which is able to simultaneously optimize several objective functions. Moreover, to enhance the uniqueness of the final solution, we improve the original multi-objective framework by integrating the bands correlation information which is considered as a regularizer in the optimizing process. IRMoBS method could be considered as a framework, which is easy to be extended by different objective functions.

References

[1] C.-I. Chang, *Hyperspectral data exploitation: theory and applications*. John Wiley & Sons, 2007.

[2] M. D. Farrell and R. M. Mersereau, "On the impact of pca dimension reduction for hyperspectral detection of difficult targets," *IEEE Geoscience and Remote Sensing Letters*, vol. 2, no. 2, pp. 192–195, 2005.

[3] N. Falco, J. A. Benediktsson, and L. Bruzzone, "A study on the effectiveness of different independent component analysis algorithms for hyperspectral image classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 2183–2199, 2014.

[4] X. Geng, K. Sun, L. Ji, and Y. Zhao, "A fast volume-gradient-based band selection method for hyperspectral image," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 11, pp. 7111–7119, 2014.

[5] K. Sun, X. Geng, and L. Ji, "Exemplar component analysis: A fast band selection method for hyperspectral imagery," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 5, pp. 998–1002, 2015.

[6] W. Sun, L. Zhang, B. Du, W. Li, and Y. M. Lai, "Band selection using improved sparse subspace clustering for hyperspectral imagery classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 6, pp. 2784–2797, 2015.

[7] W. Sun, L. Zhang, L. Zhang, and Y. M. Lai, "A dissimilarity-weighted sparse self-representation method for band selection in hyperspectral imagery classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 9, pp. 4374–4388, 2016.

[8] G. Zhu, Y. Huang, J. Lei, Z. Bi, and F. Xu, "Unsupervised hyperspectral band selection by dominant set extraction," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 1, pp. 227–239, 2016.

[9] M. G. Asl, M. R. Mobasheri, and B. Mojaradi, "Unsupervised feature selection using geometrical measures in prototype space for hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 7, pp. 3774–3787, 2014.

[10] C. Sui, Y. Tian, Y. Xu, and Y. Xie, "Unsupervised band selection by integrating the overall accuracy and redundancy," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 1, pp. 185–189, 2015.

[11] R. Y. Nakamura, L. M. G. Fonseca, J. A. Dos Santos, R. d. S. Torres, X.-S. Yang, and J. P. Papa, "Nature-inspired framework for hyperspectral band selection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 4, pp. 2126–2137, 2014.

[12] H. Su, B. Yong, and Q. Du, "Hyperspectral band selection using improved firefly algorithm," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 1, pp. 68–72, 2016.

[13] M. Zhang, J. Ma, and M. Gong, "Unsupervised hyperspectral band selection by fuzzy clustering with particle swarm optimization," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 5, pp. 773–777, 2017.

[14] C. Qian, Y. Yu, and Z.-H. Zhou, "Subset selection by pareto optimization," in *Advances in Neural Information Processing Systems*, 2015, pp. 1774–1782.

[15] M. Gong, M. Zhang, and Y. Yuan, "Unsupervised band selection based on evolutionary multiobjective optimization for hyperspectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 1, pp. 544–557, 2016.

[16] X. Xu and Z. Shi, "Multi-objective based spectral unmixing for hyperspectral images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 124, pp. 54–69, 2017.

[17] Q. Zhang and H. Li, "MOEA/D: A multi-objective evolutionary algorithm based on decomposition," *IEEE Transactions on Evolutionary Computation*, vol. 11, no. 6, pp. 712–731, 2007.

[18] B. Pan, Z. Shi, and X. Xu, "R-VCANet: A new deep-learning-based hyperspectral image classification method," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 5, pp. 1975–1986, 2017.

[19] ——, "Hierarchical guidance filtering-based ensemble classification for hyperspectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 7, pp. 4177–4189, 2017.