Hyperspectral Image Unmixing Method Based on Multiple Kernel Graph Non-negative Matrix Factorization

Junyi Yao, Hongqiao Wang, Guangyuan Fu, Ling Wang
Xi’an Research Institute of High-Tech, Xi’an 710025, China
epgcyaojunyi@163.com, *whq05@mails.tsinghua.edu.cn

Abstract. In order to improve the accuracy of non-negative matrix factorization (NMF) in hyperspectral images unmixing, this paper proposes a hyperspectral image unmixing method based on multiple kernel graph NMF. Using multiple kernel learning (MKL), the hyperspectral data is mapped to the high-dimensional feature space, so that the linearly inseparable hyperspectral data in the low-dimensional space is linearly separable. Introducing the graph regular terms that characterizes the spatial structure relationship between pixels, the internal manifold structure of the hyperspectral image is effectively expressed. The results of both simulated dataset and real dataset show that compared with other method, the proposed method can improve the unmixing accuracy.

1. Introduction

As the characteristics of high spectral resolution, good expression of ground information and wide spectral response range, hyperspectral images have widely used in agriculture, minerals, military and other aspects[1-4]. However, due to the limitation of imaging principle and optical devices, the spatial resolution of hyperspectral images is not high, which leads to the common existence of mixed pixels in hyperspectral images. How to improve the accuracy of hyperspectral image unmixing has a direct impact on the application of hyperspectral images in high precision classification and image resolution improvement.

Non-negative Matrix Factorization (NMF) can be used for unsupervised unmixing of hyperspectral image[5-6]. When the spectral information is unknown, the NMF method approximathely decomposes the hyperspectral data into the product form of the basic matrix and the sparse matrix, namely the endmember matrix and the endabundance matrix. More recently, the NMF method has been widely used in hyperspectral image unmixing. Introducing the partition matrix theory to reduce the calculation scale of NMF method, Huang et. al. proposed a hyperspectral image unmixing method based on incremental kernel NMF (IKNMF)[5]. For no pure pixels exist in the hyperspectral images, Gan et. al. introduced a sparse constrained graph regularized NMF method[7]. Sun et. al. proposed an orthogonal NMF method that imposes orthogonality constraints on the factor matrices can improve the clustering performance[8], since it ensures the independent of the endmember spectra. However, all the above improved NMF algorithms have some shortcomings.

In recent years, because the kernel method can map the data which is linearly inseparable in the original space into a high-dimensional feature space to make it linearly separable and effectively improve the unmixing accuracy, the kernel method has been widely used in hyperspectral image unmixing[9-11]. Using the projection iterative method for endmember abstraction from the hyperspectral image and simulating the nonlinear training and testing dataset, Wu et. al. proposed an unmixing hyperspectral imagery based on support vector nonlinear approximating regression[12]. The
MKL method adopts the form of combining multiple kernel functions. By setting weight parameters for different kernel functions, the problem of kernel function selection is avoided\[13-15\]. Compared with the single kernel method, the MKL method can realize the automatic selection and weighting of the kernel function and obtain better mapping ability. Tan et. al. introduced the multiple kernel support vector machine (SVM) in nonlinear mixed pixel decomposition\[16\], and got higher unmixing accuracy than traditional single kernel method.

To improve the unmixing accuracy of the NMF method, this paper improves the graph NMF (GNMF) method by introducing MKL method, and proposes a hyperspectral image unmixing method based on multiple kernel graph non-negative matrix factorization. And this paper verifies the accuracy of the algorithm through simulation data and real hyperspectral image data experiments.

2. Related Work

2.1. NMF

NMF is a basis decomposition method. For non-negative matrix $V$, NMF method aims to find two non-negative matrixes $W$ and $H$

$$V \approx WH$$

The objective function of NMF is defined as

$$f(W, H) = \frac{1}{2} \|V - WH\|^2$$

where $V$ is the raw data, $W$ and $H$ are the decomposed base matrix and sparse matrix respectively. For the NMF method, the number of the required parameters $W$ and $H$ is more than the number of functions. As a result, the NMF method has obvious non-convexity, and converges on local optimum easily which leads to low accuracy of unmixing results\[17\].

2.2. GNMF

GNMF introduces the spatial relationship between adjacent pixels into constraints by constructing the nearest neighbour graph\[18\]. If two pixels belong to the same type of features, we can consider that the Euclidean distance of their spectral vectors is also close, and their corresponding abundance values are also the closest. For two pixels $v_i$ and $v_j$, if they correspond to the same type of features, their spectral curves will also be similar, which means the corresponding abundance values $H_i$ and $H_j$ are also similar in the subspace. Constructed the nearest neighbour graph of the hyperspectral data $\{v_1, v_2, \ldots, v_n\} \subset \mathbb{R}^{m \times n}$ with all pixels, and get the weight matrix $\Omega$. The objective function of GNMF is defined as

$$f(W, H) = \frac{1}{2} \|V - WH\|^2 + \frac{\mu}{2} Tr(HLH^T)$$

where $\mu$ is the graph regularization parameter, and $L$ is the Laplacian matrix based on manifold constraint? $L = D - \Omega$, $D = \text{diag}(d_1, d_2, \ldots, d_n)$, $d_i = \sum_{j=1}^{n} \Omega_{ij}$.

3. Approach

3.1. Determination of the Graph Regularization Weight

Using the homogeneity and difference between adjacent pixels of the hyperspectral image, the graph regularization weight function is constructed. The common weight definitions include 0-1 weighting, heat kernel weighting and dot product weighting\[18\].

In this paper, the Euclidean distance between two adjacent pixels is used to calculate their similarity\[19\]. For the pixel $v_i$ and $v_j$ in the hyperspectral image, the weight of their relationship can be calculated by 0-1 weighting as follows
\[
\Omega_{ij} = \begin{cases} 
1 & \text{if } v_i \in N_p(v_j) \text{ or } v_j \in N_p(v_i) \\
0 & \text{else} 
\end{cases} 
\] (4)

where \( N_p(v_i) \) is the \( p \times p \) neighbourhood centred on pixel \( v_i \).

Because the hyperspectral data has convex geometry characteristic, we use the heat kernel function to calculate the weight between the pixel and other pixels in the neighbourhood

\[
\Omega_{ij} = e^{-\frac{\|v_i - v_j\|^2}{\sigma}} 
\] (5)

Use (4) and (5), the graph regularization weight can be defined as

\[
\Omega_{ij} = \begin{cases} 
e^{-\frac{\|v_i - v_j\|^2}{\sigma}} & \text{if } v_i \in N_p(v_j) \text{ or } v_j \in N_p(v_i) \\
0 & \text{else} 
\end{cases} 
\] (6)

### 3.2. Multiple Kernel Graph Non-Negative Matrix Factorization

For the hyperspectral data \( V = [v_1, v_2, \cdots, v_n] \in \mathbb{R}^{m \times n} \), we can define a nonlinear map from the original input space \( V \) to the feature space \( F \) as follows

\[
\phi: v \in V \rightarrow \phi(v) \in F^m 
\] (7)

For the \( m \) objects, we can denote

\[
\phi(V) = [\phi(v_1), \phi(v_2), \cdots, \phi(v_n)] 
\] (8)

In the feature space, we use (1) and denote

\[
\phi(V) = W_\phi H_\phi + \varepsilon 
\] (9)

where \( W_\phi \in \mathbb{R}^{n \times r} \) and \( H_\phi \in \mathbb{R}^{r \times m} \) are the decomposed base matrix and sparse matrix respectively in the feature space. \( W_\phi \in \mathbb{R}^{n \times r} \) is a convex combination of \( \phi(V) \). So, there is \( U \in \mathbb{R}^{m \times r} \), and \( \sum U_{ij} = 1 \), where \( W_\phi = \phi(V)U \). The objective function is defined as

\[
f(U, H_\phi) = \frac{1}{2} \| \phi(V) - W_\phi H_\phi \|^2_F 
\]

\[
= \frac{1}{2} \| \phi(V) - \phi(V)U H_\phi \|^2_F 
\]

\[
= \frac{1}{2} \text{Tr} \left\{ \phi^\top(V)\phi(V) - 2\phi^\top(V)\phi(V)U H_\phi + H_\phi U^\top \phi^\top(V)\phi(V)U H_\phi \right\} 
\] (10)

We can define a convex combination of multiple kernel matrix \( K = \sum_{j=1}^{n} \beta_j K_j = \phi(V)^\top \phi(V) \), where \( \beta_j \geq 0 \), and \( \sum_{j=1}^{n} \beta_j = 1 \). With (3) and (10), the objective function of the multiple kernel graph non-negative matrix factorization can be defined as

\[
f(U, H_\phi) = \frac{1}{2} \text{Tr} \left\{ \sum_{j=1}^{n} \beta_j K_j \| E - U H_\phi \|^2_F \right\} + \frac{\mu}{2} \text{Tr} \left( H_\phi L H_\phi^\top \right) 
\] (11)

Fixing \( U \) and \( H_\phi \), the objective function (11) becomes
\[
f(U, H_{\phi}) = \frac{1}{2} \sum_{j=1}^{s} \beta_j \text{Tr} \left( k_j \| E - U H_{\phi}^T \|^2 \|_{F} \right) + \frac{\mu}{2} \text{Tr} \left( H_{\phi} L H_{\phi}^T \right)
\]

subject to \( \beta^T = 1 \), and \( \beta \geq 0 \). \( t_j = \text{Tr} \left( k_j \| E - U H_{\phi}^T \|^2 \|_{F} \right) \), and \( R = \text{Tr} \left( H_{\phi} L H_{\phi}^T \right) \). We can solve \( \beta \) of the multiple kernel matrix by using linear programming.

Define \( \xi = (\xi_y) \) and \( \tau = (\tau_y) \) to be the Lagrange multipliers with constraints \( U_{ij} \geq 0 \) and \( H_{\phi,ij} \geq 0 \).

Fixing \( \beta \), the objective function (11) becomes

\[
f(U, H_{\phi}) = \frac{1}{2} \text{Tr} \left( k \| E - U H_{\phi}^T \|^2 \|_{F} \right) + \frac{\mu}{2} \text{Tr} \left( H_{\phi} L H_{\phi}^T \right) + \text{Tr}(\xi U) + \text{Tr}(\tau H_{\phi})
\]

Then the unconstrained gradient \( \frac{\partial f}{\partial U} \) and \( \frac{\partial f}{\partial H_{\phi}} \) are given by

\[
\begin{align*}
\frac{\partial f}{\partial U} &= K U H_{\phi}^T H_{\phi}^T - K H_{\phi}^T + \xi = 0 \\
\frac{\partial f}{\partial H_{\phi}} &= \phi^T U^T K U - K U + \mu \phi \left( L + L^T \right) + \tau = 0
\end{align*}
\]

According to the Karush-Kuhn-Tucker (KKT) conditions, \( \zeta_y U_{ij} = 0 \), \( \tau_y (H_{\phi})_{ij} = 0 \), then

\[
\begin{align*}
\left( K U H_{\phi}^T H_{\phi}^T \right) U_{ij} - \left( K H_{\phi}^T \right) U_{ij} &= 0 \\
\left( H_{\phi} U^T K U \right) (H_{\phi})_{ij} - \left( K U \right) (H_{\phi})_{ij} + \left[ \mu \phi \left( L + L^T \right) \right] (H_{\phi})_{ij} &= 0
\end{align*}
\]

Using the gradient, we can construct the update rules as

\[
\begin{align*}
U_{ij} \leftarrow \frac{K H_{\phi}^T}{K U H_{\phi}^T} U_{ij} \\
(H_{\phi})_{ij} \leftarrow \frac{K U}{H_{\phi} U^T K U + \mu \phi \left( L + L^T \right)} (H_{\phi})_{ij}
\end{align*}
\]

where \( \epsilon \) is the changing the threshold. When \( \| \phi(V) - \phi(V) U H_{\phi} \| \leq \epsilon \), the update iteration is completed.

The algorithm of multiple kernel graph non-negative matrix factorization (MKGNMF) is summarized in Algorithm 1.

Table 1 Code of hyperspectral image unmixing method based on multiple kernel graph non-negative matrix factorization

| Algorithm 1: |
|--------------|
| Input: \( V \in \mathbb{R}^{m \times r}, k = [k_1, k_2, \ldots, k_s] \) |
| Output: \( W_{\phi} \in \mathbb{R}^{m \times r'}, H_{\phi} \in \mathbb{R}^{r' \times m} \) |
| 1: Initialize \( W \) by using equation (6) |
| 2: Initialize \( \beta = [\beta_1, \beta_2, \ldots, \beta_s] \) and \( \beta_i = \frac{1}{s} \) |
| 3: Initialize \( U \) and \( H_{\phi} \) |
| 4: for \( t=1 \) to \( t=100 \) go (a) to (c) steps |
| (a) Update \( K \) by using equation (12) |
4. Experiment Results and Discussions

4.1. Experiment settings
To of different hyperspectral image unmixing methods, ONMF, GNMF and the proposed method, experiments are developed using simulated dataset and real hyperspectral image.

The simulated dataset is synthesized by the method of reference [20]. From splib07a, the United States Geological Survey (USGS) seventh edition of the spectral library, a series of the spectrum are selected to synthesize simulated hyperspectral images. Firstly, extract the spectral features of all minerals from the AVIRIS 1995 spectral library in splib07a. Remove the severely interfered spectral features, and arrange them according to the spectral angle distance between any two spectra from large to small to form an experimental spectral library. Then, extract the spectra of any six materials from the experimental spectrum library, as shown in Fig.1. Finally, generate the simulated hyperspectral image based on the linear spectral mixture model. The size of the simulated hyperspectral image is $65 \times 65$, including 224 spectral features, which satisfies abundance non-negativity constraint (ANC) and abundance sum-to-one constraint (ASC). Fig.2 shows the abundance map corresponding to the six spectral endmembers. The simulated hyperspectral image includes 36 square regions, each of which has a size of $13 \times 13$. Among them, the 6 squares in the first row are pure pixel areas, which means each of them only includes one spectral end element. The second row to the sixth row are mixed pixel areas, which means each of them are mixed by two to six end elements with different numbers in equal proportions. The background area is mixed by six endmembers in random proportions. To simulate sensor noise and other errors, independent identically distributed Gaussian white noise is added, and the signal to noise ratio (SNR) is 40dB.
The real hyperspectral dataset, Urban, is collected by the HYDICE sensor in 1997. The size of the hyperspectral image after cropping is 36, and each pixel corresponds to an area of one square meter. There are 210 bands in total, ranging from 400 to 2500nm, with a spectral resolution of 10nm. After removing the bands (1~4, 76, 87, 101~111, 136~153, 198~210) affected by dense water vapor and atmospheric effects, 162 bands are retained. In this area, there are mainly six types of ground features: asphalt road, grass, tree, roof, metal, and dirt.

In the experiment, spectral angel distance (SAD) and root mean square error (RMSE) are used to evaluate the unmixing accuracy. SAD is used to measure the degree of similarity between the real endmembers and the endmembers extracted by unmixing method. The smaller the SAD value, the smaller the difference between the spectrum obtained by the method and the real ground object spectrum. RMSE is used to measure the similarity between the true abundance of ground features and the calculated abundance by unmixing method. The smaller the RMSE value, the closer the ground feature abundance obtained by the method is to the real abundance of ground features.

The SAD is defined as

\[
SAD(A, A') = \arccos \left( \frac{A^T A'}{\|A\| \|A'\|} \right)
\]

(17)

where \(A\) is the real endmember spectrum in the spectral library, and \(A'\) is the endmember spectrum extracted by unmixing method.

The RMSE is defined as

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (X_i - X'_i)^2}
\]

(18)

where \(X_i\) is the true abundance of features, \(X'_i\) is the estimated abundance by unmixing method.

4.2. Simulation data experiment results

Fig.3 shows the comparison between the abundance map obtained by unmixing methods and the real abundance map by the three algorithms of ONMF, GNMF, and MKGNMF. The first row corresponds to the real abundance map of the simulated hyperspectral map, and the second to fourth rows are the abundance maps obtained by ONMF, GNMF, and MKPNMF. It can be seen from the Fig.3 that compared with the ONMF algorithm, the GNMF algorithm can extract the end members of six ground features and perform abundance estimation, but the result of abundance estimation is more affected by the noise. The noise in the abundance map obtained by the MKGNMF algorithm is significantly less than that of the ONMF and GNMF algorithms, indicating that the introduction of MKL method can improve the robustness of the method, making the abundance map obtained by the method closer to the true abundance map, which is significantly improve the accuracy of unmixing.
In order to show the unmixing accuracy of the three unmixing methods, Table 2 evaluates the unmixing results from SAD, mean SAD, and RMSE, where the optimal value of each result is bold. It can be seen from Table 2 that the SAD value and the mean SAD value of each end member, and the RMSE value of MKGNMF are significantly smaller than the ONMF and GNMF algorithms, which means MKGNMF method has clearly outperformed other methods in endmember estimation and abundance estimation.

### Table 2 Results of 3 algorithms for simulated hyperspectral image unmixing

| Class | SAD | ONMF | GNMF | MKGNMF |
|-------|-----|------|------|---------|
| 1     | 1   | 0.4596 | 0.2451 | **0.2243** |
| 2     | 2   | 0.2272 | 0.2523 | **0.0427** |
| 3     | 3   | 0.4727 | 0.2130 | **0.0794** |
| 4     | 4   | 0.2690 | 0.2017 | **0.1533** |
| 5     | 5   | 0.3260 | 0.2293 | **0.0119** |
| 6     | 6   | 0.7191 | 0.2939 | **0.1540** |
| Average |     | **0.4123** | **0.2392** | **0.1109** |
| RMSE  |     | **0.3172** | **0.2546** | **0.1521** |

### 4.3. Real data experiment results

Fig.4 shows the comparison of the reference abundance map and the abundance map obtained by ONMF, GNMF and MKGNMF for the Urban image. The first row is the reference abundance map of real features, the second to fourth rows correspond to the abundance map obtained by ONMF, GNMF and MKGNMF. And the first to sixth columns correspond to the six types of ground features: asphalt road, grass, tree, roof, metal, and dirt. It can be seen from Fig.4 that MKGNMF can effectively separate the six types of ground features. The abundance map obtained by MKGNMF has a great similarity to the reference abundance map, which means the unmixing effect is better than that of ONMF and GNMF. MKGNMF can better identify the characteristics corresponding to the real ground objects, and the same ground objects have better continuity.
Fig. 4 The reference abundance map on Urban dataset and the abundance maps obtained by unmixing methods

Table 3 shows the results of experiments on the Urban dataset. It can be seen from the Table 3 that, compared with ONMF and GNMF, the SAD value and RMSE value of MKGNMF are both the best for the unmixing accuracy. This verifies the experiment results of simulated dataset and shows that MKGNMF also has good unmixing accuracy for real hyperspectral datasets.

| Class    | SAD (ONMF) | SAD (GNMF) | SAD (MKGNMF) |
|----------|------------|------------|--------------|
| Asphalt Road | 0.5631  | 0.4431     | **0.0790**   |
| Grass    | 0.5485  | 0.2610     | **0.1995**   |
| Tree     | 0.1476  | 0.2942     | **0.0743**   |
| Roof     | 0.2701  | 0.3675     | **0.1068**   |
| Metal    | 0.6250  | 0.4067     | **0.3638**   |
| Dirt     | 0.8402  | 0.4185     | **0.1118**   |
| Average  | **0.4991** | **0.3652** | **0.1559**   |
| **RMSE** | **0.3906** | **0.2838** | **0.1625**   |

5. Conclusion
Based on the NMF method, this paper has introduced the spectral structure information of the hyperspectral image by constructing the nearest neighbour graph, and has proposed a hyperspectral image unmixing method based on multiple kernel graph non-negative matrix factorization. This method is validated on both simulated dataset and real hyperspectral dataset, and achieved better performance when compared against the ONMF and GNMF method. In the future, we will explore other effective way to improve the time complexity of the algorithm to get faster unmixing efficiency.

Acknowledgments
This work was jointly supported by the National Natural Science Foundation of China (Grant No:61403397) and Natural science basic research plan in Shaanxi province of china (Grant No. 2020JM-358, 2015JM6313).

References
[1] Zhu F, Halimi A, Honeine P, et al. Correntropy Maximization Via Admm: Application to Robust Hyperspectral Unmixing[J]. Ieee Transactions on Geoscience and Remote Sensing, 2017, 55(9): 4944-4955.
[2] Li J, Bioucas-dias JM, Plaza A, et al. Robust Collaborative Nonnegative Matrix Factorization for Hyperspectral Unmixing[C]//Ieee Transactions on Geoscience and Remote Sensing. New York: Ieee, 2016: 6076-6090.

[3] Jing Yuan, Zhang Yujin, Gao Fangping. An Overview on Linear Hyperspectral Unmixing[J]. Journal of Infrared and Millimeter Waves, 2018, 37(5): 43-61.

[4] Bin Yang, Bin Wang. Review of Nonlinear Unmixing for Hyperspectral Remote Sensing Imagery[J]. Journal of Infrared and Millimeter Waves, 2017, 36(2): 173-185.

[5] Huang R, Li X, Zhao L. Incremental Kernel Non-negative Matrix Factorization for Hyperspectral Unmixing[C]//2016 Ieee International Geoscience and Remote Sensing Symposium (igarss), New York: Ieee, 2016: 6569-6572.

[6] Lee D, Seung H. Algorithms for Non-negative Matrix Factorization[J]. Adv. Neural Inform. Process. Syst., 2001, 13(6): 556-562.

[7] Yuquan Gan. Research and Application of Spectral Unmixing Method for Hyperspectral Remote Sensing Imagery[D]. University of Chinese Academy of Sciences in Partial Fulfillment of the Requirement, 2018.

[8] Li Sun, Gengxin Zhao. Decomposition of Mixed Pixels of Hyperspectral Remote Sensing Images Based on Orthogonal Non-negative Matrix Factorization[J], 2017, 48(2): 107-110.

[9] Renfeng Zhang. The Research and Application of Multiple Kernel Learning Method[D]. Jiangnan University, 2015.

[10] An S, Yun J, Choi S. Multiple Kernel Nonnegative Matrix Factorization[C]//2011 Ieee International Conference on Acoustics, Speech and Signal Processing (icassp), [S.l.]: [s.n.], 2011: 1976-1979.

[11] Fang B, Li Y, Zhang P, et al. Kernel Sparse Nmf for Hyperspectral Unmixing[J]. Ieee International Conference on Orange Technologies, Icot 2014, 2014: 41-44.

[12] Bo Wu, Liangpei Zhang, Pingxiang Li. Unmixing Hyperspectral Imagery Based on Support Vector Nonlinear Approximating Regression[J], 2006, 10(3): 26-32.

[13] Wang Qingchao, Fu Guangyuan, Wang Hongqiao, et al. Local Variable Sparsity Based Multiple Kernel Learning Algorithm[J]. Acta Electronica Sinica, 2018, 46(4): 930-937.

[14] Wang Qingchao, Fu Guangyuan, Wang Hongqiao, et al. Fusion of Multi-scale Feature Using Multiple Kernel Learning Forhyperspectral Image Land Cover Classification[J]. Optics and Precision Engineering, 2018, 26(4): 245-253.

[15] Yao J, Wang H, Wang L, et al. High Precision Classification of Hyperspectral Image Based on a Hierarchical Localized Multiple Kernel Learning Method[C]//2019 12th International Congress on Image and Signal Processing, Biomedical Engineering and Informatics (cisp-bmei), New York: Ieee, 2019: 1-5.

[16] Xiong Tan, Xuchu Yu, Pengqiang Zhang, et al. Nonlinear Mixed Pixel Decomposition of Hyperspectral Imagery Based on Multiple Kernel Svm[J], 2014, 22(7): 212-220.

[17] Yang Z, Oja E. Linear and Nonlinear Projective Nonnegative Matrix Factorization[J]. Ieee Transactions on Neural Networks, 2010, 21(5): 734-749.

[18] Cai D, He X, Han J, et al. Graph Regularized Nonnegative Matrix Factorization for Data Representation[J]. Ieee Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(8): 1548-1560.

[19] Eason DT, Andrews M. Total Variation Regularization Via Continuation to Recover Compressed Hyperspectral Images[J]. Ieee Transactions on Image Processing, 2014, 24(1): 284-293.

[20] Wu Jiankang. A Study of Sparse Representation for Hyperspectral Unmixing[D]. Xidian University, 2014.