Multiscale Low-Rank Spatial Features for Hyperspectral Image Classification

Weiwei Sun, Member, IEEE, Wenjing Shao, Jiangtao Peng, Member, IEEE, Gang Yang, Member, IEEE, Xiangchao Meng, Member, IEEE, and Qian Du, Fellow, IEEE

Abstract—This letter presents a multiscale low-rank decomposition (MSLRD) method to extract multiscale spatial structures from hyperspectral images. The MSLRD assumes that ground objects have divergent characteristics in changing spatial scales. It decomposes each band image into a series of block-wise matrices, where these low-rank blocks take detailed spatial structures at multiple scales. It formulates the low-rank matrix decomposition problem into minimizing the ranks of all block matrices and adopts the alternative direction of the multiplier method to optimize it. Experiments on Indian Pines and Pavia University data sets show that the MSLRD can greatly improve the classification performance of regular classification on spectral features (i.e., all bands) and perform better than five state-of-the-art spatial feature extraction methods.

Index Terms—Classification, hyperspectral imagery (HSI), multiscale low-rank decomposition (MSLRD), spatial features.

I. INTRODUCTION

HYPERSPECTRAL imaging technique simultaneously measures spectral signals and collect images of ground objects using hundreds of narrow bands [1]. With the classification operation, the obtained hyperspectral imagery (HSI) can be used to recognize diverse materials with subtle spectral differences, which benefits many practical mapping applications, for example, coastal wetlands, geological exploration, and precision agriculture [2].

In an image scene, each HSI pixel corresponds to a high-dimensional spectral vector, and classification is to categorize all the spectral pixels into one of the several classes of main ground objects. Spectral features (i.e., spectral vectors) are usually used by a pretrained classifier, for example, support vector machine (SVM), random forest (RF), and collaborative representation classification [3]. However, spectral variations across different spatial locations render that the same ground objects may have divergent spectral responses, and different ground objects may exhibit similar spectral curves. That brings about a great challenge for discriminating different classes from spectral features and easily leads to noisy and unsatisfactory classification maps with pixel-wise classifiers.

To alleviate the above problem, spatial features were adopted to promote the HSI classification [4]. Many spectral feature extraction methods were then proposed to describe local dependence or spatial correlations of nearby pixels in a spatial window. The local binary pattern (LBP) method quantifies the gray differences between centering pixels and their neighbors to estimate the local spatial texture features [5]. The 3-D-Gabor wavelet explores signal variances in joint spatial–spectral domains to extract spatial features from the HSI data [6]. The morphological profiles (MPs) implement a series of morphological operators of opening and closing by reconstruction to extract structure features, for example, differential MP (DMP), extended MP (EMP) [7], and extinction profile (EP) [8]. DMP was defined as the derivative of EMP to enhance the visualization of multiscale spatial feature analysis. However, DMP and EMP may not fully characterize information related to the gray-level characteristics of ground objects in the image, since they use fixed structure elements. In contrast, attribute profile (AP) makes a multilevel decomposition of the HSI data based on attribute files, for example, area of regions, measures of texture, shape, morphology, and contrast [9]. Later, extrema-oriented extinction filters were adopted into AP, and the obtained EP is proven to be less sensitive to image resolution and performs better than AP [10].

Different from preceding techniques, Markov random field and conditional random field describe the spatial dependence in the label level, where the pixel variation sensitive indices are used to adjust the prior distribution of image labels [11]. Besides, more recently, deep learning has been introduced to construct complicated networks and extract deep spatial features, for example, convolutional neural network and autoencoder networks [12]. These achievements successfully promote the classification accuracy of the HSI data, but most methods only extract single-scale spatial features with predefined parameters. That is, they may describe the spatial characteristics of ground objects with specific size, shape, and texture, whereas various features at other spatial scales cannot be detected completely. It may be argued that the MP-based methods can characterize multiscale spatial structures with a sequence of...
progressively coarser filters. However, their behaviors correlate closely with complicated parameter configurations (e.g., structure elements sizes, attribute thresholds), which severely hinder their usefulness.

In this letter, we present a multiscale low-rank decomposition (MSLRD) method to extract multiscale spatial features, aiming for improving the classification results of the HSI data. We assume that ground objects in the image scene exhibit divergent spatial structures or spatial correlations at different scales and modify the low-rank matrix factorization to obtain the MSLRD model. The MSLRD decomposes each band image into a series of low-rank block-wise matrices, where the low-rank blocks characterize the spatial structures of different ground objects at increasing scales. The MSLRD is then formulated as the rank minimization problem of all block matrices, and we use the alternating direction method of multipliers (ADMMs) [13] to solve this problem.

II. LOW-RANK DECOMPOSITION MODEL

Consider the collected hyperspectral data as a three-dimensional matrix denoted by $Y^{spe} = [Y_i]_{i=1}^D \in \mathbb{R}^{M \times N \times D}$, where $Y_i$ is the $i$th band and the length and width of a single band image respectively, and $D$ is the number of bands. In the realistic world, each band image $Y$ is always corrupted by noise or outliers caused by precision limits of the imaging spectrometer and errors in analog-to-digital conversion. Accordingly, the image $Y$ can be mathematically modeled as [14]

$$Y = X + S$$

where $X$ is the low-rank matrix, and $S$ is the sparse error matrix which takes outliers or noise. $X$ is assumed to be low-rank, since it takes almost all the information of $Y$ and is continuous and smooth. $S$ is assumed to be sparse, since the noise or outliers exist in the image with low probability. Finding the solution of $X$ and $S$ can be formulated as a famous mathematical problem called robust principal component (PC) analysis, which can be expressed as

$$\min_{X,S} \|Y - X\|_1 + \lambda \|S\|_1, \quad \text{s.t., } Y = X + S$$

(2)

where $\|\cdot\|_1$ denotes the nuclear norm operator, $\|\cdot\|_1$ denotes the sum of absolute value of the matrix entries, and $\lambda$ is the regularization term to balance the two terms. The first term $\|X\|_1$ guarantees the low-rank property of the matrix $X$, and the second term $\|S\|_1$ ensures the sparsity of $S$. The inexact augmented Lagrange multipliers (IALMs) algorithm can be used to solve the optimization problem in (2).

III. MSLRD MODEL

A. MSLRD Model

In the image scene, the ground objects often exhibit divergent spatial structures or spatial correlations at different scales. Multiscale analysis allows us to obtain a more accurate and compact understanding of the spatial structure for each ground object and helps us identify them from others. The low-rank decomposition model in (1) can only estimate the general information of each band image with the low-rank data matrix $X$. Therefore, we revise it into an MSLRD model to extract more detailed spatial structure.

Fig. 1 illustrates the idea of the MSLRD model. Consider a multiscale partition $\{P_l\}_{l=0}^L$ for the $M \times N$ data matrix of each band image, where each block $B_l$ in $P_l$ is low-rank and has an order-magnitude larger than those in its previous scale $P_{l-1}$. Meanwhile, consider a block reshape operator $R_b(Y)$ to extract each block $B_l$ from the full data matrix $X_l$ and reshape it into an $m_l \times n_l$ matrix, where $m_l \times n_l$ is the $l$th block matrix size. In this letter, for simplicity, we set the block size $m_l = n_l$ and set them as the powers of 2, for example, $m_l = 2^l$, and the largest block size is $M \times N$. Accordingly, each $Y$ can be decomposed into a series of data matrices $\{X_l\}_{l=0}^L \in \mathbb{R}^{M \times N}$, in which each $X_l$ is the block-wise low-rank with respect to its partition $P_l$, with $L = \lceil \log_2 \min(M, N) \rceil$ + 1. The mathematical expression of the MSLRD model is [15]

$$Y = \sum_{l=0}^L X_l, \quad \text{s.t., } X_l = \sum_{B_l \in P_l} R_b(B_l), B_l = U_b S_b V_b^T$$

(3)

where $U_b \in \mathbb{R}^{m_l \times n_l}, S_b \in \mathbb{R}^{n_l \times m_l}$, and $V_b \in \mathbb{R}^{n_l \times m_l}$ come from the singular value decomposition (SVD) of $B_l \in \mathbb{R}^{m_l \times m_l}$, which guarantees the low-rank property as $r_{bl}$. $R_b(\cdot)$ is a block embedding operator that embeds the block $B_l$ into the full-size zero matrix and helps formulate the $l$th partitioned matrix $X_l$. It is expected that each partition matrix $X_l$ will illustrate detailed spatial structures of ground objects at the $l$th scale. Meanwhile, the largest block matrix $X_L$ has the low-rank feature that can be regarded as taking the background information of $Y$. The model (3) can be optimized by minimizing the ranks of all block matrices

$$\min_{X_0, X_1, \ldots, X_L} \sum_{l=0}^L \sum_{B_l \in P_l} \text{rank}(R_b(X_l)), \quad \text{s.t., } Y = \sum_{l=0}^L X_l$$

(4)

The nuclear norm has good approximation to the rank optimization. For each scale $l$, we implement the block-wise
nuclear norm [15] to relax the block-wise ranks and formulate (4) into a convex program (5) accordingly as
\[
\min_{X_0, X_1, \ldots, X_{t-1}} \sum_{l=0}^{L} \lambda_l \|X_l\|_{(s)} , \quad \text{s.t., } Z_l = \sum_{l=0}^{L} X_l \quad (5)
\]
where \(\|X_l\|_{(s)} = \sum_{B \in P} \|R_B(X_l)\|_s\) is the block-wise nuclear norm for the \(l\)th scale, \(\|\cdot\|_s\) is the nuclear norm operator, and \(\lambda_l\) is the regularization parameter.

Previous studies suggest the regularization parameter \(\lambda_l\) is the Gaussian complexity of each norm \(\|\cdot\|_{(s)}\) [15]. Meanwhile, the resulting expression for the Gaussian complexity has been proven to be the maximum singular value of a random Gaussian matrix. Therefore, we follow the studies by Bandeira and Handel [16] and set the regularization parameter \(\lambda_l \sim \sqrt{m_l + \sqrt{m_l} + (\log(MN/\max(m_l, n_j)))} / 2\). For the first partitioned matrix with 1 \(\times\) 1 block size, \(\lambda_0 = (\log(MN))^{1/2}\), and for the last partitioned matrix with \(M \times N\) block size, \(\lambda_L \sim \sqrt{M + \sqrt{N}}\).

**B. Optimization Solution**

The program (5) is convex, which can be solved by the ADMM [17]. Equation (6) can be written as
\[
\min_{X_l, Z_l} \begin{cases} 
I \left( Y = \sum_{l=0}^{L} X_l \right) + \sum_{l=0}^{L} \lambda_l \|Z_l\|_{(s)} , \quad \text{s.t., } Z_l = X_l 
\end{cases} \quad (6)
\]
where \(I(\cdot)\) is the indicator function, and \(Z_l\) is an auxiliary variable to assist in joint minimization. By introducing a Lagrangian multiplier \(\Delta\), (6) can be equivalently transformed as
\[
\min_{X_l, Z_l, \Delta} \begin{cases} 
I \left( Y = \sum_{l=0}^{L} X_l \right) + \sum_{l=0}^{L} \lambda_l \|Z_l\|_{(s)} + \rho/2 \|Z_l - X_l\|^2_F \\
+ \text{tr}(\Delta^T (Z_l - X_l)) 
\end{cases} \quad (7)
\]
where \(\rho\) is the penalty coefficient, and \(\text{tr}(\cdot)\) is the trace operator. The updating rules for \(X_l\) are devised using the proximal operator for the indicator function, which is simply the projection operator to the set. At the \((l+1)\)th iteration, with other variables fixed, the primal variable \(X_l^{(l+1)}\) can be updated by minimizing
\[
\min_{X_l^{(l+1)}} \begin{cases} 
I \left( Y = \sum_{l=0}^{L} X_l^{(l+1)} \right) + \rho/2 \|Z_l^{(l)} - X_l^{(l+1)}\|^2_F \\
+ \text{tr}(\Delta_l^{(l)T} (Z_l^{(l)} - X_l^{(l+1)})) 
\end{cases} \quad (8)
\]
and \(X_l^{(l+1)}\) has a closed-form solution \(X_l^{(l+1)} = (Z_l^{(l)} - \Delta_l^{(l)}) / (1/L + 1)(Y - \sum_{l=0}^{L} Z_l^{(l)} - \Delta_l^{(l)})\) [15]. Moreover, singular value soft-thresholding operator (SVT) [18] is adopted to derive the updating rules for \(Z_l\). Consider the block-wise nuclear norm \(\|X_l\|_{(s)}\) for each scale is separate with respect to each block, its proximal function with \(\lambda_l\) can be given by the block-wise SVT operator [15], that is, \(\text{BLOCKSVT}_{\lambda_l} (\cdot) = \sum_{B \in P} \|R_B\|^{1/2} \text{SVT}_{\lambda_l} (R_B(\cdot))\). Accordingly, we implement the SVT for each block and then return the result to the original block. Accordingly, the variable \(Z_l\) can be updated via the SVT operator with parameter \(\lambda_l\) as
\[
Z_l^{(l+1)} = \text{BLOCKSVT}_{\lambda_l} \left( X_l^{(l+1)} + \Delta_l^{(l)} \right) . \quad (9)
\]

The Lagrangian multiplier \(\Delta_l^{(l+1)}\) can be updated as
\[
\Delta_l^{(l+1)} = \Delta_l^{(l)} + (X_l^{(l+1)} - Z_l^{(l+1)}). \quad (10)
\]

The above iterations are repeated until satisfying the convergence conditions \(\|X_l^{(l+1)} - X_l^{(l)}\|_{(s)} < \tau\) and \(\|Z_l^{(l+1)} - Z_l^{(l)}\|_{\infty} < \tau\) or the maximum iteration time \(t_{\text{max}} > 500\). The proper multiscale low-rank matrix \(X_l\) can be obtained after convergence. The multiscale matrices \(X_l^{\text{spa}} = \{X_l\}_{l=0}^{L-1}\) reflect detailed spatial features of the \(l\)th band image \(Y_l\), and the largest low-rank matrix \(X_L\) is discarded because of the background term. We then stack the multiscale matrices estimated from all \(D\) bands and generate three-order tensor-like spatial features \(Y^{\text{spa}} = \{X_l^{\text{spa}}\}_{l=1}^{D}\) at different scales.

**IV. EXPERIMENTAL RESULTS AND ANALYSIS**

**A. Experimental Data Sets**

The first data are the Indian Pines collected by the AVIRIS sensor from JPL. It was downloaded from the Multispectral Image Data Analysis System Group at Purdue University, West Lafayette, IN, USA. It has 200 spectral resolutions, 10-\(\text{nm}\) spectral resolutions ranging from 400 to 2500 nm, and 145 \(\times\) 145 pixels. It has 200 spectral bands after bad band removal and 16 classes of ground objects to be classified.

The second data are the Pavia University (PaviaU) by the ROSIS sensor over Pavia, North Italy. It was downloaded from the Group of Intelligent Computational, Basque University. It has 1.3-\(\text{m}\) spatial resolutions and 103 bands, covering nine classes of main ground objects. The implemented data have 610 \(\times\) 340 pixels and 102 bands, after removing the noisy bands.

**B. Parameter Settings**

The AMDD of MSLRD is initialized with \(Z_0^{(0)} = Y_0^{(0)} = \Delta_0^{(0)} = 0\), \(\rho = 0.01\), \(\tau = 10^{-6}\). The SVM is used as the classifier for the spatial feature \(Y^{\text{spa}}\). It implements the radial basis functions as the kernel function, where the variance parameter and the penalization factor are obtained via cross-validation. We quantify the classification accuracy with overall classification accuracy (OCA), average classification accuracy (ACA), and kappa coefficient (KC). In the experiments, the training samples on each class of Indian Pines and PaviaU are randomly selected.

**C. Classification Results**

We compare the classification accuracies of MSLRD with those of 3-D-Gabor [6], rILBP [5], EMP [7], AP [9], and EP [10], where the later three are the state-of-the-art multiscale MP methods. In 3-D-Gabor, the orientations are manually set to \(\{0, \pi/4, \pi/2, -\pi/2, -\pi/4\}\), and the frequencies are set to be \(\{1/2, 1/4, 1/6, 1/8\}\); the patch sizes of Indian Pines and PaviaU are \(11 \times 11\) and \(7 \times 7\) respectively. In rILBP, using cross-validation, the radius, the neighborhood size, and the number of PCs are chosen as 2, 8, and 4, respectively; the local patch sizes in the histogram statistics on Indian Pines and PaviaU are 17 and 21 respectively. In EMP, with the first four PCs, the type of structure element is set to be “disk,” and the sizes on Indian Pines and PaviaU are manually set as \(\{4, 7, 10, 13, 16, 19, 22, 25\}\) and \(\{1, 3, 5, 7, 9, 11, 13, 15\}\), respectively.
Fig. 2. Classification maps of different methods on the Indian Pines. (a) 3-D-Gabor (OCA = 76.01%). (b) riLBP (OCA = 98.00%). (c) EMP (OCA = 97.82%). (d) AP (OCA = 94.42%). (e) EP (OCA = 94.52%). (f) MSLRD (OCA = 98.77%).

Fig. 3. Classification maps of different methods on the PaviaU. (a) 3-D-Gabor (OCA = 86.45%). (b) RiLBP (OCA = 92.26%). (c) EMP (OCA = 97.05%). (d) AP (OCA = 96.04%). (e) EP (OCA = 97.56%). (f) MSLRD (OCA = 99.39%).

TABLE I

<table>
<thead>
<tr>
<th>Class</th>
<th>Name</th>
<th>Train</th>
<th>Test</th>
<th>All bands</th>
<th>3D-Gabor</th>
<th>riLBP</th>
<th>EMP</th>
<th>AP</th>
<th>EP</th>
<th>MSLRD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Alfalfa</td>
<td>5</td>
<td>41</td>
<td>36.59</td>
<td>26.83</td>
<td>97.56</td>
<td>90.24</td>
<td>95.12</td>
<td>97.56</td>
<td>96.88</td>
</tr>
<tr>
<td>2</td>
<td>Corn-notill</td>
<td>143</td>
<td>1285</td>
<td>75.41</td>
<td>67.00</td>
<td>97.20</td>
<td>97.82</td>
<td>88.64</td>
<td>95.68</td>
<td>97.50</td>
</tr>
<tr>
<td>3</td>
<td>Corn-mintill</td>
<td>83</td>
<td>747</td>
<td>66.80</td>
<td>64.66</td>
<td>97.86</td>
<td>98.13</td>
<td>94.11</td>
<td>95.72</td>
<td>96.90</td>
</tr>
<tr>
<td>4</td>
<td>Corn</td>
<td>24</td>
<td>213</td>
<td>59.14</td>
<td>42.25</td>
<td>96.24</td>
<td>97.65</td>
<td>87.32</td>
<td>95.88</td>
<td>99.40</td>
</tr>
<tr>
<td>5</td>
<td>Grass-pasture</td>
<td>48</td>
<td>435</td>
<td>82.53</td>
<td>74.48</td>
<td>98.39</td>
<td>95.40</td>
<td>93.79</td>
<td>91.95</td>
<td>98.82</td>
</tr>
<tr>
<td>6</td>
<td>Grass-trees</td>
<td>73</td>
<td>657</td>
<td>96.04</td>
<td>93.76</td>
<td>98.02</td>
<td>98.17</td>
<td>99.70</td>
<td>97.56</td>
<td>100</td>
</tr>
<tr>
<td>7</td>
<td>Grass-pasture-mowed</td>
<td>25</td>
<td>56.00</td>
<td>20.00</td>
<td>96.00</td>
<td>96.00</td>
<td>92.00</td>
<td>92.00</td>
<td>95.00</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Hay-windrowed</td>
<td>48</td>
<td>430</td>
<td>98.61</td>
<td>95.58</td>
<td>99.77</td>
<td>100</td>
<td>99.77</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>9</td>
<td>Oats</td>
<td>2</td>
<td>18</td>
<td>38.89</td>
<td>22.22</td>
<td>88.89</td>
<td>77.78</td>
<td>38.89</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>10</td>
<td>Soybean-notill</td>
<td>97</td>
<td>875</td>
<td>66.40</td>
<td>67.77</td>
<td>95.31</td>
<td>94.17</td>
<td>88.57</td>
<td>90.06</td>
<td>97.79</td>
</tr>
<tr>
<td>11</td>
<td>Soybean-mintill</td>
<td>246</td>
<td>2209</td>
<td>80.76</td>
<td>82.57</td>
<td>99.14</td>
<td>99.86</td>
<td>97.06</td>
<td>98.78</td>
<td>99.30</td>
</tr>
<tr>
<td>12</td>
<td>Soybean-clean</td>
<td>59</td>
<td>534</td>
<td>69.85</td>
<td>58.24</td>
<td>97.00</td>
<td>91.95</td>
<td>88.01</td>
<td>88.95</td>
<td>98.55</td>
</tr>
<tr>
<td>13</td>
<td>Wheat</td>
<td>21</td>
<td>184</td>
<td>98.91</td>
<td>95.11</td>
<td>97.83</td>
<td>99.46</td>
<td>98.91</td>
<td>99.46</td>
<td>100</td>
</tr>
<tr>
<td>14</td>
<td>Woods</td>
<td>127</td>
<td>1138</td>
<td>94.38</td>
<td>93.59</td>
<td>99.82</td>
<td>100</td>
<td>99.21</td>
<td>99.74</td>
<td>99.55</td>
</tr>
<tr>
<td>15</td>
<td>Buildings-Grass-Trees-Drives</td>
<td>39</td>
<td>347</td>
<td>52.74</td>
<td>54.18</td>
<td>97.69</td>
<td>97.98</td>
<td>96.25</td>
<td>94.52</td>
<td>99.63</td>
</tr>
<tr>
<td>16</td>
<td>Stone-Steel-Towers</td>
<td>9</td>
<td>84</td>
<td>89.29</td>
<td>58.33</td>
<td>89.29</td>
<td>90.48</td>
<td>100</td>
<td>94.05</td>
<td>100</td>
</tr>
</tbody>
</table>

Using the first four PCs, the attribute file and threshold of AP on Indian Pines are “area” and [49, 169, 361, 625, 961, 1369, 1849, 2401] and those of PaviaU are “inertia” and [0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8] respectively. Using cross-validation, the attribute file, the threshold, and the number of PCs in EP on both data sets are “area,” 2, and 4 respectively.

Figs. 2 and 3 illustrate the classification maps of the six methods on both data sets, and Tables I and II list detailed
classification accuracies of each class of using all bands and spatial features obtained from these methods. 3-D-Gabor performs the worst, which may be because it only extracts coarse texture features with specific frequencies and directions and may not describe complicated features. In contrast, the other five methods clearly outperform that using spectral features of all bands, boosting over 17.3% and 4.6% of OCAs on Indian Pines and PaviaU, respectively. EMP achieves comparable accuracies to EP, and they both behave better than AP, especially in Indian Pines. The advantage of EP over AP results from that the extinction filters in EP can preserve more regions, especially in Indian Pines. The advantage of EP over AP results from that the extinction filters in EP can preserve more regions, especially in Indian Pines. The MSLRD performs the best among all the methods and produces at least 11.7% higher accuracies than MP-class features (i.e., EMP, AP, and EP) on Indian Pines but worse on PaviaU. The MSLRD performs the best among all the methods and produces at least 11.7% higher OCAs than all the bands on both data sets. Particularly, it can greatly improve the classification accuracy of classes with smaller training sample sizes, for example, Alfalfa, corn, oats, and Woods in Indian Pines. The explanation is that the block-wise matrices can characterize spatial structures of ground objects with different spatial scales, which helps the MSLRD better distinguish them from each other. Besides, the MSLRD performs better than all its single-scale low-rank features, with at least 2.9% and 1.1% higher OCAs on Indian Pines and PaviaU data sets, respectively.

V. CONCLUSION

This letter proposes an MSLRD method to promote the classification accuracy of HSI classification. The MSLRD decomposes each band image into a series of low-rank block-wise matrices, in which each block characterizes the spatial structures of ground objects at different spatial scales. It formulates the model into a convex program and optimizes it via ADMM. The experimental results on two HSI data sets show that the MSLRD is superior to five state-of-the-art methods without the need of fine parameter configuration. Moreover, the proposed MSLRD performs better than all single-scale low-rank features.

REFERENCES