IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 4, 2018, accepted November 19, 2018, date of publication December 3, 2018,
date of current version December 31, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2884501

A Novel Regularized Nonnegative Matrix Factorization for Spectral-Spatial Dimension Reduction of Hyperspectral Imagery

WEI YAN[1], BOB ZHANG[1], (Member, IEEE), AND ZUYUAN YANG[2], (Member, IEEE)
[1]Department of Computer and Information Science, University of Macau, Macau, China
[2]Guangdong Key Laboratory of IoT Information Technology, School of Automation, Guangdong University of Technology, Guangzhou 510006, China

Corresponding author: Bob Zhang (bobzhang@umac.mo)

This work was supported by in part by the National Natural Science Foundation of China under Grant 61602540 and Grant 61722304, in part by the Pearl River S&T Nova Program of Guangzhou under Grant 201610010196, and in part by the Guangdong Natural Science Funds for Distinguished Young Scholar under Grant 2014A030306037.

ABSTRACT Dimension reduction (DR) is an essential preprocessing for hyperspectral image (HSI) classification. Recently, nonnegative matrix factorization (NMF) has been shown as an effective tool for the DR of hyperspectral data given the fact that it provides interpretable results. However, the basic NMF ignores the geometric structure information of the HSI data, thus limiting its performance. To this end, a novel regularized NMF method, termed NMF with adaptive graph regularizer (NMFAGR), is proposed for the spectral-spatial dimension reduction of hyperspectral data in this paper. Specifically, to enhance the preservation ability of the geometric structure information, the NMFAGR performs the dimension reduction and graph learning simultaneously. Regarding the mutual correlation between these two tasks, a graph regularizer is added as an interaction. Moreover, to effectively utilize complementary information among spectral-spatial features, the NMFAGR allocates feature weight factors automatically without requiring any additional parameters. An efficient algorithm is utilized to solve the optimization problem. The effectiveness of the proposed method is demonstrated on three benchmark hyperspectral data sets through experimentation.

INDEX TERMS Hyperspectral images, feature extraction, pattern recognition.

I. INTRODUCTION

Hyperspectral images (HSIs) are widely used in earth monitoring, agriculture research, and mineral detection [1]–[3]. These applications often need the classification of each pixel in the image. Due to the fact that there are hundreds of spectral bands for sampling, HSI data is high-dimensional and contains rich spectral and spatial information for accurate classification of various materials in the observed scene. The performance of classifiers deteriorates as the dimension increases because of a lack of training samples (Hughes phenomenon [4]). Meanwhile, high-dimensional data always result in high computation costs and huge storage capacity [5]. Thus, to obtain a more accurate classification performance, a dimension reduction (DR) procedure is often taken as a preprocessing step.

As a preprocessing technique for classification, DR aims to find a low-dimensional representation for high-dimensional data without losing desirable information relative to the original data. DR contributes to alleviate the small-sample-size problem and to improve the performance of the classifier [6]–[11]. The widely used unsupervised DR methods include: principal component analysis (PCA) [12], independent component analysis (ICA) [13], and non-negative matrix factorization (NMF). PCA seeks an orthogonal projection to maximize the global data variance. ICA assumes that the sources are statistically independent and recovers them from the observed signals. Both PCA and ICA cannot ensure that the data in the low-dimensional representation space are all non-negative [14]. In contrast, nonnegative matrix factorization (NMF) [15]–[19] can preserve nonnegativity by imposing the nonnegative constraint on the low-dimensional representation. The negative values lack any physical meaning since HSI data are nonnegative [14], [20], [21]. More importantly, NMF yields additive combinations. The additivity has a close relationship to HSI. This is because each pixel measurement is often modeled as a positive mixing of reflectance values of the materials in the scene measured by the pixel. From a geometric view, NMF aims to determine a convex cone that ''well describes'' the data. Based on this view, many geometric-based NMF

method for the analysis of HSI have been proposed in the last decade [17], [22], [23].

Recently, various researchers have shown that high-dimensional data in the Euclidean space cannot be uniformly filled up by the HSIs. These image data can be regarded as sampled data from or near a submanifold of an ambient space [24]–[27]. To increase the learning performance by considering the underlying manifold structure of HSI data, many graph-based DR methods have been proposed for HSI processing [28]–[30]. These methods mainly employ two consecutive independent steps: the first step encodes the manifold structure information by constructing a neighborhood weights (similarities) graph matrix from the given HSI data, and the second step performs DR on the HSI data based on the generated graph matrix. Although the performance of these graph-based DR approaches is prominent in HSI processing, its efficiencies can be improved since the performance of DR methods is mostly governed by how effective the graph is constructed. At the same time, these methods utilize fixed graphs without a learning mechanism which is proved to be powerful and widely utilized in machine learning areas [31].

The above DR methods are spectral-based methods. They use the spectral information to represent each pixel. In fact, for HSIs, the observed pixels in the images are spatially related. In terms of the classification of HSIs, nearby pixels in HSIs are measured from spatially closed area, very likely with the same labels [32]. The low-dimensional representation based on only the spectral information may be insufficient and can lead to under-classification [1].

Many methods that adopt spectral-spatial-based DR have been developed in recent years [32]–[35]. Among such methods, Zhou *et al.* proposed a spatial and spectral regularized local discriminant embedding (SSRLDE) method for the dimension reduction process of hyperspectral data [5]. Zhang *et al.* proposed the multiple-features-combining (MFC) approach [33], and the multiple-features t-distributed stochastic neighbor embedding (MSNE) [34]. Both incorporated spectral feature and spatial features (such as spectral, shape, texture and morphological features) into the dimension reduction process. Wen *et al.* devised an orthogonal NMF-based method which can not only achieve a non-negative factorization but also exploit the complementary information that arises among heterogeneous features [35]. However, most of the above mentioned spectral-spatial based methods requires learning an additional parameter to thoroughly explore the complementary properties of multiple HSI features.

In this paper, we propose a novel regularized NMF model called NMF with Adaptive Graph Regularizer (NMFAGR) for spectral-spatial dimension reduction of HSI data. It is different from most of the existing graph-based DR methods for HSI data in two significant aspects. The first one is that NMFAGR performs dimension reduction and learns a shared graph weights matrix for spectral-spatial features simultaneously. The second is NMFAGR learns optimal weights for spectral-spatial features automatically without requiring an additional parameter. Specifically, NMFAGR extends the standard NMF by integrating an adaptive graph regularizer into the basic NMF model. The proposed method improves the neighborhood preserving property of the low-dimensional representation matrix. Therefore, the classifiers obtain better performance when taking such optimal representation as input.

The following parts of this paper will be organized as follows. In Section 2, a brief review of the basic NMF and the graph regularizer is made. Section 3 introduces Auto-weighted Graph Learning for Multi-feature Learning and our proposed NMFAGR algorithm, followed by its optimization algorithm. Section 4 provides the results of the experimentation on three hyperspectral data sets that are publicly available. Finally, we provide some concluding remarks in Section 5.

II. NMF AND GRAPH REGULARIZER

Dimension reduction is made by finding a low-dimensional representation $\mathbf{V}$. Specifically, given a data matrix $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_n] \in \mathcal{R}_+^{m\times n}$, in which m is the feature dimension and n denotes the number of samples. NMF seeks a nonnegative basis $\mathbf{U} \in \mathcal{R}_+^{m\times d}$ and an encoding matrix $\mathbf{V} \in \mathcal{R}_+^{n\times d}$, whose product can approximate well the original matrix $\mathbf{X}$. As $d \ll m$ and $d \ll n$, NMF yields low-dimensional representation of the original data. $\mathbf{V}$ can be regarded as the new representation of the input data. The optimal value of $\mathbf{U}$ and $\mathbf{V}$ can be found by solving the following optimization problem:

$$\mathcal{O}_F = ||\mathbf{X} - \mathbf{U}\mathbf{V}^{\mathrm{T}}||_F^2, \quad \text{s.t. } \mathbf{U} \geq 0,\ \mathbf{V} \geq 0. \tag{1}$$

where $||\cdot||_F$ is Frobenius norm.

Due to the existing manifold structure of HSI data, manifold-based dimension reduction methods have been successfully used for remote sensing [24]. The goal of these methods is to find the low-dimensional representation with respect to the manifold structure of HSI data. To this end, a graph regularizer is added into the objective function of dimension reduction. This regularizer is described in [36] and [37] as follows:

$$\begin{aligned}\mathcal{O} &= \sum_{i,j=1} \left\| \mathbf{V}_{i.}^{\mathrm{T}} - \mathbf{V}_{j.}^{\mathrm{T}} \right\|_F^2 \mathbf{W}_{ij} \\ &= \mathrm{Tr}(\mathbf{V}^{\mathrm{T}}\mathbf{D}\mathbf{V}) - \mathrm{Tr}(\mathbf{V}^{\mathrm{T}}\mathbf{W}\mathbf{V}) \\ &= \mathrm{Tr}(\mathbf{V}^{\mathrm{T}}\mathbf{L}\mathbf{V}).\end{aligned} \tag{2}$$

where $\mathrm{Tr}(\cdot)$ denotes the trace of a matrix and $\mathbf{V}_{j.}^{\mathrm{T}} = (v_{j1}, v_{j2}, \ldots, v_{jr})^{\mathrm{T}}$. $\mathbf{W}$ is the graph weights matrix and $\mathbf{D}$ is a diagonal matrix whose entries are column sums of $\mathbf{W}$, $\mathbf{D}_{jj} = \sum_j \mathbf{W}_{ij}$. $\mathbf{L} = \mathbf{D} - \mathbf{W}$, which is a graph Laplacian matrix. Please see [24] for the details of constructing the graph weights matrix.

By minimizing $\mathcal{O}$, we expect that if two pixels x_i and x_j are close (i.e., $\mathbf{W}_{ij}$ is big), then $\mathbf{V}_{i.}$ and $\mathbf{V}_{j.}$, the low-dimensional embedded vectors of these two data samples are also close to each other.

III. PROPOSED APPROACH

Consider that we obtain S types of spectral-spatial features (e.g., texture, spectral, and morphological features, S = 3), the sth feature matrix is represented as $\mathbf{X}(s) \in \mathcal{R}^{m_s \times n}$, $(s = 1, 2, \ldots, \mathrm{S})$, in which s denotes a specific feature within S features in total, m_s is the dimension of the sth feature, and n represents the number of samples. After stacking all the feature vectors, we have the multiple-feature data set as $\mathbf{X}^{m \times n}$, in which $\sum_{s=1}^{\mathrm{S}} m_s = m$.

Most of existing graph-based DR methods for HSI data find low-dimensional representations based on a fixed graph matrix. The corresponding drawbacks are: (1) the quality of such representations is sensitive to that of the graph construction as it lacks a learning mechanism; (2) there is no interaction between the low-dimensional representation learning process and that of the graph construction. (3) Regarding the complementary information among spectral-spatial features, an extra weight parameter is required to learn which may be impractical.

To address the above-mentioned limitations, we proposed a novel regularized NMF model for dimension reduction of HSI data. The proposed method finds an optimal low-dimensional feature representation of the high-dimensional multiple-feature data set $\mathbf{X}$ with respect to the geometrical structure of HSI data. Specifically, the process of graph learning is combined with that of the dimension reduction for better capturing the geometrical information of HSI data. Moreover, to differentiate the importance among spectral-spatial features, a novel strategy is adopted to learn feature weights during the process of graph learning. It is nearly parameter-free so that be more practical. Following this, we will first introduce the auto-weighted graph learning method from [38]. Afterwards, the optimization strategy of the proposed algorithm will be introduced in detail.

A. AUTO-WEIGHTED GRAPH LEARNING FOR MULTI-FEATURE LEARNING (AGL)

The aim of AGL [38] is to learn a common weights matrix shared by different types of features. In addition, this model can automatically assign an ideal weight for each feature without additional weight and penalty parameters. The optimal graph weights matrix can be learned by solving the following optimization problem:

$$\begin{aligned} &\min_{\mathbf{W}} \sum_{s} \theta_s \sum_{i,j} \|x_i^s - x_j^s\|_2^2 \, w_{ij} + \gamma \|\mathbf{W}\|_F^2, \\ &\text{s.t. } w_i^{\mathrm{T}} 1 = 1, \quad 0 \le w_{ij} \le 1. \end{aligned} \tag{3}$$

where

$$\theta_s = 1 \Big/ 2\sqrt{\sum_{i,j} \|x_i^s - x_j^s\|_2^2 w_{ij}}, \tag{4}$$

x_i^s is the sth feature of the sample i, and γ is used to keep the weights distribution smooth. $\mathbf{W}$ denotes the learned weight matrix shared by each feature.

As θ_s is dependent on the target variable $\mathbf{W}$, that is to say the value of θ_s can be updated correspondingly. If feature s is optimal, then $\sum_{i,j} \|x_i^s - x_j^s\|_2^2 \, w_{ij}$ should be small, and thus the learned weight for feature s is large according to the definition of θ_s and vice versa. This method optimizes the weights meaningfully and can obtain better results than the classical combination approach which assigns equal weights to all the features [33]–[35].

B. NMFAGR

To better preserve the geometric structure information, we performs the dimension reduction and the graph learning simultaneously. Specifically, we incorporate the AGL as a regularizer into the basic NMF model. Considering the mutual correlation between the low-dimensional representations and the learned graph, a graph regularizer is used as an interaction. The obtained objective function can be considered as:

$$\begin{aligned} \min f(\mathbf{U}, \mathbf{V}, \mathbf{W}) &= ||\mathbf{X} - \mathbf{U}\mathbf{V}^{\mathrm{T}}||^2 + \alpha \mathrm{Tr}(\mathbf{V}^{\mathrm{T}}\mathbf{L}\mathbf{V}) \\ &+ \sum_{s} \theta_s \sum_{i,j} ||x_i^s - x_j^s||_2^2 w_{ij} + \gamma ||\mathbf{W}||_F^2, \\ \text{s.t. } \mathbf{U} \ge 0, \quad \mathbf{V} \ge 0, \mathbf{V}^{\mathrm{T}}\mathbf{V} &= \mathbf{I}, w_i^{\mathrm{T}} = 1, 0 \le w_{ij} \le 1. \end{aligned} \tag{5}$$

where the first term is used to learn the low-dimensional representation. The second term can be regarded as an interaction between the low-dimensional representation and the learned graph. The third and fourth terms are used to learn the graph matrix.

C. OPTIMIZATION FOR NMFAGR

Recently the Optimal Gradient Method (OGM) method has been applied to solve the sub-problems of NMF and its variants [39]–[41]. Here we extend this idea to NMFAGR. For notational simplicity, we have $f(\mathbf{U}, \mathbf{V}, \mathbf{W})$ as $f_1(\mathbf{U})$ when $\mathbf{V}$ and $\mathbf{W}$ are fixed, and $f_2(\mathbf{V})$ when $\mathbf{U}$ and $\mathbf{W}$ are fixed.

1) Fix $\mathbf{V}$ and $\mathbf{W}$, and solve for $\mathbf{U}$. The optimization reduces to

$$\min f_1(\mathbf{U}) = ||\mathbf{X} - \mathbf{U}\mathbf{V}^{\mathrm{T}}||_F^2, \quad \text{s.t.} \quad \mathbf{U} \ge 0. \tag{6}$$

The OGM algorithm can be used to resolve the above constrained optimization problem.

2) Fix $\mathbf{U}$ and $\mathbf{W}$, and solve for $\mathbf{V}$. The orthogonal constraint can be reformulated as [41]:

$$\sum_{i \ne j} v_i^{\mathrm{T}} v_j = \mathrm{Tr}(\mathbf{V}\mathbf{Q}\mathbf{V}^{\mathrm{T}}), \quad \mathbf{Q} = \overline{1} - \mathbf{I} \tag{7}$$

where $\overline{1}$ represents the matrix whose elements are all one and $\mathbf{I}$ is the identity matrix. Then, (5) can be reformulated to:

$$\begin{aligned} \min f_2(\mathbf{V}) = ||\mathbf{X} - \mathbf{U}\mathbf{V}^{\mathrm{T}}||_F^2 + \alpha \mathrm{Tr}(\mathbf{V}^{\mathrm{T}}\mathbf{L}\mathbf{V}) \\ + \beta \mathrm{Tr}(\mathbf{V}\mathbf{Q}\mathbf{V}^{\mathrm{T}}). \end{aligned} \tag{8}$$

To extend the OGM algorithm to solve the above constrained problem (8), we have the following lemma.

Lemma 1: The cost function $f_2(\mathbf{V})$ in (8) is convex with respect to $\mathbf{V}$ on the convex set $\mathbf{V} \in \mathcal{R}_+^{n \times d}$.

Proof: As the function $f_2(\mathbf{V})$ is continuous, twice differentiable, it is convex with respect to $\mathbf{V} \in \mathcal{R}_+^{n\times d}$ if the Hessian matrix

$$\nabla_{\mathbf{V}}^2 f = (\mathbf{U}^{\mathrm{T}}\mathbf{U} + \beta\mathbf{Q}) \otimes \mathbf{I}_n + \mathbf{I}_d \otimes \alpha\mathbf{L}$$

is semidefinite on the convex set $\mathcal{R}_+^{n\times d}$, which is quite obvious since $\nabla_{\mathbf{V}}^2 f$ is nonnegative, implying that $\mathrm{vec}(\mathbf{V})^{\mathrm{T}}(\nabla_{\mathbf{V}}^2 f)\mathrm{vec}(\mathbf{V}) \geq 0$ for any $\mathbf{V} \in \mathcal{R}_+^{n\times d}$, where $\mathrm{vec}(\mathbf{V})$ is the vectorization of $\mathbf{V} \in \mathcal{R}_+^{n\times d}$. As a result, once the matrix $\mathbf{U}$ is fixed, the minima of $f_2(\mathbf{V})$ will always exist and any local minimum of $f_2(\mathbf{V})$ is also a global one. □

Lemma 2: The gradient of $f_2(\mathbf{V})$ is Lipschitz continuous and the Lipschitz constant is $\mathbf{L}_v = \left\|\mathbf{U}^{\mathrm{T}}\mathbf{U}\right\|_2 + \|\alpha\mathbf{L}\|_2 + \|\beta\mathbf{Q}\|_2$.

Proof: The gradient of $f_2(\mathbf{V})$ is: $\nabla_{\mathbf{V}} f_2(\mathbf{V}) = \alpha\mathbf{L}\mathbf{V} + \mathbf{V}(\mathbf{U}^{\mathrm{T}}\mathbf{U} + \beta\mathbf{Q}) - \mathbf{X}^{\mathrm{T}}\mathbf{U}$. Since (8) is a linear combination of $f(\mathbf{U},\mathbf{V}) = ||\mathbf{X} - \mathbf{U}\mathbf{V}^{\mathrm{T}}||_F^2$, $\mu_1(\mathbf{V}) = \mathrm{tr}(\mathbf{V}^{\mathrm{T}}\mathbf{L}\mathbf{V})$, and $\mu_2(\mathbf{V}) = tr(\mathbf{V}\mathbf{Q}\mathbf{V}^{\mathrm{T}})$. The Lipschitz constant of $\nabla_{\mathbf{V}} f_2(\mathbf{V})$ can be calculated as a linear combination of the Lipschitz constants of the $\nabla f(\mathbf{U},\mathbf{V})$, $\nabla\mu_1(\mathbf{V})$, and $\nabla\mu_2(\mathbf{V})$, wherein $\nabla f(\mathbf{U},\mathbf{V})$, $\nabla\mu_1(\mathbf{V})$, and $\nabla\mu_2(\mathbf{V})$ are the gradient of $f(\mathbf{U},\mathbf{V})$, $\mu_1(\mathbf{V})$, and $\mu_2(\mathbf{V})$, respectively. According to [39], the Lipschitz constant of $\nabla f(\mathbf{U},\mathbf{V})$ and $\nabla\mu_1(\mathbf{V})$ are $\left\|\mathbf{U}^{\mathrm{T}}\mathbf{U}\right\|_2$ and $\alpha\|\mathbf{L}\|_2$, respectively.

Then, for any two matrices $\mathbf{V}_1, \mathbf{V}_2 \in \mathcal{R}_+^{n\times k}$, one can derive that

$$\begin{aligned}&\|\nabla\mu_2(\mathbf{V}_1) - \nabla\mu_2(\mathbf{V}_2)\|_F^2 \\ &\quad= \|(\mathbf{V}_1 - \mathbf{V}_2)(\beta\mathbf{Q})\|_F^2 \\ &\quad= \mathrm{tr}\left((\mathbf{P}\Sigma\mathbf{P}^{\mathrm{T}}(\mathbf{V}_1 - \mathbf{V}_2))^{\mathrm{T}}(\mathbf{P}\Sigma\mathbf{P}^{\mathrm{T}}(\mathbf{V}_1 - \mathbf{V}_2))\right) \quad (9)\end{aligned}$$

where $\mathbf{P}\Sigma\mathbf{P}^{\mathrm{T}}$ is the singular value decomposition of $\beta\mathbf{Q}$, and tr denotes trace. Let the largest singular value be δ. It can be derived from (9) that

$$\begin{aligned}&\|\nabla\mu_2(\mathbf{V}_1) - \nabla\mu_2(\mathbf{V}_2)\|_F^2 \\ &\quad= \mathrm{tr}\left(\mathbf{P}^{\mathrm{T}}(\mathbf{V}_1 - \mathbf{V}_2)(\mathbf{V}_1 - \mathbf{V}_2)^{\mathrm{T}}\mathbf{P}\Sigma^2\right) \\ &\quad\leq \delta^2\mathrm{tr}\left(\mathbf{P}^{\mathrm{T}}(\mathbf{V}_1 - \mathbf{V}_2)(\mathbf{V}_1 - \mathbf{V}_2)^{\mathrm{T}}\mathbf{P}\right) \\ &\quad= \delta^2\|\mathbf{V}_1 - \mathbf{V}_2\|_F^2 \quad (10)\end{aligned}$$

where the remaining two equations come be derived from the fact that $\mathbf{P}^{\mathrm{T}}\mathbf{P}$ and $\mathbf{P}\mathbf{P}^{\mathrm{T}}$ are identity matrices. From (10), we can find a constant $\mathbf{L}_q$ (e.g., $\mathbf{L}_q = \delta$), such that

$$\|\nabla_{\mathbf{V}} f_2(\mathbf{V}_1) - \nabla_{\mathbf{V}} f_2(\mathbf{V}_2)\|_F \leq \mathbf{L}_q\|\mathbf{V}_1 - \mathbf{V}_2\|_F .$$

Therefore, $\nabla_{\mathbf{V}} f_2(\mathbf{V})$ is Lipschitz continuous and the Lipschitz constant is the largest singular value of $\mathbf{U}^{\mathrm{T}}\mathbf{U} + \alpha\mathbf{L} + \beta\mathbf{Q}$, i.e., $\mathbf{L}_c = \mathbf{U}^{\mathrm{T}}\mathbf{U} + \alpha\mathbf{L} + \beta\mathbf{Q}$. This completes the proof. □

By replacing $\mathbf{L}_c$ and $\nabla_{\mathbf{V}} f_2(\mathbf{V})$ with the step size (e.g., $\mathbf{L}_u$) and gradient (e.g., $\nabla_{\mathbf{U}} f_1(\mathbf{U})$), respectively, we can obtain the optimal solution of $\mathbf{U}$.

3) Fix $\mathbf{U}$ and $\mathbf{V}$, and solve for $\mathbf{W}$. Denote $d_{ij}^x = \sum_s \theta_s\|x_i^s - x_j^s\|_2^2$, which represents the weighted distance between data points x_i and x_j. Then, the problem (5) is reformulated as:

$$\begin{aligned}&\min_{\mathbf{W}} \sum_{i,j}(d_{ij}^x w_{ij} + \gamma w_{ij}^2) + \alpha\sum_{i,j}\|v_i - v_j\|_2^2 w_{ij} \\ &\text{s.t. } w_i^{\mathrm{T}}1 = 1, \quad 0 \leq w_{ij} \leq 1. \quad (11)\end{aligned}$$

Denote $d_{ij}^v = \|v_i - v_j\|_2^2$, the optimal solution for $\mathbf{W}$ in problem (11) can be obtained by an efficient iterative algorithm proposed in [38]:

$$\begin{aligned}w_{ij} &= \frac{\eta - d_{ij}}{\gamma} \quad \text{for } j = 1, 2, \ldots k \\ w_{ij} &= 0 \quad \text{for } j = k+1, \ldots n \quad (12)\end{aligned}$$

where $d_{ij} = d_{ij}^x + \alpha d_{ij}^v$, $\eta = ((\gamma + \sum_{j=1}^k d_{ij})/k)$, and $k = \{j|\eta - d_{ij} > 0, j = 1, 2, \ldots, n\}$.

Algorithm 1 summarizes the proposed method.

Algorithm 1 The Framework of NMFAGR for HSI Dimension Reduction

Input:

A hyperspectral remote-sensing image;

Output:

The low-dimensional representation of the input hyperspectral remote-sensing image;

1: Construct spectral-spatial feature data set $\mathbf{X} = \{\mathbf{X}_{(s)} \in \mathcal{R}^{m_s\times n}\}_{s=1}^{\mathbf{S}}$ by extracting spectral-spatial features from the input hyperspectral remote-sensing image;
2: Randomly select a subset of samples from $\mathbf{X}$ for learning projection matrix $\mathbf{U}$;
3: Initialize $\theta_s = 1/\mathbf{S}$, $\mathbf{U}_t$,$\mathbf{V}_t$,$\mathbf{Y}_0 = \mathbf{U}_t$,$\mathbf{Z}_0 = \mathbf{V}_t$, $\delta_0 = 1$, $\mathbf{L}_u = \left\|\mathbf{U}^{\mathrm{T}}\mathbf{U}\right\|_2$, $t = 0$;
4: Update $\mathbf{U}$, $\mathbf{V}$;
   1) Update $\mathbf{U}_t = \mathcal{P}_+(\mathbf{Y}_t - \frac{1}{\mathbf{L}_u}\nabla_{\mathbf{U}} f_1(\mathbf{Y}_t))$, where $\mathcal{P}_+$ represents the non-negative projection;
   2) Update $\delta_{t+1} = \frac{1+\sqrt{4\delta_t^2+1}}{2}$;
   3) Update $\mathbf{Y}_{t+1} = \mathbf{U}_t + \frac{\delta_t - 1}{\delta_{t+1}}(\mathbf{U}_t - \mathbf{U}_{t-1})$;
   4) $t \leftarrow t + 1$;
   5) Repeat steps 1) – 4) until the convergence criterion is met;
   6) Update $\mathbf{V}$ by replacing $\mathbf{Y}_t$, $\mathbf{L}_u$ and $\nabla_{\mathbf{U}} f_1(\mathbf{U})$ with $\mathbf{Z}_t$, the step size (e.g., $\mathbf{L}_v$) and gradient (e.g., $\nabla_{\mathbf{V}} f_2(\mathbf{V})$), respectively;
   7) Update the graph matrix $\mathbf{W}$ by solving problem (11);
5: For test data, the low-dimensional feature representation is obtained by $\mathbf{D} = \mathbf{U}^{\dagger}\mathbf{X}$, where $\mathbf{U}^{\dagger}$ represents the pseudo-inverse of $\mathbf{U}$.
6: **return** The new representation of the input hyperspectral remote-sensing image in low-dimensional subspace;

TABLE 1. Class–specific accuracies in percentages on Indian Pines.

Class	PCA(28)	LPP(28)	NPE(28)	MFC(28)	MONMF(28)	NMFAGR(28)
1	90.22	78.00	**97.22**	88.89	77.78	**97.22**
2	74.24	73.13	55.52	79.68	77.50	**81.42**
3	65.00	83.18	80.51	82.44	83.59	**90.38**
4	86.79	93.51	93.05	95.19	88.77	**99.47**
5	90.61	93.84	94.23	92.84	90.90	**94.46**
6	94.97	97.65	95.68	96.08	94.56	**99.56**
7	90.00	**94.44**	**94.44**	88.89	83.33	88.89
8	99.30	99.00	99.00	98.60	99.07	**99.53**
9	98.00	99.00	**100.00**	99.00	95.00	**100.00**
10	75.90	**86.44**	73.43	79.93	79.72	83.95
11	75.11	76.17	73.63	72.72	**78.30**	71.89
12	84.48	89.69	**90.42**	90.27	89.50	**90.42**
13	**99.35**	**99.35**	**99.35**	98.15	**99.35**	**99.35**
14	88.93	**97.53**	93.95	96.20	96.30	95.56
15	90.62	93.45	94.88	97.12	98.21	**98.81**
16	71.57	77.47	91.39	93.45	90.36	**95.18**
OA	81.18	83.97	80.29	85.07	85.75	**87.79**
κ	78.45	82.70	77.89	83.03	83.15	**86.12**

TABLE 2. Class–specific accuracies in percentages on the University of Pavia.

Class	PCA(33)	LPP(33)	NPE(33)	MFC(33)	MONMF(33)	NMFAGR(33)
1	53.67	53.75	49.65	83.66	81.85	**84.61**
2	51.24	52.72	52.43	73.45	80.15	**87.00**
3	72.42	73.29	53.71	**85.65**	84.20	82.19
4	85.17	80.12	86.48	78.50	87.39	**98.20**
5	96.28	99.03	99.18	**99.85**	97.83	**99.85**
6	63.67	64.34	66.75	82.25	85.08	**90.58**
7	95.67	99.09	94.85	99.00	**99.32**	91.74
8	79.16	80.36	85.81	95.10	96.24	**96.32**
9	95.34	97.23	99.36	93.45	97.87	**99.89**
OA	67.56	68.89	67.85	81.22	85.07	**89.25**
κ	58.87	60.96	59.82	76.06	80.58	**86.06**

TABLE 3. Class–specific accuracies in percentages on the Pavia Centre.

Class	PCA(35)	LPP(35)	NPE(35)	MFC(35)	MONMF(35)	NMFAGR(35)
1	96.83	93.83	94.57	90.93	99.36	**99.95**
2	51.00	74.41	54.93	84.74	84.01	**87.22**
3	59.16	46.72	77.31	84.87	79.22	**93.21**
4	59.14	51.59	41.98	**76.26**	47.51	60.86
5	77.33	67.80	60.86	88.97	74.92	**93.44**
6	92.31	82.15	80.06	90.89	92.09	**96.07**
7	78.69	57.26	78.60	84.65	85.83	**86.02**
8	95.93	88.06	90.04	96.24	95.58	**97.21**
9	81.53	91.48	97.41	86.75	95.55	**100.00**
OA	90.42	85.70	86.79	91.28	93.85	**96.45**
κ	86.58	80.24	81.65	87.97	91.32	**94.98**

IV. EXPERIMENTS AND DISCUSSION

In this section two public hyperspectral data sets are used to conducted experiments. The task of classification is utilized to evaluate the performance of the proposed method. In our experiments, we first apply dimension reduction methods on the spectral and spatial features to find their low dimensional representation, and then employed the SVM classifier for classifying the test samples in that new feature representation [42].

A. DATA SETS

Indian Pines Data Set: The data set was collected by using the AVIRIS sensor over the Indian Pines region, northwest Indiana, USA, in 1992. The image scene consists of 145 $\times$ 145 pixels. The total number of spectral bands for sampling is 220. In experiments, 20 spectral bands were discarded because of noise and the water-absorption phenomena. There are 16 classes and 10,249 samples in total, with varying from 20 to 2,455 in each class.

University of Pavia Data Set: This data set was acquired by using the Reflective Optics System Imaging Spectrometer sensor (ROSIS) during a flight campaign in 2003 over the urban area of Pavia University, northern Italy. The collected image contains 610 $\times$ 340 pixels. The total number of spectral bands for sampling is 103. The data used for experiments has nine classes with a total of 42,776 samples.

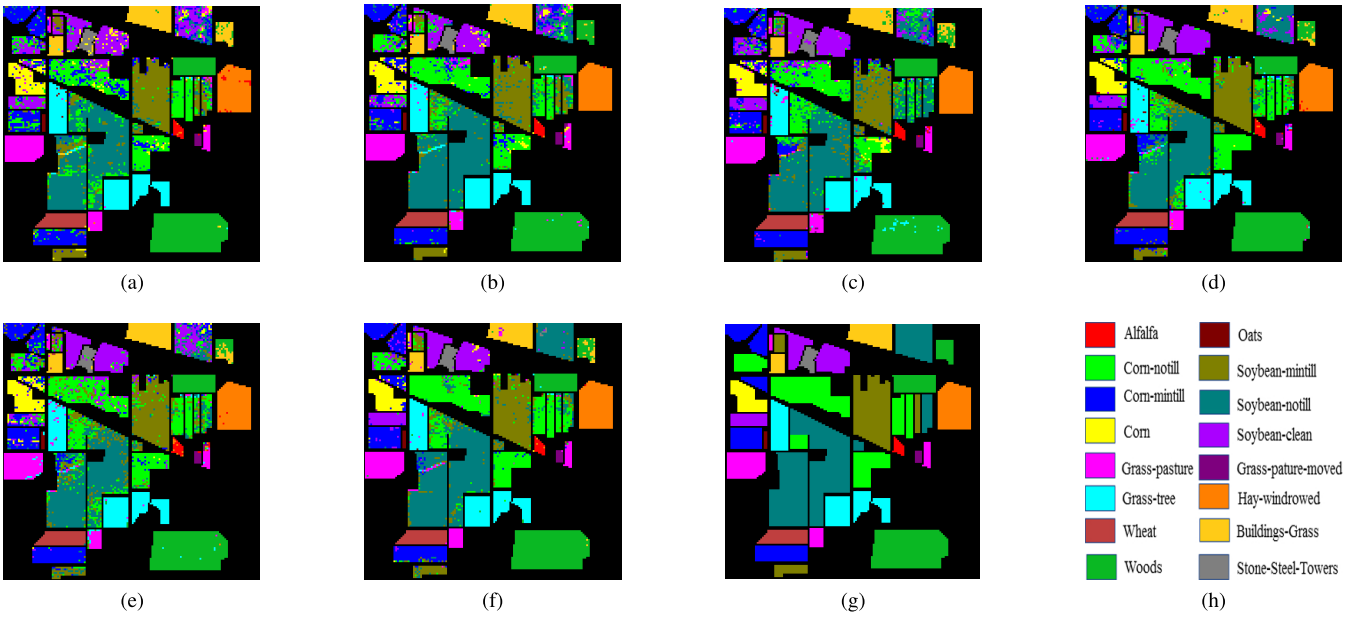


FIGURE 1. Classification maps for Indian Pines: (a) PCA features with SVM classifier, (b) LPP features with SVM classifier, (c) NPE features with SVM classifier, (d) MFC features with SVM classifier, (e) MONMF features with SVM classifier, (f) NMFAGR features with SVM classifier, (g) Color map of ground truth, and (h) the class legends for the Indian Pines data set.

Pavia Center Data Set: This data set was also acquired by the ROSIS sensor during a flight campaign over the city center of Pavia, Italy. The number of spectral bands in the acquired image is 102, and the geometric resolution is 1.3 meters. The data used for experiments has nine classes with a total of 7456 samples. This data was used in the 2008 IEEE Geoscience and Remote Sensing Data Fusion Technical Committee contest.

B. SPECTRAL AND SPATIAL FEATURES

In our experiments, we employed three types of features, i.e., spectral feature, texture feature, and morphological feature, as the raw representation of the input HSI data.

1) *The spectral feature:* The reflectance value in every spectral band is used to represent the spectral feature of a pixel.

$$\mathbf{X}_{spectral} = [x_1, x_2, \ldots, x_e]^T \tag{13}$$

where x_i is the reflectance value of the pixel in spectral channel i.

2) *The texture feature:* The gray-level co-occurrence matrix (GLCM) is employed to collect the texture feature. We firstly use the PCA transformation to obtain the principal component images from the raw image [43]. After that, the GLCM is employed to these images. In our experiments, three nonnegative values: energy, contrast, and homogeneity were used as measures with which to obtain the texture feature. The texture feature of a pixel is obtained by:

$$X_{texture} = [x_{ene}, x_{con}, x_{hom}]^T \tag{14}$$

These three measurements are measured by employing an inter-pixel distance in four directions (0°, 45°, 90°, and 135°); Then, we averaged these obtained measurements over the four directions. Following [35], the moving window size and the quantitative level are set to be 27 and 64, respectively.

3) *The morphological feature:* An extended attribute profile (EAP) is employed to describe other spatial attributes of the image [44], [45]. EAP is extended based on the attribute profiles (APs). APs are collected by employing a sequence of morphological attribute filters (AFs) to a scalar image. Given a sequence of ordered criteria $\{B\} = \{B_1, B_1, B_1, \ldots, B_c\}$, an AP is extracted by using a sequence of attribute−thinning and attribute−thickening transformations to the input image f

$$AP_f = \{\phi_b(f), \ldots, \phi_1, f, r_1(f), \ldots, r_b(f)\} \tag{15}$$

where ϕ_i and r_i denote the thickening and thinning operation with respect to criterion B_i, respectively. The EAP is collected by obtaining an AP on each of the first q features (F_i) calculated employing a feature extraction technique on the hyperspectral image

$$EAP = \{AP(F_1), AP(F_2), \ldots, AP(F_q)\}. \tag{16}$$

Following [35], the area attribute is employed to build an EAP. We also perform PCA on the hyperspectral image to obtain the first few principal component (PC) images prior to extracting the morphological features by employing EAP on these images.

Nevertheless, it must be pointed out that our proposed dimension reduction method is actually a general

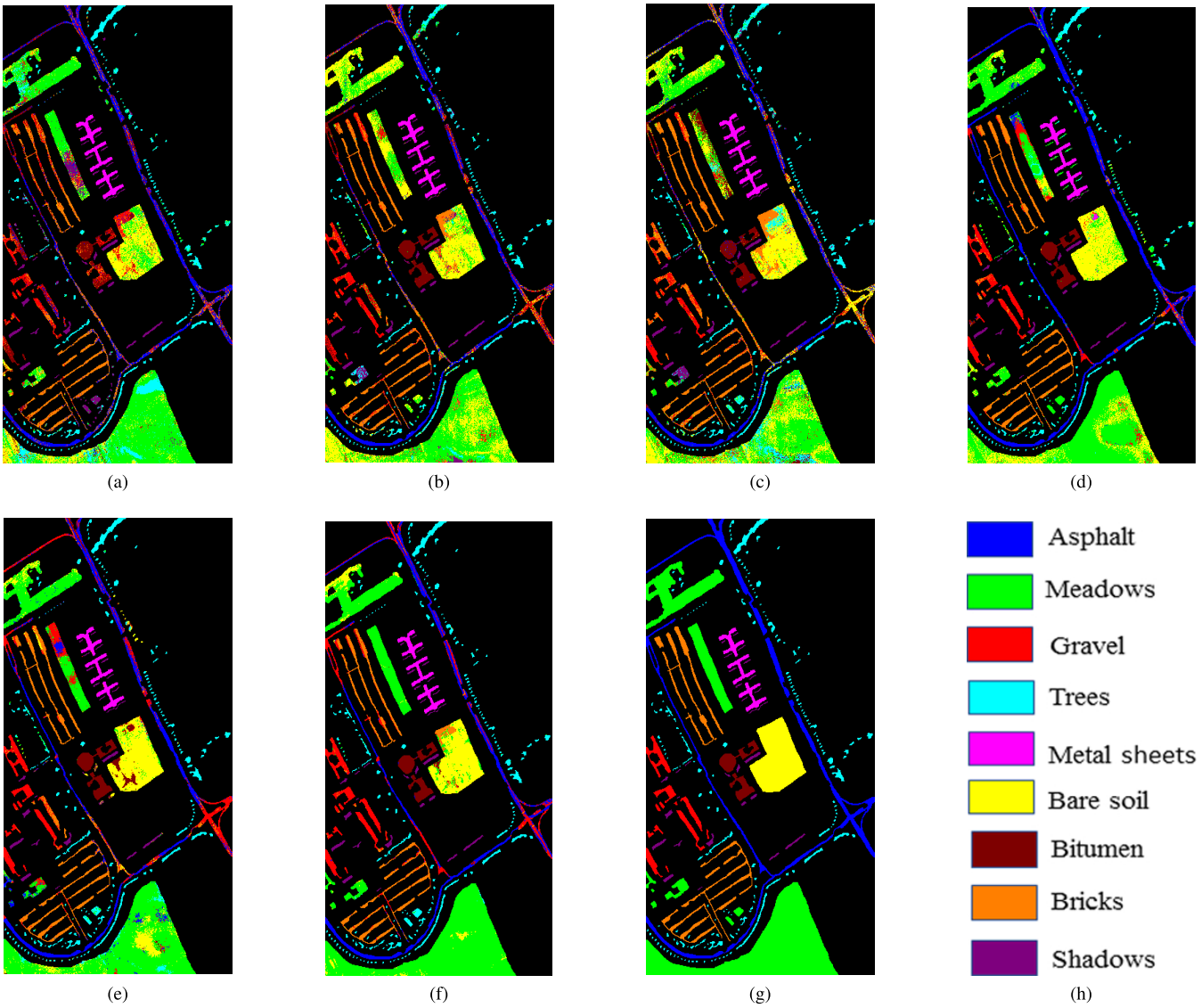


FIGURE 2. Classification maps for the University of Pavia: (a) PCA features with SVM classifier, (b) LPP features with SVM classifier, (c) NPE features with SVM classifier, (d) MFC features with SVM classifier, (e) MONMF features with SVM classifier, (f) NMFAGR features with SVM classifier, (g) Color map of ground truth, and (h) the class legends for the University of Pavia data set.

framework that is capable of dealing with any types of spectral-spatial features as input.

C. EXPERIMENTAL DESIGN

In order to validate the effectiveness of our proposed method, we compared it with several state-of-the-art dimension reduction methods. In our experiments, all of these methods were performed on long vectors, which were concatenations of spectral and spatial features. Moreover, we addressed two state-of-the-art multiple-feature learning methods, i.e., the MFC [33], and the MONMF [35]. The comparison methods were listed as follows:

1) Traditional dimension reduction method PCA;
2) Graph-based dimension reduction NPE;
3) Graph-based dimension reduction LPP;
4) Spectral-spatial dimension reduction method MFC [33];
5) Spectral-spatial dimension reduction method MONMF [35];
6) Proposed method NMFAGR.

Each HSI data set was randomly split into training and test sets. Ten samples per class were selected for training and the remaining for testing. The implementations of PCA, NPE, and LPP are publicly available.[1] The MFC and MONMF implementations were provided courtesy of their authors. For fairly comparison, we adopted cross-validation to obtain the optimal parameters of each method. The regularization parameter setting of MFONMF was the same as the original paper [35]. There are three parameters in the proposed method. The analysis of parameters will be introduced later.

[1] http://www.cad.zju.edu.cn/home/dengcai/

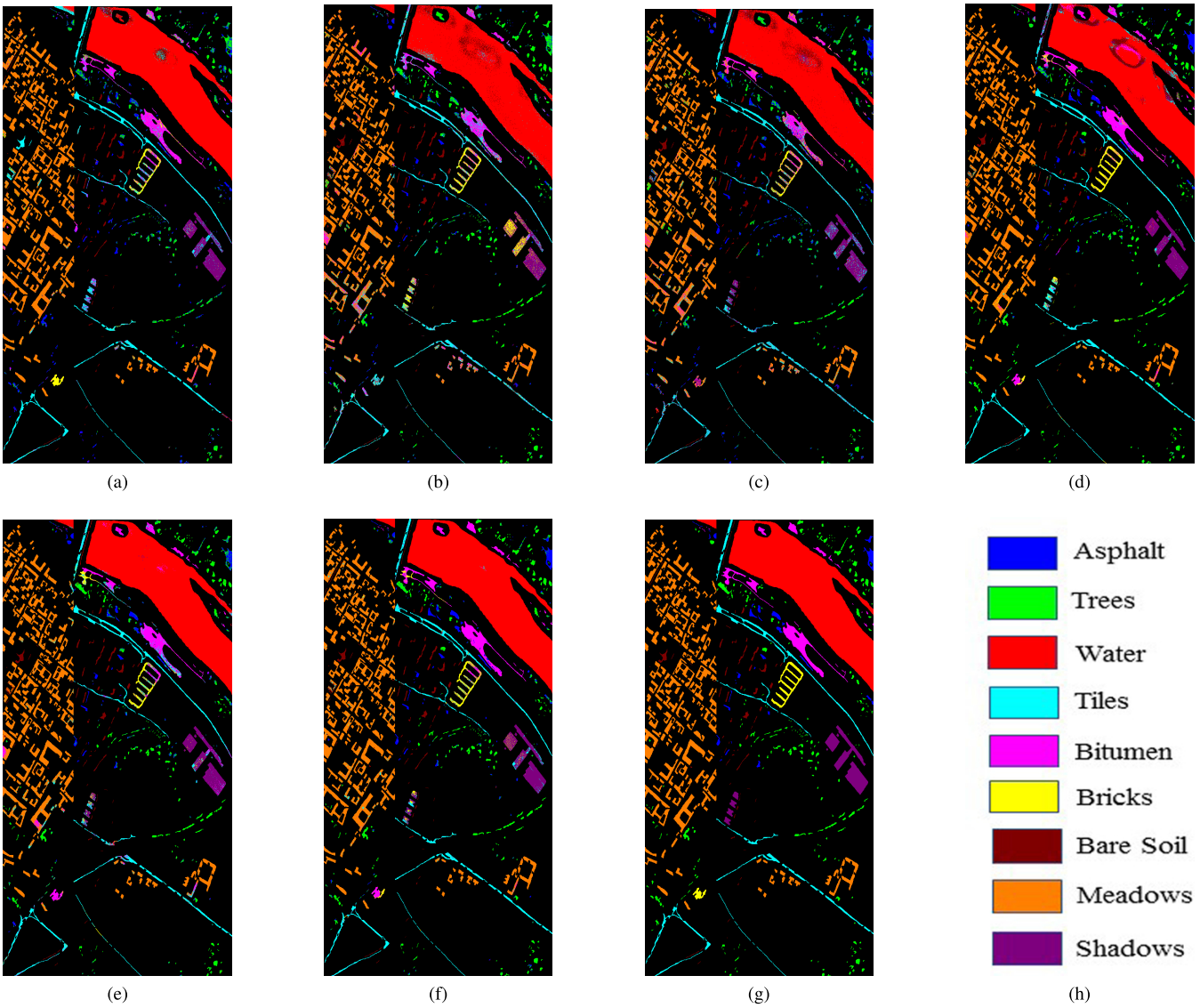


FIGURE 3. Classification maps for the Pavia centre: (a) PCA features with SVM classifier, (b) LPP features with SVM classifier, (c) NPE features with SVM classifier, (d) MFC features with SVM classifier, (e) MONMF features with SVM classifier, (f) NMFAGR features with SVM classifier, (g) Color map of ground truth, and (h) the class legends for the Pavia Center data set.

D. EXPERIMENTAL RESULTS

Table 1, Table 2, and Table 3 show the class-specific accuracies, the overall accuracies (OAs), and kappa coefficients (κ) in percentages. The number in each bracket represents the optimal dimension of the reduced features. The six dimension reduction-based classification maps of these three data sets are presented in Fig. 1, Fig. 2 and Fig. 3. From the results, NMFAGR outperforms other methods in terms of OA and kappa coefficient. On the India Pines data set, the proposed method achieved a 2.04% and a 2.97% improvement in terms of OA and kappa coefficient, compared with the second best result. On the Pavia University data set, the proposed method produced a 4.18% and a 5.48% improvement in terms of OA and kappa coefficient, when measured with the second highest result. On the Pavia center data set, the proposed method achieved a 2.60% and a 3.66% improvement in terms of OA and kappa coefficient, compared with the second best result.

Compared with the traditional dimension reduction method (e.g., PCA), the proposed method NMFAGR achieves an improvement. This is because NMFAGR can effectively and efficiently make use of the geometric information of HSI data to guide the dimension reduction process. When evaluated with graph-based methods (e.g., LPP and NPE), the proposed method NMFAGR obtains a better performance. This could be attributed to NMFAGR performing dimension reduction and graph learning simultaneously. Compared with multi-feature learning methods (e.g., MFC and MFONMF), NMFAGR obtains better performance. This is because NMFAGR learns the weight factors automatically such that it effectively make use of complementary information among different types of features.

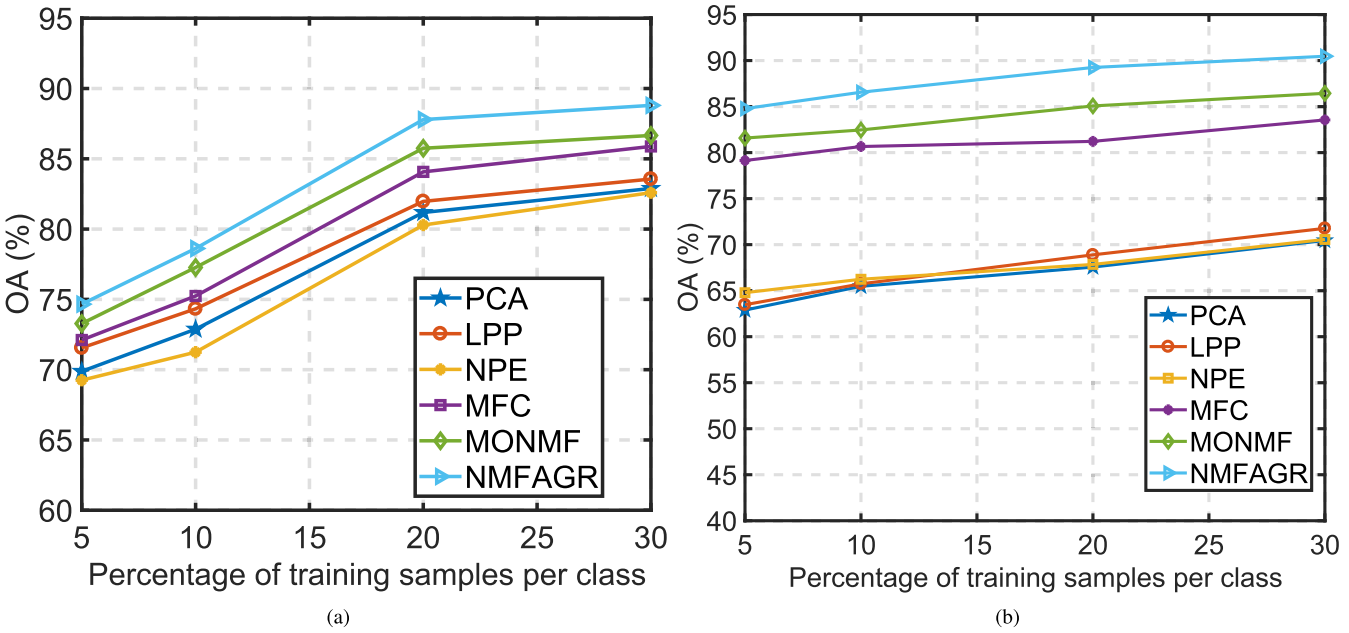


FIGURE 4. The changing proportions of training samples with respect to OA for data sets (a) Indian Pines, (b) and University of Pavia.

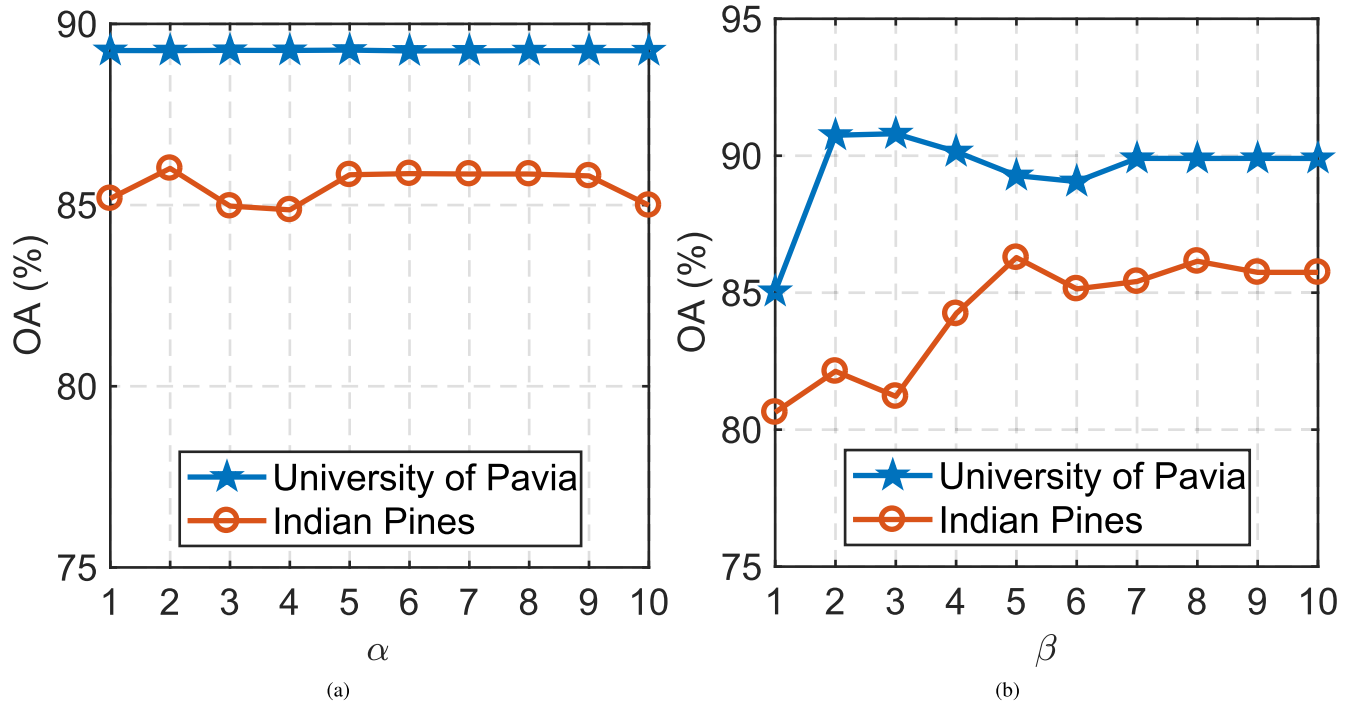


FIGURE 5. Parameter sensitivity of Indian Pines and University of Pavia data sets. (a) Regularization parameter α with respect to OA. (b) Parameter β with respect to OA.

Fig. 4 illustrates the OA of all methods on the Indian Pines, and PaviaU data sets with varying proportions of training samples. The proposed method NMFAGR is superior to other five methods in all cases. Graph-based dimension reduced methods perform better than traditional dimension reduced method PCA.

For the proposed method, there are three parameters, e.g., two regularization parameters α, and the number of subspace dimension d. In the experiment, α and β are chosen from 1 to 10, where the interval is 1. Then classification OA with respect to these two parameters is reported in Fig. 5. From the results, we can see that the proposed method is stable with

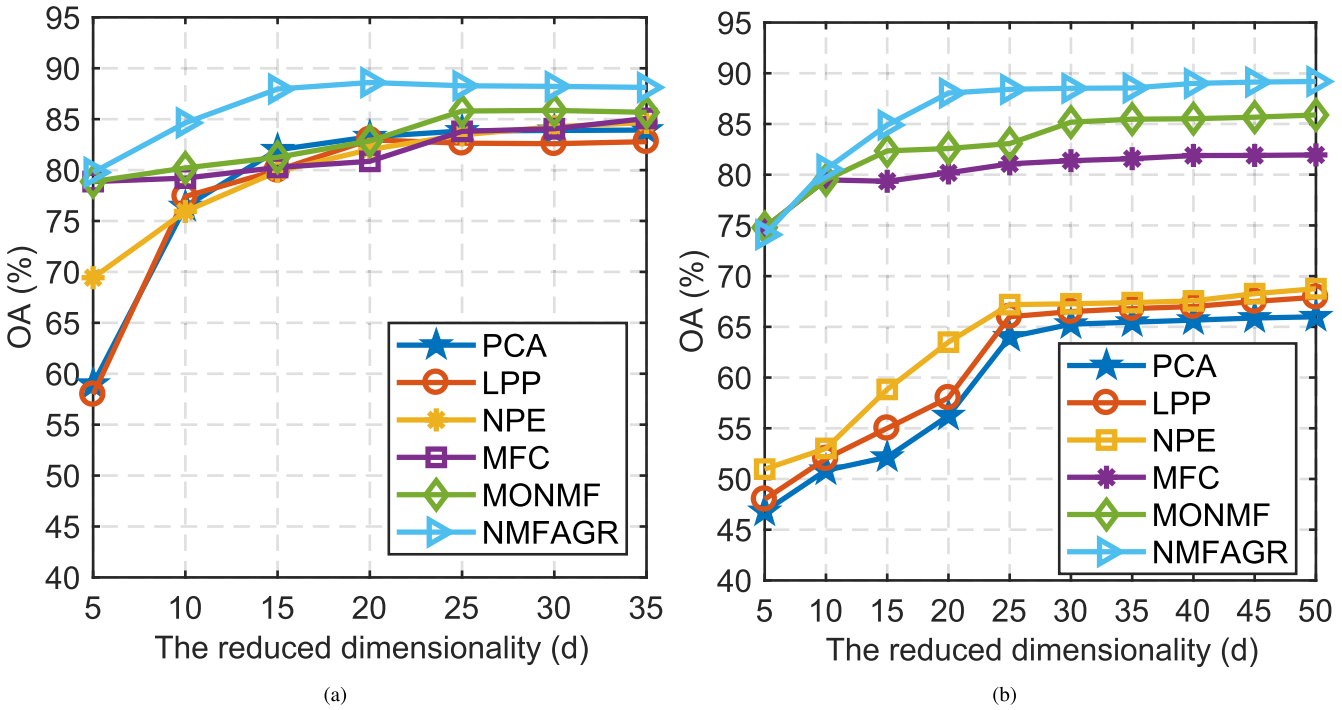


FIGURE 6. The reduced dimension d with respect to OA for datasets (a) Indian Pines and (b) University of Pavia.

TABLE 4. Computational cost of different dimension reduction methods (in seconds).

Data set	PCA	LPP	NPE	MFC	MONMF	NMFAGR
India Pines	0.058	19.15	65.26	1.89	1250.49	97.22
University of Pavia	0.165	690.13	6000.28	2.38	1077.50	81.42
Pavia Centre	0.046	30.27	50.49	5.58	120.50	61.42

respect to α. In addition, NMFAGR is stable when the β is set to be bigger than 6. For the India Pines dataset, the reduced dimension d is selected from 5 to 35. For the University of Pavia, d is chosen from 5 to 50. The corresponding results are shown in Fig. 6. We can see that the inflection point from the classification results is around the dimension of 25 for the India Pines and 30 for the University of Pavia.

Table 4 shows the computation time for the various algorithms on the three data sets. Only the computational time in the process of dimension reduction is recorded. Each method is implemented using MATLAB R2017a installed in a personal computer with Intel i7–6700 Quad Core Processor and 16 GB of RAM, and the Microsoft Windows 7 operational system. From Table 4, We observed that PCA is the fastest while MONMF is the slowest. This is because the MONMF method requires many iterations to guarantee convergence. In contrast, the proposed method is much faster since it adopts the optimal gradient method to accelerate the convergence speed.

V. CONCLUSIONS

We proposed a new graph-based NMF method in this paper known as NMFAGR by incorporating an AGL regularization constraint into the NMF model. A major advantage of this method is that NMFAGR learns both the low-dimensional representation and an optimal graph simultaneously. To improve the locality preservation of the encoding matrix, NMFAGR establishes an interaction between the obtained low-dimensional representation and the learned graph. Moreover, to effectively differentiate the importance of spatial-spectral features, we used a novel way to lean feature weight factors without adding any parameter. Experimentation on three hyperspetral imagery data sets show the superiority of our algorithm as a dimension reduction method.

REFERENCES

[1] J. C. Harsanyi and C.-I. Chang, "Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 4, pp. 779–785, Jul. 1994.
[2] H. Yuan and Y. Y. Tang, "Learning with hypergraph for hyperspectral image feature extraction," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 8, pp. 1695–1699, Aug. 2015.
[3] S. Prasad, J. Chanussot, J. E. Fowler, J. Bioucas-Dias, and C. D. Creuserre, "Introduction to the issue on advances in hyperspectral data processing and analysis," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 6, pp. 961–963, Sep. 2015.
[4] G. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. Inf. Theory*, vol. IT-14, no. 1, pp. 55–63, Jan. 1968.

[5] Y. Zhou, J. Peng, and C. L. P. Chen, "Dimension reduction using spatial and spectral regularized local discriminant embedding for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 2, pp. 1082–1095, Feb. 2015.

[6] H. Yuan and Y. Y. Tang, "Spectral–spatial shared linear regression for hyperspectral image classification," *IEEE Trans. Cybern.*, vol. 47, no. 4, pp. 934–945, Apr. 2017.

[7] H. Luo *et al.*, "Hyperspectral image classification using metric learning in one-dimensional embedding framework," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 5, pp. 1987–2001, May 2017.

[8] L. Zhang, Y. Zhong, B. Huang, J. Gong, and P. Li, "Dimensionality reduction based on clonal selection for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 12, pp. 4172–4186, Dec. 2007.

[9] W. Sun and Q. Du, "Graph-regularized fast and robust principal component analysis for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3185–3195, Jun. 2018.

[10] W. Sun, L. Tian, Y. Xu, D. Zhang, and Q. Du, "Fast and robust self-representation method for hyperspectral band selection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 11, pp. 5087–5098, Nov. 2017.

[11] W. Sun, G. Yang, B. Du, L. Zhang, and L. Zhang, "A sparse and low-rank near-isometric linear embedding method for feature extraction in hyperspectral imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 4032–4046, Jul. 2017.

[12] C. Rodarmel and J. Shan, "Principal component analysis for hyperspectral image classification," *Surv. Land Inf. Sci.*, vol. 62, no. 2, p. 115, 2002.

[13] J. Wang and C.-I. Chang, "Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 6, pp. 1586–1600, Jun. 2006.

[14] E. Esser, M. Moller, S. Osher, G. Sapiro, and J. Xin, "A convex model for nonnegative matrix factorization and dimensionality reduction on physical space," *IEEE Trans. Image Process.*, vol. 21, no. 7, pp. 3239–3252, Jul. 2012.

[15] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, pp. 788–791, Oct. 1999.

[16] Z. Yang, Y. Xiang, K. Xie, and Y. Lai, "Adaptive method for nonsmooth nonnegative matrix factorization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 4, pp. 948–960, Apr. 2017.

[17] Z. Yang, G. Zhou, S. Xie, S. Ding, J.-M. Yang, and J. Zhang, "Blind spectral unmixing based on sparse nonnegative matrix factorization," *IEEE Trans. Image Process.*, vol. 20, no. 4, pp. 1112–1125, Apr. 2011.

[18] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2001, pp. 556–562.

[19] D. Wang, X. Gao, and X. Wang, "Semi-supervised nonnegative matrix factorization via constraint propagation," *IEEE Trans. Cybern.*, vol. 46, no. 1, pp. 233–244, Jan. 2016.

[20] X. Wang, Y. Zhong, L. Zhang, and Y. Xu, "Spatial group sparsity regularized nonnegative matrix factorization for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 11, pp. 6287–6304, Nov. 2017.

[21] Y. Zhong, A. Ma, Y. S. Ong, Z. Zhu, and L. Zhang, "Computational intelligence in optical remote sensing image processing," *Appl. Soft Comput.*, vol. 64, pp. 75–93, Mar. 2018.

[22] C.-H. Lin, W.-K. Ma, W.-C. Li, C.-Y. Chi, and A. Ambikapathi, "Identifiability of the simplex volume minimization criterion for blind hyperspectral unmixing: The no-pure-pixel case," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 10, pp. 5530–5546, Oct. 2015.

[23] C.-H. Lin, C.-Y. Chi, Y.-H. Wang, and T.-H. Chan, "A fast hyperplane-based minimum-volume enclosing simplex algorithm for blind hyperspectral unmixing," *IEEE Trans. Signal Process.*, vol. 64, no. 8, pp. 1946–1961, Apr. 2016.

[24] D. Lunga, S. Prasad, M. M. Crawford, and O. Ersoy, "Manifold-learning-based feature extraction for classification of hyperspectral data: A review of advances in manifold learning," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 55–66, Jan. 2014.

[25] S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: A general framework for dimensionality reduction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 1, pp. 40–51, Jan. 2007.

[26] W. Li and Q. Du, "Laplacian regularized collaborative graph for discriminant analysis of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7066–7076, Dec. 2016.

[27] H. Luo *et al.*, "Hyperspectral image classification based on spectral–spatial one-dimensional manifold embedding," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 9, pp. 5319–5340, Sep. 2016.

[28] W. Li, S. Prasad, J. E. Fowler, and L. M. Bruce, "Locality-preserving dimensionality reduction and classification for hyperspectral image analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 4, pp. 1185–1198, Apr. 2012.

[29] J. Wen, Z. Tian, X. Liu, and W. Lin, "Neighborhood preserving orthogonal PNMF feature extraction for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 759–768, Apr. 2013.

[30] Y. Zhai *et al.*, "A modified locality-preserving projection approach for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 8, pp. 1059–1063, Aug. 2016.

[31] W. Zhuge, F. Nie, C. Hou, and D. Yi, "Unsupervised single and multiple views feature extraction with structured graph," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 10, pp. 2347–2359, Oct. 2017.

[32] R. Ji, Y. Gao, R. Hong, Q. Liu, D. Tao, and X. Li, "Spectral-spatial constraint hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 3, pp. 1811–1824, Mar. 2014.

[33] L. Zhang, L. Zhang, D. Tao, and X. Huang, "On combining multiple features for hyperspectral remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, pp. 879–893, Mar. 2012.

[34] L. Zhang, L. Zhang, D. Tao, and X. Huang, "A modified stochastic neighbor embedding for multi-feature dimension reduction of remote sensing images," *ISPRS J. Photogramm. Remote Sens.*, vol. 83, pp. 30–39, Sep. 2013.

[35] J. Wen, J. E. Fowler, M. He, Y.-Q. Zhao, C. Deng, and V. Menon, "Orthogonal nonnegative matrix factorization combining multiple features for spectral–spatial dimensionality reduction of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 7, pp. 4272–4286, Jul. 2016.

[36] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Comput.*, vol. 15, no. 6, pp. 1373–1396, 2003.

[37] B. Mikhail and N. Partha, "Laplacian eigenmaps and spectral techniques for embedding and clustering," in *Proc. NIPS*, 2002, pp. 585–591.

[38] F. Nie, G. Cai, J. Li, and X. Li, "Auto-weighted multi-view learning for image clustering and semi-supervised classification," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1501–1511, Mar. 2017.

[39] N. Guan, D. Tao, Z. Luo, and B. Yuan, "NeNMF: An optimal gradient method for nonnegative matrix factorization," *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 2882–2898, Jun. 2012.

[40] Z. Yang, Y. Zhang, W. Yan, Y. Xiang, and S. Xie, "A fast non-smooth nonnegative matrix factorization for learning sparse representation," *IEEE Access*, vol. 4, pp. 5161–5168, 2016.

[41] B. Li, G. Zhou, and A. Cichocki, "Two efficient algorithms for approximately orthogonal nonnegative matrix factorization," *IEEE Signal Process. Lett.*, vol. 22, no. 7, pp. 843–846, Jul. 2015.

[42] L. Zhang, Q. Zhang, B. Du, X. Huang, Y. Y. Tang, and D. Tao, "Simultaneous spectral-spatial feature selection and extraction for hyperspectral images," *IEEE Trans. Cybern.*, vol. 48, no. 1, pp. 16–28, Jan. 2016.

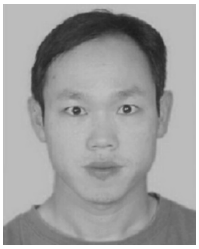
[43] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973.

[44] M. Dalla Mura, J. A. Benediktsson, B. Waske, and L. Bruzzone, "Morphological attribute profiles for the analysis of very high resolution images," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 10, pp. 3747–3762, Oct. 2010.

[45] M. D. Mura, J. A. Benediktsson, B. Waske, and L. Bruzzone, "Extended profiles with morphological attribute filters for the analysis of hyperspectral data," *Int. J. Remote Sens.*, vol. 31, no. 22, pp. 5975–5991, Dec. 2010.

WEI YAN was born in Ji'an, Jiangxi, China. He received the master's degree from the School of Automation, Guangdong University of Technology, Guangzhou, China, in 2016. He is currently pursuing the Ph.D. degree with the Faculty of Science and Technology, University of Macau. His research interests include machine learning, nonnegative signal processing, and blind signal processing.

BOB ZHANG (M'11) received the B.A. degree in computer science from York University, Toronto, ON, Canada, in 2006, the M.A.Sc. degree in information systems security from Concordia University, Montréal, QC, Canada, in 2007, and the Ph.D. degree in electrical and computer engineering from the University of Waterloo, Waterloo, ON, Canada, in 2011. He was with the Center for Pattern Recognition and Machine Intelligence, and later was a Post-Doctoral Researcher with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, USA. He is currently an Assistant Professor with the Department of Computer and Information Science, University of Macau. His current research interests include medical biometrics, pattern recognition, and image processing.



ZUYUAN YANG (M'15) received the B.E. degree from the University of Hunan University of Science and Technology, Xiangtan, China, in 2003, and the Ph.D. degree from the South China University of Technology, Guangzhou, China, in 2010. He received the Excellent Ph.D. Thesis Award Nomination of China. He joined the National Program for New Century Excellent Talents in University and received the Guangdong Distinguished Young Scholar. He is currently a Researcher with the School of Automation, Guangdong University of Technology, Guangzhou. His research interests include machine learning, nonnegative matrix factorization, and image processing.