Hyperspectral Image Clustering Based on Unsupervised Broad Learning

Yi Kong, Yuhu Cheng, *Member, IEEE*, C. L. Philip Chen, *Fellow, IEEE*, and Xuesong Wang, *Member, IEEE*

***Abstract*— Due to the difficulty of labeling a large number of training samples of a hyperspectral image (HSI), unsupervised clustering methods have drawn great attention. The recently proposed broad learning (BL) can implement both linear and nonlinear mappings. However, the original BL is a supervised model. In this letter, a novel method named unsupervised BL (UBL) is introduced for HSI clustering. First, a graph-regularized sparse autoencoder is performed on the input and mapped feature of UBL in order to maintain the intrinsic manifold structure of origin HSI. Then, the objective function of UBL composed of an l_2-norm of output-layer weights and a graph regularization term is designed, which can be easily solved by choosing eigenvectors corresponding to the smallest eigenvalues. Finally, the HSI clustering results can be obtained by applying spectral clustering on the output of UBL. Experiments on three popular real HSI data sets demonstrate that, compared with several competitive methods, UBL can achieve better clustering performance.**

***Index Terms*— Clustering, hyperspectral image, manifold, unsupervised broad learning (UBL).**

I. Introduction

HYPERSPECTRAL image (HSI) acquired by hyperspectral sensors provides a rich spectral and spatial information; thus, it has become a powerful tool for earth surface monitoring [1]. HSI classification and clustering are two commonly used methods in the field of earth surface monitoring. In general, as a supervised method, HSI classification can only deal with labeled data. However, in real-world applications, labeled HSI pixels are very limited. Furthermore, labeling a large amount of HSI pixels is time-consuming and requires considerable expertise. HSI clustering methods are unsupervised, which can determine the class label of HSI pixels without employing any label information. However, due to the high spectral resolution and complex spatial structure of HSI, it is not easy for HSI clustering.

Manuscript received September 12, 2018; revised November 29, 2018; accepted March 20, 2019. Date of publication April 18, 2019; date of current version October 30, 2019. This work was supported by the Outstanding Innovation Scholarship for Doctoral Candidate of "Double First Rate" Construction Disciplines of the China University of Mining and Technology. *(Corresponding author: Xuesong Wang.)*

Y. Kong, Y. Cheng, and X. Wang are with the School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116, China, and also with the Xuzhou Key Laboratory of Artificial Intelligence and Big Data, Xuzhou 221116, China (e-mail: kongyicumt@163.com; chengyuhu@163.com; wangxuesongcumt@163.com).

C. L. P. Chen is with the Navigation College, Dalian Maritime University, Dalian 116026, China, also with the Unmanned System Research Institute, Northwestern Polytechnical University, Xi'an 710072, China, and also with the School of Intelligent Engineering, South China University of Technology, Guangzhou 510640, China (e-mail: philip.chen@ieee.org).

Color versions of one or more of the figures in this letter are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LGRS.2019.2907598

There are four types of frequently used clustering methods, which are centroid-based methods [2], density-based methods, biological methods, as well as graph-based methods. However, these methods suffer from errors due to the uniformly distributed feature point within the feature space, which results from the large spectral variability observed in HSI [3]. Recently, sparse subspace clustering (SSC) [4] was applied to HSI due to its robustness. SSC groups data points into different subspaces, with each data point being represented by the data points residing in its own subspace. However, it focuses on the spectral characteristic only, whereas ignores the abundant spatial information, which degraded the HSI clustering performance. To address this issue, Yan *et al.* [5] combined two similarity matrix construction methods with SSC, so that the sparse coefficient matrices can cover both spectral and spatial information of HSI. Zhang *et al.* [6] proposed two spectral–spatial SSC methods, SSC-S and S4C, to obtain more accurate sparse coefficient matrices based on both spectral and spatial information. However, S4C is based on the local mean constraint, which cannot explore the detailed spatial information. Therefore, Zhai *et al.* [3] proposed the L2-SSC, which promoted the spatial smoothness by adding the l_2-norm total variation regularization term to the SSC.

More recently, Chen and Liu [7] proposed the broad learning (BL), a three-layer feedforward flat network, which mainly consists of two parts. The first one is "mapped feature (MF)," which maps the input data with random weights [or weights fine-tuned with linear sparse autoencoder (SAE)] to a feature space. The second one is "enhancement node (EN)" that can expand the features in a wide sense by mapping MF to EN with random weights. Both the MF and EN are connected to the output layer, and the connecting weights can be easily obtained by solving a ridge regression problem. To further improve the classification accuracy of BL, Liu and Chen [8] added a feature extraction process with k-means before BL. By incorporating the class-probability structure into BL, Kong *et al.* [9] proposed a semisupervised BL (SBL) method for HSI classification. Experiments on three HSI data sets show that SBL can achieve higher classification accuracy than BL. However, all of the above are supervised or semisupervised methods, which ask for at least one labeled sample for each surface object in the training data set.

Since it is practically difficult to manually label HSI pixels for each surface object, the applicability of supervised methods is limited. Therefore, in this letter, a novel hyperspectral image clustering method based on unsupervised BL (UBL) is introduced. The main contributions can be summarized as:

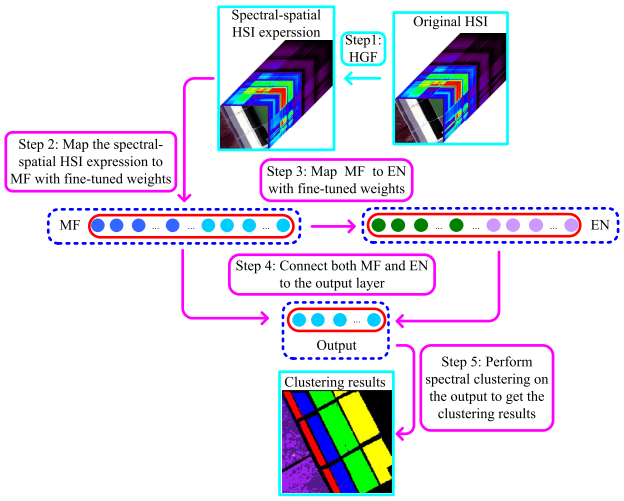

Fig. 1. Flowchart of HSI clustering based on UBL.

1) an unsupervised version of BL is proposed for HSI clustering; 2) the original BL uses the random weights to map MF to EN; thus, the learning performance of BL is affected by the random condition. To address this issue, the weights of EN are fine-tuned by using the graph regularized SAE (GRSAE); and 3) to preserve the manifold structure of original HSI data, three graph regularization terms are, respectively, added to the fine-tuned process of MF and EN weights and the final objective function of UBL.

II. HSI Clustering Based on UBL

As is shown in Fig. 1, the flowchart of UBL-based HSI clustering includes five steps: 1) preprocess the original HSI with hierarchical guidance filtering (HGF) to obtain the spectral–spatial representation of HSI; 2) map the spectral–spatial HSI expression to MF with weights fine-tuned by GRSAE; 3) expand the broad of MF with EN and the weights of EN are also fine-tuned with GRSAE; 4) connect both the MF and EN to the output layer to get the output of UBL, and the output-layer weights for UBL can be easily obtained by solving the generalized eigenvalue decomposition problem; and 5) perform spectral clustering on the output to get the clustering results.

The original HSI data are presented in the form of 3-D tensor and contains a wealth of spectral and spatial information. Directly vectorizing the 3-D tensor into a 1-D vector will result in the huge increment of data dimension and the loss of inherent structure information. Recently, HGF was introduced to get the spectral–spatial representation of HSI, which can remove the noise and local details while preserving the overall structure of HSI. More details can be found in [10]. Here, the original HSI is processed with HGF to get the spectral–spatial representation.

A. Fine-Tune Weights of MF and EN With GRSAE

For the original BL, the connection weights of MF and EN are randomly generated. To achieve a better feature representation, Chen and Liu [7] adopted linear SAE to fine-tune them. Whether the input data are mapped with weights randomly generated or fine-tuned with linear SAE, the resulting features often ignore the intrinsic structure (e.g., geometric manifold structure) of the input data. Manifold learning aims to construct a graph in which vertices representing data points are connected by edges. It is shown that it can better represent the relationship between data. Therefore, a large number of graph embedding-based methods are used for HSI clustering. In addition, graph regularization is also a common regularization technique. By adding the graph regularization term, the manifold structure of the input data can be maintained. In this letter, we apply the graph regularization technique to the fine-tune process of weights of MF and EN; thus, the MFs and ENs can maintain the same manifold structure with input data.

Given the HGF-based HSI samples $\boldsymbol{X} = \{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n\} \in R^{n\times d}$, where n is the number of samples and d is the dimension of samples. In terms of UBL, the input is first mapped to MF via G^{M} groups of weights $\tilde{\boldsymbol{W}}_i^{\mathrm{M}}$

$$\tilde{\boldsymbol{E}}_{\boldsymbol{i}} = \boldsymbol{X}\tilde{\boldsymbol{W}}_{\boldsymbol{i}}^{\mathbf{M}} \tag{1}$$

where $\tilde{\boldsymbol{E}}_{\boldsymbol{i}} \in R^{n\times d^{\mathrm{M}}}$ represents the ith group of MFs, and d^{M} is the dimension of features in each group. Assume the desired weights and features are $\boldsymbol{W}_i^{\mathrm{M}}$ and $\boldsymbol{E}_i$, similar to SAE, we optimize the following objective function:

$$\arg\min_{\boldsymbol{W}_i^{\mathrm{M}}} \left\|\boldsymbol{X}\boldsymbol{W}_i^{\mathrm{M}} - \boldsymbol{E}_i\right\|_2^2 + \lambda\left\|\boldsymbol{W}_i^{\mathrm{M}}\right\|_1 \tag{2}$$

where λ is the sparse regularization parameter. Suppose that two data points $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are close to each other in the original space, the corresponding MFs $\boldsymbol{e}_i$ and $\boldsymbol{e}_j$ should also be close to each other according to the manifold assumption. By adding this constraint to (2), we can get the objective function of GRSAE

$$\begin{aligned}&\arg\min_{\boldsymbol{W}_i^{\mathrm{M}}} \left\|\boldsymbol{X}\boldsymbol{W}_i^{\mathrm{M}} - \boldsymbol{E}_i\right\|_2^2 + \lambda\left\|\boldsymbol{W}_i^{\mathrm{M}}\right\|_1 + \alpha\sum_{i=1}^{n}\sum_{j=1}^{n}\boldsymbol{w}_{ij}\|\boldsymbol{e}_i - \boldsymbol{e}_j\|_2^2\\ &= \arg\min_{\boldsymbol{W}_i^{\mathrm{M}}} \left\|\boldsymbol{X}\boldsymbol{W}_i^{\mathrm{M}} - \boldsymbol{E}_i\right\|_2^2 + \lambda\left\|\boldsymbol{W}_i^{\mathrm{M}}\right\|_1 + \alpha\mathrm{tr}\left(\boldsymbol{E}_i{}^{\mathrm{T}}\boldsymbol{L}\boldsymbol{E}_i\right)\end{aligned} \tag{3}$$

where α is the graph regularization parameter of GRSAE, $\mathrm{tr}(\cdot)$ represents the trace operation, and $\boldsymbol{L}$ is the Laplace matrix that can be obtained by constructing a k-nearest neighbor graph. Equation (3) can be solved by the alternating direction method of multipliers (ADMM) [11] and an auxiliary variable $\boldsymbol{O}$ is introduced. Then, (3) can be rewritten as

$$\begin{aligned}&\arg\min_{\boldsymbol{W}_i^{\mathrm{M}}} \left\|\boldsymbol{X}\boldsymbol{W}_i^{\mathrm{M}} - \boldsymbol{E}_i\right\|_2^2 + \lambda\|\boldsymbol{O}\|_1 + \alpha\mathrm{tr}\left(\boldsymbol{E}_i{}^{\mathrm{T}}\boldsymbol{L}\boldsymbol{E}_i\right)\\ &\text{s.t.}\ \ \boldsymbol{W}_i^{\mathrm{M}} - \boldsymbol{O} = \boldsymbol{0}.\end{aligned} \tag{4}$$

The Lagrangian expression of (4) is

$$\begin{aligned}J = \arg\min_{\boldsymbol{W}_i^{\mathrm{M}}} &\left\|\boldsymbol{X}\boldsymbol{W}_i^{\mathrm{M}} - \boldsymbol{E}_i\right\|_2^2 + \lambda\|\boldsymbol{O}\|_1 + \alpha\mathrm{tr}\left(\boldsymbol{E}_i{}^{\mathrm{T}}\boldsymbol{L}\boldsymbol{E}_i\right)\\ &+\rho\boldsymbol{u}^{\mathrm{T}}\left(\boldsymbol{W}_i^{\mathrm{M}} - \boldsymbol{O}\right) + \frac{\rho}{2}\left\|\boldsymbol{W}_i^{\mathrm{M}} - \boldsymbol{O}\right\|_2^2\end{aligned} \tag{5}$$

where $\rho > 0$ is a constant. According to the ADMM optimization method, $\boldsymbol{W}_i^{\mathrm{M}}$, $\boldsymbol{O}$, and $\boldsymbol{u}$ are updated alternatively while keeping the other variables fixed.

1) Update $\boldsymbol{W}_i^{\mathrm{M}}$*:* The update process of $\boldsymbol{W}_i^{\mathrm{M}}$ can be carried out by solving the following function:

$$\boldsymbol{W}_i^{\mathrm{M}(k+1)} = \arg\min_{\boldsymbol{W}_i^{\mathrm{M}}} J\left(\boldsymbol{W}_i^{\mathrm{M}}, \boldsymbol{O}, \boldsymbol{u}\right). \tag{6}$$

Calculate the derivative of J about $\boldsymbol{W}_i^{\mathrm{M}}$ and set it to 0, we can get

$$\boldsymbol{W}_i^{\mathrm{M}(k+1)} = \frac{\boldsymbol{X}^{\mathrm{T}}\boldsymbol{E}_i - \rho(\boldsymbol{O}^{(k)} - \boldsymbol{u}^{(k)})}{\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} + \rho\boldsymbol{I} + \alpha\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{L} + \boldsymbol{L}^{\mathrm{T}})\boldsymbol{X}}. \tag{7}$$

2) Update $\boldsymbol{O}$*:* Here, a soft threshold operation $S_\kappa(\cdot)$ is used

$$S_\kappa(\upsilon) = \begin{cases} \upsilon - \eta, & \upsilon > \eta \\ 0, & |\upsilon| \le \eta \\ \upsilon + \eta, & \upsilon < -\eta \end{cases} \tag{8}$$

where η is the predefined threshold, such as 10^{-3}. Then, the update process of $\boldsymbol{O}$ is

$$\boldsymbol{O}^{(k+1)} = S_{\lambda/\rho}\left(\boldsymbol{W}_i^{\mathrm{M},(k+1)} + \boldsymbol{u}^{(k)}\right). \tag{9}$$

3) Update $\boldsymbol{u}$*:* The update formula for $\boldsymbol{u}$ is

$$\boldsymbol{u}^{(k+1)} = \boldsymbol{u}^{(k)} + \left(\boldsymbol{W}_i^{\mathrm{M},(k+1)} - \boldsymbol{O}^{(k+1)}\right). \tag{10}$$

The above-mentioned three steps are alternated until convergence or predetermined number of iterations is reached. Then, we can get the desired $\boldsymbol{W}_i^{\mathrm{M}}$, and $\boldsymbol{E}_i$ can be calculated by

$$\boldsymbol{E}_i = \boldsymbol{X}\boldsymbol{W}_i^{\mathrm{M}}. \tag{11}$$

The weights from MF to EN in the original BL are randomly generated, and the performance was affected by the random condition. Therefore, the stability of BL cannot be guaranteed. To address this issue, the weights from MF to EN are also fine-tuned with GRSAE, which can be easily realized by replacing $\boldsymbol{X}$ with $\boldsymbol{E} = [\boldsymbol{E}_1, \boldsymbol{E}_2, \ldots, \boldsymbol{E}_{G^{\mathrm{M}}}]$. Then, the fine-tuned weights $\boldsymbol{W}^{\mathrm{E}}$ can be obtained; furthermore, the features of EN can be calculated by

$$\boldsymbol{F} = \sigma(\boldsymbol{E}\boldsymbol{W}^{\mathrm{E}}) \tag{12}$$

where $\sigma(\cdot)$ is a nonlinear function, and tansig is selected here. $\boldsymbol{F} \in R^{n \times d^{\mathrm{E}}}$ are the features of EN, and d^{E} represents the number of nodes in EN.

B. UBL

For BL, after obtaining the features of MF and EN, the output can be calculated as follows:

$$\tilde{\boldsymbol{Y}} = [\boldsymbol{E}|\boldsymbol{F}]\tilde{\boldsymbol{W}}^m \tag{13}$$

where $\tilde{\boldsymbol{Y}} \in R^{n \times c}$ is the output vector, and $\tilde{\boldsymbol{W}}^m$ is the output-layer weights that can be got by solving the following problem:

$$\arg\min_{\tilde{\boldsymbol{W}}^m} \|[\boldsymbol{E}|\boldsymbol{F}]\tilde{\boldsymbol{W}}^m - \tilde{\boldsymbol{Y}}'\|_2^2 + \delta\|\tilde{\boldsymbol{W}}^m\|_2^2 \tag{14}$$

where $\tilde{\boldsymbol{Y}}' \in R^{n \times c}$ is the true label vector of samples, and c is the number of classes. Equation (14) can be solved by giving an approximation to the Moore–Penrose generalized inverse. In BL, $\delta \to 0$. Then, the solution is

$$\begin{aligned} \tilde{\boldsymbol{W}}^m &= [\boldsymbol{E}|\boldsymbol{F}]^+\tilde{\boldsymbol{Y}}' \\ &= \lim_{\delta \to 0}(\delta\boldsymbol{I} + [\boldsymbol{E}|\boldsymbol{F}][\boldsymbol{E}|\boldsymbol{F}]^{\mathrm{T}})^{-1}[\boldsymbol{E}|\boldsymbol{F}]^{\mathrm{T}}\tilde{\boldsymbol{Y}}' \end{aligned} \tag{15}$$

where $(\cdot)^+$ is the generalized inverse operation, and $\boldsymbol{I}$ is the identity matrix.

In the case of unsupervised clustering, there is no prior knowledge or label information. Therefore, when constructing the objective function of UBL, the empirical risk term should be removed from (14), and only the l_2-norm term is retained. In addition, in order to pass the manifold structure of original HSI data to the output layer, the graph regularization term needs to be added to the objective function. In summary, the model of UBL can be expressed as

$$\begin{aligned} &\arg\min_{\boldsymbol{W}^m} \frac{1}{2}(\|\boldsymbol{W}^m\|_2^2 + \zeta\mathrm{Tr}([\boldsymbol{E}|\boldsymbol{F}]^{\mathrm{T}}(\boldsymbol{W}^m)^{\mathrm{T}}\boldsymbol{L}^m\boldsymbol{W}^m[\boldsymbol{E}|\boldsymbol{F}])) \\ &\text{s.t. } (\boldsymbol{W}^m)^{\mathrm{T}}\boldsymbol{W}^m = \boldsymbol{I} \end{aligned} \tag{16}$$

where ζ is the graph regularization parameter of UBL, and $\boldsymbol{L}^m$ is the Laplace matrix constructed by $[\boldsymbol{E}|\boldsymbol{F}]$ through k-nearest neighbor graph. In addition, to avoid the rank deficit phenomenon of generalized eigenvalue decomposition, a constraint term $(\boldsymbol{W}^m)^{\mathrm{T}}\boldsymbol{W}^m = \boldsymbol{I}$ is added. Equation (16) can be solved by calculating the eigenvectors corresponding to the c smallest eigenvalue of the following function:

$$(\boldsymbol{I} + \zeta[\boldsymbol{E}|\boldsymbol{F}]^{\mathrm{T}}\boldsymbol{L}^m[\boldsymbol{E}|\boldsymbol{F}])\boldsymbol{W}^m = \xi[\boldsymbol{E}|\boldsymbol{F}]^{\mathrm{T}}[\boldsymbol{E}|\boldsymbol{F}]\boldsymbol{W}^m. \tag{17}$$

After obtaining $\boldsymbol{W}^m$, the output of UBL can be obtained by

$$\boldsymbol{Y} = [\boldsymbol{E}|\boldsymbol{F}]\boldsymbol{W}^m. \tag{18}$$

Furthermore, the clustering result $\boldsymbol{Y}^c$ can be obtained by applying spectral clustering on $\boldsymbol{Y}$.

III. Experiments

The proposed UBL is evaluated on three real hyperspectral data sets including Indian Pines, Pavia University, and Salinas. The Indian Pines data set was acquired in 1992 by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor in the Northwestern Agricultural Region of Indiana, with 0.4–2.45-μm spectral range, size of 145×145 pixels, and 220 bands. According to the setting in [6], a typical subset of $75 \times 82 \times 220$ is selected as the experimental data with six classes. The Pavia University data set was obtained from the Reflective Optics System Imaging Spectrometer (ROSIS) sensors at Pavia University Italy in 2003, with 0.43–0.86-μm spectral range, size of 610×340 pixels, and 103 spectral bands. A typical subset of $170 \times 160 \times 103$ is selected as the experimental data, with nine classes [6]. The Salinas data set was acquired by the AVIRIS sensor over the Salinas Valley, with 512×217 pixels, 204 bands (after removing 20 water absorption bands). A typical subset of $140 \times 150 \times 204$ is selected as the experimental data, with six classes [6]. The ground-truth maps of the above-mentioned data sets are shown in Fig. 2.

A. Parameter Settings

According to the description of UBL, the adjustable parameters are: regularization parameters λ, α, and ζ, number of groups of MF G^{M}, number of nodes in MF per group d^{M}, and number of nodes in EN d^{E}. These parameters are settled by trial and error. Given $d^{\mathrm{M}} = G^{\mathrm{M}} \in \{3, 5, 10, 15, 20\}$, $d^{\mathrm{E}} \in \{3^2, 5^2, 10^2, 15^2, 20^2\}$, $\lambda = \alpha \in \{0.01, 0.1, 1, 5, 10\}$,

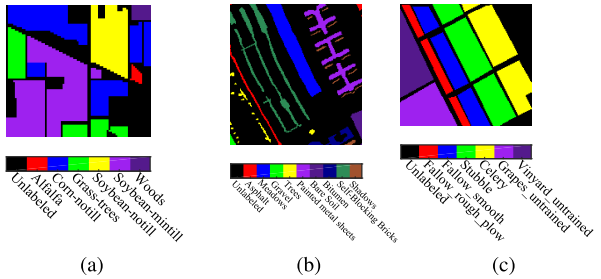

Fig. 2. Ground-truth maps of HSI data sets. (a) Indian Pines. (b) Pavia university. (c) Salinas.

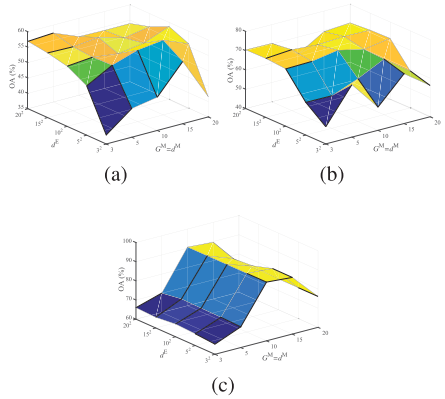

Fig. 3. OA versus G^{M}, d^{M}, and d^{E}. (a) Indian Pines. (b) Pavia university. (c) Salinas.

and $\zeta \in \{0.01, 0.1, 1, 5, 10\}$, the relationship between the overall accuracy (OA) and these parameters is shown in Figs. 3 and 4. It is demonstrated as follows.

1) On the one hand, small d^{M} and G^{M} means that the dimension of features in MF is low; thus, the input HSI data cannot be fully expressed. On the other hand, large d^{M} and G^{M} may result in redundant features in MF. Therefore, in the following experiments, we set $d^{\mathrm{M}} = G^{\mathrm{M}} = 15$ for the Indian Pines data set and $d^{\mathrm{M}} = G^{\mathrm{M}} = 10$ for both Pavia University and Salinas data sets.
2) When d^{E} is set to 25, 100, and 100 on the three HSI data sets, respectively, OA gets the highest.
3) In order to achieve higher OA, we set $\lambda = \alpha = 1$ and $\zeta = 10$ for the Indian Pines data set, $\lambda = \alpha = 10$ and $\zeta = 0.01$ for the Pavia University data set, and $\lambda = \alpha = 0.1$ and $\zeta = 0.01$ for the Salinas data set.

B. Comparative Experiments

To demonstrate the clustering performance of the proposed UBL, the following nine methods, which are classified into four types, are investigated for comparison.

1) Traditional clustering methods: fuzzy C-means (FCM) [2] and SSC [4]. In addition, in order to realize spectral–spatial clustering, HGF-FCM and HGF-SSC are also selected for comparison, where the original HSI data are preprocessed with HGF.
2) Spectral–spatial clustering methods: SSC-S [6], S4C [6], and L2-SSC [3].
3) Unsupervised extreme learning machine (UELM) [12].
4) UBL without graph regularization terms (UBL1).

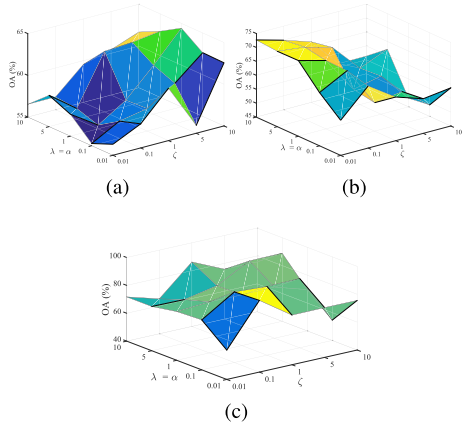

Fig. 4. OA versus λ, α, and ζ. (a) Indian Pines. (b) Pavia university. (c) Salinas.

The input of UBL1 is HGF-filtered HSI data, and the weights of MF and EN are optimized with SAE, but no graph regularization terms are added.

The proposed UBL and the above-mentioned nine comparative methods are performed on the three HSI data sets for clustering, respectively. Due to the stochastic characteristic in the experiments, we repeat each experiment five times to get the average value. Table I gives the comparison of clustering performance on the three data sets, where three evaluating indexes, including OA (%), Kappa coefficient, and consumed time (t, s), are considered. Fig. 5 shows the clustering maps on the Indian Pines data set by using different clustering methods.

It can be observed from Table I and Fig. 5 that:

1) Compared with SSC-based methods such as SSC, HGF-SSC, SSC-S, S4C, and L2-SSC, UBL yields higher OA and Kappa coefficient. The reasons are twofold: on the one hand, all of the SSC-based methods are linear, which cannot exploit the nonlinear structure of HSI, where UBL is the nonlinear method, which can achieve more complex mapping than linear methods. On the other hand, by incorporating the manifold regularization technique, UBL can preserve the manifold structure of input data to the output.
2) Compared with nonlinear methods such as UELM and UBL1, UBL shows the superiority on OA and Kappa coefficient. This is because the manifold regularization technique is only used in the construction of the final objective function in UELM. In addition, UBL can realize width expanding while the structure of UELM is fixed. Furthermore, UBL benefits from the appropriate graph regularization technique; thus, the manifold structure can be preserved after several mapping processes.
3) Among the ten methods, spectral–spatial clustering methods (HGF-FCM, HGF-SSC, SSC-S, S4C, L2-SSC, UELM, UBL1, and UBL) can achieve higher OAs and Kappa coefficients than spectral-based clustering methods (FCM and SSC). This is because samples with different classes may have the same or similar spectral characteristics. If only the spectral information is taken into consideration, more misclassification may occur. Fortunately, samples shared similar spectral

TABLE I
COMPARISON OF CLUSTERING PERFORMANCE

Dataset	Index	FCM [2]	HGF -FCM	SSC [4]	HGF -SSC	SSC-S [6]	S4C [6]	L2-SSC [3]	UELM [12]	UBL1	UBL
Indian Pines	OA (%)	46.91	52.27	48.83	51.73	49.89	50.05	51.19	56.89	52.30	**62.58**
	Kappa	0.3436	0.3682	0.3400	0.4134	0.3727	0.3740	0.3944	0.4362	0.4175	**0.4690**
	t (s)	5.74	**3.29**	445.02	231.71	1261.13	3024.90	718.50	77.09	94.50	104.90
Pavia University	OA (%)	47.35	54.17	53.67	59.59	69.62	68.39	61.84	61.17	64.93	**70.83**
	Kappa	0.3986	0.4658	0.4539	0.5151	0.6322	0.6210	0.5412	0.5473	0.5892	**0.6533**
	t (s)	6.07	**4.57**	630.74	650.28	42104.05	45995.34	22520.97	236.50	226.34	237.44
Salinas	OA (%)	72.32	74.43	72.76	85.38	86.71	86.72	86.51	80.30	75.07	**91.42**
	Kappa	0.6635	0.6894	0.6667	0.8208	0.8359	0.8360	0.8334	0.7621	0.6952	**0.8943**
	t (s)	16.38	**14.26**	10185.87	9760.61	12638.15	22992.69	2311.99	3403.52	2492.91	2509.68

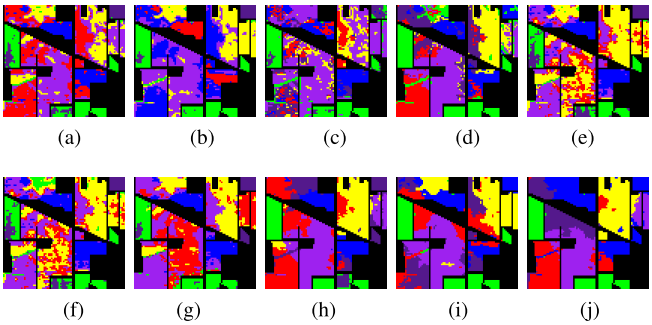

Fig. 5. Clustering maps on Indian Pines data set. (a) FCM. (b) HGF-FCM. (c) SSC. (d) HGF-SSC. (e) SSC-S. (f) S4C. (g) L2-SSC. (h) UELM. (i) UBL1. (j) UBL.

characteristics may have different spatial information. Therefore, by introducing spatial information, the discriminant ability of methods can be enhanced.

4) The least time-consuming methods are FCM and HGF-FCM. Excluding these two methods, UELM and UBL1 have shorter consumed time comparatively. UBL consumed longer time than ULEM and UBL1, but it can get higher OAs than UELM and UBL1.
5) Taking the Indian Pines data set as an example, the clustering map obtained by UBL is smoother and more detailed. On the other hand, other methods may misclassify more Alfalfa into Corn-notill, Soybean-notill, or Soybean-mintill, more Soybean-notill into Corn-notill or Soybean-mintill, and more Soybean-mintill into Corn-notill or Grass-trees.

IV. CONCLUSION

A novel HSI clustering method, named UBL, was proposed in this letter. First, by incorporating the graph regularization technique into the weights fine-tuning process of MF, the manifold structure of the input data can be maintained into the features of the MF. Then, by adopting GRSAE to fine-tune the EN weights, the influence of randomness on UBL clustering performance is avoided, and the stability of UBL is enhanced. Furthermore, both MF and EN are connected to the output layer with the output layer weights. By combining the l_2-norm regularization term and graph regularization term in the objective function, the manifold structure of the input data is maintained to the output layer. The objective function can be easily obtained by solving the generalized eigenvalue decomposition problem. Finally, the clustering results can be obtained by performing spectral clustering on the output vectors of UBL. Experimental results on three real hyperspectral image data sets demonstrate that the proposed method can obtain higher clustering accuracy than several commonly used clustering methods such as FCM-based, SSC-based, and UELM. Since anomaly detection is a very promising and meaningful application in HSI areas [13], extending the UBL to HSI anomaly detection is our future work.

REFERENCES

[1] L. Zhang, L. Zhang, and B. Du, "Deep learning for remote sensing data: A technical tutorial on the state of the art," *IEEE Geosci. Remote Sens. Mag.*, vol. 4, no. 2, pp. 22–40, Jun. 2016.
[2] A. G. S. Filho *et al.*, "Hyperspectral images clustering on reconfigurable hardware using the k-means algorithm," in *Proc. 16th Symp. Integr. Circuits Syst. Design (SBCCI)*, Sep. 2003, pp. 99–104.
[3] H. Zhai, H. Zhang, L. Zhang, P. Li, and A. Plaza, "A new sparse subspace clustering algorithm for hyperspectral remote sensing imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 1, pp. 43–47, Jan. 2017.
[4] W. Sun, L. Zhang, B. Du, W. Li, and Y. M. Lai, "Band selection using improved sparse subspace clustering for hyperspectral imagery classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2784–2797, Jun. 2015.
[5] Q. Yan, Y. Ding, J.-J. Zhang, Y. Xia, and C. H. Zheng, "A discriminated similarity matrix construction based on sparse subspace clustering algorithm for hyperspectral imagery," *Cogn. Syst. Res.*, vol. 53, pp. 98–110, Jan. 2019.
[6] H. Zhang, H. Zhai, L. Zhang, and P. Li, "Spectral–spatial sparse subspace clustering for hyperspectral remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3672–3684, Jun. 2016.
[7] C. L. P. Chen and Z. L. Liu, "Broad learning system: An effective and efficient incremental learning system without the need for deep architecture," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 1, pp. 10–24, Jan. 2018.
[8] Z. Liu and C. L. P. Chen, "Broad learning system: Structural extensions on single-layer and multi-layer neural networks," in *Proc. Int. Conf. Secur., Pattern Anal., Cybern. (SPAC)*, Dec. 2017, pp. 136–141.
[9] Y. Kong, X. Wang, Y. Cheng, and C. L. P. Chen, "Hyperspectral imagery classification based on semi-supervised broad learning system," *Remote Sens.*, vol. 10, no. 5, p. 685, Apr. 2018.
[10] B. Pan, Z. Shi, and X. Xu, "Hierarchical guidance filtering-based ensemble classification for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 4177–4189, Jul. 2017.
[11] J. F. C. Mota, J. W. F. Xavier, P. M. Q. Aguiar, and M. Püschel, "D-ADMM: A communication-efficient distributed algorithm for separable optimization," *IEEE Trans. Signal Process.*, vol. 61, no. 10, pp. 2718–2723, May 2013.
[12] G. Huang, S. Song, J. N. D. Gupta, and C. Wu, "Semi-supervised and unsupervised extreme learning machines," *IEEE Trans. Cybern.*, vol. 44, no. 12, pp. 2405–2417, Dec. 2017.
[13] X. Kang, X. Zhang, S. Li, K. Li, J. Li, and J. A. Benediktsson, "Hyperspectral anomaly detection with attribute and edge-preserving filters," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 10, pp. 5600–5611, Oct. 2017.