Multiple features learning via rotation strategy

Junshi Xia, Lionel Bombrun, Yannick Berthoumieu, Christian Germain

MULTIPLE FEATURES LEARNING VIA ROTATION STRATEGY

Junshi Xia, Lionel Bombrun, Yannick Berthoumieu,and Christian Germain

Université de Bordeaux and CNRS, IMS, UMR 5218, F-33405 Talence, France

ABSTRACT

Images are usually represented by different groups of features, such as color, shape and texture attributes. In this paper, we propose a classification approach that integrates multiple features, such as spectral and spatial information. We refer this approach to multiple feature learning via rotation (MFL-R) strategy, which adopt a rotation-based ensemble method by using a data transformation approach. Five data transformation methods, including principal component analysis (PCA), neighborhood preserving embedding (NPE), linear local tangent space alignment (LLTSA), linearity preserving projection (LPP) and multiple feature combination via manifold learning and patch alignment (MLPA) are used in the MFL-R framework. Experimental results over two hyperspectral remote sensing images demonstrate that MFL-R with MLPA gains better performances and is not sensitive to the tuning parameters.

Index Terms— Multiple feature learning, Rotation strategy, Ensemble learning, Classification

1. INTRODUCTION

Supervised classification is a task of primary importance for a wide range of practical applications in various image datasets (e.g., remote sensing and medical images) [1–4]. Recently, ensemble learning or multiple classifier systems (MCSs) have shown to be of the great capacity in improving the classification performances [5–7]. One of the current state-of-the-art MCSs method is the Rotation Forest (RoF), which is an ensemble of decision trees (DTs) [8]. RoF applies data transformation technique (i.e., principal component analysis) into the disjoint subsets split by the original features to obtain the sparse rotation matrix, which is used to create the new features of DT. Since the sparse rotation matrix is very diverse among the different DTs, RoF improves both accuracy and diversity within the ensemble [8]. Although RoF provides excellent performance in the previous studies [8–12], they focus only on a single kind of feature.

Supervised classification with only one kind of features might not be sufficient to obtain reliable results. Instead, the combination of multiple features (e.g., texture and shape attributes) can contribute to a more comprehensive interpretation of the scene. For instance, texture of signatures cannot differentiate between objects made of the same materials (e.g. roofs and roads made with the same asphalt), while they may be easily distinguished by their shapes [13].

Two well-known strategies to combine multiple features are the parallel and concatenated combinations [14, 15]. The former fuses the results generated by the independent classifiers with different features, while the latter integrates the features into one vector and then obtains the result via a classifier. The parallel combination keeps low the dimensionality of the input features. However, its performance depends on the diversity among the features. In the concatenated combination, stacking the features may yield to redundant information, making it difficult to select an optimal combination of features. Furthermore, the high dimensionality of the stacked features, as well as the limited number of training samples, provokes the curse of dimensionality, thus resulting in lower classification accuracy [14, 15].

In order to alleviate these problems, Fauvel *et al.* [16] and Ghamisi *et al.* [17] applied supervised dimensionality reduction techniques on both spectral and spatial features, and then stacked the reduced spectral and spatial features together. However, supervised dimensionality reduction methods suffer from the limited number of available training samples. To overcome this, Liao *et al.* [13] proposed a joint data fusion and dimension reduction, namely graph-based data fusion (GDF), for the classification of multi-sensors data. Zhang *et al.* [18] developed a unified framework based on manifold learning and patch alignment (MLPA) to linearly combine multiple features. However, the number of extracted features in [13] and [18] is hard to determine, which needs to be further investigated.

Inspired by the idea of the RoF, the main contribution of the paper is to propose a robust classification framework that combines multiple features based on rotation-based ensemble. We refer it to multiple feature learning via rotation strategy (MFL-R). In this framework, we first split different kinds of features into several disjoint subsets and apply the data transformation method to each subset. Five data transformation methods, including principal component analysis (PCA), neighborhood preserving embedding (NPE) [19], linear local tangent space alignment (LLTSA) [20], linearity preserving projection (LPP) [21] and multiple features combination via manifold learning and patch alignment (MLPA) [18] are considered. Second, new training set for the decision tree (DT)

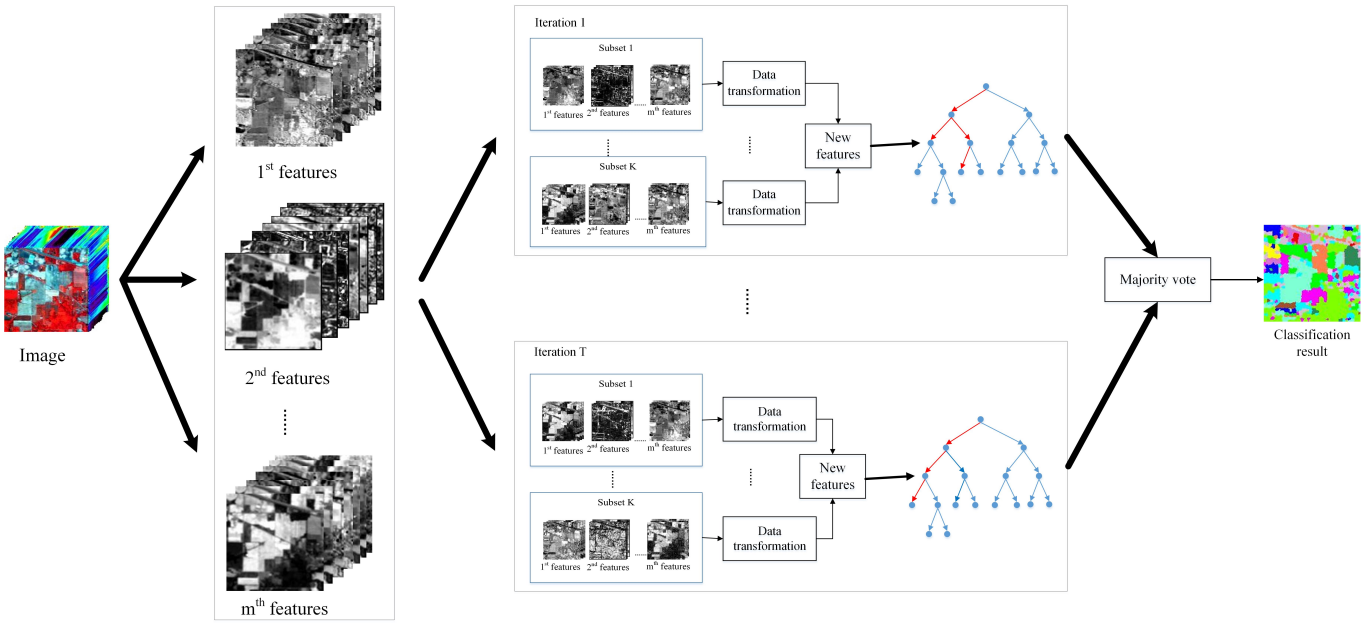

Fig. 1. Flowchart of the proposed framework.

classifier are formed by concatenating the extracted features contained in each subset. Furthermore, two hyperspectral remote sensing images are used to validate the performances of the proposed approach. Experiments suggest the effectiveness of the proposed framework, especially with MLPA.

The remaining of this paper is organized as follows. The proposed MFL-R is described in Section 2. Section 3 presents an application to hyperspectral remote sensing image classification. Conclusions and perspectives are drawn in Section 4.

2. PROPOSED CLASSIFICATION FRAMEWORK

The proposed MFL-R framework is based on the rotation-based ensemble that aims at generating diverse individual classification results using random feature selection and data transformation technique, which improves individual accuracy and diversity within the ensemble simultaneously [8–10].

Let $\mathbf{X} = \left\{\mathbf{X}^j\right\}_{j=1}^{m} \left(\mathbf{X}^j = \left\{\mathbf{x}_i^j\right\}_{i=1}^{n}\right)$ denote m kinds of features of training samples with the corresponding label $\mathbf{Y} = \{y_i\}_{i=1}^{n}$, where $\mathbf{x}_i^j \in \mathbb{R}^{D^j}$ and $y_i \in \{1, ..., C\}$ denotes the label information, where C is the total number of classes.

The steps of MFL-R (as shown in Fig. 1) are detailed in the following.

- The first step consists in splitting each kind of feature space into K disjoint subsets. A subset of each kind of features contain $\lfloor D^j/K \rfloor, j = 1, .., m$ features. In this case, there are different features in each subset, resulting in generating different new features by using data transformation in the next step.
- In the second step, the data transformation is applied to each subset to create new features. This step aims at producing discriminant and decorrelate features, which are beneficial with the classifier. In this work, PCA, LLTSA, LPP, NPE and MLPA are considered.
- In the third step, the new features are formed by concatenating the extracted components (the size of d) in each subset and then are used to train an individual DT classifier. Classification and regression tree (CART) with *Gini* index as the split function is considered.
- The final result is produced by integrating the individual DT classifiers that are generated by repeating the above steps T times.

Since PCA, LLTSA [20], LPP [21] and NPE [19] are well-known approaches, interested reader can find more details in their original reference. In the following, we give the general description of MLPA. MLPA aims at preserving the local geometric properties of the features and produces the new features by simultaneously optimizing the weights in the following objective function constructed by multiple features [18].

$$\arg\min_{\mathbf{Z},\omega} \sum_{i=1}^{m} \omega_i^r \mathrm{tr}\left(\mathbf{Z}\mathbf{M}_i\mathbf{Z}^\top\right)$$

$$s.t. \, \omega_i > 0, \sum_{i=1}^{m} \omega_i = 1 \quad (1)$$

$$\mathbf{M} = \mathbf{Q} - \mathbf{W} \quad (2)$$

$$\mathbf{Q}_{ii} = \sum_{j=1}^{n} \mathbf{W}(i,j) \quad (3)$$

Table 1. Overall, average, κ and class-specific accuracies obtained for the Indian Pines AVIRIS image

Class	No of Samples		Single feature			Combination		MFL-R				
	Train	Test	Spec	SG	EMPs	Concatenated	Parallel	PCA	LPP	LLTSA	NPE	MLPA
Alfalfa	10	44	71.67	67.04	89.63	88.70	89.63	92.78	93.89	92.41	93.52	**95.74**
Corn-no till	10	1514	27.75	20.31	62.20	57.36	64.27	64.06	65.70	**68.69**	67.79	66.70
Corn-min till	10	824	26.86	26.75	63.25	58.05	52.88	71.13	70.25	68.74	70.29	**75.22**
Bldg-Grass-Tree-Drives	10	224	36.41	41.28	84.10	72.22	62.35	89.79	93.63	93.68	91.58	**94.02**
Grass/pasture	10	487	55.79	39.13	83.32	74.97	77.18	83.58	82.31	82.07	80.10	**84.39**
Grass/trees	10	737	44.94	46.93	76.12	73.43	67.82	**89.29**	82.85	88.30	82.05	82.30
Grass/pasture-mowed	10	16	75.77	79.62	92.69	94.23	88.46	97.69	99.23	**99.62**	98.08	98.08
Corn	10	479	52.35	44.81	89.65	87.71	66.99	96.50	98.20	98.75	**99.02**	98.73
Oats	10	10	83.00	82.00	95.50	93.00	94.50	**100.00**	98.50	99.50	**100.00**	99.50
Soybeans-no till	10	958	34.16	23.26	68.21	60.96	42.60	74.40	72.13	73.14	71.67	**76.25**
Soybeans-min till	10	2458	28.58	16.34	64.04	62.22	29.57	64.47	68.60	70.27	66.97	**73.15**
Soybeans-clean till	10	604	17.67	20.59	**73.66**	68.29	26.47	65.60	66.14	66.27	71.21	70.33
Wheat	10	202	82.97	58.54	91.37	91.60	87.74	96.13	98.58	**99.20**	99.01	99.10
Woods	10	1284	72.70	46.29	74.38	71.40	70.73	**90.53**	88.65	86.17	89.68	**90.53**
Hay-windrowed	10	370	37.39	35.47	87.21	85.08	51.79	92.82	95.24	**96.05**	92.08	95.71
Stone-steel towers	10	85	92.63	75.79	96.42	95.16	94.63	**98.53**	96.53	97.26	96.63	97.16
Overall accuracy (OA)			39.96	30.23	71.40	67.46	53.15	76.28	76.78	77.71	76.78	**79.40**
Average accuracy (AA)			52.54	45.26	80.74	77.15	66.73	85.46	85.65	86.26	85.60	**87.31**
kappa coefficients (κ)			33.70	23.77	68.01	63.65	48.15	73.40	73.86	74.90	73.90	**76.83**

Table 2. OAs, AAs, κ obtained for the Salinas AVIRIS image.

	Single feature			Combination		MFL-R				
	Spec	SG	EMPs	Concatenated	Parallel	PCA	LPP	LLTSA	NPE	MLPA
OA	73.94	68.71	82.40	75.69	81.64	87.19	89.56	88.41	88.19	**91.87**
AA	83.78	76.58	90.18	83.13	87.80	92.21	94.08	93.42	92.96	**95.49**
κ	71.90	66.62	80.28	75.18	79.29	86.51	88.48	87.87	86.77	**91.22**

$$\mathbf{W}(i,j) = \exp(-\left\|\mathbf{x}_i - \mathbf{x}_j\right\|^2 / t) \tag{4}$$

where, ω_i and $\mathbf{M}_i$ are respectively the weight and the alignment matrix of the ith kind of features. m is the number of kinds of features. $r > 1$ is a control parameter. t is a radius parameter.

Since it is impossible to find a global optimal solution of $\mathbf{Z}$ and ω in (1), an alternating optimization is used to obtain a local optimal solution by iteratively updating $\mathbf{Z}$ and ω [18]. Thus, $\mathbf{Z}$ is calculated by initializing the same weights to each kind of feature and repeating updating $\mathbf{Z}$ and ω until convergence. Then, $\mathbf{V}$ is achieved by using the linear regression model

$$\mathbf{V} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Z} \tag{5}$$

In this work, the proposed MFL-R framework with PCA, LLTSA, LPP, NPE and MLPA data transformation are named respectively MFL-R-PCA, MFL-R-LLTSA, MFL-R-LPP, MFL-R-NPE and MFL-R-MLPA.

3. APPLICATIONS TO HYPERSPECTRAL IMAGE CLASSIFICATION

In this section, two hyperspectral remote sensing images are used to evaluate the proposed framework. They are recorded by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor over the Indian Pines in Northwestern Indiana and Salinas Valley, California, USA. The former is composed of 145 $\times$ 145 pixels with 220 spectral bands (spatial resolution: 20 m), while the latter comprises 512 $\times$ 217 samples with 224 spectral bands (spatial resolution: 3.7 m)

In the proposed framework, three kinds of features, including spectral, spectral gradient (SG) and extended morphological profiles (EMPs) [22] are utilized. SG is a surface reflectance descriptor which is obtained by subtracting the neighbor features. EMPs is formed by stacking all the computed morphological profiles (MPs), which perform opening and closing by reconstruction with different sizes of structural element (SE) on the first few components extracted from the image [22]. For Indian Pines, the dimensionality of spectral, SG and EMPs are 220, 219 and 183. For Salinas, the dimensionality of spectral, SG and EMPs are 224, 223 and 204.

Ten samples per class are randomly selected to form the training set and the rest of the pixels are used for testing. It is emphasized that the classification problem is very challenging with a very limited number of training samples. The results are obtained after ten repetitions. In this experiment, the number of classifiers (T), subsets (K) and extracted features in each subset (d) are set to be 20, 5 and 10, respectively. Tables 1 and 2 presents the classification accuracies of the proposed method, as well as the ones obtained for the single kind of features, parallel and concatenated combination. In particular, parallel combination integrates the classification results obtained for three kinds of features. Concatenated combination stacks the three kinds of features together and then classifies via the DT classifier. It can be seen from the two

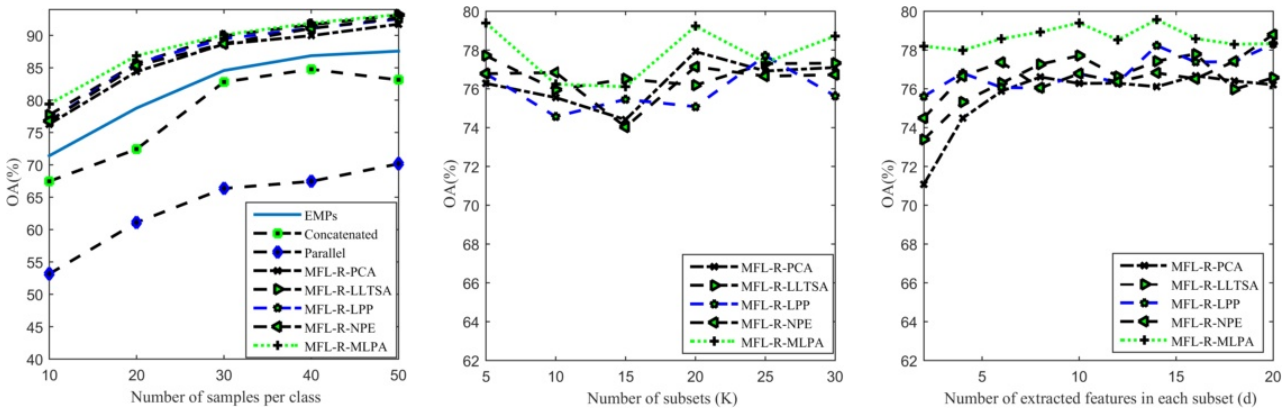

Fig. 2. Indian Pines AVIRIS image. Effects of (a) different numbers of training samples. (b) Number of subsets (K). (c) Number of components (d) kept in each subset.

Table 3. OAs (in Percent), AOAs (in Percent) and Diversities obtained for Indian Pines AVIRIS Image

Methods	MFL-R				
	PCA	LLTSA	LPP	NPE	MLPA
OA	76.28	76.78	77.71	76.78	**79.40**
AOA	68.71	68.87	69.01	68.52	**70.47**
Diversity	0.51	0.51	0.53	0.52	**0.54**

tables that the parallel and concatenated combination cannot improve the classification performance and gain worse performance than EMPs since they suffer from the limited number of training samples. The proposed method reports significantly better performance than other approaches. MFL-R-LLTSA, MFL-R-LPP and MFL-NPE perform slightly better than MFL-R-PCA because they consider local information in the data transformation process [10]. MFL-R-MLPA yields to the best results due to the fact that MLPA considers the statistical properties of each feature to achieve a physically meaningful unified low-dimensional representation of multiple features, which is beneficial for improving the performance [18]. Moreover, we use the percentage average overall accuracies of the individual DT classifier, "AOA (%)", and the *coincident failure diversity (CFD)* [14] to analyze the proposed method (seen in Table 3). A higher value of *CFD* means a stronger ensemble. MFL-R-MLPA outperforms MFL-R with other data transformation approaches due to the higher accuracies and diversity among the member classifiers.

The last part of our experiments focuses on the sensitivity analysis of the number of training samples and the tuning parameters in the proposed method. Fig. 2(a) shows the effect of the number of training samples. For the minor classes (Grass/pasture-mowed and Oats), the number of training samples remains constant (10 samples). From Fig. 2(a), the overall accuracy increases monotonically as the size of the training set increases and the MFL-R-MLPA gives the best results in all cases. The number of subsets (K) and extracted features in each subset (d) are the key parameters of the proposed framework. The effects of K and d are depicted in Fig. 2(b) and (c). There is no pattern of dependency between K and the accuracy. Larger values of d result in higher accuracies of MFL-R-PCA, MFL-R-LLTSA, MFL-R-LPP and MFL-R-NPE. However, MFL-R-MLPA is stable with this parameter. Hence, selection of parameters is not very critical for the MFL-R-MLPA, which is an important added advantage.

4. CONCLUSIONS AND PERSPECTIVES

In this paper, we propose a novel classification framework with multiple features, namely multiple feature learning via rotation strategy (MFL-R). The main idea of MFL-R is based on rotation-based ensemble with data transformation. Five data transformation technqiues, including PCA, LLTSA, LPP, NPE and MLPA are considered. Experimental results on two hyperspectral remote sensing images show that the proposed method leads to significant improvement than traditional parallel and concatenated combination. Among them, MFL-R-MLPA achieves the best performances and is not sensitive to the parameters. Thus, we promote the use of MLPA in the proposed MFL-R framework.

In the future, we would like to further investigate the potential of MFL-R to classify other kinds of features.

5. ACKNOWLEDGMENT

The authors would like to thank Prof D. Landgrebe from Purdue University for providing the data set. This study has been carried out with financial support from the French State, managed by the French National Research Agency (ANR) in the frame of the "Investments for the future" Programme IdEx Bordeaux-CPU (ANR-10-IDEX-03-02).

6. REFERENCES

[1] D. A. Landgrebe, *Signal Theory Methods in Multispectral Remote Sensing*, New York: Wiley, 1984.

[2] W. Hu, R. Hu, N. Xie, H. Ling, and S. Maybank, "Image classification using multiscale information fusion based on saliency driven nonlinear diffusion filtering," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1513–1526, 2014.

[3] P. Foggia, G. Percannella, P. Soda, and M. Vento, "Benchmarking HEp-2 cells classification methods," *IEEE Trans. Med. Imag.*, vol. 32, no. 10, pp. 1878–1889, 2013.

[4] I. Hedhli, G. Moser, J. Zerubia, and S.B. Serpico, "New cascade model for hierarchical joint classification of multitemporal, multiresolution and multisensor remote sensing data," in *International Conference on Image Processing (ICIP)*, 2014, pp. 5247–5251.

[5] J. A. Benediktsson, J. Chanussot, and M. Fauvel, "Multiple classifier systems in remote sensing: from basics to recent developments," in *Proceedings of the 7th International Workshop on Multiple Classifier Systems, Prague, Czech Republic, May 23-25*, 2007, pp. 501–512.

[6] Y. Su, S. Shan, X. Chen, and W. Gao, "Hierarchical ensemble of global and local classifiers for face recognition," *IEEE Trans. Image Process.*, vol. 18, no. 8, pp. 1885–1896, 2009.

[7] S. Boukir, G. Li, and N. Chehata, "Classification of remote sensing data using margin-based ensemble methods," in *IEEE International Conference on Image Processing (ICIP)*, 2013, pp. 2602–2606.

[8] J. J. Rodriguez, L.I. Kuncheva, and C.J. Alonso, "Rotation forest: A new classifier ensemble method.," *IEEE Trans. Pattern Anal. Mach. Intell*, vol. 28, no. 10, pp. 1619–1630, 2006.

[9] J. Xia, P. Du, X. He, and J. Chanussot, "Hyperspectral remote sensing image classification based on rotation forest," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 1, pp. 239 – 243, 2014.

[10] J. Xia, J. Chanussot, P. Du, and X. He, "Spectral-spatial classification for hyperspectral data using rotation forests with local feature extraction and Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2532–2546, 2015.

[11] T. Kavzoglu, I. Colkesen, and T. Yomralioglu, "Object-based classification with rotation forest ensemble learning algorithm using very-high-resolution Worldview-2 image," *Remote Sens. Letters*, vol. 6, no. 11, pp. 834–843, 2015.

[12] P. Du, A. Samat, B. Waske, S. Liu, and Z. Li, "Random forest and rotation forest for fully polarized SAR image classification using polarimetric and spatial features," *ISPRS J. Photogramm Remote. Sens.*, vol. 105, pp. 38 – 53, 2015.

[13] W. Liao, R. Bellens, A. Pizurica, S. Gautama, and W. Philips, "Combining feature fusion and decision fusion for classification of hyperspectral and lidar data," in *IEEE Geoscience and Remote Sensing Symposium, Quebec City, QC, Canada, July 13-18*, 2014, pp. 1241–1244.

[14] L.I. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*, Wiley-Interscience, 2004.

[15] P. Du, J. Xia, W. Zhang, K. Tan, Y. Liu, and S. Liu, "Multiple classifier system for remote sensing image classification: A review," *Sensors*, vol. 12, no. 4, pp. 4764–4792, 2012.

[16] M. Fauvel, J. A. Benediktsson, J. Chanussot, and J. R. Sveinsson, "Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 11, pp. 3804–3814, 2008.

[17] P. Ghamisi, J. A. Benediktsson, and J. R. Sveinsson, "Automatic spectral-spatial classification framework based on attribute profiles and supervised feature extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 9, pp. 5771–5782, 2014.

[18] 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, 2012.

[19] X. He and P. Niyogi, "Locality preserving projections," in *Advances in Neural Information Processing Systems, MIT Press.*, 2003.

[20] T. Zhang, J. Yang, D. Zhao, and X. Ge, "Linear local tangent space alignment and application to face recognition.," *Neurocomputing*, vol. 70, no. 7-9, pp. 1547–1553, 2007.

[21] X. He, D. Cai, S. Yan, and H. Zhang, "Neighborhood preserving embedding," in *Proceedings of IEEE International Conference on Computer Vision (ICCV) , 17-20 October, Beijing, China*, 2005, pp. 1208–1213.

[22] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE T. Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 480–491, 2005.