A NOVEL ENSEMBLE CLASSIFIER OF HYPERSPECTRAL AND LIDAR DATA USING MORPHOLOGICAL FEATURES

Junshi Xia, Naoto Yokoya, and Akira Iwasaki

Research Center for Advanced Science and Technology, The University of Tokyo,
4-6-1 Komaba, Meguro-ku, Tokyo 153-8904, Japan

ABSTRACT

Due to the benefits and limitation of different remote sensing sensors, fusion of the features from multiple sensors, such as hyperspectral and light detection and ranging (LiDAR) is an effective method for land cover mapping. In this paper, we propose a novel ensemble classifier to fuse hyperspectral and LiDAR datasets for classification. First, morphological features are used to model spatial and elevation information from the first few principal components (PCs) of the original hyperspetcral (HS) image and LiDAR data. Second, we split different kinds of features (i.e., spectral bands, morphological features of hyperspectral and LiDAR), into several disjoint subsets and apply the data transformation method to each subset. In particular, three data transformation methods, including principal component analysis (PCA), linearity preserving projection (LPP) and unsupervised graph fusion (UGF) are considered. Third, the features extracted in each subset are concatenated to classify by a random forest (RF) classifier. Experimental results on a co-registered HS and LiDAR data provide the effectiveness and potentialities of the proposed ensemble classifier.

***Index Terms*—** Ensemble classifier, morphological features, hyperspectral, LiDAR

1. INTRODUCTION

Recently, the fusion of information derived from different sensors, such as hyperspectral (HS), multispectral (MS) and LiDAR, provides a better understanding of the same area, when compared to a single sensor [1]. For instance, LiDAR data provides the height information of different objects [2], whereas hyperspectral imaging acquires hundreds or thousands of narrow spectral bands, which give a high discrimination capacity between the various land cover classes [3, 4]. Thanks to the complementary information provided by MS, HS and LiDAR data, many promising techniques are proposed to fuse these datasets in a classification task [5–10]. A natural and straightforward way is to stack elevation information of LiDAR as additional features to spectral bands from optical sensors. A typical example is investigated in [5, 8] with the aim of classification of complex forest. They indicated that LiDAR can distinguish different classes with similar spectral signatures. Lemp and Weidner [6] used the LiDAR data to generate the segmentation map of the scene and then classified the segmentation regions of the HS data. Pedergnana *et al.* [7] computed extended attribute profiles (EAPs) for both HS (MS) and LiDAR, and then stacked them with spectral and elevation information for the classification of a rural and an urban area. Gu *et al.* [9] proposed a novel multiple-kernel learning (MKL) model for urban classification to integrate heterogeneous features from MS and LiDAR data. Liao *et al.* [10–12] proposed a series of graph-based dimensionality reduction for the classification of HS and LiDAR data with morphological features.

From the preceding literature review, it can be seen that three strategies are often adopted to fuse hyperspetral and LiDAR data:

- simply concatenating several kinds of feature sources (i.e., spectral, spatial and elevation information).
- applying dimensionality reduction techniques to the stacked features.
- integrating the features from HS and LiDAR data to formulate the multiple kernels for kernel-based methods.

However, stacking several kinds of feature sources contains redundant information and increase the dimensionality as well as the limited training samples, leading to a dissatisfied result [13]. Selection of components in dimensionality reduction, and kernels and parameters in kernel-based methods is still an open question, which needs to be further investigated. In order to tackle the above drawbacks, we proposed a novel ensemble classifier of hyperspectral and LiDAR data using morphological features. This work is inspired by our previous work: rotation-based ensemble [14, 15]. More specifically, we use random subspace and data transformation techniques, including principal component analysis (PCA), linearity preserving projection (LPP) [16] and unsupervised graph fusion (UGF) [11] to construct the ensemble. The rest of this paper is organized as follows. Hyperspectral and LiDAR datasets are described in Section 2, as well as morphological features. The ensemble classifier is proposed in Sec-

This study has been carried out with financial support from JSPS KAKENHI Grant Number 16F16053.

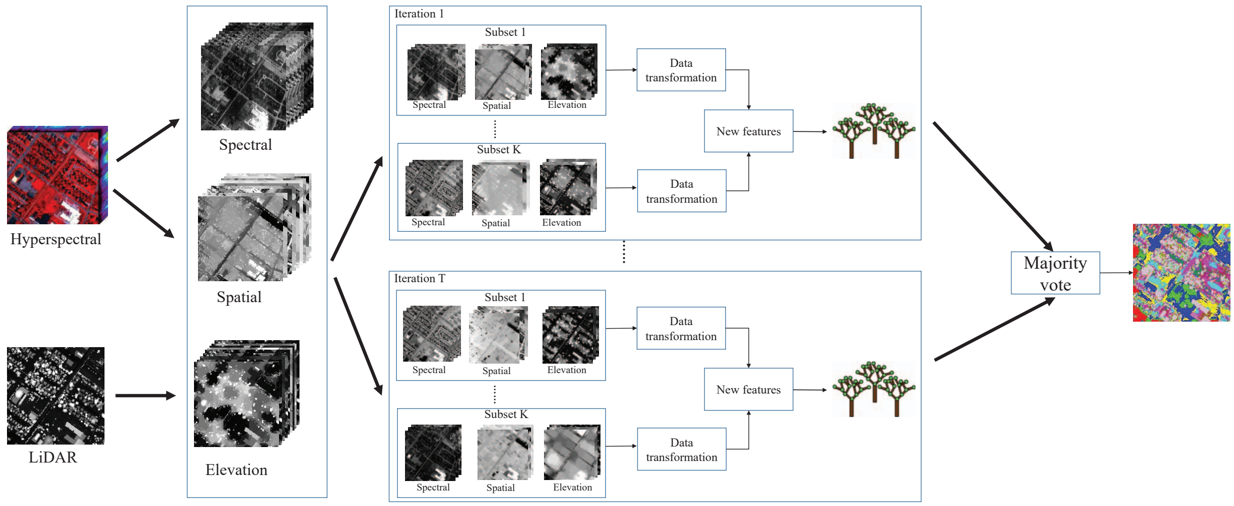


Fig. 1. Flowchart of the proposed ensemble classifier.

tion 3. Section 4 presents the results and analysis. Conclusions are drawn in Section 5.

2. DATASETS AND MORPHOLOGICAL FEATURES

An HS and LiDAR data were acquired by the NSF-funded Center for Airborne Laser Mapping (NCALM) on June 2012 over the University of Houston and the neighboring area with the same ground sampling distance (2.5 m). HS data has 144 spectral bands with a wavelength range from 380 to 1050 nm. The whole scene, consisting of 349 $\times$ 1905 pixels, contains 15 classes. Number of training and test samples are shown in Table 1 and Fig. 4(a) shows the false color image of HS data.

Morphological openings and closings with partial reconstruction [17] are used to produce morphological features for both LiDAR data and the first three PCs of the original HS. For linear structuring elements (SE), the openings or closings over every 10° and use 10% of the length of the SE is adopted. Then, MPs with 20 openings and closings (range: 5-100, step size:5) are generated. For a disk-shaped SE, MPs with 15 openings and closings (range: 1-15, step size:1) are computed. Thus, the size of MPs on HS and LiDAR are 210 and 70, respectively.

3. PROPOSED ENSEMBLE CLASSIFIER

To make all the features sources with the same dimension, we use kernel principal component analysis (KPCA) [18] to normalize their dimensions and reduce the noise. Moreover, the dimension of each feature source is normalized to the smallest dimension of all the feature sources $D = 70$.

Let $\mathbf{X}^{Sta} = \left\{\mathbf{X}^{Spe}, \mathbf{X}^{Spa}, \mathbf{X}^{Elv}\right\} = \left\{\mathbf{x}_i^{Spe}, \mathbf{x}_i^{Spa}, \mathbf{x}_i^{Ele}\right\}_{i=1}^{n}$ denote three kinds of features (i.e., spectral, spatial and elevation information) of training samples with the corresponding label $\mathbf{Y} = \{y_i\}_{i=1}^{n}$, where $\mathbf{x}_i^{Spe} \in \mathbb{R}^D$, $\mathbf{x}_i^{Spa} \in \mathbb{R}^D$, $\mathbf{x}_i^{Ele} \in \mathbb{R}^D$ and $y_i \in \{1, ..., C\}$ denotes the label information, where C is the total number of classes.

The flowchart of the ensemble classifier is shown in Fig. 1 and the main steps are presented as follows.

- First, we split each kind of feature space into K disjoint subsets. A subset of each kind of features contain $\lfloor D/K \rfloor$ features.
- Second, we apply data transformation to each subset to produce the new features. In this work, PCA, LPP, and UGF are adopted.
- Third, we concatenate the extracted components in each subset to generate the new features, which is used to train an individual Random Forest (RF) [19] classifier.
- Finally, we integrate the RF classifiers, which are generated by repeating the above steps T times, to achieve the classification result.

Data transformation plays a important role for the proposed ensemble classifier. Generally, the objective of linear data transformation is to find a transformation matrix $\mathbf{W}$, which can be obtained from the following eigenvalue decomposition problem:

$$\mathbf{S}_1\mathbf{w} = \lambda\mathbf{S}_2\mathbf{w} \tag{1}$$

where, $\mathbf{S}_1$ and $\mathbf{S}_2$ are the specific matrices.

In PCA, $\mathbf{S}_1$ and $\mathbf{S}_2$ are defined as the covariance and identify matrices. LPP [16] aims at preserving the local neighborhood information the process, in which $\mathbf{S}_1$ and $\mathbf{S}_2$ are formulated as:

$$\mathbf{S}_1 = \mathbf{X}^{Sta}\mathbf{L}(\mathbf{X}^{Sta})^\top \tag{2}$$

$$\mathbf{S}_2 = \mathbf{X}^{Sta}\mathbf{A}(\mathbf{X}^{Sta})^\top \tag{3}$$

Table 1. Overall, average, κ and class-specific accuracies.

Class	No of Samples		Ori$_{HS}$	EMPs$_{HS}$	EMPs$_{Li}$	EMPs$_{HSLi}$	Stack	Ensemble		
	Train	Test						PCA	LPP	UGF
Grass Healthy	20	1053	80.25	76.92	44.06	81.20	80.34	**82.91**	80.44	80.82
Grass Stressed	20	1064	79.98	73.31	40.98	80.45	54.51	84.30	**88.63**	80.83
Grass Synthetis	20	505	97.43	99.60	94.65	99.60	99.60	**100.00**	**100.00**	**100.00**
Tree	20	1056	**94.03**	89.30	54.45	93.18	90.63	91.76	89.58	87.41
Soil	20	1056	93.94	99.43	77.18	93.09	96.59	**99.91**	95.55	99.34
Water	20	143	89.51	84.62	73.43	79.02	82.52	94.41	95.80	**97.20**
Residential	20	1072	47.48	73.51	69.50	68.47	86.38	88.25	**89.09**	86.66
Commercial	20	1053	27.16	35.04	62.20	67.62	86.32	86.13	86.13	**88.79**
Road	20	1059	67.14	62.89	41.93	66.01	71.39	89.61	83.95	**90.56**
Highway	20	1036	38.03	48.65	36.68	42.86	42.66	42.95	53.57	**56.66**
Railway	20	1054	63.38	75.71	87.86	79.70	91.37	94.40	82.64	93.93
Parking Lot 1	20	1041	37.18	79.54	74.16	68.40	76.37	65.90	**80.50**	77.23
Parking Lot 2	20	285	31.23	**64.91**	52.28	57.89	52.28	52.98	57.89	62.11
Tennis Court	20	247	97.57	100.00	97.57	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**
Running Track	20	473	86.68	95.56	17.97	98.94	90.91	**100.00**	92.60	99.79
Overall accuracy (OA)			65.56	74.15	59.60	76.37	77.05	83.91	84.05	**85.48**
Average accuracy (AA)			68.73	77.27	61.66	78.43	78.59	84.91	85.09	**86.75**
kappa coefficients (κ)			62.95	72.15	56.22	74.39	75.11	82.56	82.74	**84.05**

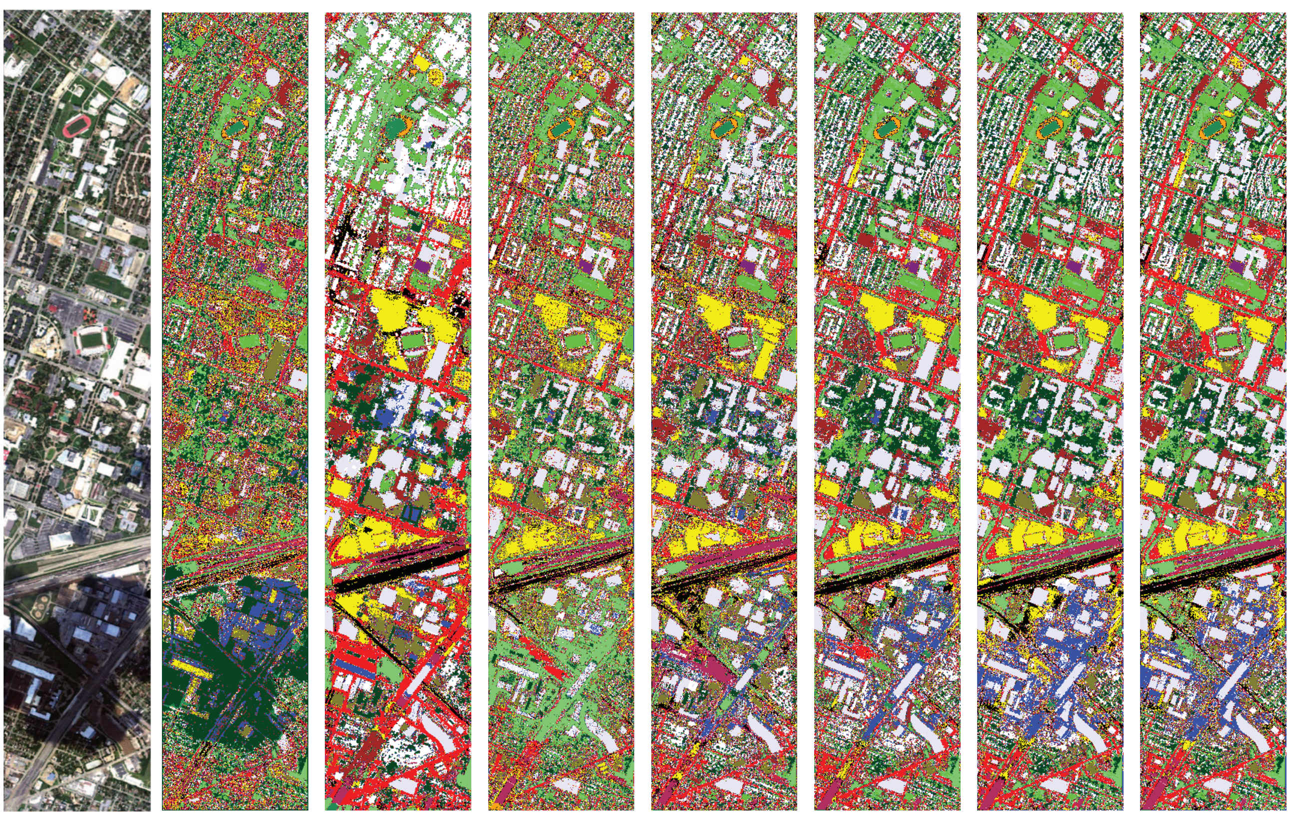

(a) (b) (c) (d) (e) (f) (g) (h)

Fig. 2. Classification maps produced by the different schemes. (a) False color compsite image of hyperspectral data. Thematic maps using (b) Origial spectral bands. (c) EMPs$_{Li}$. (d) EMPs$_{HSLi}$. (e) Stack. (f)E-PCA. (g) E-LPP. (h) E-UGF.

where, $\mathbf{L} = \mathbf{D} - \mathbf{A}$. $\mathbf{A}$ is a symmetric matrix with $A_{ij} = 1$ if $\mathbf{x}_i^{Sta}$ and $\mathbf{x}_j^{Sta}$ are close, and $A_{ij} = 0$ if $\mathbf{x}_i^{Sta}$ and $\mathbf{x}_j^{Sta}$ are far apart. $\mathbf{D}$ is a diagonal matrix whose entries are column sums of $\mathbf{A}$.

In UGF [11], $\mathbf{S}_1$ and $\mathbf{S}_2$ are defined as follows:

$$\mathbf{S}_1 = \mathbf{X}^{Sta}\mathbf{L}^{Fus}(\mathbf{X}^{Sta})^\top \tag{4}$$

$$\mathbf{S}_2 = \mathbf{X}^{Sta}\mathbf{A}^{Fus}(\mathbf{X}^{Sta})^\top \tag{5}$$

$$\mathbf{A}^{Fus} = \mathbf{A}^{Spe} \odot \mathbf{A}^{Spa} \odot \mathbf{A}^{Ele} \tag{6}$$

where, $\odot$ denotes element-wise multiplication, i.e., $A_{ij}^{Fus} = A_{ij}^{Spe}A_{ij}^{Spa}A_{ij}^{Ele}$. It should be noted that $A_{ij}^{Fus} = 1$ only if $A_{ij}^{Spe} = 1$, $A_{ij}^{Spa} = 1$, and $A_{ij}^{Ele} = 1$. It implies that $\mathbf{x}_i^{Sta}$ is close to $\mathbf{x}_j^{Sta}$ only if all individual features points $\mathbf{x}_i^{Ind}$ ($Ind \in Spe, Spa, Ele$) are close to $\mathbf{x}_j^{Ind}$, which indicates that $\mathbf{x}_i^{Sta}$ and $\mathbf{x}_j^{Sta}$ have similar spectral, spatial, and evaluation characteristics.

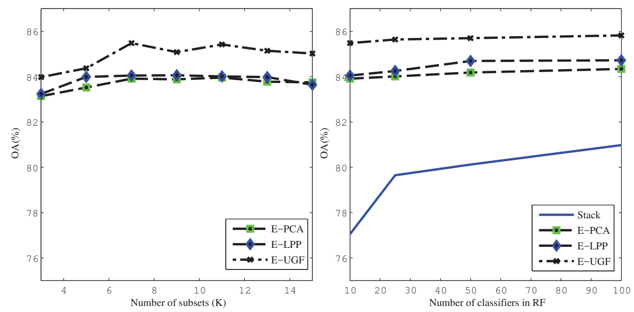
In this paper, data transformation is applied in each subset, and all components are kept. In this situation, the numbers of input and output features are both set to be $3 \times \lfloor D/K \rfloor$. The excellent performance of the proposed ensemble attributes to simultaneous improvements in two aspects: 1) promote the diversity by the use of random subspace and data transformation on training set [20, 21]; 2) improve the accuracies of RF classifiers by keeping all extracted components [21]. The proposed ensemble classifiers with PCA, LPP, and UGF are named respectively as E-PCA, E-LPP, and E-UGF.

4. EXPERIMENTAL RESULTS AND ANALYSIS

In our experiments, T and K are set to 10 and 7, respectively. For the RF, the number of classifiers is set to be 10, and the number of features in a subset is set to be the default value (square root of the number of the used features). Different feature sources are scaled to $[0, 1]$ before classification.

We compare our proposed ensemble classifier with the following schemes: 1) using the original HS (Ori$_{HS}$); 2) using the MPs computed on the first three PCs of the original HS (MPs$_{HS}$); 3) using the MPs computed on the LiDAR (MPs$_{Li}$); 4) stacking MPs computed from both LiDAR and the first three PCs of the original HS (MPs$_{HSLi}$); 5) stacking all normalized dimensional features (Stack). The classification results are investigated by measuring the overall accuracy (OA), the average accuracy (AA), the Kappa coefficient (κ), and the class-specific accuracies.

Table 1 reports the accuracy values generated by the proposed ensemble classifier along with the compared methods. From Table 1, we can find that the classification results by using single feature source are not accurate. However, each feature source shows its efficiency on various classes. For instance, spectral features provide discrimination on trees, grass, and water, whereas LiDAR shows excellent capacity on separating on man-made objects, e.g., commercial and railway. Stacking all normalized dimensional features, and MPs computed from both LiDAR and the first three PCs of the original HS slightly improve the performance. By using an ensemble strategy, our proposed methods yield the better performance than other compared methods. In this case, the OAs of E-PCA, E-LPP and E-UGF are 83.91%, 84.05%, and 85.48%, with 9%-24% and 6% that using single feature source and stacked features. Among the ensemble classifiers, E-UGF obtains the best result. The main reason is that UGF aims at combining multiple feature sources to generate the transformation matrix that can be used to improve the accuracy of member RF classifiers and the diversity, which is beneficial for the ensemble. In addition, our proposed method is competitive with the previous study [10], which won the Best Paper Challenge of the 2014 IEEE GRSS Data Fusion Contest.



(a) (b)

Fig. 3. Sensitivity to the change of (a) number of subsets (K) and (b) number of classifiers in RF.

Effects of the number of subsets (K) and the number of classifiers in RF are shown in Fig. 3. It can be seen that the proposed ensemble classifier is not sensitive to the two parameters, which is viewed as the added advantage.

5. CONCLUSION

The main contribution of this paper is to develop a framework to fuse spectral, spatial, and elevation from multi-sensor data in an ensemble strategy for the classification task. Morphological features with partial reconstruction are used to capture the spatial and elevation from HS and LiDAR data. Experimental results demonstrated its superiority.

Acknowledgment

The authors would like to thank the NCALM and the Hyperspectral Image Analysis group from the University of Houston for providing the data sets, and the IEEE GRSS Data Fusion Technical Committee for organizing the 2013 Data Fusion Contest

6. REFERENCES

[1] C. Debes et al., "Hyperspectral and LiDAR data fusion: Outcome of the 2013 GRSS Data Fusion Contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2405–2418, June 2014.

[2] J. Castorena, U. S. Kamilov, and P. T. Boufounos, "Autocalibration of LiDAR and optical cameras via edge alignment," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2016, pp. 2862–2866.

[3] J. Xia, L. Bombrun, T. Adali, Y. Berthoumieu, and C. Germain, "Classification of hyperspectral data with ensemble of subspace ICA and EPF," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2016, pp. 1422–1426.

[4] S. Bahrampour, N. M. Nasrabadi, A. Ray, and K. W. Jenkins, "Kernel task-driven dictionary learning for hyperspectral image classification," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2015, pp. 1324–1328.

[5] M. Dalponte, L. Bruzzone, and D. Gianelle, "Fusion of hyperspectral and LiDAR remote sensing data for classification of complex forest areas," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1416–1427, May 2008.

[6] D. Lemp and U. Weidner, "Improvements of roof surface classification using hyperspectral and laser scanning data," in *Proc. ISPRS Joint Conf- 3rd Int. Symp. Remote Sens. Data Fusion Over URBAN Areas5th Int. Symp. URS Areas, Tempe, AZ, USA, Mar 14-16*, 2005.

[7] M. Pedergnana, P. R. Marpu, M. Dalla Mura, J. A. Benediktsson, and L. Bruzzone, "Classification of remote sensing optical and LiDAR data using extended attribute profiles," *IEEE J. Sel. Top. Sign. Proces.*, vol. 6, no. 7, pp. 856–865, Nov 2012.

[8] T. Matsuki, N. Yokoya, and A. Iwasaki, "Hyperspectral tree species classification of Japanese complex mixed forest with the aid of LiDAR data,," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 5, pp. 2177–2187, May 2015.

[9] Y. Gu, Q. Wang, X. Jia, and J. A. Benediktsson, "A novel MKL model of integrating LiDAR data and MSI for urban area classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 10, pp. 5312–5326, Oct 2015.

[10] W. Liao, J. Xia, P. Du, and W. Philips, "Semi-supervised graph fusion of hyperspectral and LiDAR data for classification," in *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, July 2015, pp. 53–56.

[11] 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 *2014 IEEE Geoscience and Remote Sensing Symposium*, July 2014, pp. 1241–1244.

[12] W. Liao, A. Pizurica, R. Bellens, S. Gautama, and W. Philips, "Generalized graph-based fusion of hyperspectral and LiDAR data using morphological features," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 3, pp. 552–556, March 2015.

[13] M. Dalla Mura, A. Villa, J. A. Benediktsson, J. Chanussot, and L. Bruzzone, "Classification of hyperspectral images by using extended morphological attribute profiles and ICA,," *IEEE Geosci. Remote Sens. Lett.*

[14] J. Xia, J. Chanussot, P. Du, and X. He, "Rotation-Based Ensemble Classifiers for High-Dimensional Data," in *Fusion in Computer Vision*, Bogdan Ionescu, Jenny Benois-Pineau, Tomas Piatrik, and Georges Quénot, Eds., pp. 135–160. Springer, 2014.

[15] 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.

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

[17] W. Liao, R. Bellens, A. Pizurica, W. Philips, and Y. Pi, "Classification of hyperspectral data over urban areas using directional morphological profiles and semi-supervised feature extraction," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 6, pp. 11771190, April 2012.

[18] B. Schölkopf, A. Smola, and K. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Comput.*, vol. 10, no. 5, pp. 1299–1319, July 1998.

[19] L. Breiman, "Random forest," *Machine Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[20] J. Xia, M. Dalla Mura, J. Chanussot, P. Du, and X. He, "Random subspace ensembles for hyperspectral image classification with extended morphological attribute profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 9, pp. 4768–4786, 2015.

[21] L. Kuncheva and J. Rodriguez, "An experimental study on rotation forest ensembles," in *Proceedings of the 7th International Workshop on Multiple Classifier Systems*, Prague, Czech Republic, 2007, pp. 459–468.