

# MULTIPLE COMPOSITE KERNEL LEARNING FOR HYPERSPECTRAL IMAGE CLASSIFICATION

Peijun Du<sup>1</sup>, Junshi Xia<sup>2</sup>, Pedram Ghamisi<sup>3</sup>, Akira Iwasaki<sup>2</sup>, Jón Atli Benediktsson<sup>4</sup>

<sup>1</sup> Key Laboratory for Satellite Mapping Technology and Applications of State Administration of Surveying, Mapping and Geoinformation of China, Nanjing University, 210093 Nanjing, China

<sup>2</sup> Research Center for Advanced Science and Technology, The University of Tokyo, 153-0041 Tokyo, Japan

<sup>3</sup> Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), 82234 Weßling, Germany

<sup>4</sup> Faculty of Electrical and Computer Engineering, University of Iceland, 101 Reykjavik, Iceland

## ABSTRACT

In this work, we develop a new framework to combine ensemble learning and composite kernel learning for hyperspectral image classification. We refer it as the multiple composite kernel learning, which is based on an iterative architecture. More specifically, in each iteration, we use the rotation-based ensemble to create rotation matrix, which is used to generate rotated features for both spectral and spatial information (e.g., extinction profiles). Then, the new spectral and spatial features are integrated into the composite kernels based on support vector machines classifier. Different rotation matrices will lead to obtaining various newly spectral and spatial characteristics, thereby they further increase the diversity and the classification performance. Experimental results on Indian Pines benchmark hyperspectral dataset demonstrate the excellent performance of the proposed method.

**Index Terms**— Composite kernel learning, Ensemble learning, Classification, Hyperspectral image

## 1. INTRODUCTION

With the significant development of hyperspectral sensors, hyperspectral images can be acquired to monitor the Earth's surface with both high spectral and spatial resolution [1]. In order to obtain high precision classification result and smooth classification maps, the spectral and spatial information should be integrated into the classification process [2]. The composite kernels (CK) based on support vector machines (SVMs) is the commonly used technique to achieve this task [3].

Another development for providing enhanced classification performance is the combination of single learning algorithms, which is identified as multiple classifier systems (MCSs) or ensemble learning [4, 5]. The main component for the success of MCSs is the diversity within the ensemble [4]. The greatest diversity can be provided by the joint use of different training samples and features. Such an approach

is exploited in the rotation-based ensemble architecture [6–9], which uses data transformation and random feature selection to construct diverse individual classification results.

Since CK and rotation-based ensemble are two powerful techniques for the classification of hyperspectral images, it seems to be valuable to combine them to further improve the classification performance. In our previous work [9], we proposed rotation based SVMs (RoSVM) for the pixel-wise classification of hyperspectral data, which had shown better performance than Random subspace SVMs (RSSVM) [10].

In this paper, extended by the idea of RoSVM [9], we propose multiple composite kernels (MCK) learning that combines ensemble learning and composite kernel learning, for the classification of hyperspectral images. We adopt Multi-Attribute Extinction Profiles (EMEPs) to extract spatial information since it can provide highly accurate classification results and be efficient in terms of automatic parameter selection compared to attribute profiles (APs) [11].

## 2. RELATED WORKS

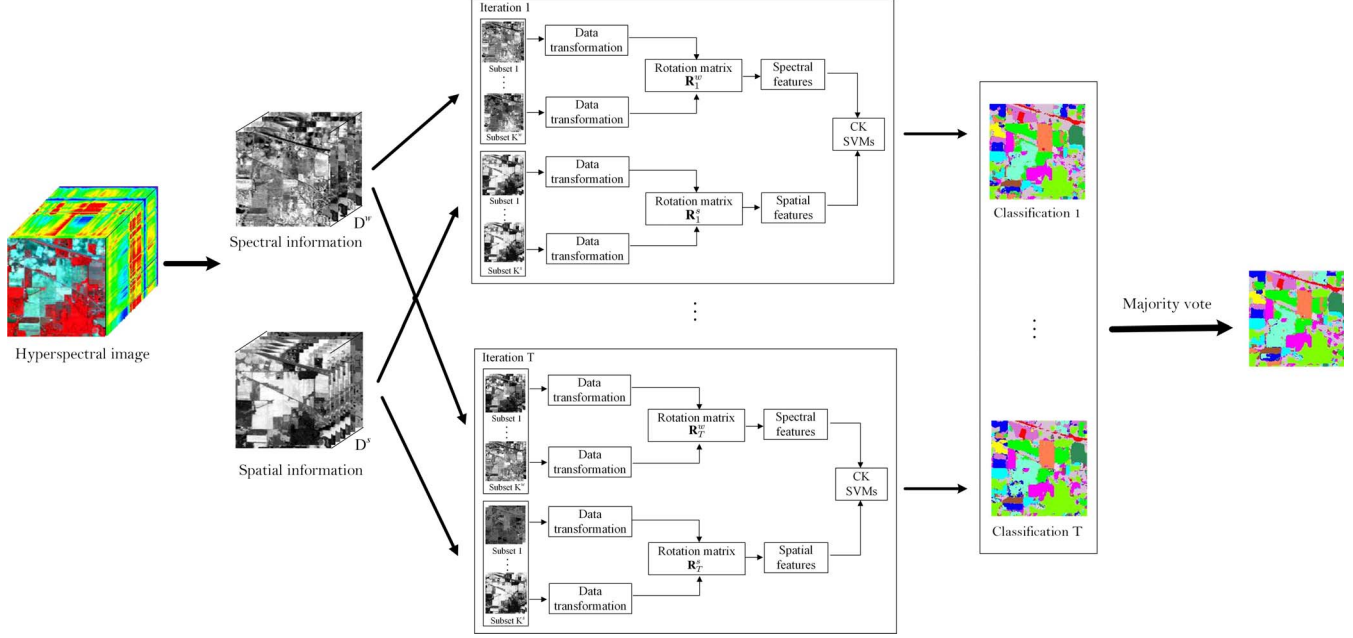
### 2.1. Multi-Attribute Extinction Profiles (EMEPs)

Extinction filters (EFs) are based on the concept of extinction values that measures the persistence of the extrema in the image [12, 13]. The Extinction Profiles (EPs) can be produced composing of a sequence of thinning and thickening transformations defined with a sequence of progressively stricter criteria. An EP for the input gray scale image,  $\mathbf{F}$ , is given by:

$$EP(\mathbf{F}) = \underbrace{\{\phi^{P_{\lambda_L}}(\mathbf{F}), \phi^{P_{\lambda_{L-1}}}(\mathbf{F}), \dots, \phi^{P_{\lambda_1}}(\mathbf{F})\}}_{\text{thickening profile}}, \quad (1)$$

$$\underbrace{\{\mathbf{F}, \gamma^{P_{\lambda_1}}(\mathbf{F}), \dots, \gamma^{P_{\lambda_{L-1}}}(\mathbf{F}), \gamma^{P_{\lambda_L}}(\mathbf{F})\}}_{\text{thinning profile}},$$

where  $P_{\lambda} : \{P_{\lambda_i}\} (i = 1, \dots, L)$  IS a set of  $L$  ordered predicates (i.e.,  $P_{\lambda_i} \subseteq P_{\lambda_k}, i \leq k$ ).  $\gamma$  and  $\phi$  are thinning and thickening operators, respectively [11].



**Fig. 1.** Flowchart of the proposed ensemble framework.

In [14], the extension of EPs, *extended extinction profiles* (EEP), has been generalized and adapted for spatial information extraction from hyperspectral data. To achieve this purpose, the most informative features extracted by principal component analysis (PCA) or independent component analysis (ICA) [15] can be preserved as base images [ $\mathbf{Q}_i$  ( $i = 1, \dots, m$ )] to produce the profiles. By concatenation of all the aforementioned information into a single stacked vector, one can come up with the concept of *extended multi-EPs* (EMEPs):

$$EMEP(\mathbf{Q}) = \{EEP_{\mathcal{A}_1}(\mathbf{Q}), EEP'_{\mathcal{A}_2}(\mathbf{Q}), \dots, EEP'_{\mathcal{A}_k}(\mathbf{Q})\}, \quad (2)$$

$$EEP(\mathbf{Q}) = \{EP(\mathbf{Q}_1), EP(\mathbf{Q}_2), \dots, EP(\mathbf{Q}_m)\}. \quad (3)$$

where  $EEP_{\mathcal{A}_i}$  is an EEP built with a set of predicates evaluating the attribute  $\mathcal{A}_i$ .  $EEP' = EEP \setminus \{\mathbf{Q}_i\}_{i=1, \dots, m}$  is considered to avoid including the input image  $\mathbf{Q}_i$  in the EMEPs.

## 2.2. Composite kernel learning

For a given training set  $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)$ ,  $y_i \in \{-1, +1\}$ , the SVMs solves the following program [3, 16]:

$$\begin{aligned} \min_{\mathbf{w}, b, \xi} : & \quad \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_i \xi_i \\ \text{s.t.} : & \quad \forall_{i=1}^n : y_i [\mathbf{w}'\phi(\mathbf{x}_i) + b] \geq 1 - \xi_i \\ & \quad \forall_{i=1}^n : \xi_i > 0 \end{aligned} \quad (4)$$

where,  $\phi$  is the nonlinear mapping function.  $\mathbf{w}$  and  $b$  are the parameters that define the linear classifier.  $C$  is the regularization parameter.  $\xi_i$  is the slack variable.

Thanks to the kernel function, the decision boundary of the classifier for test pixel  $\mathbf{x}$  is given by

$$f(x) = \text{sgn} \left( \sum_i^n y_i a_i \mathcal{K}(\mathbf{x}_i, \mathbf{x}) + b \right), \quad (5)$$

where,  $\mathcal{K}$  is a kernel function,  $\mathcal{K}(\mathbf{x}_i, \mathbf{x}) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}) \rangle$ .

Let us denote  $\mathbf{x}_i^w \in \mathbb{R}^{N^w}$  and  $\mathbf{x}_i^s \in \mathbb{R}^{N^s}$  of spectral and spatial information of a pixel  $\mathbf{x}_i$ .  $D^w$  and  $D^s$  are the numbers of spectral and spatial features. The spectral kernel  $\mathcal{K}_w$  and spatial kernel  $\mathcal{K}_s$  are computed as follows:

$$\mathcal{K}_w(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\|\mathbf{x}_i^w - \mathbf{x}_j^w\|^2 / 2\sigma_w^2) \quad (6)$$

$$\mathcal{K}_s(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\|\mathbf{x}_i^s - \mathbf{x}_j^s\|^2 / 2\sigma_s^2) \quad (7)$$

where,  $\sigma_w$  and  $\sigma_s$  are the widths of the spectral and spatial radius basis function (RBF) kernels. Thus, the composite kernel is represented as:

$$\mathcal{K} = \mu \mathcal{K}_w + (1 - \mu) \mathcal{K}_s \quad (8)$$

where,  $\mu$  is the control parameter balancing the spectral and spatial information.

## 3. MULTIPLE COMPOSITE KERNEL LEARNING

The proposed MCK framework is based on the rotation-based ensemble that aims at generating diverse individual CK classifiers using random feature selection and data transformation, which improves individual accuracy and diversity within the ensemble simultaneously [6, 7].

Let us denote the spectral  $\mathbf{X}^w$  and spatial features (e.g., EMEPs)  $\mathbf{X}^s$  of the training set with the dimensions of  $D^w$  and  $D^s$ , respectively.  $T$  is the number of iterations.  $K^w$  and  $K^s$  are the number of subsets, and  $M^w$  and  $M^s$  are the number of subsets, for respectively spectral and spatial features.

Fig. 1 presents the main procedure of the proposed MKL framework. The main steps of the MKL method can be summarized as follows:

1. the spectral and spatial features are divided into  $K^w$  and  $K^s$  groups, and each group has  $M^w$  and  $M^s$  features. It should be noted that the number of features of the last group is lower than  $M^w$  ( $M^s$ ) if  $D^w$  ( $D^s$ ) is indivisible by  $K^w$  ( $K^s$ ).
2. the new training features  $\tilde{\mathbf{X}}_{i,j}^w$  ( $\tilde{\mathbf{X}}_{i,j}^s$ ) is selected from  $\mathbf{X}_{i,j}^w$  ( $\mathbf{X}_{i,j}^s$ ) using the bootstrap technique with 75% size, where  $\mathbf{X}_{i,j}^w$  represents the  $j^{th}$  subset of the  $i^{th}$  ( $i = 1, \dots, T$ ) iterations.
3.  $\tilde{\mathbf{X}}_{i,j}^w$  is transformed by a data transformation (e.g., principal components analysis) algorithm to get the coefficients  $v_{i,j}^{(1)}, \dots, v_{i,j}^{(M_j^w)}$ , the size of  $v_{i,j}^{(\cdot)}$  is  $M^w \times 1$ . We treat  $\tilde{\mathbf{X}}_{i,j}^s$  in the same way.
4. sparse rotation matrices  $\mathbf{R}_i^w$  and  $\mathbf{R}_i^s$  are constructed by the above coefficients:

$$\mathbf{R}_i^w = \begin{bmatrix} v_{i,1}^{(1)}, \dots, v_{i,1}^{(M_1^w)} & 0 & \dots & 0 \\ 0 & v_{i,2}^{(1)}, \dots, v_{i,2}^{(M_2^w)} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & v_{i,K^w}^{(1)}, \dots, v_{i,K^w}^{(M_{K^w}^w)} \end{bmatrix}$$

$$\mathbf{R}_i^s = \begin{bmatrix} v_{i,1}^{(1)}, \dots, v_{i,1}^{(M_1^s)} & 0 & \dots & 0 \\ 0 & v_{i,2}^{(1)}, \dots, v_{i,2}^{(M_2^s)} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & v_{i,K^s}^{(1)}, \dots, v_{i,K^s}^{(M_{K^s}^s)} \end{bmatrix}$$

5. the columns of  $\mathbf{R}_i^w$  and  $\mathbf{R}_i^s$  are rearranged to the feature order of  $\mathbf{X}^w$  and  $\mathbf{X}^s$ .
6. the new spectral and spatial features are  $\mathbf{X}^w \mathbf{R}_i^w$  and  $\mathbf{X}^s \mathbf{R}_i^s$ , which are used as the input of the  $i^{th}$  CK-SVMs classifier and each CK-SVMs classifier is trained in parallel topology.
7. the final result is produced by combining the individual classification results by repeating above steps  $T$  times using the majority vote rule.

#### 4. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, the hyperspectral image recorded by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor over the Indian Pines in Northwestern Indiana, USA, is used to evaluate the performance. This scene is composed of

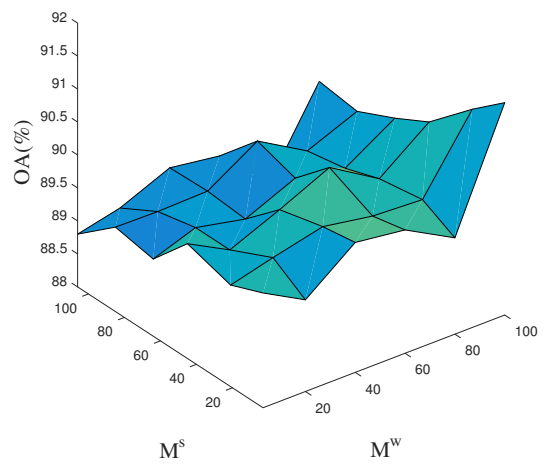


Fig. 2. Sensitivity to the change of the number features in a subset of spectral ( $M^w$ ) and spatial ( $M^s$ ) features.

145 × 145 pixels with 200 spectral bands after removing 20 noisy and water absorption bands (spatial resolution: 20 m). This dataset has 16 classes of interest.

EPs have been applied to the first three independent components extracted by ICA. Five different extinction profiles, including area ( $a$ ), diagonal of the bounding box, volume, height, and standard deviation have been taken into account in the EPs [11]. The settings of these profiles can be detailed in [11] and [14].  $D^w$  and  $D^s$  are 200 and 213, respectively.  $\mathcal{K}^w$  and  $\mathcal{K}^s$  refer to the method obtained using spectral and EMEPs, respectively. The optimal parameters for the SVMs are achieved by the five-fold cross validation [17].

$T$  is empirically set to be 20 [8]. For the CK and MCK,  $\mu$  is varied from 0.1 to 0.9, and only best results are shown. We apply ten Monte Carlo runs and mean values are presented. Fig. 2 lists the performances with different values of the two parameters (30 samples per class). The optimized  $M^w$  and  $M^s$  are 100 and 3, respectively.

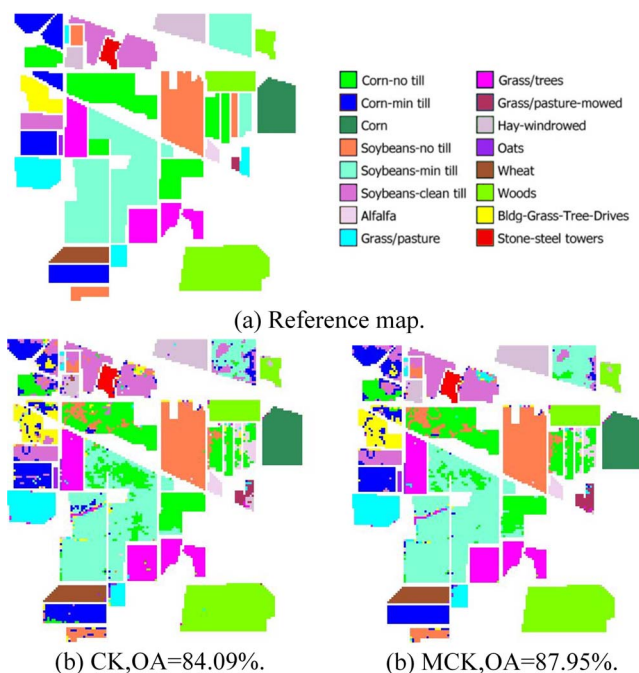
Table 1 presents the overall accuracies and average accuracies obtained for different kernel methods for different sizes of the training set. It can be observed that CK gains the better performance than  $\mathcal{K}^w$  and  $\mathcal{K}^s$ , and the ensemble strategy also improve the performance. By combing ensemble and CK learning, our proposed framework yields the best classification results in terms of accuracies in all cases. Fig.3 shows the classification maps of CK and MCK corresponding to one of ten Monte Carlo runs (20 samples per class) in Table 1.

#### 5. CONCLUSION

This paper presents a novel framework for hyperspectral image classification by combing CK and Rotation-based ensemble. Experimental results on the public Indian Pines

**Table 1.** Overall and average accuracies (in parentheses) obtained for different algorithms using different sizes of training set.

Samples per class	$\mathcal{K}_w$	$\mathcal{K}_s$	Ensemble		CK	MCK
			$\mathcal{K}_w$	$\mathcal{K}_s$		
5	47.81±3.29(58.71)	58.97±4.97(69.03)	55.37±3.55(65.83)	66.06±4.88(75.22)	61.14±4.15(71.27)	<b>67.29±4.76(76.29)</b>
10	55.92±3.38(68.93)	71.37±4.00(80.22)	65.12±2.54(76.64)	76.17±2.10(84.52)	73.39±4.26(82.25)	<b>78.82±1.73(86.64)</b>
15	63.63±1.45(74.80)	77.10±3.09(84.67)	73.83±1.00(84.67)	82.47±2.35(89.03)	80.45±3.31(87.54)	<b>84.73±2.12(90.90)</b>
20	66.26±1.64(77.21)	80.58±2.24(87.62)	75.39±2.13(85.16)	85.18±1.84(90.88)	83.43±2.15(90.05)	<b>87.68±2.26(92.92)</b>
25	70.17±1.40(81.20)	84.61±1.43(90.22)	77.40±1.31(87.02)	86.88±1.75(91.66)	86.53±1.73(91.89)	<b>89.79±1.62(94.02)</b>
30	71.64±1.47(82.18)	85.38±1.80(90.42)	79.50±1.59(87.85)	88.83±1.78(92.82)	87.81±1.96(92.67)	<b>91.22±2.03(94.69)</b>
35	72.98±1.37(83.52)	86.44±1.31(92.03)	80.64±1.89(88.79)	88.89±1.69(93.16)	88.65±1.81(93.43)	<b>91.52±1.96(95.32)</b>
40	74.59±0.99(84.33)	88.02±1.47(92.07)	82.14±1.12(89.57)	90.17±1.06(93.64)	89.92±1.37(94.17)	<b>92.68±1.15(95.80)</b>
45	76.05±2.14(85.57)	88.62±1.76(92.71)	82.13±1.38(89.52)	90.65±0.75(94.22)	90.62±1.30(94.49)	<b>92.66±0.81(95.83)</b>
50	76.78±0.65(85.84)	90.22±1.09(93.32)	82.45±1.79(89.96)	91.30±0.99(93.83)	91.87±1.14(95.14)	<b>93.38±1.15(96.19)</b>



**Fig. 3.** Classification maps of CK and MCK with OAs.

image demonstrate the excellent performance of the proposed framework, which captures the advantages of both aforementioned techniques.

## 6. ACKNOWLEDGMENT

The authors would like to thank Prof D. Landgrebe from Purdue University for providing the image. This work was supported in part by the National Natural Science Foundation of China (Grant No. 41631176).

## 7. REFERENCES

[1] C. I. Chang, *Hyperspectral Data Exploitation: Theory and Applications*, Wiley-Interscience, Hoboken, NJ, 2007.

[2] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Advances in spectral-spatial classification of hyperspectral images," *Proceedings of the IEEE*, vol. 101, no. 3, pp. 652–675, 2013.

[3] G. Camps-Valls, L. Gomez-Chova, J. Munoz-Mari, and J. Vila-Frances, "Composite kernels for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 1, pp. 93–97, 2006.

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

[5] 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.

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

[7] 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.

[8] 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.

[9] J. Xia, J. Chanussot, P. Du, and X. He, "Rotation-based support vector machine ensemble in classification of hyperspectral data with limited training samples," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1519–1531, March 2016.

[10] B. Waske, S. Van Der Linden, J. A. Benediktsson, A. Rabe, and P. Hostert, "Sensitivity of support vector machines to random feature selection in classification of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 7, pp. 2880–2889, 2010.

[11] P. Ghamisi, R. Souza, J. A. Benediktsson, X. X. Zhu, L. Rittner, and R. Lotufo, "Extinction profiles for the classification of remote sensing data," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 5631–5645, 2016.

[12] C. Vachier, "Extinction value: a new measurement of persistence," in *IEEE Work. Nonlinear Sig. Image Proc.*, 1995, vol. I, pp. 254–257.

[13] M. Grimaud, "New measure of contrast: the dynamics," 1992, vol. 1769, pp. 292–305.

[14] P. Ghamisi, R. Souza, J. A. Benediktsson, L. Rittner, R. Lotufo, and X. X. Zhu, "Hyperspectral data classification using extended extinction profiles," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 11, pp. 1641–1645, 2016.

[15] J. A. Benediktsson and P. Ghamisi, *Spectral-Spatial Classification of Hyperspectral Remote Sensing Images*, Artech House Publishers, INC, Boston, USA, 2015.

[16] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, 2004.

[17] C. C. Chang and C. J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.