

Lawrence Berkeley National Laboratory

Climate & Ecosystems

Title

Hyperspectral Image Classification With Rotation Random Forest Via KPCA

Permalink

<https://escholarship.org/uc/item/4ft944dx>

Journal

IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 10(4)

ISSN

1939-1404

Authors

Xia, Junshi
Falco, Nicola
Benediktsson, Jn Atli
[et al.](#)

Publication Date

2017

DOI

10.1109/jstars.2016.2636877

Peer reviewed

Hyperspectral Image Classification With Rotation Random Forest Via KPCA

[View Document](#)

10

Paper

Citations

427

Full

Text Views

Related Articles

[Semisupervised One-Class Support Vector Machines for Classification of Remote Se...](#)

[Rotation-Based Support Vector Machine Ensemble in Classification of Hyperspectra...](#)

[\(Semi-\) Supervised Probabilistic Principal Component Analysis for Hyperspectral ...](#)

[View All](#)

5

Author(s)

[Junshi Xia](#); [Nicola Falco](#); [Jón Atli Benediktsson](#); [Peijun Du](#); [Jocelyn Chanusot](#)

[View All Authors](#)

Abstract

[Authors](#)

[Figures](#)

[References](#)

[Citations](#)

[Keywords](#)

[Metrics](#)

[Media](#)

Abstract:

Random Forest (RF) is a widely used classifier to show a good performance of hyperspectral data classification. However, such performance could be improved by increasing the diversity that characterizes the ensemble architecture. In this paper, we propose a novel ensemble approach, namely rotation random forest via kernel principal component analysis (RoRF-KPCA). In particular, the original feature space is first randomly split into several subsets, and KPCA is performed on each subset to extract high order statistics. The obtained feature sets are merged and used as input to an RF classifier. Finally, the results achieved at each step are fused by a majority vote. Experimental analysis is conducted using real hyperspectral remote sensing images to evaluate the performance of the proposed method in comparison with RF, rotation forest, support vector machines, and RoRF-PCA. The obtained results demonstrate the effectiveness of the proposed method.

ISSN Information:

INSPEC Accession Number: 16773852

DOI: [10.1109/JSTARS.2016.2636877](https://doi.org/10.1109/JSTARS.2016.2636877)

Sponsored by: [IEEE Geoscience & Remote Sensing Society](#)

Funding Agency:

 Contents

• [Download PDF](#)

- [Download Citation](#)
- [View References](#)
- [Email](#)
- [Print](#)
- [Request Permissions](#)
- [Export to Collabratec](#)

-
- [Alerts](#)

SECTION I.

Introduction

Hyperspectral sensors capture images in a few tens or hundreds of narrow spectral channels which provide discriminant information for different materials in a given scene [1], [2]. The capability of a classifier can be affected when a small number of training samples is available on a high number of spectral channels, provoking the Hughes phenomenon [3].

Over the last two decades, researchers have investigated a variety of approaches to alleviate the Hughes phenomenon [4]–[7]. Recently, kernel-based methods, e.g., support vector machine (SVMs), have attracted increasing attention due to their encouraging performance in handling high-dimensional data [4]–[6]. Melgani and Bruzzone [5] investigated the effectiveness of strategies based on ensembles of binary SVMs used to solve multiclass problems in hyperspectral data and found that one against one is the most effective strategy [5], [8]. Campus-Valls and Bruzzone [6] assess the performance of several kernel-based classifiers, such as regularized radial basis function neural networks (Reg-RBFNN), SVMs, kernel Fisher discriminant, and regularized AdaBoost (Reg-AB). Experimental results demonstrated that SVMs yield the best results with a much lower computational cost. However, one of the main issues of such technique is related to the selection of the kernel parameters, which is usually addressed by performing a high computationally expensive cross-validation.

Another effective strategy in classifying high-dimensional data is the combination of simple classifiers, which is identified as multiple classifier systems (MCSs) [9], [10]. In this context, random forest (RF), which has been shown to be effective in dealing with high-dimensional datasets, is proposed by Breiman [11]. Ham *et al.* [12] and Gialson *et al.* [13] investigated the performance of RF in hyperspectral data and multisource remote sensing and geographic data. Two tree-based ensemble classification algorithms, AdaBoost and RF, were compared regarding classification accuracy, computational time, and classification stability, in addressing the classification of ecotopes task by analyzing hyperspectral imagery in [14]. Although both AdaBoost and RF attained similar classification results in terms of accuracy, and both outperformed a neural network-based classifier, RF was shown to be faster and more stable than AdaBoost. Waske *et al.* [15] investigated the performance of

SVMs and RF, comparing them with a maximum likelihood classifier and the spectral angle mapper. Although SVMs and RF had shown some diversity in the classification results, the global performance is quite similar. The parameter selection for RF is less expensive than the ones required by SVMs.

Since then, a considerable amount of literature has been published on remote sensing applications, in which RF is exploited in other typology of remote sensing data, such as high spatial resolution data [16] and LiDAR datasets [17]. Recently, many extensions of the standard RF algorithm have been suggested in the following three main strategies:

1. reduction of the features prior the classification via RF;
2. definition of the efficient aggregation strategies in RF; and
3. definition of an ensemble of RFs.

Considering the first strategy, Tuv *et al.* [18] generated a compact subset of nonredundant features for RF. Representative works that follow the second strategy are the alternating decision forest [19] and oblique random forests [20], [21]. The former method extends RF to any given differentiable loss function without losing the main characteristics and benefits of the original RF [19]. The latter method projects the original data at each node into another feature space that calculate the best split [20]. An oblique decision tree ensemble in which each decision hyperplane in the internal node of tree was calculated by multisurface proximal SVMs in [21]. An example of the third strategy is represented by the work in [22], in which a novel ensemble of the RF classifier, namely rotation random forest (RoRF), was developed. The main idea of this approach is related to the data transformation, which in this case was performed via principal component analysis (PCA), applied at node level to identify the best split. Such approach was exploited in [23] for hyperspectral image classification, obtaining promising results.

PCA aims at finding a representative lower-dimension projection that keeps most of the variance of the original data. The main theoretical limitation of PCA is the assumption of a linear relationship between spectral features, which does not reflect the real data behavior [1]. To circumvent this limitation, a nonlinear version of PCA, namely Kernel PCA (KPCA), which is capable of capturing part of the higher order statistics, thus, extracting nonlinear features, was introduced [24]. Fauvel *et al.* [25] demonstrated the effectiveness of the extracted features using KPCA in providing better class separability when compared to those obtained via PCA.

Inspired by the idea of RoRF and the aforementioned considerations on KPCA, the main contribution of this paper is to propose a novel classification scheme

based on the RoRF integrated with KPCA, namely RoRF-KPCA. The integration of the KPCA will permit the extraction of informative features at the node level, providing a more representative subset of features on those extracted by exploiting the original algorithm. The effectiveness of the proposed approach is assessed concerning accuracy and diversity in the ensemble via experimental analysis, which is carried on two real hyperspectral datasets. In the analysis, the impact of different sizes of the training set and the use of various kernel functions are investigated. It should be emphasized that, although RoRF-KPCA can be combined with spatial information, i.e., Markov random fields (MRFs) [26] and extended multiattribute profiles (EMAPs) [23], we focus on pixel-wise classifier.

SECTION II.

Background

This section provides an introduction to the background on both RF and KPCA. Let us denote $\mathcal{D} = \{\mathbf{x}_i, y_i\}_{i=1}^n$ as the training set, where $\mathbf{x}_i \in \mathbb{R}^D$ is a D -dimensional pixel vector, $y_i = 1, \dots, C$ is the corresponding class label, C is the number of classes and n is the number of training samples.

A. Random Forest (RF)

RF is an ensemble of T decision trees. During the training stage, decision trees are independently constructed on a bootstrap training set with randomly chosen features using two steps bagging techniques [11]. Each decision tree is constructed by the following steps:

1. selecting subset training samples from the training set \mathcal{D} ;
2. randomly selecting $D_{\text{try}} \leq D$ features and consequent identification of the best split; and
3. growing of the tree to the maximum depth without pruning.

During the classification stage, a given sample \mathbf{x}^* is classified by going through each decision tree until a leaf node is reached. A classification result (the decision function h) is assigned to each leaf node. The class label y^* is determined by taking the class having the most votes.

B. Kernel Principal Component Analysis

KPCA maps the original input space into a high dimensional feature space using a kernel trick [24]. For a given nonlinear mapping, the *input data space* \mathbb{R}^D can be mapped into the *feature space* \mathcal{H}

$$\Phi: \mathbb{R}^D \rightarrow \mathcal{H}, \mathbf{x} \mapsto \Phi(\mathbf{x}). \quad (8)(9)$$

[View Source](#) 

Given a set of n training samples $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ in \mathbb{R}^D , the covariance operator, \mathbf{D}_Φ , of the *feature space* \mathcal{H} can be constructed by

$$\mathbf{D}_\Phi \mathbf{m}_\Phi = \frac{1}{n} \sum_{i=1}^n (\Phi(\mathbf{x}_i) - \mathbf{m}_\Phi)(\Phi(\mathbf{x}_i) - \mathbf{m}_\Phi)^T = \frac{1}{n} \sum_{i=1}^n (\Phi(\mathbf{x}_i)). \quad (10)(11)$$

[View Source](#) 

Since each eigenvalue of a positive operator is nonnegative in a Hilbert space, it follows that all nonzero eigenvalues of \mathbf{D}_Φ are positive. In this way, the

eigenvector is linearly expressed as $\beta = \sum_{i=1}^n \alpha_i \Phi(\mathbf{x}_i)$, where, β is the eigenvector of \mathbf{R} .

In order to obtain the expansion coefficients, let us denote $\mathbf{Q} = [\Phi(\mathbf{x}_1), \dots, \Phi(\mathbf{x}_n)]$ and form an $M \times M$ Gram matrix, $\mathbf{R} = \mathbf{Q}^T \mathbf{Q}$, whose elements can be determined by virtue of kernel tricks

$$\mathbf{R} = \Phi(\mathbf{x}_i)^T \Phi(\mathbf{x}_j) = \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j) = \langle \mathbf{x}_i, \mathbf{x}_j \rangle. \quad (12)$$

[View Source](#)

\mathbf{R} is centralized by $\mathbf{R} = \mathbf{R} - \mathbf{1}_n \mathbf{R} - \mathbf{R} \mathbf{1}_n + \mathbf{1}_n \mathbf{R} \mathbf{1}_n$, where $\mathbf{1}_n = (1/n) \mathbf{1}_n \times \mathbf{1}_n$.

Thus, the orthogonal eigenvectors $\gamma_1, \dots, \gamma_n$ of \mathbf{R} corresponding to the n largest positive eigenvalues, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ are calculated. The orthogonal eigenvectors β_1, \dots, β_n corresponding to the n largest positive eigenvalues, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$, are $\beta_i = \frac{1}{\sqrt{\lambda_i}} \mathbf{Q} \gamma_i, i = 1, \dots, n$.

After the projection of the mapped sample $\Phi(\mathbf{x})$ onto the eigenvector system β_1, \dots, β_n , we can obtain the KPCA transformed feature vector $\mathbf{z} = (z_1, \dots, z_n)^T$ by $\mathbf{z} = \mathbf{P}^T \Phi(\mathbf{x})$, where $\mathbf{P} = \{\beta_1, \dots, \beta_n\}$.

Specifically, the i th KPCA component z_i is obtained by

$$z_i = \beta_i^T \Phi(\mathbf{x}) = \frac{1}{\sqrt{\lambda_i}} \gamma_i^T \mathbf{Q}^T \Phi(\mathbf{x}) = \frac{1}{\sqrt{\lambda_i}} \gamma_i^T [\langle \mathbf{x}_1, \mathbf{x} \rangle, \dots, \langle \mathbf{x}_n, \mathbf{x} \rangle], i = 1, \dots, n. \quad (13)$$

[View Source](#)

In this paper, the following three well-known kernels for the implementation of KPCA are used and investigated:

1. *Linear*: $\langle \mathbf{x}_i, \mathbf{x}_j \rangle = \mathbf{x}_i \cdot \mathbf{x}_j$
2. *Polynomial*: $\langle \mathbf{x}_i, \mathbf{x}_j \rangle = (\mathbf{x}_i \cdot \mathbf{x}_j + 1)^d, d \in \mathbb{R}$
3. *RBF*: $\langle \mathbf{x}_i, \mathbf{x}_j \rangle = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|_2^2 / 2\sigma^2), \sigma \in \mathbb{R}$

SECTION III.

Rotation Random Forest-Kernel Principal Component Analysis

The success of MCSs depends not only on the choice of the base classifier but also on the diversity within the ensemble [9]. The use of different training samples and features provides the greatest diversity [9]. To achieve this task, rotation-based ensemble architecture was designed considering multiple DTs, denote as rotation forest (RoF) [27].

Although RF is considered as a robust classifier, it shows low diversity, leading to propagation of errors. Aiming at improving both the diversity and classification accuracy of the RF classifiers within the ensemble, we propose a new rotation-based architecture as an ensemble of RF classifiers coupled with KPCA (RoRF-KPCA), which is exploited to handle nonlinearity in the hyperspectral data. Since rotation-based ensemble and KPCA are both considered powerful techniques in different ways, combining them to benefit from the capabilities of both seems desirable.

SECTION Algorithm 1

RoRF-KPCA

Training phase

Input: $\{\mathbf{X}, \mathbf{Y}\} = \{\mathbf{x}_i, y_i\}_{i=1}^n$: training samples, T : number of classifiers, K : number of subsets, M : number of features extracted in a subset, L : base classifier. The ensemble $\geq \emptyset$. \mathbb{F} : Feature set

Output: The ensemble \geq

for $i = 1: T$ **do**

randomly split the features \mathbb{F} into K subsets \mathbb{F}_{ij}

for $j = 1: K$ **do**

form the new training set $\mathbf{X}_{i,j}$ with \mathbb{F}_{ij}

randomly select the 75% of the initial training samples to generate $\mathbf{X}_{i,j}$

calculate the kernel matrices by $\mathbf{X}_{i,j}$, $\mathbf{K}_{\text{train}_{i,j}} = \langle \mathbf{X}_{i,j}, \mathbf{X}_{i,j} \rangle$

perform KPCA to transform $\mathbf{X}_{i,j}$ with the aim of getting the

coefficients $\mathbf{P}_{i,j} = [\beta^{(1)}_{i,j}, \dots, \beta_{M_{i,j}}]$

end for

the features extracted will be given by: $\mathbb{F}_{\text{new}_{i,j}} = [\mathbf{K}_{\text{train}_{i,j,1}} \mathbf{P}_{i,1}, \dots, \mathbf{K}_{\text{train}_{i,j,k}} \mathbf{P}_{i,k}]$

train an RF classifier L_i using $\{\mathbb{F}_{\text{new}_{i,j}}, \mathbf{Y}\}$

Add the classifier to the current ensemble, $\geq \Rightarrow \cup L_i$.

end for

Prediction phase

Input: The ensemble $\geq = \{L_i\}_{i=1}^T$. A new sample \mathbf{x}_* . Transformation matrix: \mathbf{P} .

Output: class label y_*

for $i = 1: T$ **do**

for $j = 1: K$ **do**

generate the kernel matrices between $\mathbf{X}_{i,j}$ and \mathbf{x}_* , $\mathbf{K}_{\text{test}_{i,j}} = \langle \mathbf{X}_{i,j}, \mathbf{x}_* \rangle$

generate the test features of \mathbf{x}_* , $\mathbb{F}_{\text{test}_{i,j}} = [\mathbf{K}_{\text{test}_{i,j,1}} \mathbf{P}_{i,1}, \dots, \mathbf{K}_{\text{test}_{i,j,k}} \mathbf{P}_{i,k}]$

end for

run the RF classifier L_i using $\mathbb{F}_{\text{test}_{i,j}}$ as input

end for

calculate the confidence \mathbf{x}_* for each class and assign the class label y_* to the class with the largest confidence.

The main train steps are summarized as follows:

1. the original feature space is divided into K disjoint subspaces, each one composed of M features;
2. about 75% size of the original training set are randomly selected via Bagging technique;

3. KPCA is performed on each K subset to obtain the coefficients and kernel matrices for each subspace;
4. the new training data set is formed by concatenating M extracted features in each subset that are achieved by rotating the original training set using the obtained coefficients and kernel matrices;
5. an RF classifier is trained with the new training dataset; and

6. the process is repeated T times. The final classification result is produced by integrating the results obtained by each step via majority voting.

The detailed training and prediction steps are presented in Algorithm 1. The reason for selecting 75% size of $\mathbf{X}_{i,j}$ is to avoid obtaining the same coefficients when the same features are chosen and hence promote the diversity [27]. Diversity in RoRF-KPCA is promoted in two aspects: 1) random selection of features; and 2) KPCA data transformation applied to the selected features using bootstrap sampling technique.

SECTION IV.

Experimental Results and Analysis

A. Hyperspectral Datasets

Two well known public hyperspectral datasets are used to evaluate the performance of the proposed method.

1) Indian Pines AVIRIS

The first dataset was captured by Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) over northwestern Indiana in June 1992, with 220 spectral bands and a spatial resolution of 20 m/pixel. The whole scene (145×145) consists of 16 classes, ranging in size from 20 to 2468 pixels (seen in Table I).

TABLE I Indian Pines AVIRIS and University of Pavia ROSIS Images: Class Name and Number of Reference Samples

No.	AVIRIS		ROSIS	
	Name	Ref. samples	Name	Ref. samples
1	Alfalfa	54	Bricks	3682
2	Corn-no till	1434	Shadows	947
3	Corn-min till	834	Metal Sheets	1345
4	Bldg-Grass-Tree-Drives	234	Bare Soil	5029
5	Grass/pasture	497	Trees	3064
6	Grass/trees	747	Meadows	18649
7	Grass/pasture-mowed	26	Gravel	2099
8	Corn	489	Asphalt	6631
9	Oats	20	Bitumen	1330
10	Soybeans-no till	968		
11	Soybeans-min till	2468		
12	Soybeans-clean till	614		
13	Wheat	212		
14	Woods	1294		
15	Hay-windrowed	380		
16	Stone-steel towers	95		

2) University of Pavia ROSIS

The second dataset, which was acquired over an university area in the city of Pavia, Italy, was collected by the Reflective Optics System Imaging Spectrometer (ROSIS) with 115 bands (wavelength range from 430 to 860 nm) and a very high spatial resolution of 1.3 m/pixel. It consists of 610×340 pixels. About 12 noisy bands have been removed, and the remaining 103 bands have been used in the classification. The reference data with nine classes of interest is composed of 42 776 pixels (seen in Table I).

B. Experimental Setup

Here, we present the experimental setup used in this paper.

1. Number of classifiers in the ensemble: $T=10$.
2. Number of features in a subset: $M=10$.
3. Base classifier: RF. The number of trees is set to 10 and the number of features in a subset is set to the default value (square root of the number of the used features).
4. Kernels in KPCA: linear, RBF, and Polynomial kernels. In RBF, σ is estimated with the median distance between all used samples. In Polynomial, d is set to 2.


For comparison, several classifiers have been used as baseline (ten repetitions were done for which a new training set has been randomly generated).

1. Random Forest [11].
2. Rotation Forest [27], [28].
3. SVMs [29]. SVMs with a RBF Kernel is used.
4. RF/SVMs+PCA/KPCA. PCA/KPCA is used as the feature extraction method prior to RF/SVMs.
5. RoRF-PCA [23].

For the cases of RF/SVMs+PCA/KPCA, the range of extracted components is from 2 to 30. Only best results are reported.

The performances are evaluated by the overall accuracy (OA), average accuracy (AA) Kappa coefficient (κ), average of overall accuracies (AOA) of the individual classifiers, and *coincident failure diversity (CFD)* [30]

$$CFD = \{0, 1 - p_0 \sum_{i=1}^T (1 - p_i)^{i-1} p_i\} \quad p_0 = 0, p_0 < 1 \quad (14)$$

[View Source](#)  where p_i is the probability of B taking the value i . $B = \{1, 1 - i, \dots, 0\}$ is defined as a random variable that denotes the proportion of classifiers that are incorrect on a randomly selected input pattern. Higher values lead to stronger diversity.

C. Experimental Results and Analysis

Tables II and III display the global accuracies (OA, AA, and κ) for both the AVIRIS and ROSIS datasets.¹ As we expected, the proposed RoRF-KPCA significantly outperforms the other methods regarding OA using the McNemar's test with a 95% confidence interval.. In particular, both RF and SVM offer poorer classification results when the AVIRIS image is considered (see Table II). In this case, both RF and SVM with PCA/KPCA can increase the classification performance. RoRF-PCA performs better than RF, SVM, and RF/SVM+PCA/KPCA. In the case of ROSIS image(see Table III), SVM provides better results than RF. RoRF-PCA performs better than RF and RoF+PCA/KPCA and provides slightly higher classification accuracies than SVM and SVM+PCA. Moreover, we can see that the choice of using different kernel functions in the RoRF-KPCA is not critical in this context, and the approach provides an absolute stability in providing classification results very close each other.

TABLE II Classification Accuracies for Indian Pines AVIRIS Using Different Training Set

Samples per class		RF	RoF	RF+KPCA	RF+PCA	SVM	SVM+PCA	SVM+KPCA	RoRF-PCA	RoRF-KPCA		
										Linear	RBF	Poly
10 samples	OA	47.73	50.30	47.57	47.70	46.60	54.88	48.77	53.86	65.08	66.12	63.6
	AA	59.53	61.92	59.40	59.31	57.56	65.46	60.80	65.46	75.94	77.13	75.1
	κ	41.76	44.74	41.77	41.74	40.22	49.53	43.11	48.37	60.86	62.09	59.
20 samples	OA	54.34	57.02	54.76	54.68	54.44	60.21	57.26	60.40	71.05	71.12	71.3
	AA	67.33	67.73	66.55	66.16	64.68	70.36	67.40	72.27	81.81	81.98	82.0
	κ	49.06	51.99	49.53	49.53	48.92	55.45	52.25	55.81	67.61	67.65	67.8
30 samples	OA	58.72	61.05	57.96	58.62	58.96	63.88	60.29	64.82	74.92	75.56	75.3
	AA	70.70	71.20	70.02	70.18	67.36	74.29	70.41	75.84	84.70	84.75	84.9
	κ	53.94	56.42	53.18	53.86	53.85	59.66	55.64	60.63	71.90	72.58	72.3
40 samples	OA	62.75	63.98	60.95	61.21	60.99	66.12	62.96	68.06	77.71	77.66	76.6
	AA	73.37	75.19	71.57	71.69	68.84	75.55	73.04	77.89	86.43	86.49	86.0
	κ	58.34	59.63	56.33	56.65	56.09	62.10	58.66	64.23	74.95	74.92	73.8
50 samples	OA	64.00	65.01	62.41	62.42	62.62	68.06	63.68	68.68	78.38	78.65	78.6
	AA	74.89	75.78	74.00	73.45	70.44	77.36	73.66	79.11	87.37	87.11	86.9
	κ	59.80	60.73	58.07	58.04	57.95	64.25	59.42	64.94	75.74	76.00	75.9

TABLE III Classification Accuracies for University of Pavia ROSIS Images Using Different Training Set

Samples per class		RF	RoF	RF+KPCA	RF+PCA	SVM	SVM+PCA	SVM+KPCA	RoRF-PCA	RoRF-KPCA		
										Linear	RBF	Poly
10 samples	OA	62.30	68.86	63.97	63.57	62.78	68.02	60.97	70.74	71.96	70.66	69.5
	AA	71.84	75.82	72.92	73.10	74.36	77.73	72.52	78.64	78.63	79.38	78.9
	κ	53.30	61.23	55.18	54.76	54.63	60.25	52.08	63.54	64.70	63.49	62.2
20 samples	OA	64.89	74.77	69.64	68.35	73.73	74.20	63.90	75.19	76.16	76.17	77.3
	AA	75.08	80.83	76.46	76.17	81.22	81.69	75.36	82.01	82.13	82.35	83.2
	κ	56.61	68.31	61.63	60.26	66.94	67.43	55.53	68.73	69.79	69.79	71.1
30 samples	OA	68.61	78.99	71.55	71.61	76.41	78.30	67.58	79.05	79.23	79.42	78.7
	AA	77.60	83.87	78.05	78.01	83.55	84.11	78.29	84.92	84.75	84.88	84.8
	κ	60.68	73.44	64.04	64.00	70.21	72.27	59.62	73.42	73.52	73.79	73.0
40 samples	OA	70.68	79.82	71.81	72.39	80.21	77.72	69.84	80.18	80.68	80.77	79.9
	AA	79.04	85.02	78.99	79.24	85.55	85.26	80.10	86.10	86.08	85.81	85.7
	κ	63.15	74.38	64.47	65.05	74.80	71.80	62.28	74.90	75.34	75.40	74.5
50 samples	OA	70.94	80.62	73.55	73.07	81.08	81.50	69.70	81.87	82.75	82.22	82.8
	AA	79.79	96.06	80.11	79.83	86.86	86.75	81.21	87.37	87.11	87.14	87.6
	κ	63.54	75.38	66.49	65.91	75.82	76.26	62.45	76.93	77.86	77.22	77.9

Tables IV and V report the class-specific accuracies using a fixed training set (20 samples per class). RoRF-KPCA shows a clear and consistent improvement OA with respect to RF of about +17% and +12%, RF+PCA of about +17% and +6%, RF+KPCA of about +17% and + 7%, SVM of about +17% and +3%, SVM+PCA of about +11% and +2%, SVM+KPCA of about +14% and + 11%, RoRF-PCA of about +11% and +1% for AVIRIS and ROSIS data, respectively. It can be noticed that the proposed approach can achieve better class-specific accuracies on the other approaches. In particular, for the classes “Corn-no till,” “Soybeans-no till,” “Soybeans-clean till,” and “Hay-windrowed,” the proposed approach increases the classification accuracies of at least ten percentage points.

TABLE IV Indian Pines AVIRIS Image

Class	RF	RoF	RF+PCA	RF+KPCA	SVM	SVM+PCA	SVM+KPCA	RoRF-PCA	RoRF-KPCA		
									Linear	RBF	Poly
1	84.81	75.29	83.33	85.19	76.47	85.19	88.33	87.41	89.63	92.22	91.3
2	32.57	46.55	36.40	44.66	36.41	41.35	39.43	34.89	58.65	57.03	61.9
3	42.01	50.31	52.52	45.50	39.02	51.20	50.07	48.19	51.97	51.43	51.4
4	53.12	66.96	58.12	52.31	69.91	82.91	62.39	65.09	83.12	85.17	82.8
5	76.66	76.79	61.97	60.44	74.55	76.26	70.52	77.20	86.38	84.85	85.0
6	67.95	75.25	68.81	68.13	78.06	73.90	70.25	80.47	91.34	92.36	92.0
7	98.08	78.33	100.00	95.00	93.33	96.15	94.23	99.62	99.23	99.23	99.2
8	80.57	78.83	84.46	78.59	66.35	79.14	77.26	85.56	92.70	92.68	93.6
9	90.00	78.00	90.00	84.50	60.00	75.00	77.00	96.00	98.00	99.00	98.0
10	48.61	58.39	56.30	56.10	49.35	66.12	65.89	56.18	76.03	76.17	73.8
11	44.04	37.71	42.34	40.01	44.67	52.23	50.26	51.71	53.16	54.57	52.0
12	42.56	47.86	40.23	46.91	40.19	50.98	38.78	51.66	80.70	79.06	81.9
13	93.73	92.97	95.28	90.75	96.30	95.28	92.12	94.01	99.06	98.87	98.7
14	80.00	81.03	71.25	76.89	81.73	68.08	73.05	86.00	89.10	89.46	91.1
15	46.08	44.08	45.00	38.63	34.47	25.53	31.42	44.58	62.87	62.47	61.4
16	96.42	95.33	96.84	94.63	94.13	97.89	94.74	97.79	97.05	97.05	98.3
OA	54.34	57.02	54.76	54.68	54.44	60.21	57.26	60.40	71.05	71.12	71.3
AA	67.33	67.73	66.55	66.16	64.68	70.36	67.40	72.27	81.81	81.98	82.0
κ	49.06	51.99	49.53	49.53	48.92	55.45	52.25	55.81	67.61	67.65	67.8

TABLE V University of Pavia ROSIS Image

Class	RF	RoF	RF+PCA	RF+KPCA	SVM	SVM+PCA	SVM+KPCA	RoRF-PCA	RoRF-KPCA		
									Linear	RBF	Poly
1	65.56	72.45	58.99	60.75	81.04	74.00	69.91	77.59	79.01	75.72	76.0
2	99.99	99.98	99.71	99.93	99.92	99.84	99.87	100.00	99.97	99.97	99.9
3	99.15	99.56	98.21	98.05	99.19	99.31	99.08	99.50	99.37	99.12	99.3
4	65.81	81.88	58.63	63.21	71.41	71.99	58.75	80.63	78.27	77.69	77.4
5	82.25	90.67	91.48	86.77	93.33	90.79	91.86	88.38	89.88	89.93	91.9
6	56.11	70.80	67.34	63.70	69.28	70.38	55.42	70.52	72.91	72.90	74.7
7	54.02	60.30	57.22	57.57	59.27	72.68	51.34	61.53	58.33	66.88	68.0
8	68.35	68.96	71.57	70.78	67.50	68.25	62.23	69.78	70.55	70.72	71.1
9	84.51	82.83	83.26	84.76	90.07	87.93	88.65	90.13	90.89	88.18	90.7
OA	64.89	74.77	69.64	68.35	73.73	74.20	63.90	75.19	76.16	76.17	77.3
AA	75.08	80.83	76.46	76.17	81.22	81.69	75.36	82.01	82.13	82.35	83.2
κ	56.61	68.31	61.63	60.26	66.94	67.43	55.53	68.73	69.79	69.79	71.1

The accuracies of member classifiers and diversity within the ensemble are two essential components in MCSs. Here, we present OAs, AOs, and diversities in Tables VI and VII. AOs and diversities of RoRF-KPCA are significantly better than the ones of RF and RoF for AVIRIS image. For ROSIS image, RoRF-KPCA get slightly higher values of AOs than RoF-PCA, although RoRF-KPCA are

significantly better RoRF-PCA. The main reason could be related to a higher presence of nonlinear data structures in AVIRIS image due to the lower spatial resolution, and thus a higher presence of mixtures (i.e., vegetation and soil) and more variability (i.e., vegetation with different level of hydric stress) in the scene. Hence, the use of KPCA in RoRF can better capture the information, leading to high precision classification result.

TABLE VI Indian Pines AVIRIS Image

Classifiers	RF	RoF	RoRF-PCA	RoRF-KPCA		
				Linear	RBF	Polynomial
OA	54.34	57.02	60.40	71.05	71.12	71.30
AOA	45.21	47.86	54.13	62.81	62.83	63.11
Diversity	0.34	0.42	0.44	0.55	0.55	0.56

TABLE VII University of Pavia ROSIS

Classifiers	RF	RoF	RoRF-PCA	RoRF-KPCA		
				Linear	RBF	Polynomial
OA	64.89	74.77	75.19	76.16	76.17	77.32
AOA	57.14	70.82	71.74	72.14	72.03	72.68
Diversity	0.46	0.57	0.59	0.59	0.60	0.60

To investigate the impact of the number of features in a subset (M) on the classification accuracy (see Fig. 1), tests are performed considering different values of M (10, 25, 50, and 100 for AVIRIS image, and 10, 25, and 50 for ROSIS image). For AVIRIS image, a larger M gain higher OAs of RoRF-PCA. In contrast, RoRF-KPCA provides greater accuracy considering fewer features. For ROSIS image, when M becomes larger, RoRF-PCA and RoRF-KPCA tends to have worse performances. However, for both images, RoRF-KPCA yield the best classification results for small values of M .

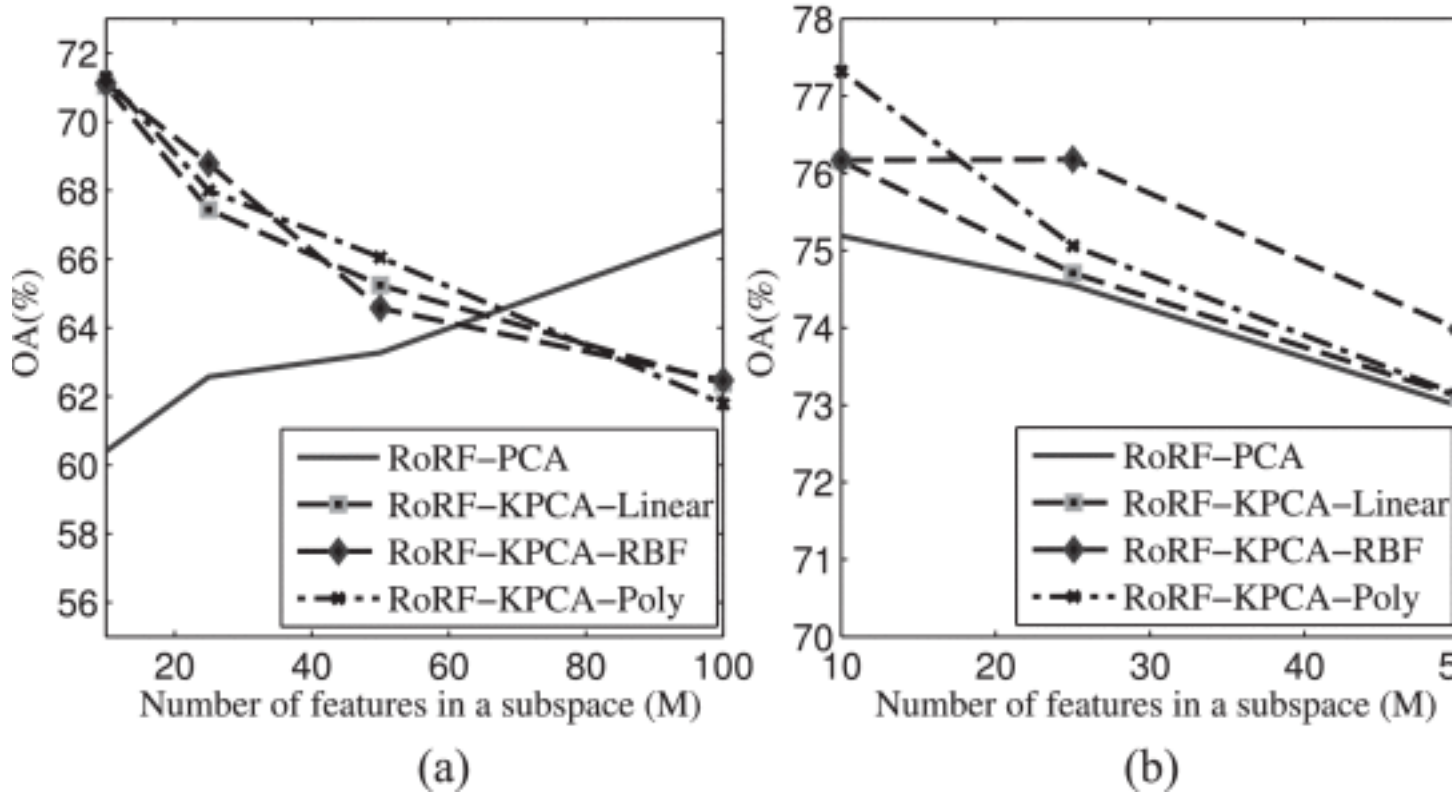


Fig. 1.

Sensitivity to the change of number of features in a subset (M). (a) Indian Pines AVIRIS. (b) University of ROSIS.

[View All](#)

Furthermore, we analyzed the sensitivity analysis of different bootstrap sizes in RoF. When we use a small size, there is insufficient training sample for constructing reliable rotation matrix to generate the accurate result but the diversity is promoted. When we use the total size, we can get the accurate result, but the diversity is reduced. Therefore, we select a moderate size (75%) to balance the diversity and accuracy.

Other parameters, such as the number of trees (T) in RF and number of RFs (T), have a small and nonsignificant effect on the OA. Hence, parameter selection is not very critical for the proposed method, which is a significant added advantage. In practice, the users might select a small value of M as in our cases to improve the diversity and the accuracy of member classifier to increase the classification capability.

D. Discussion

Rotation-based is the effective strategy to construct the ensemble because it enhances both the accuracy of the member classifiers and the diversity. The first rotation-based ensemble is the RoF [27]. We investigated RoF with PCA, MNF, ICA, and LFDA for hyperspectral classification and RoF-PCA obtained the best results in [31]. Then, we integrated RoF with local feature extraction and MRFs in [26], and rotation-based ELM and RoRF with EMAPs in [23]. SVM was

investigated in a rotation-based architecture with random projection, to deal with limited training set in [32]. We proposed RoFcs to alleviate the traditional RoF that performs data transformation on the training samples of each subset in [33].

Although RoF and its improvements achieved remarkable performance, they just consider the linear data transformation. However, linear relationship between spectral features does not reflect the real data behavior. A nonlinear kernel version (e.g., KPCA) was introduced to circumvent this limitation. Based on this assumption, we aim at improving the performance of RF by using a rotation-based ensemble, where the KPCA is introduced as the data transformation method. Experimental results indicated that the proposed RoRF-KPCA could gain the better performance than other compared methods. An additional advantage is that the selection of the parameter is not critical. The users can select a small value of the number of features in a subset (M) to increase the classification performance. In RoF [26] and RoRF-PCA [23], the optimal value of M depends on the datasets. The main limitation of the proposed method is the high computation time, due to the high computational complexity of calculating the kernel matrices.

SECTION V.

Conclusion

Hyperspectral image classification was addressed by exploiting a rotation strategy applied to RF (base classifier) and KPCA (data transformation). Two hyperspectral datasets characterized by a different context (urban and agricultural areas) were used for validation assessment. The obtained results confirmed the high capability of the proposed method, in particular for Indian Pine dataset, which is characterized by a low resolution and a high presence of mixed classes.

For the proposed approach, parameter tuning is needed, however, different kernel functions (linear, RBF, and Polynomial) provide very similar results, making this choice not critical in this context. The only parameter that affects the performance is the number of features in a subset (M). In this case, a small value of M would be preferable.

In future research, we will utilize semisupervised methods [34] and spatial features in our classification process. Furthermore, the optimization to reduce the computational burden and classification problems will be investigated.

ACKNOWLEDGMENT

The authors would like to thank Prof D. Landgrebe from Purdue University, USA and Prof P. Gamba for providing the hyperspectral remote sensing images.