Three-dimensional discrete cosine transform-based feature extraction for hyperspectral image classification

Manoharan Prabukumar
Shrutika Sawant
Sathishkumar Samiappan
Loganathan Agilandeeswari

Three-dimensional discrete cosine transform-based feature extraction for hyperspectral image classification

**Manoharan Prabukumar,[a,][*] Shrutika Sawant,[a]
Sathishkumar Samiappan,[b] and Loganathan Agilandeeswari[a]**
[a]Vellore Institute of Technology, School of Information Technology and Engineering,
Vellore, India
[b]Mississippi State University, Geosystems Research Institute,
Starkville, Mississippi, United States

Abstract. The hyperspectral remote sensor acquires hundreds of contiguous spectral images, resulting in large data that contain a significant amount of redundant information. This high-dimensional and redundant data always influence the efficiency of the data processing. Therefore, feature extraction becomes one of the critical tasks in hyperspectral image classification. A transform-domain-based feature extraction technique, three-dimensional discrete cosine transform (3-D DCT), is proposed. The reason behind the transform domains is that, generally, an invertible linear transform reconstructs the image data to provide the independent information about the spectra or more separable transformation coefficients. Moreover, DCT has excellent energy compaction properties for highly correlated images, such as hyperspectral images, which reduces the complexity of the separation significantly. Unlike the discrete wavelet transform that requires sequential transform to obtain the approximation and detailed coefficients, DCT extracts all coefficients simultaneously. As a result, computation time in the feature extraction can be reduced. The experimental results on three benchmark datasets (Indian Pines, Pavia University, and Salinas) show that the proposed approach produces a good classification in terms of overall accuracy, average accuracy as well as Cohen's kappa coefficient (κ) when compared with some traditional as well as transform-based feature extraction algorithms. Experimental result also shows that the proposed method requires less computational time than the transform-based feature extraction method. © *2018 Society of Photo-Optical Instrumentation Engineers (SPIE)* [DOI: 10.1117/1.JRS.12.046010]

Keywords: hyperspectral image; feature extraction; three-dimensional discrete cosine transform; image classification.

Paper 180631 received Jul. 27, 2018; accepted for publication Sep. 25, 2018; published online Oct. 23, 2018.

1 Introduction

Over the past decade, hyperspectral imaging is of widespread interest among the remote sensing research community due to its ability to discriminate between the variety of ground objects.[1] The hyperspectral data consist of rich information in both spectral and spatial domains, which has opened opportunities in numerous diverse field applications, such as land cover classification,[2] target detection,[3] tree species classification,[4] food technology,[5] and medical imaging.[6] The hyperspectral data present a very difficult challenge caused by a large number of narrow spectral bands with a small number of available labeled training samples. This problem along with other difficulties, such as high variations of the spectral signature from identical material, high similarities of the spectral signatures between some different materials, and noise from the sensors and environment, will significantly decrease the classification accuracy. Therefore, feature

---

*Address all correspondence to: Manoharan Prabukumar, E-mail: mprabukumar@vit.ac.in

extraction is an essential task in hyperspectral image processing to explore the hidden discriminant features of hyperspectral data that are useful for the classification and in turn increases the classification accuracy.

Researchers have proposed various feature extraction techniques in the past few years for the extraction of features from the hyperspectral images. Feature extraction is the transformation of the original feature space into a new set of coordinates or features.[7] The feature extraction process preserves the most informative contents of the original high-dimensional feature space. Principle component analysis (PCA) is one of the most commonly used feature extraction techniques.[8,9] This is because the PCA is an invertible transformation, which makes easy to interpret the extracted features. PCA finds the projections with lower reconstruction error for the whole data. It works on the global features and ignores the local information. Hence, segmented PCA (SPCA) is proposed as an extended version of PCA where PCA is applied to the blocks that are composed of correlation between bands to use local information.[10] As an extension to PCA, modified algorithms are proposed, such as maximum noise fraction (MNF)[11] and kernel PCA (KPCA).[12] Probabilistic PCA (PPCA) is a generative latent variable model in connection with maximum likelihood function, which is also used to extract features.[13] Independent component analysis[14] is then proposed to extract class discriminant features. Another best-known feature extraction approach is linear discriminant analysis (LDA).[15] LDA finds the projections that preserve the most discriminative information. Many other extensions to the above-mentioned two approaches have been developed, such as regularized LDA,[15] nonparametric weighted feature extraction (NWFE),[16] and kernel NWFE.[17] The PCA- and LDA-based methods assume that the distribution of the samples in a class is Gaussian; however, sample distribution is not always Gaussian, and sometimes it may have complex multimodal structure. Therefore, locality preserving feature extraction methods have emerged, which includes local Fisher's discriminant analysis (LFDA) and locality-preserving projection (LPP). In Refs. 18 and 19, random feature selection (RFS)-based methods are developed to explore the diverse feature set that leads to the higher classification performance. Clustering-based feature extraction techniques are also widely used for feature extraction, which removes the redundancies and the correlated features;[20,21] however, most of the clustering methods focus on only spectral features rather than exploring the hidden discriminant features.

The above-mentioned approaches are nearly matrix-based approaches or vector-based approaches. However, the original hyperspectral data are represented in a three-dimensional (3-D) volumetric array, which includes two spatial dimensions and one spectral dimension. Therefore, it is more obvious to represent the hyperspectral data as a 3-D cube or tensor[22] to preserve the higher-order statistical structure. The transform-domain method, 3-D discrete wavelet transform (DWT), is used to extract the texture feature at different scales and frequencies and has achieved the significant classification performance.[23,24] Recently, deep learning techniques have emerged too as the most powerful methods for feature extraction of the hyperspectral data. Deep learning techniques that are extensively used for feature extraction includes deep belief network,[25] stacked autoencoder,[26] convolutional neural network (CNN),[27] and recurrent neural network.[28] A 3-D convolutional neural network (3-D CNN) framework is proposed in Ref. 29 to extract the deep spectral–spatial features. It is observed that the tensor-based or 3-D methods provide the significant performance as the joint spectral–spatial structure is adequately preserved. Although the deep learning methods provide the significant deep feature representation of the high-dimensional data that can improve the classification performance of the system, it also increases the computation time and the complexity of the algorithm.

From the study of various existing feature extraction techniques, we found the following challenges such as:

a. The existing feature extraction methods fail to explore the hidden discriminant features as well as to provide the more complementary features while reducing the redundant information.

b. Most of the existing feature extraction methods fail to provide the promising results when the number of labeled samples is limited.

c. When dealing with the high dimensional data, some existing methods demand high computational cost.[30]

d. Even though the existing transform domain methods have achieved the significant classification performance, it takes more computational time.[23,31]

In this work, three-dimensional discrete cosine transform (3-D DCT) for classification of hyperspectral images is proposed. DCT exhibits excellent energy compaction properties, and large DCT coefficients are located in the low-frequency region. Therefore, DCT is chosen for feature extraction in hyperspectral image classification. DCT extracts highly discriminative and informative features from the hyperspectral images. The proposed method transforms the hyperspectral image into a DCT coefficient matrix and looks for a signature pattern in the DCT domain for classifying different land cover classes. Further, support vector machine (SVM) classifier is used to obtain labels of unknown samples of the hyperspectral images. To the best our knowledge, this is the first time where DCT is used for feature extraction in hyperspectral image classification. This technique has shown very distinct features that are more suitable for hyperspectral classification, including high classification accuracy and computational efficiency. The main contribution of this paper can be summarized as follows:

a. Distinct features are extracted from hyperspectral image data as DCT captures local variation present in the hyperspectral data, which increases discrimination among the different land cover classes.

b. DCT involves computation of real data only. Hence, the proposed method significantly reduces the computational load without compromising the overall classification accuracy.

c. The proposed method has shown the distinct properties that are extremely suitable for hyperspectral image classification including exploration of extrinsic discriminant features, high computational efficiency, and very high classification accuracy.

The rest of the paper is arranged as follows: overview of feature extraction using 3-D DCT is given in Sec. 2. Section 3 deals with experimentation on standard benchmark datasets and discusses the findings of the experiments, and finally, Sec. 4 presents the conclusion and future directions.

2 Proposed Three-Dimensional Discrete Cosine Transform-Based Feature Extraction Framework

In this section, the proposed feature extraction framework for hyperspectral image classification is explained in detail. As shown in Fig. 1, the proposed approach consists of two stages such as feature extraction and classification. The following subsection deals with the detailed explanation of the various stages present in the proposed system.

Consider the hyperspectral image dataset that is represented as, $X \in R^{H \times W \times N}$, where H and W be the height and width of hyperspectral image and N is a total number of spectral bands or the feature dimension. Assume the training samples of hyperspectral image data as $x = [x_1, x_1, \ldots, x_M]$, $y = [y_1, y_1, \ldots, y_M]$, are the labels of training samples that belong to the k classes in the data that are denoted as $\Omega = [\Omega_1, \Omega_2, \ldots, \Omega_k]$.
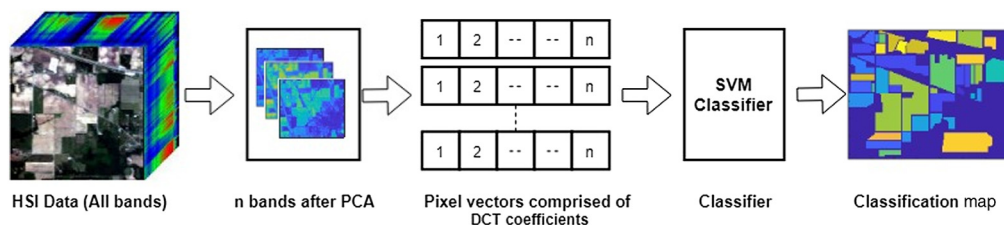

HSI Data (All bands)   n bands after PCA   Pixel vectors comprised of DCT coefficients   Classifier   Classification map

Fig. 1 Block diagram of the proposed feature extraction framework.

2.1 Principal Component Analysis

Principal component analysis (PCA) is widely used in the image preprocessing step to reduce the dimension and redundancy. PCA reduces the dimension of the image by choosing information only from the significant bands. It uses a vector space transform, which reduces the dimensionality of the original dataset and can be interpreted as a dataset with fewer variables, called principal components (PCs).[32]

Let us consider the hyperspectral image $X \in R^{H \times W \times N}$, where H and W be the height and width of the hyperspectral image and N is the total number of spectral bands or the feature dimension, respectively. The pixel vector of a hyperspectral image is represented as follows:

$$\mathbf{x_i} = [x_1, x_2, \ldots x_N]^T. \tag{1}$$

The mean μ of all image pixel vectors can be written as follows:

$$\mu = \frac{1}{M} \sum_{i=1}^{M} \mathbf{x_i}, \tag{2}$$

where $M = H \times W$ denotes the total number of pixels in a spectral band.

The covariance matrix can be given as follows:

$$\mathbf{C_M} = \frac{1}{M} \sum_{i=1}^{M} (\mathbf{x_i} - \mu)(\mathbf{x_i} - \mu)^T. \tag{3}$$

The eigenvalue decomposition of the covariance matrix is follows:

$$\mathbf{C_M} = \mathbf{BDB^T}, \tag{4}$$

where $\mathbf{D}$ is the diagonal matrix composed of eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_N$ of covariance matrix $\mathbf{C_M}$ and $\mathbf{B}$ is the orthonormal matrix comprised of N eigenvectors and is given as follows:

$$\mathbf{B} = (b_1, b_2, \ldots, b_N). \tag{5}$$

Then linear transformation can be calculated as follows:

$$\chi_\mathbf{i} = \mathbf{B_j^T x_i}, \qquad i = 1, 2, \ldots, M, \tag{6}$$

which is the PCA pixel vector and j is the number of principal components with the highest eigenvalues. Each pixel vector is mapped similarly, and the new low-dimensional image is obtained.

2.2 Three-Dimensional Discrete Cosine Transform

Discrete cosine transform (DCT) is one of the most widely used techniques in numerous areas of image processing including the denoising and compression.[33] Due to energy compaction property of DCT, the image information is represented using a few DCT coefficients. Thus, making DCT more suitable for image compression applications. As DCT is a linear and invertible transformation, it can provide easy separation of the transformation coefficients. The extracted independent transformation coefficients give a meaningful data structure that allows extracting information at a finer level of precision. The favorable outcome of such transformation is the removal of the interpixel redundancy as well as interband redundancy. In this paper, 3-D DCT is applied to the hyperspectral cube, which encodes the information

in the form of DCT coefficients. It should also be noted that 3-D DCT can be achieved by applying two-dimensional (2-D) DCT to a pixel vector. The 2-D DCT $f(m, n)$ of size $M \times N$ is given as follows:

$$f(u, v) = \propto_u \propto_v \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(m, n) \cos\left[\frac{\pi(2m+1)u}{2M}\right] \cos\left[\frac{\pi(2n+1)v}{2N}\right], \qquad (7)$$

where $0 \leq u \leq M - 1$, $0 \leq v \leq N - 1$ and

$$\propto_u = \begin{cases} \sqrt{1/M} & \text{if } u = 0 \\ \sqrt{2/M} & \text{otherwise} \end{cases}, \qquad \propto_v = \begin{cases} \sqrt{1/N} & \text{if } v = 0 \\ \sqrt{2/N} & \text{otherwise} \end{cases}.$$

To extract the features from the low-dimensional hyperspectral image, 2-D DCT is applied to each pixel vector of low-dimensional image and coefficients of DCT are computed. One of the main characteristics of DCT is its ability to convert the energy of the image into a few coefficients. Thus, in the field of pattern recognition, DCT coefficients are widely used as features.

The DCT coefficients of each pixel at position (m, n) of the low-dimensional hyperspectral image can be directly concatenated to form its feature vector:

$$\mathbf{x_{m,n}} = [f_1(m, n, \cdot), f_2(m, n, \cdot), \ldots, f_j(m, n, \cdot)], \qquad (8)$$

$$\hat{\mathbf{f}}_\mathbf{j} = E(\mathbf{x_{m,n}}) = E[f_1(m, n, \cdot), f_2(m, n, \cdot), \ldots, f_j(m, n, \cdot)], \qquad (9)$$

where $E(\cdot)$ is the expectation operator.

Let $\hat{\mathbf{f}} \in R^{H \times W \times j}$ be the final concatenated cube of the 3-D DCT-based feature vector is given as

$$\hat{\mathbf{f}} = (\hat{\mathbf{f}}_1, \hat{\mathbf{f}}_2, \ldots, \hat{\mathbf{f}}_\mathbf{j}). \qquad (10)$$

2.3 Support Vector Machine Classifier

SVM has been widely used in the classification of hyperspectral image because of its particular advantages in solving problems about small-sized samples training, nonlinear, and high dimensions.[34,35] In this paper, the SVM classifier is employed to get the final classification map. Assume the training samples of hyperspectral image data as $x = [x_1, x_1, \ldots, x_M]$, $y = [y_1, y_1, \ldots, y_M]$, are the labels of training samples that belong to the k classes in the data that are denoted as, $\Omega = [\Omega_1, \Omega_2, \ldots, \Omega_k]$ and a nonlinear kernel mapping $(\cdot)$, x_i is a pixel vector with j-dimensional spectrum. The SVM technique solves

$$\min_{W, \xi, b} \left\{ \frac{1}{2} \|W\|_2^2 + C \sum_i \xi_i \right\}. \qquad (11)$$

Constrained to

$$y_i[\phi^T(x_i).w + b] \geq 1 - \xi_i, \qquad \forall \, i = 1, \ldots, l, \qquad (12)$$

where w is normal to the decision hyperplane, b represents the closest distance to the origin of the coordinate system, and l denotes the number of samples. Parameter C is the regularization parameter, which controls the generalization capabilities of the classifier, and ξ_i are the positive slack variables.

Figure 1 shows a flowchart of the proposed technique, and the entire process is summarized in Algorithm 1.

Algorithm 1 3-D DCT-based hyperspectral image classification.

Input: Hyperspectral image $X \in R^{H \times W \times N}$, k number of classes.

Output: Labels y.

1. Obtain low-dimensional image χ by applying PCA to hyperspectral image data X, $j < N$ [Eq. (6)];

2. Apply 3-D DCT to low-dimensional image χ_j and obtain the DCT coefficients for each pixel vector [Eq. (7)] and obtain DCT coefficient pixel vector [Eq. (8)];

3. Obtain the mean of DCT coefficient of each pixel vector as

$$\hat{\mathbf{f}}_{\mathbf{j}} = E[f_1(m, n, \cdot), f_2(m, n, \cdot), \dots f_j(m, n, \cdot)];$$

4. Obtain final feature vector of DCT coefficient as, $\hat{\mathbf{f}} = (\hat{\mathbf{f}}_1, \hat{\mathbf{f}}_2, \dots \hat{\mathbf{f}}_{\mathbf{j}}) \epsilon R^{H \times W \times j}$, [Eq. (10)];

5. Randomly select some samples in $\hat{\mathbf{f}}$ as training samples and use the remaining samples as testing samples;

6. Train the SVM classifier using training samples;

7. Predict class labels for testing samples and get the classification map.

---

3 Experimentation

In this section, to assess the effectiveness of the proposed method, a series of experiments on three standard datasets were conducted, namely Indian pines, Pavia University, and Salinas dataset.[36] All the experiments are conducted using MATLAB 2018a on PC with 16 GB RAM and 2.70 GHz CPU. To verify the efficacy of the proposed method, few traditional feature extraction methods were considered for comparison. Widely studied methods such as SVM,[34] SVM-PCA,[9] ICDA,[37] and LDA[38] were compared. The 3-D DWT, a transform-based feature extraction method, is also considered.[23] For the SVM method, the original hyperspectral image is directly used for the classification without any feature extraction step.

3.1 Dataset Description

a. The first dataset is Indian Pine dataset, which is captured by Airborne Visible Infrared Imaging Spectrometer (AVIRIS) over North–Western Indiana region in June 1992. This dataset contains 16 classes of agriculture as well as vegetation species. The size of the dataset is 145×145 pixels with 20-m spatial resolution and 10-nm spectral resolution over the range of 400 to 2500 nm. This scene contains 224 spectral reflectance bands, where only 204 bands will remain for experimentation after the removal of water absorption bands.

b. The second dataset is University of Pavia dataset, which is captured by Reflective Optical System Imaging Spectrometer (ROSIS) over Pavia, Northern Italy in July 2002. This dataset contains nine different classes. The size of the dataset is 610×340 pixels with 1.3-m spatial resolution over the range of 430 to 860 nm. This scene contains 103 spectral reflectance bands.

c. The third dataset is Salinas dataset, which is captured by Airborne Visible Infrared Imaging Spectrometer (AVIRIS) over Salinas Valley, California. This dataset contains 16 different classes. The size of the dataset is 512×217 pixels with 3.7-m spatial resolution over the range of 400- to 2500-nm range. This scene contains 224 spectral reflectance bands.

3.2 Performance Metrics

The performance of the proposed method is compared with other competing methods using three widely used quality metrics, i.e., overall accuracy, average accuracy, class-wise accuracy, and kappa coefficient. Overall accuracy (OA) is the percentage of correctly classified pixels in the

whole scene. Average accuracy (AA) is the mean of percentage of correctly labeled pixels for each class. Classwise accuracy is also known as producer's accuracy. Kappa coefficient is a robust measure of the degree of agreement, which integrates diagonal and off-diagonal entries of the confusion matrix.

3.3 *Parameter Setting*

In the beginning, to evaluate the effectiveness of the proposed method with less amount of labeled data, 20% samples for each class from the reference data of Indian pine dataset, Pavia University dataset, and Salinas dataset are randomly chosen as training samples, and the remaining samples in each class are used for testing purpose. This experiment is repeated for 10 times to evaluate an average of OA, AA, and κ. The training and testing samples used for conducting tests are shown in Table 1–3. Also, some parameters need to be tuned for the conduction of tests. For all the SVM-based methods, the penalty parameter C and the radial basis function (RBF) parameter γ are tuned through fivefold cross-validation ($\gamma = 2^{-3}, 2^{-2}, \ldots, 2^2, 2^3$, $C = 2^1, 2^2, \ldots, 2^8$). Also, few other parameters of these methods need to be tuned. For the proposed technique, the RBF parameter γ and the penalty parameter C are tuned as the methods above.

For SVM-PCA, 25 principal components (PCs) have obtained best classification accuracy. So, in this experiment, the number of PCs is set to 25.[10] The number of independent components (ICs) are selected such that it could give a better result and have lesser computation burden. Also,

Table 1 Details of Indian pines dataset including some classes, class name, training, testing, and the total number of samples.

Indian Pines				
Class		Samples		
No	Name	Train	Test	Total
1	Alfalfa	10	36	46
2	Corn-no till	286	1142	1428
3	Corn-min till	166	664	830
4	Corn	48	189	237
5	Grass-pasture	97	386	483
6	Grass-tree	146	584	730
7	Grass-pasture-mowed	6	22	28
8	Hay-windrowed	96	382	478
9	Oat	4	16	20
10	Soybean-no till	195	777	972
11	Soybean-min till	491	1964	2455
12	Soybean-clean	119	474	593
13	Wheat	41	164	205
14	Woods	253	1012	1265
15	Buildings-grass-trees-drives	78	308	386
16	Stone-steel-towers	19	74	93
Total		2055	8194	10,249

Table 2 Details of Pavia University dataset including number of classes, class name, training, testing, and the total number of samples.

Pavia University				
Class		Samples		
No.	Name	Train	Test	Total
1	Asphalt	1327	5304	6631
2	Meadows	3730	14,919	18,649
3	Gravel	420	1679	2099
4	Trees	613	2451	3064
5	Painted metal sheets	269	1076	1345
6	Bare soil	1006	4023	5029
7	Bitumen	266	1064	1330
8	Self-blocking bricks	737	2945	3682
9	Shadows	190	757	947
Total		8558	34,218	42,776

Table 3 Details of Salinas dataset including number of classes, class name, training, testing, and the total number of samples.

Salinas				
Class		Samples		
No	Name	Train	Test	Total
1	Broccoli-green-weeds-1	402	1607	2009
2	Broccoli-green-weeds-2	746	2980	3726
3	Fallow	396	1580	1976
4	Fallow-rough-plow	279	1115	1394
5	Fallow-smooth	536	2142	2678
6	Stubble	792	3167	3959
7	Celery	716	2863	3579
8	Grapes-untrained	2255	9016	11271
9	Soil-vinyard-develop	1241	4962	6203
10	Corn-senesced-green-weeds	656	2622	3278
11	Lettuce-romaine-4wk	214	854	1068
12	Lettuce-romaine-5wk	386	1541	1927
13	Lettuce-romaine-6wk	184	732	916
14	Lettuce-romaine-7wk	214	856	1070
15	Vinyard-untrained	1454	5814	7268
16	Vinyard-vertical-trellis	362	1445	1807
Total		10,833	43,296	54,129

it has been observed that a lesser or greater number of ICs may have redundant information. So as per Ref. 37, the number of ICs is set to 18.

3.4 Classification Results

This section discusses the classification results obtained for Indian pines dataset, Pavia University dataset, and Salinas dataset, the impact of the different proportions of training samples on overall accuracy and execution time taken by all competing method.

First, we illustrate how DCT influenced the original spectra of the real hyperspectral datasets. Here, the original spectra of hyperspectral datasets, such as Indian Pines, Pavia University, and Salinas dataset, are shown in Figs. 2(a), 2(d), and 2(g), respectively. The transformed output spectra after PCA transformation on hyperspectral datasets, such as Indian Pines, Pavia University, and Salinas dataset, are shown in Figs. 2(b), 2(e), and 2(h), respectively. The transformed output spectra after applying DCT on hyperspectral datasets, such as Indian Pines, Pavia University, and Salinas dataset, are shown in Figs. 2(c), 2(f), and 2(i), respectively. The spectral curves shown in Figs. 2(b), 2(e), and 2(h) indicate the high correlation between various classes of hyperspectral images, which influences the discrimination among the classes. Figures 2(a), 2(d), and 2(g) indicates the original spectral curves of datasets. It shows slightly more separation between the land cover classes. However, these curves are obtained by considering all available spectral bands, which lead to heavy computations. However, in Figs. 2(c), 2(f), and 2(i), the spectral responses of land cover classes look more separated from each other in DCT domain, which directly influences the performance of the hyperspectral image classification.
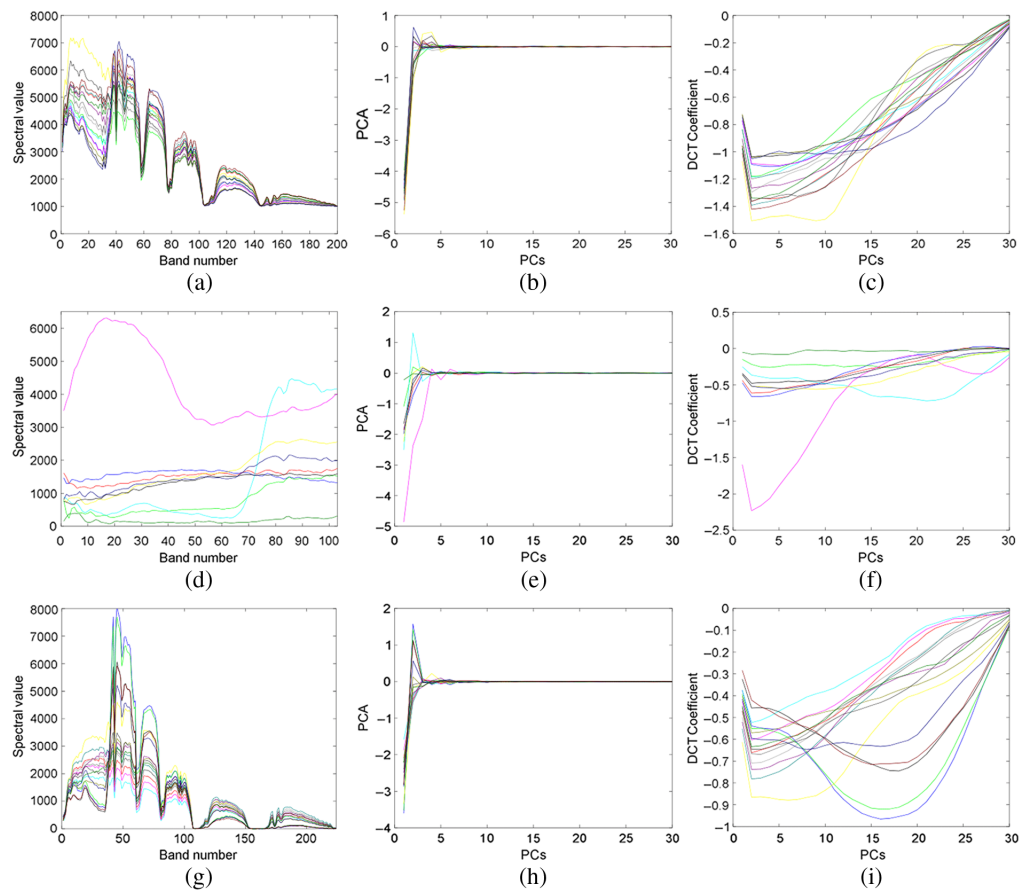

Fig. 2 Spectral response of different land cover classes of (a–c) Indian Pines dataset, (d–f) Pavia University dataset, and (g–i) Salinas dataset, before and after transformation. First column depicts original spectral response of respective datasets; the second column depicts spectral response of respective datasets after PCA transformation and third column depicts spectral response of respective datasets after discrete cosine transformation (DCT).

3.4.1 Result analysis by comparing the proposed method with different classification methods on Indian Pines dataset

The information required for experimentation, such as ground-truth data, training sample map, and testing sample map of Indian pines dataset is shown in Fig. 3. The classification map of all competing techniques on Indian pines dataset is shown Fig. 4, and the classification results (i.e., OA, classwise accuracy, AA, and κ) of all competing methods and the proposed method are shown in Table 4.

The performance of the proposed method is compared with traditional methods such as SVM, SVM-PCA, LDA, ICDA, and transform-based method such as 3-D DWT. From Table 4, it can be shown that the proposed technique attains greatest performance in terms of overall accuracy, average accuracy, classwise accuracy as well as the κ. For the PCA-based classification algorithm, the original image was reduced into few principal components that are then used for classification using SVM classifier. The PCA-based classification technique decreases the dimensionality of hyperspectral images in the spectral domain. However, it increases the discrepancy in the spatial domain (i.e., texture or shape variation). Therefore, the classification accuracies of the SVM-PCA-based method are not solely better for Indian pines dataset. By exploiting spectral–spatial features, 3-D DWT has achieved better performance in terms of OA, AA, and κ over all other competing methods, such as SVM, SVM-PCA, ICDA, and LDA. The proposed approach shows excellent and comparable classification performance due to the application of 3-D DCT features. The classification map of SVM-, SVM-PCA-, ICDA-, and LDA-based approaches have shown some salt and pepper noise that is less visible in the DWT method and proposed 3-D DCT method. This noise will disappear if the spatial information is considered for classification along with spectral information.

When compared with other competing approaches, the proposed approach improves the classification accuracy significantly as shown in Table 4 (boldface). For instance, the classification
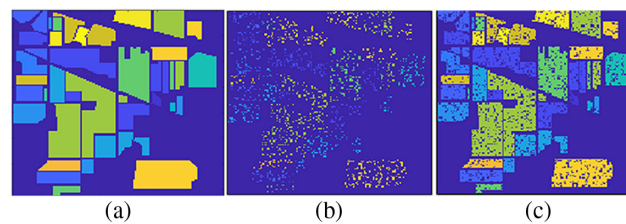

Fig. 3 Indian pine dataset information (a) ground-truth data, (b) training map, and (c) testing map.
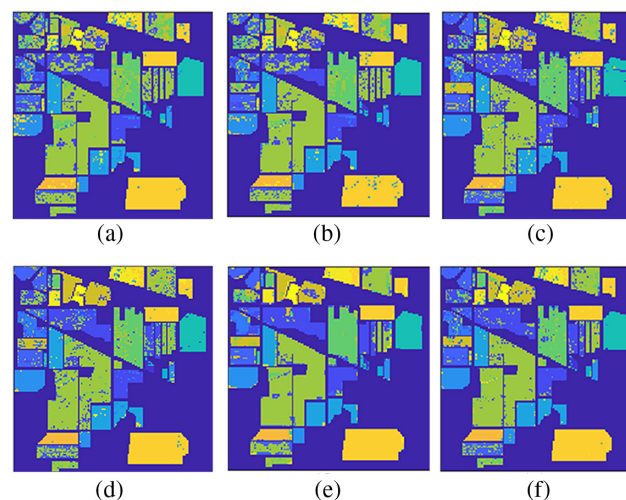

Fig. 4 Classification map of Indian pine data for all competing methods (a) SVM, (b) ICDA, (c) SVM-PCA, (d) LDA, (e) 3-D DWT, and (f) 3-D DCT.

Table 4 Comparison of classification accuracies (%) obtained by proposed method with competing methods for Indian pines dataset.

Class number	SVM[34]	SVM-PCA[9]	ICDA[37]	LDA[38]	3-D DWT[23]	3-D DCT
1	13.89	72.22	41.67	**75.00**	25.64	69.44
2	47.11	69.70	46.50	71.62	74.11	**75.83**
3	22.29	60.09	33.13	59.33	61.41	**70.63**
4	23.28	44.97	31.75	62.43	53.23	**53.96**
5	71.76	90.41	79.53	87.30	**94.39**	93.26
6	87.33	88.70	**96.06**	92.46	94.51	94.52
7	63.64	72.73	77.27	**86.36**	69.56	40.90
8	98.85	95.81	97.91	97.64	98.27	**98.95**
9	0	25.00	23.4	25.00	**94.11**	37.89
10	40.54	72.46	53.41	58.17	73.72	**79.40**
11	79.94	81.98	82.03	76.78	85.85	**86.59**
12	14.14	57.17	14.14	**78.27**	72.42	77.63
13	87.20	93.29	90.85	**99.39**	86.78	95.73
14	96.74	92.79	94.17	94.86	94.79	**97.43**
15	35.39	54.55	43.51	63.31	**71.64**	59.41
16	83.78	64.86	85.14	83.78	**88.60**	81.08
OA	65.96	77.02	66.84	77.39	81.47	**83.15**
AA	54.12	71.05	60.44	75.73	77.44	**78.36**
K	0.5667	0.7364	0.6131	0.7411	0.7876	**0.8071**

accuracy of classes "Corn-no till" and "Corn-min till" increases from 46.50% to 75.83%, 22.29% to 70.63%, respectively. However, it is observed that the proposed method is not performing well in terms of the classwise accuracy of individual classes such as "Alfalfa," "Grass-pasture-moved," and "Oat" as shown in Table 4. The reason behind the lesser accuracy is that, the classes, such as "Alfalfa," "Grass-pasture-mowed," and "Oats," have a limited number of samples (also called small classes) as shown in Table 1. By selecting 20% samples per class as training samples, these classes are represented by only a few samples in the training set, which probably do not provide a fair-enough representation of the class. For pixel-wise classifier SVM, the training samples are too limited to learn an effective model. Moreover, for classes, such as, "Grass Pasture," "Oat," "Buildings-grass-trees-drives," and "Stone-steel-towers," 3-D DWT method outperforms 3-D DCT method due to the localization property of the 3-D DWT method. 3-D DCT considers only frequency content of the signal and ignores the localized information.

3.4.2 Result analysis by comparing the proposed method with different classification methods on Pavia University dataset

The information used for experimentation such as ground-truth data, training sample map, and testing sample map of Pavia University dataset is shown in Fig. 5. The classification map of all competing techniques on Pavia University dataset is shown Fig. 6, and the classification results (i.e., OA, classwise accuracy, AA, and κ) of all competing methods and the proposed method are
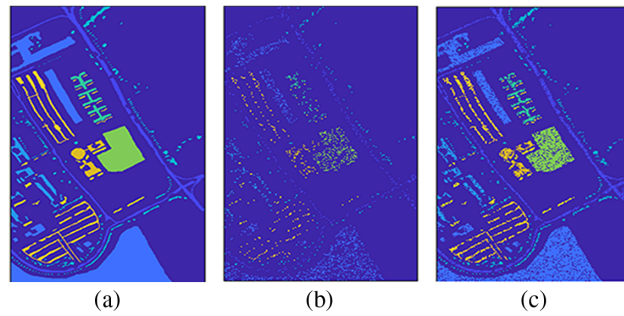
Fig. 5 Pavia University dataset information (a) ground-truth data, (b) training map, and (c) testing map.
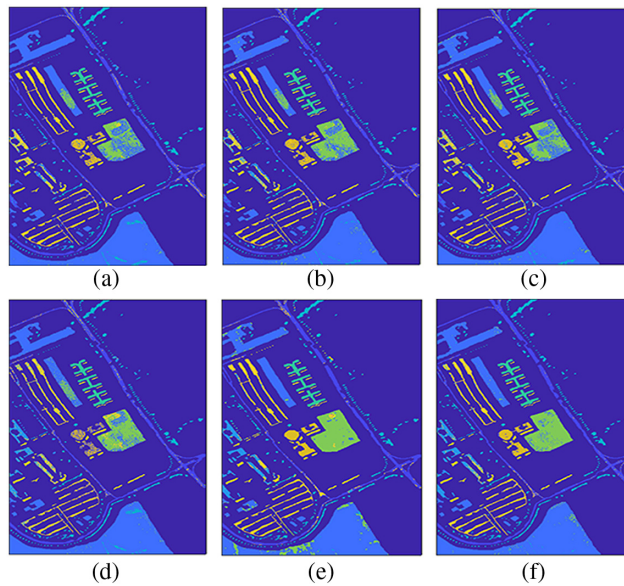

Fig. 6 Classification map of Pavia University data for all competing methods (a) SVM, (b) ICDA, (c) SVM-PCA, (d) LDA, (e) 3-D DWT, and (f) 3-D DCT.

presented in Table 5. From Fig. 6 and Table 5, it can be shown that the proposed technique attains most excellent performance in terms of OA, AA, class wise accuracy as well as κ. Also, it is noted that the traditional feature extraction methods, such as SVM, SVM-PCA, ICDA, and LDA, yield similar results.

Because of the inherent multiresolution approach to the complex data, the transform-based feature extraction method 3-D DWT shows remarkable performance in comparison with traditional feature extraction methods, SVM, SVM-PCA, ICDA, and LDA. However, the proposed methods have achieved better performance over the 3-D DWT method, which means the energy coefficients preserves more complementary information of original feature space. As shown in Fig. 6, the proposed approach can help to eliminate most of the noisy pixels generated by the other methods, and the overall classification accuracy increases by >2%. For example, misclassified pixels from other comparable methods were corrected in the green region at the center of Fig. 6, which is very close to the ground truth and also the overall classification map has become smoother. Compared with other competing approaches, the proposed approach has improved classification accuracy significantly as shown in Table 5 (boldface). For instance, the classification accuracy of class "Asphalt" increases from 88.71% to 95.12%, and the classification accuracy of class "Trees" increases from 79.63% to 95.63%. Moreover, class "Shadow" is identified with 100% accuracy. As shown in Fig. 6, for the proposed method, many pixels of the "Bare soil"
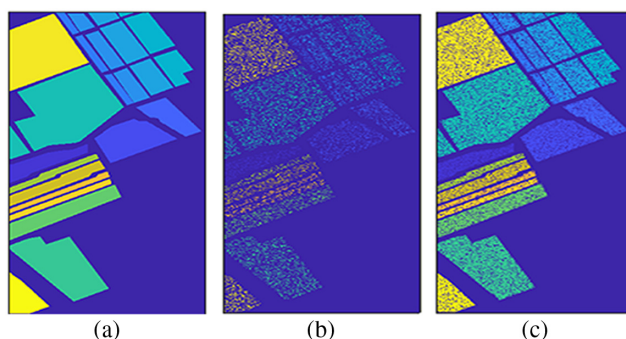
Table 5 Comparison of classification accuracies (%) obtained by proposed method with competing methods for Pavia University dataset.

Class number	SVM[34]	SVM-PCA[9]	ICDA[37]	LDA[38]	3-D DWT[23]	3-D DCT
1	90.18	88.71	89.88	88.78	91.93	**95.12**
2	94.10	95.11	94.47	93.75	92.55	**97.78**
3	14.77	24.00	31.15	65.45	**86.14**	77.96
4	79.60	81.56	82.54	86.08	92.27	**95.63**
5	98.61	98.61	98.70	99.44	98.76	**99.72**
6	43.33	47.28	62.59	63.01	**95.50**	89.71
7	73.78	82.61	78.95	43.79	**93.91**	88.44
8	87.98	87.74	87.71	78.03	**91.02**	90.66
9	99.60	99.87	**100**	99.47	**100**	**100**
OA	80.70	83.23	85.24	84.83	92.73	**94.50**
AA	75.77	78.39	80.66	79.76	93.57	**93.78**
K	0.7508	0.7721	0.8012	0.7972	0.9046	**0.9269**

class are misclassified as "Meadows" class because of the complex structure of the classes. Also, some of the pixels of "Gravel" class are misclassified as other classes, such as "Bitumen" and "Self-blocking bricks." By visual inspection, it is observed that the proposed method produces a more smooth and accurate classification map. For the classes "Gravel," "Bare soil," "Bitumen," and "Self-blocking bricks," 3-D DWT produces better accuracy than 3-D DCT. It is due to the localization property of the 3-D DWT method, whereas 3-D DCT performs a transform of frequency contents only.

3.4.3 Result analysis by comparing the proposed method with different classification methods on Salinas dataset

The information used for experimentation, such as ground-truth data, training sample map, and testing sample map of Salinas dataset, is shown in Fig. 7. Figure 8 shows the classification map of all competing techniques on Salinas dataset and the statistical results (i.e., OA, class wise accuracy, AA, and κ) of all competing methods, and the proposed method is summarized in Table 6. It is clear that the classification map of the proposed method has less noise and is

(a)   (b)   (c)

Fig. 7 Salinas dataset information (a) ground-truth data, (b) training map, and (c) testing map.
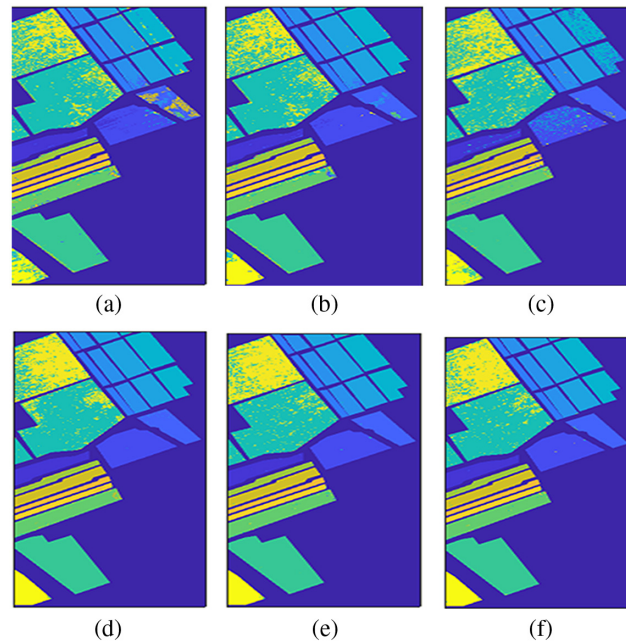
Fig. 8 Classification map of Salinas data for all competing methods (a) SVM, (b) ICDA, (c) SVM-PCA, (d) LDA, (e) 3-D DWT, and (f) 3-D DCT.

Table 6 Comparison of classification accuracies (%) obtained by proposed method with competing methods for Salinas dataset.

Class number	SVM[34]	SVM-PCA[9]	ICDA[37]	LDA[38]	3-D DWT[23]	3-D DCT
1	97.76	94.71	96.33	99.75	99.25	**99.32**
2	88.22	79.63	98.15	**99.93**	**99.93**	99.70
3	52.41	97.46	85.89	97.97	99.62	**99.75**
4	99.55	99.01	98.39	97.75	99.36	**99.46**
5	90.01	96.35	93.46	98.64	**99.07**	98.74
6	97.82	98.64	99.02	99.65	99.61	**99.81**
7	96.23	77.33	98.81	99.75	**99.93**	99.41
8	84.70	83.51	83.88	86.06	90.59	**90.68**
9	95.57	98.73	96.45	99.97	**99.94**	99.68
10	80.05	92.02	80.40	93.93	96.26	**96.91**
11	78.57	94.49	80.80	92.38	97.07	**99.06**
12	99.22	97.72	99.09	**100**	99.94	99.74
13	99.04	92.34	98.36	99.31	98.77	**99.48**
14	87.27	92.05	88.90	91.23	96.38	**98.71**
15	40.45	49.62	44.55	66.73	70.81	**77.90**
16	61.52	85.25	84.71	98.13	99.24	**99.33**
OA	81.55	84.71	85.14	91.61	93.18	**94.62**
AA	84.28	89.31	89.20	95.07	96.33	**97.33**
K	0.7937	0.8292	0.8339	0.9065	0.9239	**0.9400**

more accurate. From Fig. 8 and Table 6, it can be shown that the proposed technique attains the most significant performance in terms of OA, AA, class wise accuracy as well as the κ. Table 6 shows that ICDA and SVM-PCA methods perform better than the SVM method. Furthermore, the LDA method balances both interclass and intraclass criteria using a balancing parameter that outperforms the SVM, SVM-PCA, and ICDA.

Due to the inherent multiresolution property, the transform-based feature extraction method 3-D DWT shows remarkable performance in comparison with the traditional feature extraction methods. However, the proposed methods have achieved better performance over the 3-D DWT method, which means the energy coefficients of the DCT preserves more complementary information of original feature space. As shown in Fig. 8, the proposed approach can help to eliminate most of the noisy pixels generated by the other methods, and the overall classification accuracy increases by >2%. As shown in Fig. 8, overall classification map is very close to the ground truth.

Compared with other competing approaches, the proposed approach has improved the classification accuracy significantly as shown in Table 6 (boldface). It also is shown that the proposed method presents higher performances, especially in classes with a small number of training samples such as "Fallow-rough-plow," "Lettuce-romaine-4wk," "Lettuce-romaine-6wk," and "Lettuce-romaine-7wk." The best classwise accuracy is produced by the proposed method for most of the classes (11 out of the 16 classes). Also, the class "Lettuce-romaine" is correctly identified with 100% accuracy. However, it can be seen that the proposed approach produces slightly lesser classification accuracy for classes such as "Broccoli-green-weeds-2," "Fallow-smooth," "Celery," "Soil-vineyard-develop," and "Lettuce-romaine-5wk," which is almost negligible. The reason is that the DCT coefficients do not consider the localized information about the data.

3.4.4 *Influence of different proportions of training samples on overall accuracy*

To verify the superiority of the proposed method as the number of training samples increases, additional tests are conducted by considering randomly chosen 10%, 20%, 30%, 40%, and 50% training samples[39,40] from each class of all datasets. The remaining samples are used as testing samples. Figure 9 shows OA obtained by the proposed method for a different proportion of training samples, and it is observed that the proposed method achieves a better result as sample proportion increases. Thus, the proposed method obtains the sufficient information to divulge the discriminative features of the hyperspectral data.
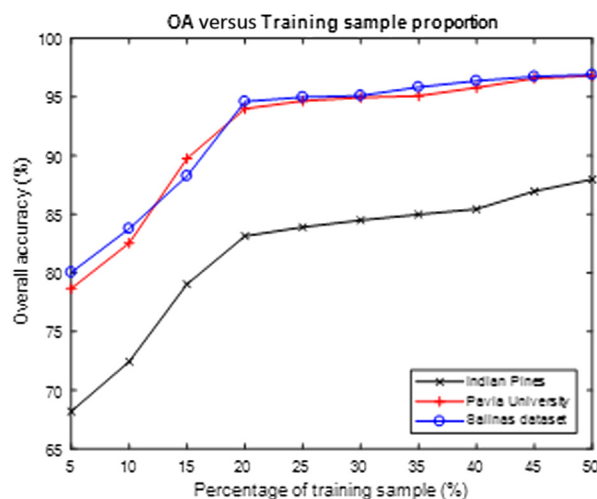

Fig. 9 Influence of different proportions of training samples on (OA) for Indian Pines, Pavia University, and Salinas dataset.
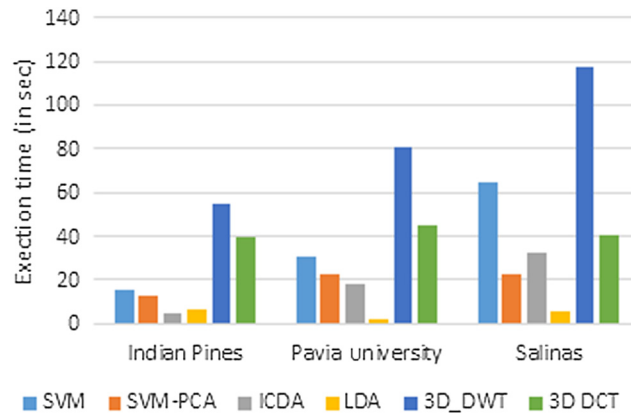
Fig. 10 Execution time (in sec) of all competing methods for all datasets, such as Indian Pines, Pavia University, and Salinas dataset.

3.4.5 *Computational time*

Figure 10 shows the computational time or the execution time (in seconds) of all competing methods for all datasets, such as Indian Pines, Pavia University, and Salinas dataset.

As shown in Fig. 10, the proposed method takes more computation time than traditional feature extraction methods but less than the transform-based feature extraction method. The 3-D DWT approach requires much more computational time, which degrades the competency of the 3-D DWT method when applied for the high dimensionality of data. The reason behind the expensive computations of the DWT method is the recursive computation of approximation and detail coefficients. In contrast, the DCT involves computation of real data only, which reduces computation burden. Also, DCT captures local variation present in the hyperspectral data that increase discrimination between different classes. So, by taking into account the overall accuracy and computational time, the proposed 3-D DCT approach significantly outperforms the other competent methods.

4 Conclusion and Future Work

In this paper, a 3-D DCT-based feature extraction technique for hyperspectral image classification is proposed. This study proved that DCT allows a more efficient representation of the hyperspectral data by removing the redundancy between the neighboring pixels and adjacent bands and provides excellent decorrelation for the hyperspectral images. This technique is beneficial to extract discriminative features from high-dimensional data and is computationally competent. The experimental results on three standard benchmark datasets demonstrate that the proposed technique is more useful in extracting informative features and removing the redundant ones. Experimental results also show that compared with popular feature extraction methods, the proposed technique has significant performance on hyperspectral image classification. The proposed method has achieved a maximum classification accuracy of 94.62% for Salinas dataset.

Although the proposed method is competitive with other state-of-the-art methods, there are still two crucial research directions deserving future attention. First, the spectral information can be integrated with the spatial information, such as edge-preserving filtering,[41] Markov random field,[42] discriminative random field method,[43] and morphological profiles[44] to improve the classification performance further. Second, the computational efficiency of the proposed method will be increased by parallel processing and graphics processing unit programming.

References

1. M. Khodadadzadeh et al., "A new framework for hyperspectral image classification using multiple spectral and spatial features," in *IEEE Geoscience and Remote Sensing Symp.*, pp. 4628–4631, Quebec City, Canada (2014).
2. B. B. Damodaran and R. R. Nidamanuri, "Dynamic linear classifier system for hyperspectral image classification for land cover mapping," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **7**(6), 2080–2093 (2014).
3. H. Ren and C. I. Chang, "Automatic spectral target recognition in hyperspectral imagery," *IEEE Trans. Aerosp. Electron. Syst.*, **39**(4), 1232–1249 (2003).
4. M. Dalponte et al., "Semi-supervised SVM for individual tree crown species classification," *ISPRS J. Photogramm. Remote Sens.* **110**, 77–87 (2015).
5. S. Marshall et al., "Hyperspectral imaging for food applications," in *23rd European Signal Processing Conf. (EUSIPCO)*, pp. 2854–2858 (2015).
6. M. A. Calin et al., "Hyperspectral imaging in the medical field: present and future," *Appl. Spectrosc. Rev.* **49**(6), 435–447 (2014).
7. C. Burges, "Dimension reduction: a guided tour," *Found. Trends Mach. Learn.* **2**(4), 275–364 (2010).
8. Keinosuke Fukunaga, "Introduction to statistical pattern recognition," 2nd ed., Academic Press Professional, Inc., San Diego, California (1990).
9. P. Dong and J. Liu, "Hyperspectral image classification using support vector machines with an efficient principal component analysis scheme," *Foundations of Intelligent Systems*, Y. Wang and T. Li, Eds., Vol. **122**, Springer (2011).
10. F. Tsai, E. K. Lin, and K. Yoshino, "Spectrally segmented principal component analysis of hyperspectral imagery for mapping invasive plant species," *Int. J. Remote Sens.* **28**(5), 1023–1039 (2007).
11. X. Liu et al., "A maximum noise fraction transform with improved noise estimation for hyperspectral images," *Sci. China Ser. F-Inf. Sci.* **52**(9), 1578–1587 (2009).
12. M. Fauvel, J. Chanussot, and J. A. Benediktsson, "Kernel principal component analysis for the classification of hyperspectral remote sensing data over urban areas," *EURASIP J. Adv. Signal Process.* **2009**(2), 783194 (2009).
13. M. E. Tipping, "Probabilistic principal component analysis," *J. R. Stat. Soc. Ser. B* **61**(3), 611–622 (1999).
14. A. Villa et al., "Hyperspectral image classification with independent component discriminant analysis," *IEEE Trans. Geosci. Remote Sens.* **49**(12), 4865–4876 (2011).
15. T. V. Bandos, L. Bruzzone, and G. Camps-Valls, "Classification of hyperspectral images with regularized linear discriminant analysis," *IEEE Trans. Geosci. Remote Sens.* **47**(3), 862–873 (2009).
16. B. C. Kuo and D. A. Landgrebe, "Nonparametric weighted feature extraction for classification," *IEEE Trans. Geosci. Remote Sens.* **42**(5), 1096–1105 (2004).
17. B. C. Kuo, C. H. Li, and J. M. Yang, "Kernel nonparametric weighted feature extraction for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.* **47**(4), 1139–1155 (2009).
18. J. Xia et al., "Improving random forest with ensemble of features and semisupervised feature extraction," *IEEE Geosci. Remote Sens. Lett.* **12**(7), 1471–1475 (2015).
19. S. Samiappan, S. Prasad, and L. Bruce, "Non-uniform random feature selection and kernel density scoring with SVM based ensemble classification for hyperspectral image analysis," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **6**(2), 792–800 (2013).
20. M. Imani and H. Ghassemian, "Band clustering-based feature extraction for classification of hyperspectral images using limited training samples," *IEEE Geosci. Remote Sens. Lett.* **11**(8), 1325–1329 (2014).
21. S. Sawant and M. Prabukumar, "Band fusion based hyper spectral image classification," *Int. J. Pure Appl. Math.* **117**(17), 71–76 (2017).
22. L. Zhang et al., "Compression of hyperspectral remote sensing images by tensor approach," *Neurocomputing* **147**(1), 358–363 (2015).

23. X. Cao et al., "Integration of 3-dimensional discrete wavelet transform and Markov random field for hyperspectral image classification," *Neurocomputing* **226**, 90–100 (2017).

24. Z. Zhu et al., "Three-dimensional Gabor feature extraction for hyperspectral imagery classification using a memetic framework," *Inf. Sci.* **298**, 274–287 (2015).

25. Y. Chen et al., "Spectral—spatial classification of hyperspectral data based on deep belief network," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **8**(6), 2381–2392 (2015).

26. X. Wan et al., "Stacked sparse autoencoder in hyperspectral data classification using spectral-spatial, higher order statistics and multifractal spectrum features," *Infrared Phys. Technol.* **86**, 77–89 (2017).

27. B. Liu et al., "Supervised deep feature extraction for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.* **56**, 1909–1921 (2018).

28. C. Shi and C. M. Pun, "Multi-scale hierarchical recurrent neural networks for hyperspectral image classification," *Neurocomputing* **294**, 82–93 (2018).

29. Y. Chen et al., "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.* **54**(10), 6232–6251 (2016).

30. S. Prasad and L. M. Bruce, "Limitations of principal components analysis for hyperspectral target recognition," *IEEE Geosci. Remote Sens. Lett.* **5**(4), 625–629 (2008).

31. Y. Zhen et al., "Classification based on 3-D DWT and decision fusion for hyperspectral image analysis," *IEEE Geosci. Remote Sens. Lett.* **11**(1), 173–177 (2014).

32. A. R. Webb, *Statistical Pattern Recognition*, Vol. **71**, No. 8, John Wiley & Sons, Ltd., England (2011).

33. S. A. Khayam, "The discrete cosine transform (DCT): theory and application," ECE 802–602 Information Theory and Coding, Michigan State University (2003).

34. F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing," *IEEE Trans. Geosci. Remote Sens.* **42**(8), 1778–1790 (2004).

35. T. A. Moughal, "Hyperspectral image classification using support vector machine," *J. Phys. Conf. Ser.* **439**, 012042 (2017).

36. "Hyperspectral Remote Sensing Scenes," Computational Intelligence Group, http://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes (2007).

37. N. Falco, J. A. Benediktsson, and L. Bruzzone, "A study on the effectiveness of different independent component analysis algorithms for hyperspectral image classification," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **7**(6), 2183–2199 (2014).

38. C.-I. Chang and H. Ren, "An experiment-based quantitative and comparative analysis of target detection and image classification algorithms for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.* **38**(2), 1044–1063 (2000).

39. X. Xu et al., "Multisource remote sensing data classification based on convolutional neural network," *IEEE Trans. Geosci. Remote Sens.* **56**(2), 937–949 (2018).

40. B. Kumar and O. Dikshit, "Hyperspectral image classification based on morphological profiles and decision fusion," *Int. J. Remote Sens.* **38**(20), 5830–5854 (2017).

41. X. Kang et al., "Spectral—spatial hyperspectral image classification with edge-preserving filtering," *IEEE Trans. Geosci. Remote Sens.* **52**(5), 2666–2677 (2014).

42. X. Cao et al., "Hyperspectral image classification with Markov random fields and a convolutional neural network," *IEEE Trans. Image Process.* **27**(5), 2354–2367 (2018).

43. A. C. Karaca et al., "Comparative evaluation of vector machine based hyperspectral classification methods," in *IEEE Int. Geoscience and Remote Sensing Symp. (IGARSS)*, pp. 4970–4973 (2012).

44. M. Fauvel and J. Benediktsson, "Spectral and spatial classification of hyperspectral data using SVMs and morphological profile," *IEEE Trans. Geosci. Remote Sens.* **46**(11), 3804–3814 (2008).

Manoharan Prabukumar received his BE degree in electronics and communication engineering from Periyar University, Tamilnadu, in 2002, his MTech degree in computer vision and image processing from Amrita school of Engineering, Coimbatore, in 2007, and his PhD in computer graphics from Vellore Institute of Technology (VIT), Tamilnadu, India, in 2014. Currently, he is working as an associate professor at the School of Information Technology

and Engineering, VIT. His research interests include hyperspectral remote sensing, image processing, computer graphics, and machine learning.

Shrutika Sawant received her BE and ME degrees in electronics and telecommunication engineering from Shivaji University, Maharashtra, India, in 2009 and 2012, respectively. Currently, she is pursuing her PhD in hyperspectral image processing from Vellore Institute of Technology (VIT), Vellore, Tamilnadu, India. She has been awarded with the senior research fellowship (SRF) from Council of Scientific and Industrial Research (CSIR), New Delhi. Her research interests include hyperspectral remote sensing, image processing, and machine learning.

Sathishkumar Samiappan received his BEngg degree in electronics and communication from Bharathiar University, Coimbatore, in 2003, his MTech degree in computer science and engineering from Amrita University, Coimbatore, India, in 2006, and his PhD in electrical and computer engineering at Mississippi State University (MSU), Starkville, Mississippi. Currently, he is an assistant research professor with the Geosystems Research Institute at MSU. His research interests include low-altitude remote sensing, pattern recognition, image processing, machine learning, and hyperspectral image classification.

Loganathan Agilandeeswari completed her PhD and is working as an associate professor at the School of Information Technology and Engineering, VIT, Vellore. She was awarded a best researcher award for year 2015 to 2016. She received her bachelor's degree in information technology and her master's degree in computer science and engineering from Anna University during 2005 and 2009, respectively. She has published 25+ papers in peer-reviewed reputed journals. She also the author of the various books, such as computer networks, mobile computing, and communication engineering.