Empirical Mode Decomposition of Hyperspectral Images for Support Vector Machine Classification

Begüm Demir, *Student Member, IEEE*, and Sarp Ertürk, *Member, IEEE*

***Abstract*—This paper presents the utilization of empirical mode decomposition (EMD) of hyperspectral images to increase the classification accuracy using support vector machine (SVM)-based classification. EMD has been shown in the literature to be particularly suitable for nonlinear and nonstationary signals and is used in this paper to decompose hyperspectral image bands into several intrinsic mode functions (IMFs) and a final residue. EMD is utilized in this paper to improve hyperspectral-image-classification accuracy by effectively exploiting the feature that EMD performs a decomposition that is spatially adaptive with respect to intrinsic features. This paper presents two different approaches for improved hyperspectral image classification making use of EMD. In the first approach, IMFs corresponding to each hyperspectral image band are obtained and the sums of lower order IMFs are used as new features for classification with SVM. In the second approach, the pieces of information contained in the first and second IMFs of each hyperspectral image band are combined using composite kernels for SVM classification with higher accuracy.**

***Index Terms*—Classification, empirical mode decomposition (EMD), hyperspectral images, support vector machines (SVMs).**

I. Introduction

HYPERSPECTRAL imaging sensors provide high-resolution spectral information from a wide range of the electromagnetic spectrum and offer improved performance for classification and detection in remote sensing applications [1].

In recent years, kernel-based hyperspectral-image-classification algorithms such as support vector machines (SVMs) [2]–[5] and relevance vector machines [6] have become very popular, because these approaches can provide comparatively high classification accuracy. Hyperspectral image classifications using different kernel-based approaches such as regularized radial basis function neural networks, SVMs, Fisher discriminant analysis, and regularized AdaBoost have, for example, been evaluated in [3].

An important research topic in the area of hyperspectral imaging in remote sensing applications comprises approaches that can provide high classification accuracy. The effectiveness of SVM-based hyperspectral image classification is shown in [3]–[5]. Although spectral characteristics are used as primary discrimination features in hyperspectral images, spatial attributes have been found very useful to increase the classification accuracy. Therefore, algorithms making use of both spectral and spatial information for classification have recently been proposed in [7]–[12]. In [7], composite kernels are used to combine spatial and spectral information of hyperspectral images to provide higher accuracy compared with that of using spectral information only. For this purpose, spatial feature vectors are obtained using either the mean only or the mean and the standard deviation together of a certain neighborhood window of the corresponding feature vector, and it is proposed to compute kernel matrices corresponding to spatial and spectral feature vectors separately and then combine these using different combination approaches. In [8], spatial feature vectors are extracted from an adaptive neighborhood (defined as a connected zone in order to reasonably process pixels that are close to the border of a structure), resulting in a self-complementary area filtering, and composite kernels are used to combine spatial and spectral information. In [9] and [10], it is proposed to use morphological profiles (MPs) for hyperspectral image classification. Initially, the principle components (PCs) of hyperspectral data are obtained, and then, MPs are constructed by applying opening and closing operations to the PCs. In [9], the MPs are used directly for classification with a neural network, while in [10], the MPs are fused with the original features for SVM-based classification. The combination of spectral and spatial information is achieved based on Markov random field in [11], and majority voting is used within regions obtained by the watershed segmentation algorithm in [12].

Spatial-filtering preprocessing methods presented in [13] and [14] have been proposed to improve the classification accuracy by decreasing the intraclass spectral variability and spatially smoothing homogeneous areas. Classification accuracy of hyperspectral images is improved in [13] by performing nonlinear filtering with anisotropic diffusion. An adaptive spatial-filtering algorithm which uses median filters with different sizes is presented in [14] and has been shown to result in improved classification accuracy. An unsharp filter is used in [15] to enhance high-frequency components (such as edges) in an image before labeling the segmented area by human supervision for SVM classification.

This paper proposes the utilization of empirical mode decomposition (EMD) of hyperspectral image bands to increase the classification accuracy by making use of the feature that EMD performs a decomposition that is spatially adaptive with respect to intrinsic features. EMD is a signal-decomposition approach proposed particularly for the analysis of nonlinear and nonstationary data [16] and decomposes the data into a finite and often small number of intrinsic mode function (IMFs). The first IMF resembles a high-pass-filtered signal, and the

Manuscript received October 30, 2009; revised March 12, 2010 and June 8, 2010. Date of publication October 4, 2010; date of current version October 27, 2010. This work was supported by the Turkish State Planning Organization under Project DPT 2008K-120800.

The authors are with the Department of Electronics and Telecommunication Engineering, Kocaeli University, Kocaeli 41040, Turkey.

Digital Object Identifier 10.1109/TGRS.2010.2070510

TABLE I
NoS FOR DIFFERENT CLASSES OF THE INDIAN PINE DATA USED IN THE EXPERIMENTS

Class	NoS
Corn-no till	1434
Corn-min till	834
Grass/Pasture	497
Grass/Trees	747
Hay-windrowed	489
Soybean-no till	968
Soybean-min till	2468
Soybean-clean till	614
Woods	1294
Total	9345

other IMFs resemble bandpass-filtered signals with their center frequency decreasing in an octave band manner like a filter bank [17]. This paper presents the detailed utilization of EMD for hyperspectral image classification. EMD is applied to individual hyperspectral image bands as a preliminary decomposition step before classification. Two different EMD-based classification approaches are presented in this paper. The first approach uses the sum of several IMFs and discards subsequent IMFs and the residue of each hyperspectral image band for classification. The second approach combines information of the first and second IMFs using composite kernels. While the first approach is not restricted to kernel-based classifiers, the second approach can only be implemented using a kernel-based classifier. SVM is used as classification method in the classification step because of its comparatively high classification accuracy. The proposed EMD-based approaches are compared with wavelet decomposition (WD)-based denoising [18] of hyperspectral image bands, which has been shown to provide an increase in hyperspectral-image-classification accuracy [19], and also with simple low-pass filtering (LPF) as well as unsharp filtering (UF) of hyperspectral image bands as in [15], spectral- and spatial-domain composite kernels as presented in [7], and MP-based classifiers presented in [9] and [10].

II. DATA SETS

Two data sets are used to present experimental results in this paper. The Indian Pine hyperspectral image [20] is a data set that is rather difficult to classify because of the high similarity between spectral signatures of different classes and heavily mixed pixels [21]. These data consist of 145×145 pixels with 220 bands. The number of bands is initially reduced to 200 by removing bands covering water absorption and noisy bands. The original ground truth has actually 16 classes, but some classes have a very small number of elements; and therefore, nine classes that have a higher number of samples (NoS) have been used, and the total NoS corresponding to each class is shown in Table I.

The second data set is the DC Mall data which have 307×1280 pixels with 210 bands [1]. The number of bands was first reduced to 191 for the DC Mall data by removing bands covering water absorption and noisy bands. Table II shows the NoS used for the DC Mall data for all classes used in the experiments.

TABLE II
NoS FOR DIFFERENT CLASSES OF THE DC MALL DATA USED IN THE EXPERIMENTS

Class	NoS
Roof	3834
Street	416
Path	175
Grass	1928
Trees	405
Water	1224
Shadow	97
Total	8079

III. EMD

The Hilbert–Huang transform (HHT) has been proposed in [16] for nonlinear and nonstationary time series analysis and includes two main steps referred to as EMD and Hilbert spectral analysis. EMD is regarded to form the key part of HHT by decomposing any complicated data into a finite and often small number of IMFs. EMD is very efficient because it is adaptive, and it is based on the local characteristic timescale of the data and, therefore, is applicable to nonlinear and nonstationary data. The IMFs typically include instantaneous frequencies as functions of time that give sharp identifications of imbedded structures. IMFs satisfy two conditions: 1) The number of zero crossings and the number of extreme points are equal or differ at most by one, and 2) the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero at any point. The advantage of EMD over other decomposition techniques is that EMD is a nonparametric data-driven method that can be used in the analysis of nonlinear and nonstationary data, and it adaptively decomposes the data into a finite number of IMFs. These IMFs, based on and derived from the data, can serve as the basis of expansion which can be linear or nonlinear as dictated by the data, and it is complete and almost orthogonal [16]. The most important properties of the IMFs obtained with EMD are that the IMFs represent local features of the signal and are obtained adaptively. It is noted in [16] that locality and adaptivity are the necessary conditions for the basis for expanding nonlinear and nonstationary signals, while orthogonality is not a necessary criterion for basis selection for a nonlinear system. Because IMFs are based on local features of the signal, it becomes reasonable to refer to instantaneous frequency content of signals. Each IMF typically contains an oscillatory mode, inherent in the data, at different (narrow) range of frequencies.

EMD has some important advantages in the case of nonlinearity and nonstationarity compared with the widely used Fourier- and wavelet-transform-based signal-decomposition techniques. Fourier transform assumes the system to be linear and the signal to be stationary [22]; but in most real systems, either natural or even man-made ones, the data are most likely to be both nonlinear and nonstationary [16]. Compared with wavelet-based decomposition techniques, there are two basic differences between EMD and the wavelet transforms [23]. First of all, different wavelet types are possible for the wavelet transform and the performance changes according to the wavelet type, while EMD has no basis functions and

decomposes the signal according to intrinsic characteristics. Second, the frequency scale in the wavelet transform is always fixed depending on sampling frequency and decomposition level; while the IMFs of EMD can have variable frequency content depending on local signal properties.

EMD has been applied in several areas of signal processing [23]–[31]. In [23], EMD and WD are used to detect human cataract using ultrasound signals, and it has been shown that the detection performance of EMD is better compared with that of WD. EMD is compared with WD in terms of the fractal-dimension estimation of a discrete sample path in [24], and presented results again show that EMD yields better estimates of fractal dimensions. EMD is initially presented for 1-D signals and then used for 2-D signals in [25]–[28]. EMD is applied to 2-D face images as a preprocess to remove illumination artifacts for a face recognition application in [25], where 2-D images are first converted to 1-D signals. EMD is extended to be applied directly in 2-D and used for image compression in [17] and [26]. In [17] and [26], 2-D spline interpolation is used to enable 2-D-EMD; however, 2-D spline interpolation results in a very high computational load for large data sizes. Fast and adaptive 2-D-EMD (FA-2-D-EMD), which uses order statistics filters instead of 2-D spline interpolation for reduced computational complexity, is proposed in [27] and [28]. EMD has been applied to hyperspectral data for preliminary dimensionality reduction in [29]. In [30], we have presented initial results showing that it is possible to use EMD to increase the classification accuracy of hyperspectral images. The normalized EMD that gives more consistent and stable instantaneous frequencies compared with [16] is presented in [31] with applications in remote sensing.

IV. EMD of Hyperspectral Image Bands

In this paper, it is proposed to use EMD to obtain the IMFs of hyperspectral image bands and to use these IMFs in the classification step to provide increased classification accuracy. In the proposed approach, EMD is applied individually to each hyperspectral image band in the spatial domain. In the proposed approach, 2-D EMD is utilized because EMD is applied to each hyperspectral band separately, and the hyperspectral image data of each spectral band is basically a 2-D signal. Both conventional 2-D-EMD [17], [26] and FA-2-D-EMD [27], [28] are evaluated for this purpose, and the utilization of these approaches for hyperspectral data is explained in the following sections.

Note that classification results for 1-D EMD applied to the spectral dimension are also provided in this paper for comparative evaluation purposes, but it is observed that 1-D EMD reduces classification accuracy.

In order to provide a common notation, $B_l(i,j)$ is used to denote the original hyperspectral image band with band index l ($l = 1, 2, \ldots, L$, with L being the total number of bands), where (i,j) shows the spatial location. Note that $IMF_{l,m}$ is used to show the values of the mth IMF (or mth order IMF) $(m = 1, 2, \ldots, M)$ of the lth hyperspectral image band.

A. 2-D-EMD of Hyperspectral Image Bands

The process used to find IMFs using 2-D-EMD [17], [26] is called sifting. Sifting is an iterative process. Here, $I_{l,m}^{(n)}(i,j)$ is used to show the present values used in the nth iteration to find the mth IMF of the lth band. This process starts from the original hyperspectral band, and the starting point can be expressed in the form of

$$I_{l,1}^{(1)}(i,j) = B_l(i,j). \tag{1}$$

The sifting process of 2-D-EMD can be summarized as follows.

1) Find all points of 2-D local maxima and all points of 2-D local minima of $I_{l,m}^{(n)}(i,j)$.
2) Create the upper envelope $(E_{\max}(i,j))$ by 2-D spline interpolation of local maxima and the lower envelope $(E_{\min}(i,j))$ by 2-D spline interpolation of local minima.
3) Calculate the mean of the upper and lower envelopes: $A_m^{(n)}(i,j) = (E_{\max}(i,j) + E_{\min}(i,j))/2$.
4) Subtract the envelope mean from the input signal: $S_m^{(n)}(i,j) = I_{l,m}^{(n)}(i,j) - A_m^{(n)}(i,j)$.
5) Check if the envelope mean fulfills the iteration stop criterion for the current IMF. The stop criterion for the current IMF is reached if the envelope mean signal is close to zero [17] such that $(\sum_{i=1}^{P}\sum_{j=1}^{R}|A_m^{(n)}(i,j)|/P \times R) < \tau$, where P and R are the dimensions of $A_m^{(n)}(i,j)$ and τ is a small threshold. This stop criterion guarantees that the IMFs capture signal features effectively. If the stop criterion is fulfilled (assume at step $n = N$), the current IMF is obtained as $IMF_{l,m}(i,j) = S_m^{(N)}(i,j)$. If the stop criterion is not fulfilled, the next iteration is started with $I_{l,m}^{(n+1)}(i,j) = S_m^{(n)}(i,j)$ and this process is repeated from step 1) to find the current IMF.
6) If the current IMF is obtained successfully, the residue signal $R_m(i,j) = I_{l,m}^{(n)}(i,j) - IMF_{l,m}(i,j)$ is computed. If the residue does not contain any more extreme points, the EMD process is stopped [17], [26]. Otherwise, the next IMF is obtained starting from step 1) using the current residue as the next input, i.e., $I_{l,m+1}^{(1)}(i,j) = R_m(i,j)$.

The original hyperspectral image bands $B_l(i,j)$ can be exactly reconstructed by adding all corresponding IMFs and the final residue. This can be formulated as

$$B_l(i,j) = \sum_{m=1}^{M} IMF_{l,m}(i,j) + R_M(i,j). \tag{2}$$

Fig. 1 shows the IMFs and residues of a sample band of the Indian Pine hyperspectral image taken over northwest Indiana's Indian Pine test site in June 1992. It is seen in Fig. 1 that the first IMF includes the highest local spatial frequency detail and the second IMF includes the next highest local spatial frequency detail and so on. Each IMF actually contains both low- and high-spatial-frequency detail at different spatial locations, which is a basic feature of EMD and is also demonstrated in [17] and [26]. It is seen that the lower order IMFs such as the first and second IMFs are much more structured and reflect the spatial structural content of the image and that higher order

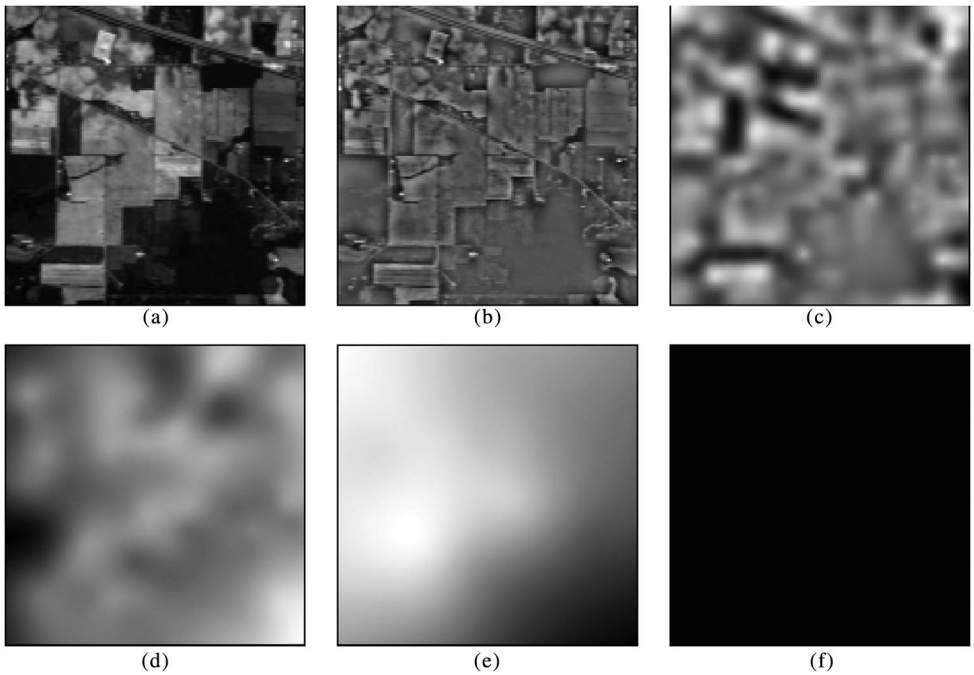


Fig. 1. 2-D-EMD of the Indian Pine image band #28. (a) Original. (b) First IMF. (c) Second IMF. (d) Third IMF. (e) Fourth IMF. (f) Final Residue.

IMFs have a rather low-pass characteristic and lack local spatial structure.

B. FA-2-D-EMD of Hyperspectral Image Bands

FA-2-D-EMD differs from 2-D-EMD in two ways [27], [28]. First, the upper and lower envelopes are obtained in FA-2-D-EMD using a direct envelope estimation method instead of the computationally intensive 2-D spline interpolation used in 2-D-EMD. Second, there is no need to apply iterations in FA-2-D-EMD, which again reduces the computational complexity.

The sifting process of FA-2-D-EMD does not require iterations, and therefore, $I_{l,m}(i,j)$ is used to show the current input values employed to find the mth IMF of the lth band. The sifting process starts from the original hyperspectral band, and the starting point can be expressed in the form of

$$I_{l,1}(i,j) = B_l(i,j). \tag{3}$$

The sifting process of FA-2-D-EMD can be summarized as follows.

1) Find all points of local maxima and all points of local minima of $I_{l,m}(i,j)$.
2) Apply MAX and MIN filters to $I_{l,m}(i,j)$ to estimate the upper envelope $E_{\max}(i,j)$ and the lower envelope $E_{\min}(i,j)$, respectively. $E_{\max}(i,j)$ and $E_{\min}(i,j)$ are obtained as

$$\begin{aligned} E_{\max}(i,j) &= \underset{(q,p)\in Z_{ij}}{MAX}\ \{I_{l,m}(q,p)\} \\ E_{\min}(i,j) &= \underset{(q,p)\in Z_{ij}}{MIN}\ \{I_{l,m}(q,p)\} \end{aligned} \tag{4}$$

where Z_{ij} denotes a square-shaped spatial window of size $s_{\max} \times s_{\max}$ for the MAX filter and size $s_{\min} \times s_{\min}$ for the MIN filter, located around the spatial location (i,j). It is proposed in [27] and [28] to obtain $s_{\max}$ and $s_{\min}$ adaptively, according to the location of local maximum and minimum points in $I_{l,m}(i,j)$. For each local maximum point, the Euclidean distance to the nearest other local maxima is determined and recorded in a corresponding array $(ary_{\max})$. The same process is carried out for each local minimum point, and the Euclidean distance to the nearest other local minima is recorded in a corresponding array $(ary_{\min})$. Different approaches have been proposed for setting the window sizes $s_{\max}$ and $s_{\min}$ in [27] and [28], all of which are evaluated in this paper, and it is found that the best results are obtained if $s_{\max}$ and $s_{\min}$ are determined as in

$$\begin{aligned} s_{\min} &= \max\{ary_{\min}\} \\ s_{\max} &= \max\{ary_{\max}\}. \end{aligned} \tag{5}$$

Finally, $s_{\max}$ and $s_{\min}$ are rounded to the nearest odd integer to set the window sizes of the MAX and MIN filters.

3) After the upper and lower envelopes are obtained as in the previous step, they have a rather stepwise characteristic as a result of the MAX and MIN filters. Therefore, LPF is applied to $E_{\max}(i,j)$ and $E_{\min}(i,j)$ to obtain smooth envelopes. For this purpose, $s_{\max} \times s_{\max}$ and $s_{\min} \times s_{\min}$ sized uniform filter kernels are used to obtain the smoother upper envelope $E_{s\max}(i,j)$ and smoother lower envelope $E_{s\min}(i,j)$, respectively. This process can be formulated in the form of

$$\begin{aligned} E_{s\max}(i,j) &= \frac{1}{s_{\max} \times s_{\max}} \sum_{(q,p)\in Z_{ij}} E_{\max}(q,p) \\ E_{s\min}(i,j) &= \frac{1}{s_{\min} \times s_{\min}} \sum_{(q,p)\in Z_{ij}} E_{\min}(q,p). \end{aligned} \tag{6}$$

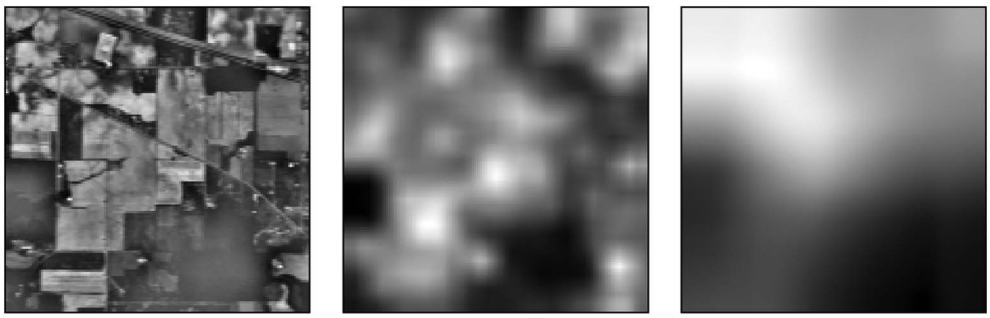


Fig. 2. FA-2-D-EMD of Indian Pine image band #28. (a) First IMF. (b) Second IMF. (c) Final residue.

4) Calculate the mean of the smoothed upper and lower envelopes: $A_m(i,j) = (E_{s\max}(i,j) + E_{s\min}(i,j))/2$.
5) Subtract the envelope mean from the input signal: $S_m(i,j) = I_{l,m}(i,j) - A_m(i,j)$.
6) The current (mth) IMF is obtained as $IMF_{l,m}(i,j) = S_m(i,j)$.
7) The residue $R_m(i,j) = I_{l,m}(i,j) - IMF_{l,m}(i,j)$ is computed, and the process is ended if the residue does not contain more than one extreme point. If the residue does contain more than one extreme point, the next IMF is computed starting with the current residue as the next input: $I_{l,m+1}(i,j) = R_m(i,j)$.

Similar to 2-D-EMD, the original hyperspectral image band $B_l(i,j)$ can exactly be reconstructed by adding all IMFs and the final residue. Fig. 2 shows example IMFs and residues of a sample band of the Indian Pine hyperspectral image. Although, 2-D-EMD provided four IMFs for this band, only two IMFs are obtained using FA-2-D-EMD. In case of FA-2-D-EMD, the first IMF is much more similar to the original hyperspectral image band itself.

The important advantage of FA-2-D-EMD compared with 2-D-EMD is the reduced computational complexity. While the 2-D spline interpolation required in 2-D-EMD can result in an extremely high computational load for hyperspectral images with large spatial resolution, FA-2-D-EMD can easily be applied to large-sized hyperspectral images.

V. SVM-Based Classification

SVMs [2] have become very popular in hyperspectral image classification because they can provide high classification performance [3]–[5]. SVM tries to find the optimal separating hyperplane that maximizes the margin between the closest training sample and the separating hyperplane and uses the boundary pixels (support vectors) to create a decision surface. SVM estimates a classification function using training data from two classes: $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_f, y_f) \in \Re^n \times \{\pm 1\}$. In kernel-based SVM classification, the original input space is mapped to a higher dimensional (Hilbert) feature space ($\phi : \Re^n \to H$) using a kernel function. A kernel function is a function that corresponds to an inner product in some expanded feature space. The classification function is obtained by solving the convex optimization problem

$$\begin{aligned} &\text{maximize} : \sum_{u=1}^{f} \alpha_u - \frac{1}{2} \sum_{u=1}^{f} \sum_{v=1}^{f} \alpha_u \alpha_v y_u y_v K(\mathbf{x}_u, \mathbf{x}_v) \\ &\text{subject to} : \sum_{u=1}^{f} \alpha_u y_u = 0 \quad \text{and} \quad 0 \le \alpha_u \le C \end{aligned} \tag{7}$$

where C controls the tradeoff between complexity (number of support vectors) and data misfit (number of nonseparable points) and is chosen by the user (i.e., set *a priori*). The kernel function $K(\mathbf{x}_u, \mathbf{x}_v) = \phi(\mathbf{x}_u)\phi(\mathbf{x}_v)$ does not actually require a direct knowledge of the transform function $\phi(\cdot)$. Note that α_u and α_v are Lagrange multipliers. Each nonzero α_u indicates that the corresponding $\mathbf{x}_u$ is a support vector. The nonlinear classifier for a sample $\mathbf{x}$ can then be expressed as

$$f = \operatorname{sgn}\left(\sum_{u=1} \alpha_u y_u K(\mathbf{x}_u, \mathbf{x}) + b \right). \tag{8}$$

Kernel functions used in SVM must satisfy Mercer's condition which requires the kernel to be a continuous symmetric kernel of a positive integral operator. Popular kernels implementing this condition are the linear kernel $K(\mathbf{x}_u, \mathbf{x}) = \mathbf{x}_u \cdot \mathbf{x}$, the polynomial kernel $K(\mathbf{x}_u, \mathbf{x}) = (\gamma \mathbf{x}_u \cdot \mathbf{x})^d$, and the radial basis function kernel $K(\mathbf{x}_u, \mathbf{x}) = \exp(-\gamma \|\mathbf{x}_u - \mathbf{x}\|^2)$, where d and γ are kernel parameters. For multiclass SVM, it is possible to combine multiple binary classifiers. The one-against-one approach is utilized in this paper for multiclass SVM because it provides fast training. In the one-against-one method, $K(K-1)/2$ binary classifiers are trained and $K(K-1)/2$ binary tests are required to make a final decision, where K is the total number of classes. Each outcome gives one vote to the winning class, and the class with the most votes is selected as the final result.

VI. Proposed EMD-Based Classification Approaches

This paper proposes to use the IMFs of hyperspectral bands for hyperspectral image classification. Two separate approaches are presented for this purpose. In both approaches, basically, only the lower order IMFs (typically the first two IMFs) are

used in the classification and higher order IMFs are disposed. In this way, it is ensured that pieces of information contained in lower order IMFs that include spatial relations are retained, while pieces of information in higher order IMFs that typically lack spatial structure are discarded.

In the first approach, several of the initial (lower order) IMFs of each band are summed to reconstruct the new data of that band by discarding remainder (higher order) IMFs and residues.

In the second approach, the first and second IMFs are treated separately and their information is combined in the classifier using composite kernels. The first approach is actually not restricted to kernel-based classifiers while the second approach is restricted to kernel-based classifiers.

SVM classification is utilized in this paper because of the high classification accuracy.

A. Using Sum of Lower Order IMFs as Features

In the first approach presented in this paper, EMD is used to extract the IMFs of each hyperspectral image band and lower order IMFs are summed to reconstruct the feature data to be used in classification, discarding higher order IMFs. Note that $IMF_{l,m}(i,j)$ is regarded to be of mth order. By this convention, lower order IMFs capture fast spatial oscillation modes while higher order IMFs typically represent slow spatial oscillation modes [32]. Therefore, if 2-D-EMD is interpreted as a spatial-scale analysis method, lower order IMFs and higher order IMFs correspond to the fine and coarse scales, respectively. Because the original hyperspectral image bands $B_l(i,j)$ can exactly be reconstructed by adding all corresponding IMFs and the final residue, as shown in (2), keeping lower order IMFs, and discarding higher order IMFs and the residues, has the effect of discarding coarse spatial scale detail. In this approach, the reconstructed new data of each hyperspectral band can be formulated as

$$B'_l(i,j) = \sum_{m=1}^{V} IMF_{l,m}(i,j) \tag{9}$$

where V shows the number of IMFs retained. Note that, in case of hyperspectral images, as EMD is applied to each hyperspectral band individually, the total number of IMFs of each band can actually be different, depending on the data itself. In case some bands have a lower number of IMFs than that used in the processing, the remaining IMFs are assumed to be equal to zero.

In the proposed approach, the new hyperspectral data of each pixel, which will be used by the classifier, comprise the spectral data corresponding to the different bands obtained as the sum of lower order IMFs and can be formulated as

$$\mathbf{x}_u = B'_l(ui, uj), \qquad l = 1, 2, \ldots, L \tag{10}$$

where (ui, uj) shows the spatial location of x_u in the hyperspectral image.

Fig. 3 shows the visual result of summing lower order IMFs. Note that 2-D-EMD is used with a stop threshold τ of 0.006 to obtain the IMFs. This figure shows the original band as well as the first IMF only (i.e., $V = 1$), the sum of the first two IMFs (i.e., $V = 2$), the sum of the first three IMFs (i.e., $V = 3$), and the sum of first four IMFs (i.e., $V = 4$) of this band. Note that 2-D-EMD is used for results presented in this figure. It is seen that the first IMF contains mainly fine spatial detail, and in case of the sum of the first two IMFs, the overall detail is quite well represented. In case of the sum of first three or first four IMFs, the reconstructed band is rather similar to the original band.

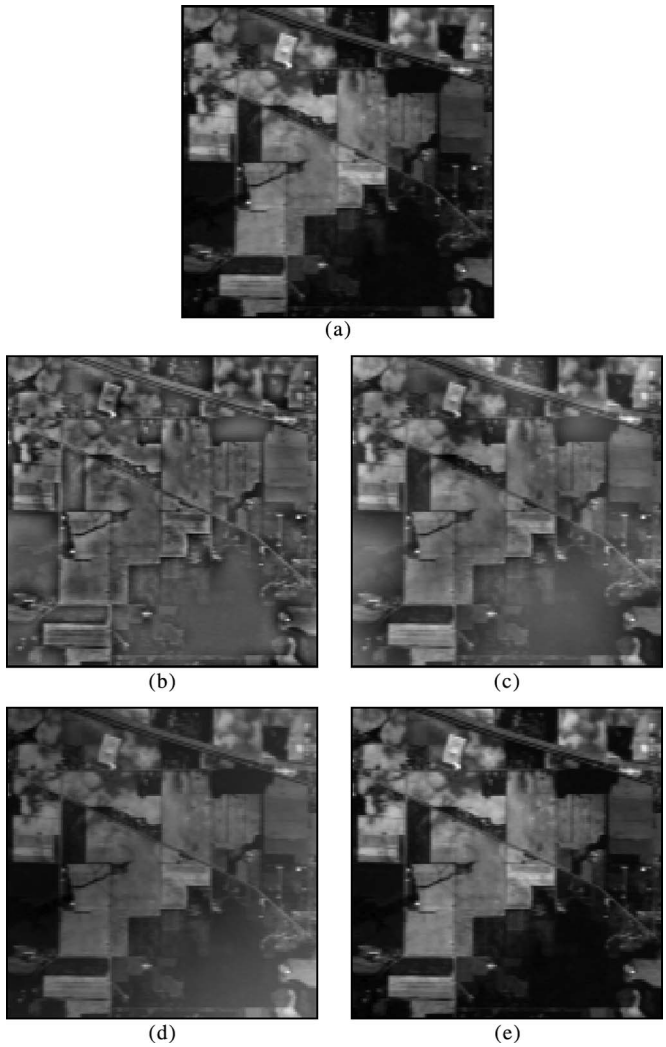


Fig. 3. Indian Pine image band # 28. (a) Original. (b) First IMF. (c) Sum of first two IMFs. (d) Sum of first three IMFs. (e) Sum of first four IMFs.

Table III shows the Bhattacharyya distance of each class to the other classes obtained in the original Indian Pine image and the new feature data reconstructed from the sum of lower order IMFs obtained for different V values. It is seen that the process of reconstructing the data as the sum of lower order IMFs increases the between-class distances. The best case is observed for $V = 2$. This is also the best case obtained in terms of classification accuracy as will be shown in the experimental results.

B. Combining IMFs Using Composite Kernels

In the second approach, it is proposed to combine the information contained in the first and second IMFs using composite kernels in the classification stage.

TABLE III
BHATTACHARYYA DISTANCE VALUES OF EACH CLASS WITH RESPECT TO OTHER CLASSES FOR ORIGINAL DATA AS WELL AS DATA RECONSTRUCTED AS THE SUM OF LOWER ORDER IMFS FOR THE INDIAN PINE DATA

		Original data	Data reconstructed from lower-order IMFs of EMD			
			V=1	**V=2**	V=3	V=4
Class Labels	1	7.45	20.23	**28.66**	18.32	15.95
	2	10.05	26.01	**38.24**	25.31	22.75
	3	18.42	32.40	**49.23**	33.61	30.28
	4	14.28	26.50	**42.19**	31.12	28.18
	5	29.83	34.05	**79.23**	61.24	56.62
	6	8.90	23.12	**33.98**	21.53	18.61
	7	5.59	18.60	**29.71**	17.46	14.96
	8	13.40	31.09	**46.26**	32.47	28.56
	9	16.09	23.29	**46.04**	39.13	34.96

Composite-kernel SVM classification that combines spectral and spatial information has been presented in [7]. In [7], initially, spatial and spectral feature vectors $(\mathbf{x}_u^s, \mathbf{x}_u^p)$ are constructed and individual kernel matrices are computed and then combined. Each spatial feature vector is basically obtained as either the mean value or the mean and standard deviation values of neighborhood pixels. Different kernel combination approaches are used in [7].

In the approach presented in this paper, it is proposed to combine the information contained in the first and second IMFs using composite kernels. Therefore, for each hyperspectral pixel, two different feature vectors are constructed, one corresponding to the spectral information obtained from the first IMF of each band and the other corresponding to the spectral information obtained from the second IMF of each band. In this case, the feature vectors corresponding to the first and second IMFs can be expressed as

$$\begin{aligned} \mathbf{x}_u^{imf1} &= IMF_{l,1}(ui, uj) \qquad l = 1, 2, \ldots, L \\ \mathbf{x}_u^{imf2} &= IMF_{l,2}(ui, uj) \qquad l = 1, 2, \ldots, L. \end{aligned} \tag{11}$$

After $\mathbf{x}_u^{imf1}$ and $\mathbf{x}_u^{imf2}$ are constructed, the individual kernel matrices are computed. In the proposed composite-kernel-based approach, the direct summation kernel, as shown in (12), and the weighted summation kernel given in (13) are used to combine these kernels

$$K(\mathbf{x}_u, \mathbf{x}_v) = K\left(\mathbf{x}_u^{\text{imf1}}, \mathbf{x}_v^{\text{imf1}}\right) + K\left(\mathbf{x}_u^{\text{imf2}}, \mathbf{x}_v^{\text{imf2}}\right) \tag{12}$$

$$K(\mathbf{x}_u, \mathbf{x}_v) = \mu K\left(\mathbf{x}_u^{\text{imf1}}, \mathbf{x}_v^{\text{imf1}}\right) + (1-\mu) K\left(\mathbf{x}_u^{\text{imf2}}, \mathbf{x}_v^{\text{imf2}}\right). \tag{13}$$

VII. EXPERIMENTAL RESULTS

For each data set, the classification performances of the proposed approaches are demonstrated using SVM classification with RBF kernel. In the experiments, the penalty parameter C of SVM is tested between 10 and 1000 with a step size increment of 20 and the γ parameter of the RBF kernel is tested between 0.1 and 2 with a step size increment of 0.1, and the best values are obtained using a fivefold cross validation approach for both data sets.

Classification results are presented with respect to different training data rates (TDRs), i.e., a TDR of 10% illustrates the case where 10% of the total data samples are used as training data and the remaining samples are used as testing data.

The classification accuracies are evaluated in terms of overall accuracy (OA) as well as kappa coefficient (k) [33]. The statistical significance of changes in the classification accuracies obtained with two different approaches is computed using McNemar's test [34]. In McNemar's test, for a significance level of 0.05, the accuracy difference is regarded to be statistically significant if $|Z| > 1.96$. If the sign of Z is positive, the first classifier outperforms the second classifier and *vice versa*. In the presented results, Z values are computed with the second classifier assigned as the direct SVM case and the first classifier assigned as the denoted classification approach.

SVM classification with 2-D-EMD is denoted as 2-D-EMD-SVM, and SVM classification with FA-2-D-EMD is denoted as FA-2-D-EMD-SVM in the experimental results. For 2-D-EMD, 2-D interpolation is carried out using thin-plate smoothing spline interpolation [35], similar to [17]. The thin-plate smoothing spline interpolation method gives a surface with continuous second derivative everywhere and turns out to successfully decompose a hyperspectral band into its IMFs and a smooth final residue with no or only a few extremum points as reported in [17].

For both, 2-D-EMD-SVM and FA-2-D-EMD-SVM, results are provided for the cases where only the first IMF is used (1 IMF), the sum of the first two IMFs is used (2 IMFs), the sum of the first three IMFs is used (3 IMFs), and the sum of the first four IMFs is used (4 IMFs). Note that, for bands that have less IMFs than that used in the summation, the absent IMFs are taken as zero.

EMD-based composite-kernel SVM classification which combines the information of the first and second IMFs is denoted as CK-2-D-EMD-SVM if 2-D-EMD is used to obtain the IMFs and CK-FA-2-D-EMD-SVM if FA-2-D-EMD is used to obtain the IMFs. These results are given in case of using direct summation kernels as well as weighted summation kernel. The parameter of μ is varied in steps of 0.1 in the range [0, 1].

For comparison purposes, results of SVM classification with WD-based denoising (denoted as WD-SVM), as presented in [18], results of SVM classification after simple LPF (denoted as LPF-SVM), as well as results of SVM classification after UF as in [15] (denoted as UF-SVM) are provided. Note that WD is implemented using symmlet orthogonal wavelets with eight vanishing moments and four decomposition levels. On the other hand, LPF is implemented using a 5×5 sized uniform filter kernel (i.e., a moving average filter). UF is implemented as the difference of the original image and the low-pass-filtered image multiplied by a constant. The value of the constant is varied within the range of [0.1, 1] with a step size increment of 0.1, and the best classification accuracy is reported in this paper. LPF is implemented using a 5×5 sized uniform filter kernel.

Furthermore, the results of MP-based SVM classification approaches [9], [10] are also provided. To obtain the MPs, the first three PCs are obtained, and four opening/closing operations are applied to each PC using a circular structural element with a

TABLE IV
COMPUTATIONAL TIME AND NUMBER OF IMFS OBTAINED USING 2-D-EMD AND FA-2-D-EMD FOR DIFFERENT τ AND W VALUES FOR THE INDIAN PINE DATA

Method	τ	W	EMD Time (min.)	Number of 1 IMF	Number with 2 IMFs	Number of 3 IMFs	Number of 4 IMFs	Number of 5 IMFs
2D-EMD	0.01	3x3	701	8	0	147	45	-
	0.008	3x3	767	3	-	142	55	-
	0.006	3x3	835	-	-	143	58	-
	0.004	3x3	924	-	-	122	78	-
	0.006	5x5	79	-	-	1	187	12
	0.006	7x7	43	-	-	43	157	-
FA-2D-EMD	-	3x3	49	-	177	23	-	-
	-	5x5	42	-	-	156	42	2
	-	7x7	34	-	-	185	15	-

step size increment of four. After this process, the number of total bands is obtained as 27. Three alternative approaches have been used in this case for comparison. First, only these 27 bands are used for SVM classification (this case is denoted as EMP). Second, the original data and MPs are directly fused together, and these fused data are used for SVM classification (this case is denoted as Spec-EMP). In the third case, the original data and MPs are fused after feature extraction (FE) is performed to each data separately (this case is denoted as FE-Spec-EMP). Although decision boundary FE and nonparametric weighted FE (NWFE) have been used in [10], FE is implemented using only NWFE in this paper. Results of Spec-EMP and FE-Spec-EMP are provided for the Indian Pine image only using a 89% variance criterion of NWFE. Thereby, the total number of original bands is reduced to ten in FE and the number of MPs is reduced to five in FE.

The results of composite-kernel SVM classification which combines spatial and spectral information [7] are provided (denoted as CK-SS-SVM). CK-SS-SVM results are given in the case of using direct summation kernels as well as weighted summation kernel. The parameter of μ is varied in steps of 0.1 in the range [0, 1]. Spatial feature vectors are obtained using the mean of the neighborhood pixels of a corresponding feature vector in a 5×5 sized window.

A. Experimental Results for the Indian Pine Data

For the Indian Pine data, the stop threshold τ of 2-D-EMD is varied between 0.004 and 0.01 with a neighborhood window of size (W) 3×3 used to obtain the local maximum and minimum extreme points. Furthermore, window sizes W of 3×3, 5×5, and 7×7 are evaluated for 2-D-EMD (with fixed $\tau = 0.006$) as well as FA-2-D-EMD for the selection of local maximum and minimum extreme points. The total number of IMFs of each band is actually different, depending on spatial content. Table IV shows the total number of IMFs and the computational load of 2-D-EMD and FA-2-D-EMD for different W and τ values. It is seen that FA-2-D-EMD is much faster than 2-D-EMD and using larger W values reduces the computational load of 2-D-EMD and FA-2-D-EMD because it provides a reduced number of extreme points.

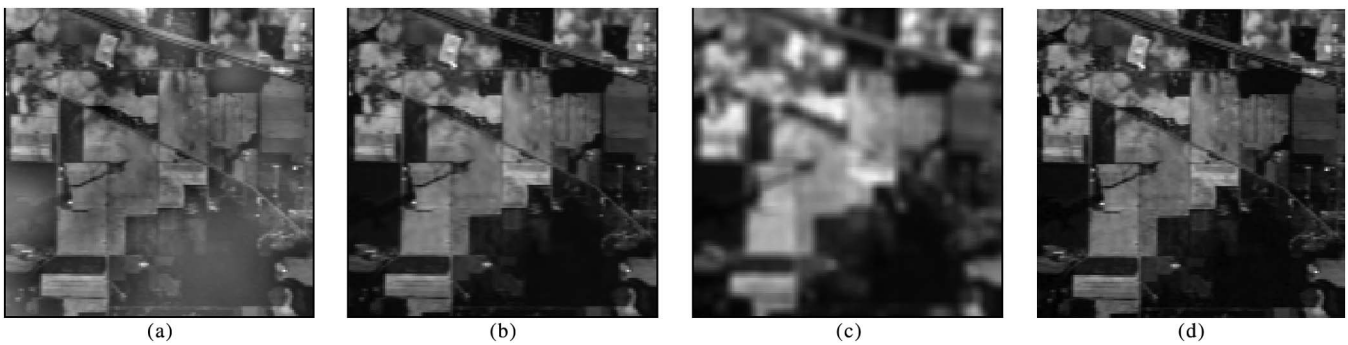
Table V shows the OAs and k values of direct SVM, 2-D-EMD-SVM, and FA-2-D-EMD-SVM, as well as WD-SVM, UF-SVM, LPF-SVM, FE-Spec-EMP, Spec-EMP, and EMP in case of different TDRs. Experimental results in Table V show that best results for 2-D-EMD-SVM are obtained in the 2 IMFs case in all TDRs. Furthermore, the best results of 2-D-EMD-SVM provide higher classification accuracies than the best results of FA-2-D-EMD-SVM in all TDRs. This shows that, although FA-2-D-EMD provides a fast alternative to 2-D-EMD, its performance is not as good in this case. The differences of classification accuracies between direct SVM and 2-D-EMD-SVM or direct SVM and FA-2-D-EMD-SVM are always obtained to be statistically significant for the best cases obtained with the two methods. Using the best results of 2-D-EMD-SVM and FA-2-D-EMD-SVM, the classification accuracies are increased by more than at least 6% and sometimes as high as 13% compared with standard SVM in all TDRs. It is seen that 2-D-EMD-SVM and FA-2-D-EMD-SVM provide higher classification accuracies than WD-SVM, UF-SVM, and LPF-SVM. Furthermore, UF preprocessing is obtained to be inefficient as it provides lower classification accuracy than conventional SVM and the differences in classification accuracies between direct SVM and UF-SVM are obtained to be statistically insignificant in all TDRs. In particular, for small training data sizes, the proposed approaches provide a significant increase in classification accuracy in case of 2-D-EMD-SVM as well as FA-2-D-EMD-SVM. For example, using a TDR of 10%, the OA of 2-D-EMD-SVM is 13% higher than that of standard SVM, 6% higher than that of WD-SVM, and 8% higher than that of LPF-SVM. In order to enable visual evaluation, Fig. 4 shows the sum of the first two IMFs, as well as the WD, LPF, and UF results for band #28 of the Indian Pine image.

EMP, which is an MP-based classifier, is obtained to give the best classification accuracies in all TDRs compared with the other MP-based classifiers such as Spec-EMP and FE-Spec-EMP for Indian pine data. If EMP is compared with 2-D-EMD-SVM and FA-2-D-EMD-SVM, it is seen that 2-D-EMD-SVM gives higher classification accuracy than EMP, but EMP provides higher classification accuracy than that of FA-2-D-EMD-SVM in all TDR cases. For example, using a TDR of 35%, the OA of 2-D-EMD-SVM is 99.67, that of EMP is 99.27, and that of FA-2-D-EMD-SVM is 98.00.

The mean class accuracy of each class for the TDR case of 35% is given in Table VI for the best results of the proposed approach, EMP, and direct SVM. It is seen that, for most of the classes, the proposed approach is better than EMP, but for a few classes, EMP provides higher accuracies.

TABLE V
OA, k, AND Z VALUES OF DIRECT SVM, WD-SVM, UF-SVM, LPF-SVM, EMP, SPEC-EMP, FE-SPEC-EMP, 2-D-EMD-SVM, AND FA-2-D-EMD-SVM USING 10% AND 35% TDRS FOR INDIAN PINE DATA

Method		10% TDR			35% TDR			65% TDR		
		OA	k	Z	OA	k	Z	OA	k	Z
SVM [2,3]		82.24	0.70	-	91.47	0.83	-	92.80	0.86	-
WD-SVM [18]		89.04	0.73	17.45	96.62	0.94	14.79	97.52	0.95	10.43
UF-SVM [15]		81.18	0.69	-4.21	90.35	0.81	-7.34	91.92	0.85	-3.83
LPF-SVM		87.47	0.72	13.68	94.38	0.89	7.92	96.02	0.93	6.68
EMP [9]		94.56	0.73	29.13	99.27	0.99	21.47	99.54	0.99	14.25
Spec-EMP [10]		92.47	0.71	24.95	98.53	0.98	19.71	99.02	0.98	13.40
FE-Spec-EMP [10]		93.19	0.78	27.65	98.33	0.97	19.46	98.30	0.97	12.16
2D-EMD-SVM ($\tau = 0.004$, $W = 3\times3$)	1 IMF	71.59	0.53	-18.40	91.31	0.87	-0.34	94.43	0.91	2.95
	2 IMFs	94.31	0.72	27.90	99.30	0.97	21.35	99.57	0.99	14.08
	3 IMFs	93.04	0.66	26.65	98.36	0.93	18.94	98.85	0.98	13.15
	4 IMFs	89.30	0.84	16.44	94.55	0.87	7.93	95.34	0.91	4.95
2D-EMD-SVM ($\tau = 0.008$, $W = 3\times3$)	1 IMF	78.86	0.59	-13.42	93.09	0.88	3.69	95.40	0.92	4.78
	2 IMFs	94.16	0.72	27.92	99.27	0.97	21.40	99.47	0.99	14.04
	3 IMFs	92.66	0.66	25.92	97.55	0.93	16.93	97.94	0.92	11.22
	4 IMFs	92.57	0.64	25.65	97.73	0.92	17.33	97.91	0.92	11.12
2D-EMD-SVM ($\tau = 0.01$, $W = 3\times3$)	1 IMF	78.01	0.60	-8.04	94.04	0.90	5.99	96.18	0.94	6.32
	2 IMFs	94.25	0.71	28.12	99.16	0.97	21.02	99.60	0.99	14.18
	3 IMFs	92.80	0.67	26.25	97.43	0.94	16.57	97.98	0.92	11.03
	4 IMFs	92.62	0.64	25.86	97.41	0.94	16.51	97.88	0.92	11.03
2D-EMD-SVM ($\tau = 0.006$, $W = 3\times3$)	1 IMF	73.96	0.54	-14.75	92.13	0.87	1.45	95.01	0.92	4.06
	2 IMFs	94.00	0.73	27.56	99.13	0.97	21.01	99.41	0.99	13.90
	3 IMFs	92.56	0.64	25.79	97.59	0.94	16.98	97.98	0.91	11.05
	4 IMFs	92.23	0.65	25.30	97.43	0.93	16.64	97.75	0.92	10.74
2D-EMD-SVM ($\tau = 0.006$, $W = 5\times5$)	1 IMF	79.73	0.63	-4.77	95.71	0.91	10.51	97.75	0.95	9.97
	2 IMFs	**95.52**	**0.81**	**30.36**	**99.67**	**0.99**	**22.30**	99.70	0.99	14.42
	3 IMFs	94.43	0.70	28.93	99.11	0.97	20.95	99.21	0.97	13.69
	4 IMFs	87.71	0.77	12.32	93.77	0.89	5.62	95.24	0.92	4.66
2D-EMD-SVM ($\tau = 0.006$, $W = 7\times7$)	1 IMF	87.87	0.61	12.71	98.33	0.96	18.53	99.08	0.98	13.16
	2 IMFs	93.14	0.67	26.41	99.40	0.98	21.87	**99.73**	**0.98**	**14.52**
	3 IMFs	93.41	0.69	27.14	98.33	0.93	19.11	99.12	0.97	13.78
	4 IMFs	90.98	0.80	21.55	95.60	0.92	11.25	96.67	0.91	8.35
FA-2D-EMD-SVM ($W = 3\times3$)	1 IMF	81.86	0.72	-0.78	92.22	0.87	1.75	94.00	0.91	2.91
	2 IMFs	91.97	0.75	23.73	97.99	0.97	18.72	98.63	0.98	12.88
	3 IMFs	91.78	0.74	23.62	98.00	0.97	18.87	98.76	0.98	13.10
FA-2D-EMD-SVM ($W = 5\times5$)	1 IMF	85.69	0.70	7.66	95.24	0.92	9.49	96.28	0.93	6.86
	2 IMFs	91.59	0.72	22.76	97.57	0.96	17.64	98.11	0.97	11.91
	3 IMFs	83.20	0.71	3.78	92.37	0.86	3.29	94.39	0.92	4.69
	4 IMFs	81.69	0.68	-3.86	90.01	0.79	-8.20	92.08	0.86	-3.56
FA-2D-EMD-SVM ($W = 7\times7$)	1 IMF	89.94	0.71	18	97.23	0.94	15.63	97.85	0.96	10.82
	2 IMFs	91.29	0.75	22.35	97.24	0.95	16.72	98.14	0.97	11.83
	3 IMFs	81.59	0.68	-3.89	89.91	0.79	-7.74	92.34	0.86	-1.90
	4 IMFs	82.24	0.69	0	91.47	0.83	0	92.80	0.86	0



(a) (b) (c) (d)

Fig. 4. Indian Pine image band #28. (a) Sum of first two IMFs. (b) WD result. (c) LPF result. (d) UF result.

The OA and k results of CK-SS-SVM, CK-2-D-EMD-SVM, and CK-FA-2-D-EMD-SVM are given in Table VII for 10% and 35% TDRs. In Table VII, it is seen that CK-2-D-EMD-SVM and CK-FA-2-D-EMD-SVM give superior performance compared with direct SVM as well as CK-SS-SVM for the Indian Pine data. For example, in a TDR case of 10%, the

TABLE VI
MEAN CLASS ACCURACY OF EACH CLASS IN CASE OF 35% TDR FOR THE INDIAN PINE DATA

Method	Class Labels								
	1	2	3	4	5	6	7	8	9
SVM	86.61	80.75	96.67	99.79	99.07	82.63	91.91	92.17	99.53
2D-EMD-SVM	**99.26**	**99.64**	99.69	**100**	99.69	**99.53**	**99.81**	**99.26**	**100**
EMP	98.95	**99.64**	**100**	**100**	**100**	97.36	99.33	**99.26**	99.76

TABLE VII
OA AND k VALUES OF DIRECT SVM, CK-SS-SVM, CK-2-D-EMD-SVM, AND CK-FA-2-D-EMD-SVM USING 10% AND 35% TDRS FOR INDIAN PINE DATA

Composite Kernel Approach	Method	10% TDR		35% TDR	
		OA	k	OA	k
-	SVM	82.24	0.70	91.47	0.83
Direct	CK-SS-SVM	88.69	0.73	96.70	0.94
	CK-2D-EMD-SVM, $\tau = 0.006$	94.98	0.77	99.75	0.99
	CK-2D-EMD-SVM, $\tau = 0.004$	95.28	0.77	99.73	0.99
	CK- FA-2D-EMD-SVM	93.21	0.65	99.74	0.99
Weighted, $\mu = 0.1$	CK-SS-SVM	87.52	0.69	95.48	0.90
	CK-2D-EMD-SVM, $\tau = 0.006$	95.43	0.77	99.77	0.99
	CK-2D-EMD-SVM, $\tau = 0.004$	**95.45**	**0.77**	99.85	0.99
	CK- FA-2D-EMD-SVM	94.30	0.65	**99.87**	**0.99**
Weighted, $\mu = 0.2$	CK-SS-SVM	87.17	0.65	95.63	0.91
	CK-2D-EMD-SVM, $\tau = 0.006$	95.24	0.76	99.80	0.99
	CK-2D-EMD-SVM, $\tau = 0.004$	95.45	0.77	99.83	0.99
	CK- FA-2D-EMD-SVM	94.13	0.65	99.83	0.99
Weighted, $\mu = 0.3$	CK-SS-SVM	86.97	0.65	95.55	0.91
	CK-2D-EMD-SVM, $\tau = 0.006$	95.11	0.76	99.77	0.99
	CK-2D-EMD-SVM, $\tau = 0.004$	95.44	0.77	99.82	0.99
	CK- FA-2D-EMD-SVM	93.88	0.65	99.77	0.99
Weighted, $\mu = 0.4$	CK-SS-SVM	86.65	0.66	95.45	0.91
	CK-2D-EMD-SVM, $\tau = 0.006$	95.11	0.76	99.79	0.99
	CK-2D-EMD-SVM, $\tau = 0.004$	95.32	0.77	99.82	0.99
	CK- FA-2D-EMD-SVM	93.57	0.65	99.69	0.99
Weighted, $\mu = 0.5$	CK-SS-SVM	86.34	0.67	95.16	0.90
	CK-2D-EMD-SVM, $\tau = 0.006$	95.26	0.77	99.75	0.99
	CK-2D-EMD-SVM, $\tau = 0.004$	95.28	0.77	99.82	0.99
	CK- FA-2D-EMD-SVM	93.21	0.65	99.67	0.99
Weighted, $\mu = 0.6$	CK-SS-SVM	85.97	0.68	94.65	0.89
	CK-2D-EMD-SVM, $\tau = 0.006$	94.77	0.75	99.71	0.99
	CK-2D-EMD-SVM, $\tau = 0.004$	93.30	0.73	99.77	0.99
	CK- FA-2D-EMD-SVM	92.75	0.67	99.64	0.99
Weighted, $\mu = 0.7$	CK-SS-SVM	85.54	0.69	94.26	0.88
	CK-2D-EMD-SVM, $\tau = 0.006$	94.54	0.76	99.67	0.99
	CK-2D-EMD-SVM, $\tau = 0.004$	95.09	0.77	99.71	0.99
	CK- FA-2D-EMD-SVM	92.45	0.67	99.53	0.99
Weighted, $\mu = 0.8$	CK-SS-SVM	85.06	0.70	93.91	0.86
	CK-2D-EMD-SVM, $\tau = 0.006$	93.99	0.76	99.51	0.98
	CK-2D-EMD-SVM, $\tau = 0.004$	94.74	0.76	99.61	0.99
	CK- FA-2D-EMD-SVM	92.05	0.68	99.35	0.99
Weighted, $\mu = 0.9$	CK-SS-SVM	84.22	0.72	93.14	0.85
	CK-2D-EMD-SVM, $\tau = 0.006$	92.62	0.75	99.13	0.98
	CK-2D-EMD-SVM, $\tau = 0.004$	93.83	0.75	99.43	0.99
	CK- FA-2D-EMD-SVM	91.17	0.71	99.13	0.98

best OA is obtained in case of using CK-2-D-EMD-SVM. In this case, the OA of CK-2-D-EMD-SVM is 95.45, that of CK-FA-2-D-EMD-SVM is 94.30, that of CK-SS-SVM is 88.69, and that of direct SVM is only 82.24. Note that the large differences in classification accuracies between direct SVM, CK-2-D-EMD-SVM, CK-FA-2-D-EMD-SVM, and CK-SS-SVM are all obtained to provide statistically significant increases in all TDRs; therefore, Z values are not given in this table.

The best results of MP-based classification are compared with the best results of the proposed approaches for different TDRs for the Indian Pine data in Table VIII. OA, k, and Z

TABLE VIII
BEST RESULTS OF MP-BASED CLASSIFIER AND PROPOSED APPROACHES FOR INDIAN PINE DATA

Method	10% TDR			35% TDR		
	OA	k	Z	OA	k	Z
EMP	94.56	0.73	-	99.27	0.99	-
Proposed	95.52	0.81	5.42	99.87	0.99	4.03

TABLE IX
AAS OF DIRECT SVM, EMD, 2-D-EMD-SVM, AND CK-2-D-EMD-SVM AVERAGED FOR TEN TRIALS WITH DIFFERENT TRAINING SETS IN CASE OF 10% TDRS FOR THE INDIAN PINE DATA

Method	AA
SVM	82.24
EMP	94.51
2D-EMD-SVM (2 IMF case)	95.27
CK-2D-EMD-SVM	95.11

TABLE X
OA VALUES OF DIRECT SVM AND 1-D-EMD-SVM (APPLIED IN SPECTRAL DOMAIN) USING 10%, 35%, AND 65% TDRS FOR INDIAN PINE DATA

Method		10%TDR	35%TDR	65%TDR
		OA	OA	OA
SVM		82.24	91.47	92.80
1D-EMD-SVM	1 IMF	73.20	83.02	82.73
	2 IMFs	55.44	77.63	80.24
	3 IMFs	59.83	78.75	81.07

values are provided in this table. Z values are computed, with the second classifier being the case that gives the best results for MP-based classification and the first classifier being the case that gives the best results for the proposed approaches. It is seen that the proposed approaches are more successful for the Indian Pine data that is rather difficult to classify.

In order to better demonstrate the performance of the proposed approaches, the average accuracies (AAs) of direct SVM, EMP, as well as the 2 IMFs cases of 2-D-EMD-SVM and CK-2-D-EMD-SVM obtained over ten trials with different training sets are provided in Table IX for the 10% TDR case. It is observed from this table that the proposed approaches increase the average classification accuracy.

For informative purposes only, 1-D-EMD-SVM results are given in Table X for the Indian Pine data set, with EMD being applied in the spectral dimension. It is seen that, in this case, classification results are below original SVM results in all TDRs and that applying 1-D-EMD in the spectral domain is found to be inefficient.

Overall, experimental results show that the EMD-based approach can significantly increase the classification accuracy of hyperspectral images. For cases with a small number of training samples, it is observed that 2-D-EMD provides higher classification accuracy than that of FA-2-D-EMD (for 10% TDR, the best classification accuracy of 2-D-EMD-SVM is 95.52%, while the best classification accuracy of FA-2-D-EMD-SVM is 91.97%) and the difference in performance is less if the composite kernel approach is used (for 10% TDR, the best classification accuracy of CK-2-D-EMD-SVM is 95.45%, while the best classification accuracy of CK-FA-2-D-EMD-SVM is 94.30%). The difference between using the sum of lower order IMFs or using composite kernels is negligible for 2-D-EMD, while the composite kernel approach is favorable for FA-2-D-EMD. For cases with a higher number of training samples (e.g., 35% TDR case), it is seen that the composite kernel EMD approach provides increased classification accuracies than the sum of lower order IMF-based EMD approach. Furthermore, in this case, CK-FA-2-D-EMD-SVM provides about the same accuracy as CK-2-D-EMD-SVM. For the composite kernel case, it is observed that a small weighting factor of $\mu = 0.1$ provides the best performance. In all cases, it is seen that the proposed approach significantly increases the classification accuracy compared with direct SVM.

TABLE XI
OA VALUES OF KNN AND 2-D-EMD-KNN USING 10%, 35%, AND 65% TDRS FOR INDIAN PINE DATA

Method		10%TDR	35%TDR	65%TDR
		OA	OA	OA
KNN		74.48	83.63	85.36
2D-EMD-KNN	1 IMF	48.92	67.48	72.92
	2 IMFs	86.04	95.28	96.15
	3 IMFs	92.18	98.17	98.28
	4 IMFs	92.32	98.26	98.48

As stated before, the first approach proposed in this paper is not restricted to kernel-based classifiers. In order to show the effectiveness of this approach for a different classifier, classification results of the k-nearest neighbor classifier [36] are provided for the Indian Pine data set. Conventional k-nearest neighbor classification results (denoted as KNN) as well as KNN with 2-D-EMD (denoted as 2-D-EMD-KNN) are shown in Table XI for 10%, 35%, and 65% TDRs. The results of 2-D-EMD-KNN are provided for the cases of 1 IMF, 2 IMFs, 3 IMFs, and 4 IMFs. The value k of KNN is varied in the range [3], [7], and the best results are obtained for $k = 3$. These results are provided in the corresponding table. The best result of 2-D-EMD-KNN is obtained in the case of 3 IMFs, and in this case, the classification accuracy is increased by at least 11% compared with KNN for all TDRs. These results confirm that the proposed algorithm provides good results independent from the classifier.

B. Experimental Results for the DC Mall Data

Because of the higher spatial size of the DC Mall data, it was not possible to implement 2-D-EMD (because of the high computational complexity), and therefore, only FA-2-D-EMD is used to demonstrate results. In this case, 182 of the bands have three IMFs and 9 of the bands have four IMFs. The window size W is fixed to 3×3.

Experimental results of the DC Mall data are provided only for TDR values of 5%, 10%, and 35%. Table XII shows the comparison results of FA-2-D-EMD-SVM with direct SVM, WD-SVM, UF-SVM, LPF-SVM, and EMP. From the table, it is seen that the best results of FA-2-D-EMD-SVM are obtained using 2 IMFs in all TDRs. The differences in classification accuracies between direct SVM and FA-2-D-EMD-SVM in case of 2 IMFs are obtained to be statistically significant. Note that, comparison with other algorithms is carried out using the best results of FA-2-D-EMD-SVM. It is seen that

TABLE XII
OA, k, AND Z VALUES OF DIRECT SVM, WD-SVM, UF-SVM, LPF-SVM, EMP, 2-D-EMD-SVM, FA-2-D-EMD-SVM, EMP, SPEC-EMP, AND FE-SPEC-EMP USING 5%, 10%, AND 35% TDRS FOR DC MALL DATA

Method		5% TDR			10% TDR			35% TDR		
		OA	k	Z	OA	k	Z	OA	k	Z
SVM [2,3]		99.17	0.96	-	99.54	0.98	-	99.66	0.98	-
WD-SVM [18]		99.17	0.96	0	99.64	0.98	2.11	99.75	0.98	1.66
UF-SVM [15]		96.21	0.92	-10.92	98.17	0.92	-6.99	99.48	0.97	-1.65
LPF-SVM		97.86	0.93	-8.98	98.74	0.93	-6.48	99.52	0.97	-1.33
EMP[9]		**99.66**	**0.98**	**4.14**	**100**	**1**	**5.74**	**100**	**1**	**4.24**
FA-2D-EMD-SVM	1 IMF	95.02	0.91	-16.17	97.85	0.95	-9.12	99.66	0.99	0
	2 IMFs	99.53	0.98	3.34	99.88	0.99	3.90	99.94	0.99	3.87
	3 IMFs	99.48	0.98	4.24	99.74	0.99	1.94	99.91	0.99	2.71
	4 IMFs	99.48	0.98	4.24	99.73	0.99	1.94	99.96	0.99	3.57

TABLE XIII
OA, k, AND Z VALUES OF DIRECT SVM, CK-SS-SVM, CK-2-D-EMD-SVM, AND CK-FA-2-D-EMD-SVM USING 5%, 10%, AND 35% TDR FOR DC MALL DATA

Composite Kernel Approach	Methods	5% TDR		10% TDR		35% TDR	
		OA	k	OA	k	OA	k
-	SVM	99.17	0.96	99.54	0.98	99.66	0.98
Direct	CK-SS-SVM	99.20	0.96	99.58	0.98	99.83	0.99
	CK-FA-2D-EMD-SVM	99.59	0.98	99.90	0.99	100	1
Weighted, $\mu = 0.1$	CK-SS-SVM	99.24	0.96	99.64	0.98	99.64	0.98
	CK- FA-2D-EMD-SVM	**99.61**	**0.98**	99.97	0.99	100	1
Weighted, $\mu = 0.2$	CK-SS-SVM	99.24	0.96	99.64	0.98	99.68	0.98
	CK- FA-2D-EMD-SVM	99.59	0.98	**100**	**1**	**100**	**1**
Weighted, $\mu = 0.3$	CK-SS-SVM	99.27	0.96	99.62	0.98	99.68	0.98
	CK- FA-2D-EMD-SVM	99.58	0.98	99.97	0.99	100	1
Weighted, $\mu = 0.4$	CK-SS-SVM	99.27	0.96	99.58	0.98	99.77	0.98
	CK- FA-2D-EMD-SVM	99.59	0.98	99.91	0.99	100	1
Weighted, $\mu = 0.5$	CK-SS-SVM	99.24	0.96	99.58	0.98	99.83	0.98
	CK- FA-2D-EMD-SVM	99.53	0.98	99.90	0.99	100	1
Weighted, $\mu = 0.6$	CK-SS-SVM	99.24	0.96	99.60	0.97	99.81	0.98
	CK- FA-2D-EMD-SVM	99.46	0.97	99.77	0.99	100	1
Weighted, $\mu = 0.7$	CK-SS-SVM	99.24	0.96	99.60	0.97	99.81	0.98
	CK- FA-2D-EMD-SVM	99.46	0.97	99.71	0.98	100	1
Weighted, $\mu = 0.8$	CK-SS-SVM	99.24	0.96	99.60	0.97	99.81	0.98
	CK- FA-2D-EMD-SVM	99.39	0.97	99.68	0.98	100	1
Weighted, $\mu = 0.9$	CK-SS-SVM	99.17	0.96	99.60	0.97	99.79	0.98
	CK- FA-2D-EMD-SVM	99.18	0.96	99.51	0.98	100	1

FA-2-D-EMD-SVM provides higher classification accuracies than those of LPF-SVM, UF-SVM, and WD-SVM in all TDRs. LPF-SVM and UF-SVM results show that LPF and UF preprocessing are not very efficient for the DC Mall data, because the differences in classification accuracies between direct SVM and LPF-SVM and between direct SVM and UF-SVM are obtained to be statistically insignificant in all TDRs. The classification accuracies of MP-based approaches are about the same as that of the proposed FA-2-D-EMD-SVM algorithm, and the difference in accuracy is statistically insignificant.

In Table XIII, it is seen that CK-FA-2-D-EMD-SVM gives better results compared with direct SVM and CK-SS-SVM. For example, in a TDR case of 10%, the OA of CK-FA-2-D-EMD-SVM is 100, that of direct SVM is 99.54, and that of CK-SS-SVM is 99.58. In addition, CK-FA-2-D-EMD-SVM provides 100% OA in at least one composite kernel case in 10% and 35% TDRs. Since the differences of classification accuracies between direct SVM and CK-FA-2-D-EMD-SVM are obtained to be statistically significant in all TDRs, Z values are not given in the tables.

Overall, experimental results of DC Mall data show that the EMD-based approaches can considerably improve the classification accuracy of hyperspectral images compared with direct SVM.

VIII. CONCLUSION

In this paper, EMD-based hyperspectral-image-classification approaches have been addressed. EMD is used to decompose each hyperspectral image band into IMFs and a final residue. Two EMD-based approaches are proposed in this paper: reconstruction of hyperspectral image bands as the sum of lower order IMFs for classification and combination of the information contained in the first and second IMFs by means of composite kernels. In order to implement EMD in this paper, two methods are used: 1) 2-D-EMD [26] (which is computationally expensive particularly for small values of W) and 2) FA-2-D-EMD [27] (which is proposed as an alternative fast implementation of 2-D-EMD). Experimental results show that EMD-based approaches can significantly improve the SVM classification

accuracy for hyperspectral images. It is important to note that the computational complexity of the proposed algorithms is much higher than that of the other compared methods if 2-D-EMD is used with small values of W. Therefore, 2-D-EMD is not suitable for hyperspectral images which have a high resolution, and in this case, it is recommended to use FA-2-D-EMD. The classification accuracy is found to be better than EMP-based approaches if 2-D-EMD is used and to be about the same if FA-2-D-EMD is used. However, EMD is a relatively recent approach and research on fast and accurate EMD techniques is still ongoing. EMD has not been exploited for utilization in hyperspectral image classification so far, and the possibility of improving the classification accuracy even further by exploiting different aspects of EMD, such as decision fusion of different EMD-based representations for example [37], seems very promising for future research.

ACKNOWLEDGMENT

The authors would like to thank Prof. D. Landgrebe for providing the Indian Pine and DC Mall data sets [1], [20] and A. Linderhed for the 2-D-EMD-SVM software.

REFERENCES

[1] D. A. Landgrebe, *Signal Theory Methods in MultiSpectral Remote Sensing*. Hoboken, NJ: Wiley, 2003.

[2] V. Vapnik, *Statistical Learning Theory*. New York: Wiley, 1998.

[3] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 6, pp. 1351–1362, Jun. 2005.

[4] 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, Aug. 2004.

[5] G. Camps-Valls, L. Gomez-Chova, J. Calpe-Maravilla, J. D. Martin-Guerrero, E. Soria-Olivas, L. Alonso-Chorda, and J. Moreno, "Robust support vector method for hyperspectral data classification and knowledge discovery," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 7, pp. 1530–1542, Jul. 2004.

[6] B. Demir and S. Ertürk, "Hyperspectral image classification using relevance vector machines," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 4, pp. 586–590, Oct. 2007.

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

[8] M. Fauvel, J. Chanussot, and J. A. Benediktsson, "Adaptive pixel neighborhood definition for the classification of hyperspectral images with support vector machines and composite kernel," in *Proc. IEEE Int. Conf. Image Process.*, San Diego, CA, 2008, pp. 1884–1887.

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

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

[11] A. Farag, R. Mohamed, and A. El-Baz, "A unified framework for map estimation in remote sensing image segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 7, pp. 1617–1634, Jul. 2005.

[12] Y. Tarabalka, J. Chanussot, J. A. Benediktsson, J. Angulo, and M. Fauvel, "Segmentation and classification of hyperspectral data using watershed," in *Proc. IGARSS*, Boston, MA, 2008, pp. III-652–III-655.

[13] M. Lennon, G. Mercier, and L. Hubert-Moy, "Nonlinear filtering of hyperspectral images with anisotropic diffusion," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Toronto, ON, Canada, 2002, pp. 2477–2479.

[14] R. D. Phillips, C. E. Blinn, L. T. Watson, and R. H. Wynne, "An adaptive noise-filtering algorithm for AVIRIS data with implications for classification accuracy," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 9, pp. 3168–3179, Sep. 2009.

[15] K. Perumal and R. Bhaskaran, "SVM based effective land use classification system for multispectral remote sensing images," *Int. J. Comput. Sci. Inf. Security*, vol. 6, no. 2, pp. 97–105, 2009.

[16] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. R. Soc. Lond. A, Math. Phys. Sci.*, vol. 454, no. 1971, pp. 903–995, Mar. 1998.

[17] A. Linderhed, "Adaptive image compression with wavelet packets and empirical mode decomposition," Ph.D. dissertation, Linköping Stud. Sci. Technol., Linköping, Sweden, 2004, Dissertation No. 909.

[18] A. Pizurica and W. Philips, "Estimating the probability of the presence of a signal of interest in multiresolution single- and multiband image denoising," *IEEE Trans. Image Process.*, vol. 15, no. 3, pp. 654–665, Mar. 2006.

[19] B. Demir and S. Ertürk, "Improved hyperspectral image classification with noise reduction pre-process," in *Proc. Eur. Signal Process. Conf.*, Lausanne, Switzerland, Aug. 2008.

[20] D. Landgrebe, AVIRIS NW Indiana's Indian Pines 1992 data set. [Online]. Available: http://dynamo.ecn.purdue.edu/~biehl/MultiSpec/documentation.html

[21] C.-I. Chang, *Recent Advances in Hyperspectral Signal and Image Processing*. Kerala, India: Transworld Res. Netw., 2006.

[22] Z. Zhidong and W. Yang, "A new method for processing end effect in empirical mode decomposition," in *Proc. Int. Conf. Commun., Circuits Syst.*, Kokura, Japan, 2007, pp. 841–845.

[23] A. Janusauskas, R. Jurkonis, A. Lukosevicius, S. Kurapkiene, and A. Paunksnis, "The empirical mode decomposition and the discrete wavelet transform for detection of human cataract in ultrasound signals," *Inf. Lith. Acad. Sci.*, vol. 16, no. 4, pp. 541–556, 2005.

[24] P. Goncalves, P. Abry, G. Rilling, and P. Flandrin, "Fractal dimension estimation: Empirical mode decomposition versus wavelets," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Honolulu, HI, 2007, pp. III-1153–III-1156.

[25] R. Bhagavatula and M. Savvides, "Analyzing facial images using empirical mode decomposition for illumination artifact removal and improved face recognition," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Honolulu, HI, 2007, pp. I-505–I-508.

[26] A. Linderhed, "Image compression based on empirical mode decomposition," in *Proc. SSAB Symp. Image Anal.*, Uppsala, Sweden, 2004, pp. 110–113.

[27] S. M. A. Bhuiyan, R. R. Adhami, and J. F. Khan, "Fast and adaptive bidimensional empirical mode decomposition using order-statistics filter based envelope estimation," *EURASIP J. Adv. Signal Process.*, vol. 2008, no. 3, pp. 1–18, Jan. 2008, Article ID 728356.

[28] S. M. A. Bhuiyan, R. R. Adhami, and J. F. Khan, "A novel approach of fast and adaptive bidimensional empirical mode decomposition," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Las Vegas, NV, 2008, pp. 1313–1316.

[29] K. L. Wu and P. F. Hsieh, "Empirical mode decomposition for dimensionality reduction of hyperspectral data," in *Proc. IGARSS*, Seoul, Korea, 2005, pp. 1241–1244.

[30] B. Demir and S. Ertürk, "Empirical mode decomposition pre-process for higher accuracy hyperspectral image classification," in *Proc. IGARSS*, Boston, MA, 2008, pp. II-939–II-941.

[31] S. R. Long and N. E. Huang, "On the normalized Hilbert transform and its applications in remote sensing," in *Signal Processing for Remote Sensing*, C. H. Chen, Ed. Boca Raton, FL: CRC Press, 2007.

[32] B. Weng and K. E. Barner, "Optimal signal reconstruction using the empirical mode decomposition," *EURASIP J. Adv. Signal Process.*, vol. 2008, pp. 1–12, 2008, Article ID 845294.

[33] J. A. Richards and X. Jia, *Remote Sensing Digital Image Analysis: An Introduction*. New York: Springer-Verlag, 1999.

[34] G. M. Foody, "Thematic map comparison: Evaluating the statistical significance of differences in classification accuracy," *Photogramm. Eng. Remote Sens.*, vol. 70, no. 5, pp. 627–633, May 2004.

[35] F. L. Bookstein, "Principal warps: Thin-plate splines and the decomposition of deformations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 6, pp. 567–585, Jun. 1989.

[36] E. Alpaydýn, *Introduction to Machine Learning*. Cambridge, MA: MIT Press, 2004.

[37] B. Demir and S. Ertürk, "Empirical mode decomposition based decision fusion for higher hyperspectral image classification accuracy," in *Proc. IGARSS*, Honolulu, HI, 2010, to be published.

Begüm Demir (S'06) received the B.S. and M.Sc. degrees in electronic and telecommunication engineering from Kocaeli University, Kocaeli, Turkey, in 2005 and 2007, respectively, where she is currently working toward the Ph.D. degree in the Department of Electronics and Telecommunication Engineering.

In 2005, she joined the Kocaeli University Laboratory of Image and Signal Processing as a Researcher. Since 2006, she has been a Research Assistant with the Department of Electronics and Telecommunication Engineering, Kocaeli University. She was a Visiting Researcher at the Remote Sensing Laboratory, University of Trento, Trento, Italy, during March–September 2009, supported by the International Research Fellowship Program (2214) of The Scientific and Technological Research Council of Turkey (TUBITAK). Her main research interests include image processing and machine learning with applications to remote sensing image analysis.

Ms. Demir is a referee for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, and the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING.

Sarp Ertürk (M'99) received the B.Sc. degree in electrical and electronics engineering from Middle East Technical University, Ankara, Turkey, in 1995 and the M.Sc. degree in telecommunication and information systems and the Ph.D. degree in electronic systems engineering from the University of Essex, Colchester, U.K., in 1996 and 1999, respectively.

From 1999 to 2001, he carried out his compulsory service at the Army Academy, Ankara. He is currently appointed as a Full Professor at Kocaeli University, Kocaeli, Turkey, where he was an Assistant Professor between 2001 and 2002 and an Associate Professor between 2002 and 2007. His research interests are in the area of digital signal and image processing, video coding, remote sensing, and digital communications.