Article

Spectral DWT Multilevel Decomposition with Spatial Filtering Enhancement Preprocessing-Based Approaches for Hyperspectral Imagery Classification

Razika Bazine , Huayi Wu and Kamel Boukhechba

The State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan 430079, China; wuhuayi@whu.edu.cn (H.W.); bk_kamel@whu.edu.cn (K.B.)

* Correspondence: bazinerazika@whu.edu.cn; Tel.: +86-27-68778311; Fax: +86-27-68778003

Received: 9 November 2019; Accepted: 3 December 2019; Published: 5 December 2019

Abstract: In this paper, spectral–spatial preprocessing using discrete wavelet transform (DWT) multilevel decomposition and spatial filtering is proposed for improving the accuracy of hyperspectral imagery classification. Specifically, spectral DWT multilevel decomposition (SDWT) is performed on the hyperspectral image to separate the approximation coefficients from the detail coefficients. For each level of decomposition, only the detail coefficients are spatially filtered instead of being discarded, as is often adopted by the wavelet-based approaches. Thus, three different spatial filters are explored, including two-dimensional DWT (2D-DWT), adaptive Wiener filter (AWF), and two-dimensional discrete cosine transform (2D-DCT). After the enhancement of the spectral information by performing the spatial filter on the detail coefficients, DWT reconstruction is carried out on both the approximation and the filtered detail coefficients. The final preprocessed image is fed into a linear support vector machine (SVM) classifier. Evaluation results on three widely used real hyperspectral datasets show that the proposed framework using spectral DWT multilevel decomposition with 2D-DCT filter (SDWT-2DCT_SVM) exhibits a significant performance and outperforms many state-of-the-art methods in terms of classification accuracy, even under the constraint of small training sample size, and execution time.

Keywords: spectral–spatial hyperspectral classification; DWT multilevel decomposition; discrete cosine transform; structural filtering; SVM; Wiener filter

1. Introduction

Hyperspectral imagery contains rich discriminative spectral and spatial characteristics about surface materials [1], which makes it a prominent source of information for a varied range of applications in areas such as agriculture, environmental planning, surveillance, target detection, medicine [2–4], etc. Most of these hyperspectral applications rely on the supervised classification task. However, supervised classification faces challenges from the Hughes phenomenon, caused by the unbalance between high dimensionality and limited availability of training samples. Furthermore, embedded noises from both the sensors and the environment decrease the classification accuracy.

Thus, to address these problems, intensive work has been devoted to provide accurate classifiers for hyperspectral images, including [5] support vector machines (SVMs) [6], random forests [7], neural networks [8,9], active learning [10,11], and sparse representation [12–14] methods. Particularly, the SVM classifier has shown high performance with regards to the classification accuracy [6]. Although these aforementioned classifiers can fully make use of the spectral information in hyperspectral imagery, they do not incorporate spatial information. Thus, the resulting classification maps are often corrupted with salt-and-pepper noise [15].

Recent advances in hyperspectral image analysis have shown that the integration of spectral and spatial information in the classification approach can abolish salt-and-pepper noise in classification maps and noticeably improve image classification accuracy [16–18]. Many spectral–spatial hyperspectral image classification approaches have been proposed to improve the accuracy further based on various methodologies including structural filtering [19–21], morphological profile (MP) [22,23], random field [24], deep learning-based approaches [25,26], sparse representation-based approaches (SRC) [27,28], and segmentation-based approaches [29,30].

The structural filtering methodology is a widely used approach for spatial preprocessing of hyperspectral images [31]. In contrast to other spatial methodologies, spatial filtering is simple and easy to implement [32]. This advantage makes it particularly suitable for practical applications. A great deal of ongoing research is focused on structural filtering to obtain the contextual features where the simplest way is to extract the spatial information based on moment criteria [33,34]. However, recent work demonstrates that transform-based filters can retain texture features, edges, and details while discarding noise [35]. At the same time, much attention has been paid to wavelet transform-based filtering [36,37]. For instance, in [19], a spectral–spatial classification framework for hyperspectral images was proposed based on using one-dimensional discrete wavelet transform (1D-DWT) to create the extended morphological profile (EMP) and two-dimensional DWT (2D-DWT) for image denoising before stacking the original data to the EMP prior to the support vector machine (SVM) classification. Another spectral–spatial classification method based on a combination of a 1D-DWT with kernel minimum noise fraction (KMNF) in the spectral domain and a 2D-DWT in the spatial domain was proposed in [38]. In this work, the combination of 1D-DWT, KMNF, and 2D-DWT is suggested to create spatial spectral Schroedinger eigenmap (SSSE) features. Similar to the approach proposed in [38], a classification framework based on the combination of 1D-DWT, kernel maximum autocorrelation factor (KMAF), and 2D-DWT with hidden Markov random fields as classification post processing was proposed in [39].

Three-dimensional filtering using three-dimensional DWT (3D-DWT) for hyperspectral imagery classification has received great attention and produced successful spectral–spatial classification approaches by considering hyperspectral image data as a 3D cube where spatial and spectral features are treated as a whole [40]. Specifically, the extracted spectral–spatial 3D-DWT coefficients are combined with other techniques to enhance the classification performance further [37]. For instance, in [40], the 3D-DWT texture extraction was combined with the sparse logistic regression for hyperspectral image classification. The 3D-DWT features have also been integrated with Markov random field in [41]. Moreover, following the recent trend of using convolutional neural networks (CNN) and deep learning-based approaches in hyperspectral image classification, in [42], the low-frequency sub-bands of the 3D redundant wavelet transform (3D-RDWT) were used to create multiple views that are integrated into a multi-view active learning approach. In [43], a framework based on combining CNN with 3D multi-resolution wavelets was proposed.

Although the existing wavelet-based spectral and spatial hyperspectral classification methods have been largely successful, there are still some issues to be solved. The main issue with wavelet-based approaches is the substitution of the original spectral signatures by only the corresponding approximation coefficients. Subsequently, when increasing the level of decomposition, the original spectral signatures get smoothed more, and they will be almost similar to each other, which affects the performance of the classifier. Nevertheless, the detail coefficients emphasizing differences between the spectral signatures must be further exploited to enhance the effectiveness of the classifier.

There is another issue related to the 3D-DWT-based approaches. Indeed, a simple observation of a hyperspectral data cube discloses that the degree of irregularity is higher in the spectral dimension than in the spatial dimensions [44]. Hence, the variation of the noise variance in the spectral dimension is, in general, more drastic than that in the spatial dimensions. However, the 3D wavelet filtering process implicitly considers that the noise variance is the same in the three dimensions and ignores the dissimilarity between the spatial and the spectral dimensions. Furthermore, for better image

representation, 3D feature extraction methods create a large number of spectral–spatial features, leading to another feature selection issue [40].

To handle these issues and fully exploit the spectral and spatial information in the original hyperspectral imagery, we propose three effective spectral–spatial hyperspectral classification frameworks based on performing the spatial filtering process in the spectral wavelet domain. The proposed approaches are based on a hybrid spectral and spatial filtering preprocessing to benefit from the signal nature dissimilarity in the spectral and the spatial dimensions. Thus, we perform a spectral multilevel DWT filter on the original dataset to separate the approximation part from the detail part that will be further filtered using a spatial filter for each level of decomposition, instead of being discarded, as it is usually adopted by wavelet-based approaches. Furthermore, the spatial filtering process is performed only on the detail part in the frequency domain to extract the remaining meaningful information effectively. Three different spatial filters are explored to filter the obtained detail information including 2D-DWT, two-dimensional discrete cosine transform (2D-DCT), and two-dimensional adaptive Wiener filter (2D-AWF). For each approach, a DWT reconstruction was performed on both the approximation and the filtered detail parts. Finally, the output of the DWT reconstruction is fed to the linear SVM classifier. As we fully exploit the spectral and spatial information, the SVM classifier is adopted in our frameworks for its robustness to the Hughes phenomenon [6] as well as for its capacity for generalization [6].

The main contribution of the paper is the application of the spatial filter in the transform domain on the detail coefficients for each level of decomposition rather than discarding this part of information, as usually adopted by the feature extraction-based approaches. The proposed frameworks make full use of the filtered spectral and spatial information of the hyperspectral image. Another important contribution of this paper concerns the simplicity of the tuning configuration. Indeed, recently, the proposed classifiers have reached very high classification accuracy where the competitive difference between them is more related to their tuning configuration simplicity. We provide effective classification frameworks involving the tuning at maximum a couple of parameters for each applied filter. The proposed frameworks fully exploit the spectral–spatial features in reasonable execution time, which is helpful to improve the classification performance. The rest of this paper is structured as follows: The principles of multilevel DWT decomposition (ML-DWT) and 2D-DWT are introduced in Section 2. The proposed frameworks are described in Section 3. Datasets and evaluation process are presented in Section 4. Section 5 gives the experimental setup and reports the results with the related analysis. In Section 6, some conclusions are given.

2. Materials and Methods

2.1. Multilevel Discrete Wavelet Transform Decomposition

Wavelet transform (WT) is constructed from small waves of varying frequency with limited duration. These waves are obtained from basic wavelet function by dilation and translation [16], where DWT is any wavelet transform for which the wavelets are discretely sampled. DWT provides a multiresolution wavelet analysis that represents and analyzes the signals at more than one resolution. Thus, features undetected at one level are easy to be detected at another level [17]. Therefore, the multilevel decomposition characteristic of DWT provides the features of an image at different resolutions.

For the application of hyperspectral data, the multilevel decomposition provides a simple hierarchical framework for interpreting the spectral information. Especially, the local spectral variation of a spectrum in different spectral bands at each scale can be detected and can provide useful information for hyperspectral image classification.

2.1.1. One-Dimensional Discrete Wavelet Transform (1D-DWT)

In the context of hyperspectral data, 1D-DWT is usually used for data reduction in the spectral dimension where the transform is applied on each pixel spectral curve x considered as a discrete

signal represented by a vector $x = [x_0, x_1, x_2, \dots, x_{N-1}]$ in N -dimensional spectral band space (N is the number of the spectral bands).

In DWT, the signal is passed through the low-pass and high-pass filters [28]. The output of the low-pass filter is the low-resolution version of the original signal, called approximation coefficients, where the output of the high-pass filter embeds the residual information of the original signal, known as detail coefficients. The most common approach to the multilevel discrete wavelet transform involves further decomposition of only the approximation coefficient at each subsequent level [45].

The best way to describe DWT is through an analysis filter bank in which the input signal $x[n]$ is decomposed into two sub-band signals. As shown in the Figure 1a, we represent the low-pass filter by $g(n)$ and the high-pass filter by $h(n)$. The output of each filter is then down-sampled by two to achieve the two sub-band signals. Thus, the low-pass filter $g(n)$ produces the approximation coefficient $x_{1,L}[n]$, which is the informative band, while high-pass filter $h(n)$ produce the detail coefficient $x_{1,H}[n]$. The original signal can be reconstructed by synthesis filters which take the upsampled $x_{1,L}[n]$ and $x_{1,H}[n]$, as shown in Figure 1b.

Figure 1. The principle of 1D-DWT. (a). DWT analysis filter bank; (b). DWT synthesis filter bank.

The multilevel DWT decomposition is obtained by the recursive application of the analysis filter bank on the low-frequency bands (the approximation coefficients) as shown in Figure 2.

Figure 2. The principle of multilevel 1D-DWT decomposition.

In our approach, the multilevel 1D-DWT decomposition is performed separately on the spectral curve of each pixel to split the approximation coefficients from the detail coefficients.

2.1.2. Two-Dimensional Discrete Wavelet Transform (2D-DWT) Filter Bank

The 2D-DWT can also be implemented using digital filters and down-samplers, similar to the 1D-DWT. For an input image $x[n,m]$, we simply take the 1D-DWT of the rows followed by the 1D-DWT of the resulting columns to obtain four frequency parts ($x_{1,LL}[n,m]$, $x_{1,LH}[n,m]$, $x_{1,HL}[n,m]$,

and $x_{1,HH}[n,m]$), as described in Figure 3. $x_{1,LL}[n,m]$ is the low frequency sub-band containing the approximation of the original image. $x_{1,LH}[n,m]$ is the high-frequency sub-band contain the vertical details of the image. $x_{1,HL}[n,m]$ is the high-frequency sub-band that contains the horizontal details of the image. $x_{1,HH}[n,m]$ is the high-frequency sub-band of the diagonal image.

Figure 3. The principle of two-dimensional DWT (2D-DWT) analysis filter bank.

In 2D-DWT multilevel decomposition, only the approximation coefficients are further decomposed. Figure 4 shows two levels of 2D-DWT decomposition.

Figure 4. Two levels of 2D-DWT decomposition.

In our approach, the 2D-DWT is explored as a spatial filter in the transform domain to extract the rest of useful information embedded in the detail part resulted from the application of spectral DWT.

Regarding the threshold technique, we opted the universal threshold of Donoho and Johnstone [46], scaled by a robust estimate of the variance. For a matrix $M(L,C)$ of size $[L,C]$, the universal threshold can be represented by the following formula:

$$\text{Thresh} = \sigma \sqrt{2 \ln(N)}, \quad (1)$$

where σ is the average variance of the noise and N is the signal length ($L \times C$).

σ is calculated using the median estimate method, as follows:

$$\sigma = \frac{\text{Median}(|M|)}{0.6745}. \quad (2)$$

3. The Proposed Approach

The proposed classification frameworks are based on using spectral DWT multilevel decomposition followed by spatial filtering of the detail coefficients for each level of decomposition. Thus, in our proposed approaches, we perform a spectral DWT multilevel decomposition on the original dataset to separate the most significant information concentrated in the approximation coefficients from the detailed and noisy information embedded in the detail coefficients. For each decomposition level, a spatial filter is carried out on the detail coefficients to filter the remaining meaningful information further. In the spatial filtering step, three spatial filters are explored, including 2D-DWT, 2D-AWF, and 2D-DCT. Therefore, regarding the type of the spatial filter, there are three different proposed frameworks: spectral DWT multilevel decomposition (SDWT)/2D-DWT, SDWT/2D-AWF, and SDWT/2D-DCT. The inverse SDWT is performed on all coefficients comprising the filtered ones to get the final preprocessed dataset that will be fed to the linear SVM classifier. The detailed flowchart of the proposed framework is illustrated in Figure 5, where the pipeline of the proposed spectral multilevel DWT decomposition-based classification approaches can be summarized in Algorithm 1.

Algorithm 1

Input: Hyperspectral imagery dataset $X \in \mathbb{R}^{M \times L \times P}$ (M rows, L columns, and P bands).

1: Flattening X to obtain a matrix $E \in \mathbb{R}^{N \times P}$ where $N = M \times L$.

2: Performing spectral multilevel DWT decomposition on the matrix E (on the P dimension) to obtain two matrices:

- $F \in \mathbb{R}^{N \times K}$ for the K approximation coefficients.
- $R \in \mathbb{R}^{N \times (P-K)}$ for the P – K detail coefficients.

3: Unflattening the matrix R to obtain the data cube $R_c \in \mathbb{R}^{M \times L \times (P-K)}$ of the detail coefficients.

4: Performing a spatial filter on each matrix $R_m \in \mathbb{R}^{M \times L}$ in the cube R_c , using one of the following filters for each approach (there are P-K matrices):

- 2D-DWT
- 2D-AWF
- 2D-DCT

5: Flattening the filtered data cube R_c to return back to the matrix $R \in \mathbb{R}^{N \times (P-K)}$.

6: Stack the two matrices F and R to obtain the matrix $D \in \mathbb{R}^{N \times P}$, with the same size as the matrix E in step 1.

7: Apply an inverse DWT on the matrix D to obtain the preprocessed matrix E.

8: Perform a linear SVM classifier on the resulting matrix E.

Output: Classification map.

Figure 5. Detailed flow chart of the proposed spectral DWT multilevel decomposition-based approaches.

4. Data and Evaluation Process

4.1. Data

Three hyperspectral scenes with variant spatial resolution downloaded at http://www.ehu.es/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes were used for the experiments including the Indian Pines, Salinas, and Pavia University datasets. The selected datasets were captured from two different sensors, including Airborne Visible Infrared Imaging Spectrometer (AVIRIS) and Reflective Optics System Imaging Spectrometer (ROSIS), capturing rural and urban areas.

The first dataset concerns the Indian Pines scene. This scene was captured over northwest Indiana's Indian Pines area by AVIRIS at a low spatial resolution of 20 m. This dataset has 220 bands of size 145×145 pixels where the bands affected by noise and water absorption were discarded, leaving 200 bands. The ground reference data contains 16 classes; among them, four classes have similar spectral signatures and are very difficult to discriminate. Besides the spectral similarity between the classes, the low spatial resolution makes this scene a challenging classification scenario [10]. Table 1 lists the related classes and describes the number of samples, the training, and the test points in this dataset.

Table 1. Classes of the Indian Pines scene with the number of testing and training samples.

Class	Type	Samples	Training	Testing
1	Alfalfa	46	23	23
2	Corn-notill	1428	100	1328
3	Corn-mintill	830	100	730
4	Corn	237	100	137
5	Grass-pasture	483	100	383
6	Grass-trees	730	100	630
7	Grass-pasture-mowed	28	14	14
8	Hay-windrowed	478	100	378
9	Oats	20	10	10
10	Soybean-notill	972	100	872
11	Soybean-mintill	2455	100	2355
12	Soybean-clean	593	100	493
13	Wheat	205	100	105
14	Woods	1265	100	1165
15	Buildings-grass-trees-drives	386	100	286
16	Stone-steel-towers	93	47	46

Figure 6 shows the original band and the ground truth data for the Indian Pines dataset.

Figure 6. The ground truth information of the Indian Pines dataset: (a) original dataset; (b) ground truth map; (c) training map; (d) test map.

The second test dataset was the Pavia University scene. This scene was captured by the ROSIS sensor over the Pavia area, northern Italy. It has 610×610 pixels with 1.3-m spatial resolution and contains 103 spectral bands. For classification purposes, the scene is divided into nine regions listed in Table 2 with the number of samples that belong to each region and the numbers of training and testing points.

Table 2. Classes of Pavia University scene with the related numbers of testing and training samples.

Class	Type	Samples	Training	Testing
1	Asphalt	6631	300	6331
2	Meadows	18,649	300	18,349
3	Gravel	2099	300	1799
4	Trees	3064	300	2764
5	Painted metal sheets	1345	300	1045
6	Bare soil	5029	300	4729
7	Bitumen	1330	300	1030
8	Self-blocking bricks	3682	300	3382
9	Shadows	947	300	647

Figure 7 illustrates the Pavia University image and its related ground truth maps.

Figure 7. The ground truth information of the Pavia University dataset: (a) original dataset; (b) ground truth map; (c) training map; (d) testing map.

The third test dataset was the Salinas scene. This dataset captured the area over Salinas Valley in California by AVIRIS at 3.7-m spatial resolution. This image comprises 220 bands of size 512×217 pixels where the bands affected by noise and water absorption were discarded, leaving 204 bands. There are 16 classes representing a variety of crops in the ground reference data. Table 3 lists these classes with the related number of samples, training, and testing points in this dataset. Figure 8 illustrates the Salinas image and its ground truth categorization maps.

Table 3. Classes of the Salinas scene with the related number of test and training samples.

Class	Type	Samples	Training	Testing
1	Brocoligreenweeds1	2009	100	1909
2	Brocoligreenweeds2	3726	100	3626
3	Fallow	1976	100	1876
4	Fallowroughplow	1394	100	1294
5	Fallowsmooth	2678	100	2578
6	Stubble	3959	100	3859
7	Celery	3579	100	3479
8	Grapesuntrained	11,271	100	11,171
9	Soilvinyarddevelop	6203	100	6103
10	Cornsenescedgreenweeds	3278	100	3178
11	Lettuceromaine4wk	1068	100	968
12	Lettuceromaine5wk	1927	100	1827
13	Lettuceromaine6wk	916	100	816
14	Lettuceromaine7wk	1070	100	970
15	Vinyarduntrained	7268	100	7168
16	Vinyardverticaltrellis	1807	100	1707

Figure 8. The ground truth information map of Salinas dataset: (a) original dataset; (b) ground truth map; (c) training map; (d) test map.

4.2. Evaluation Process

Several state-of-the-art methods were considered for comparison to validate the effectiveness of our proposed spectral DWT multilevel decomposition with 2D-DWT filter (SDWT-2DWT_SVM), spectral DWT multilevel decomposition with Wiener filter (SDWT-WF_SVM), and spectral DWT multilevel decomposition with 2D-DCT filter (SDWT-2DCT_SVM) frameworks based on the spectral DWT multilevel decomposition. These included two pixel-wise classifiers represented by the standard SVM [6], and Principal Component Analysis followed by SVM (PCA_SVM) [47]. Four closely related spectral-spatial classification methods based on the structural filtering methodology were used for comparison, including 3D-DWT (3D_SVM) [41], 3D-DWT with graph cut (3DG_SVM) [41], edge-preserving filtering (EPF) [48], and the image fusion and recursive filtering (IFRF) [49]. Moreover, two well-known denoising methods, block-matching 4D filtering (BM4D_SVM) [50] and parallel factor analysis-based approach (PARAFAC_SVM) [51], were included for comparison.

The parameter settings opted for in the compared techniques were the default settings used in the corresponding research works; the related source codes were provided by the respective authors. Specifically, for a fair comparison, linear SVM is adopted as a pixel-wise classifier for all the compared methods. The SVM classifier was implemented using MATLAB libsvm library [52]. Regarding the PCA_SVM method, the number of the principal components (PCs) fed to the linear SVM classifier was calculated using HySime criterion for the intrinsic dimension estimation for each dataset [53].

The compared methods were assessed numerically using five criteria: Average Accuracy (AA), Overall Accuracy (OA), statistical kappa coefficient (κ) [54], the accuracy of each class, and the computational time. Specifically, AA is the average of class classification accuracies, OA is calculated by dividing the number of correctly classified samples by the total number of test samples, and kappa coefficient (κ) considers errors of both omission and commission. In order to calculate the three criteria, the classification confusion matrix should be obtained first, which is defined as follows:

$$M = \begin{pmatrix} m_{11} & \dots & m_{1k} \\ \vdots & \ddots & \vdots \\ m_{k1} & \dots & m_{kk} \end{pmatrix}, \quad (3)$$

where m_{ij} shows the number of pixels that should belong to class i but are assigned to class j , and K is the number of classification classes. Then, according to the confusion matrix M , OA, AA, and kappa coefficient (κ) are calculated as follows:

$$OA = \frac{1}{N_{test}} \sum_{i=1}^K m_{ii}, \quad (4)$$

where N_{test} is the total number of test samples. Then, the AA is calculated as:

$$CA_i = \frac{m_{ii}}{N_i}, i = 1, 2, \dots, K, \quad (5)$$

$$AA = \frac{1}{K} \sum_{i=1}^K CA_i, \quad (6)$$

where CA_i is the accuracy of class i and N_i is the total test samples in class i . Finally, the kappa coefficient (κ) is calculated by

$$\kappa = \frac{OA - P_e}{1 - P_e} \quad (7)$$

where $P_e = \sum_{i=1}^K P_{i.} * P_{.i}$ is the expected agreement, $P_{i.} = \frac{R_i}{N_{test}}$, $P_{.i} = \frac{C_i}{N_{test}}$.

R_i and C_i represent the sums of i th row and i th column in the confusion matrix, respectively. For all the three criteria, a larger value indicates a better classification performance. The experiments were repeated 20 times, and the average classification accuracies (kappa, OA, AA) were retained for comparison.

In order to verify the classification robustness of our proposed frameworks in the case of a small sample size problem, we evaluated the compared methods with insufficient and sufficient training samples; 2%, 4%, 6%, 8%, and 10% training samples were randomly selected from each class to form the training set, and the rest were used as test samples. The experiments were repeated 20 times, and the average results were reported to evaluate our approaches.

To compare the computational time of the tested approaches, we calculated the time consumed during the whole classification process for an average of 20 times. The experiments were performed on a computer with an Intel(R) Xeon(R) with CPU W3680 of 3.33 GHz, and 6 GB memory; all the methods were implemented with MATLAB®.

The parameter settings of the proposed approaches will be described in the next section, followed by the experimental results and analysis.

5. Experimental Results and Analysis

In this section, the parameter estimation for our proposed approaches are described, and the experimental results are analyzed to evaluate the performance of our proposed approaches relative to the other closely related hyperspectral classification methods.

5.1. Parameter Settings

Two types of parameters related to the spectral and spatial filters must be independently fixed for each of our proposed approaches:

■ SDWT-2DWT_SVM framework:

- For both the spectral and the spatial filters, we need to fix the wavelet mother function type and its level of decomposition.
- Default global threshold is used for the spatial filter [46].

■ SDWT-WF_SVM framework:

- For the spectral filter, we need to fix the wavelet mother function type and its level of decomposition.
- For the spatial filter, we have to fix the Wiener filter patch size.

■ SDWT-2DCT_SVM framework:

- For the spectral filter, we need to fix the wavelet mother function type and its level of decomposition

- For the spatial filter, we need to fix the threshold of the 2D-DCT filter.

Regarding the Wavelet mother function type, we evaluated for each proposed approach the most common wavelet families, including 30 wavelet functions, as mentioned in Table 4.

Table 4. The wavelet families and function types investigated for parameter tuning.

Wavelet Family	Mother Function
Daubechies	haar, db2, ... , db9
Biorthogonal	bior1.1, bior1.3, bior1.5, bior2.2, bior2.4, bior2.6, bior3.1, bior3.3, bior3.5, bior3.7, bior4.4, bior5.5
Coiflets	coif1, ... , coif5
Symlets	sym2, ... , sym6

The estimation of the required parameters for each proposed approach is explained in the following three subsections independently.

5.1.1. The SDWT-2DWT_SVM Framework Parameter Estimation

The spectral filter parameters in this approach are the wavelet mother function type and its level of decomposition. These two parameters were set by comparing the classification accuracy OA with different types of wavelets for the first level of decomposition for each dataset as listed in Table 5. Then two functions with the highest accuracies were selected and further decomposed to fix the best function with its level of decomposition as listed in Table 6. The level of decomposition depends on the wavelet function type and the dimension of the selected dataset. The type of wavelet mother function and level of decomposition correspond to the most accurate classification OA for each dataset.

Table 5. Classification accuracies with respect to the spectral DWT mother function on all datasets for the SDWT-2DWT_SVM approach.

Wavelet	Indian Pines	Pavia University	Salinas	Wavelet	Indian Pines	Pavia University	Salinas
haar	82.71	98.16	94.52	bior3.3	81.66	97.82	94.45
db2	83.48	97.86	94.10	bior3.5	82.57	97.75	94.41
db3	83.47	97.65	94.96	bior3.7	82.79	97.95	94.77
db4	83.01	98.03	94.42	bior4.4	83.56	97.70	94.44
db5	84.24	97.63	94.69	bior5.5	83.22	98.03	93.94
db6	83.67	97.86	94.17	coif1	82.75	97.94	94.68
db7	84.84	97.82	93.97	coif2	82.53	98.13	94.17
db8	83.40	97.93	93.19	coif3	83.11	98.06	95.01
db9	84.82	97.74	94.02	coif4	82.78	97.85	93.88
bior1.3	83.55	98.12	94.22	coif5	83.21	98.04	94.45
bior1.5	82.87	98.09	94.12	sym2	83.36	97.71	94.52
bior2.2	82.29	97.86	94.49	sym3	83.78	97.68	94.91
bior2.4	83.03	97.81	94.63	sym4	82.83	97.79	94.90
bior2.6	83.32	97.96	94.06	sym5	83.41	97.98	94.27
bior3.1	79.20	97.81	94.37	sym6	83.54	97.94	94.14

As shown in Table 5, for Indian Pines dataset, 'db7' and 'db9' functions have the highest classification accuracies (84.84% OA and 84.82% OA, respectively). For Pavia University dataset, 'coif2' and 'bior1.3' functions have the highest classification accuracies (98.13% OA and 98.12% OA, respectively). For Salinas dataset, the highest accuracies correspond to 'coif3' with an OA of 95.01% and 'sym4' with an OA of 94.90%. For each dataset, the aforementioned functions corresponding to the first two highest accuracies are further decomposed to fix only one function with its best level of decomposition. Table 6 lists the obtained classification results using the best two functions for each dataset in the proposed SDWT-2DWT approach.

Table 6. Classification accuracies with respect to the spectral DWT level of decomposition on all datasets for the SDWT-2DWT_SVM approach.

Decomposition Level	Indian Pines		Pavia University		Salinas	
	Db7	Db9	Coif2	Bior1.3	Coif3	Sym4
1	80.34	79.83	91.98	95.26	91.69	91.13
2	82.37	82.99	96.66	96.98	93.04	93.08
3	84.12	84.70	98.11	98.11	94.47	94.24
4				98.21		94.53

As depicted from Table 6, the highest classification accuracy (84.70% OA) was obtained using ‘db9’ function at the third level of decomposition for Indian Pines dataset. For Pavia University dataset, the highest accuracy (98.21% OA) was obtained using ‘bior1.3’ function at the fourth level of decomposition. For Salinas dataset, the highest accuracy (94.53% OA) was obtained using sym4 function at its fourth level of decomposition. Thus, for the SDWT-2DWT approach, we opted the following spectral parameters (Dataset, SDWT, Wavelet function, level of decomposition):

- (Indian Pines dataset, SDWT, ‘db9’, 3);
- (Pavia University dataset, SDWT, ‘bior1.3’, 4);
- (Salinas dataset, SDWT, ‘sym4’, 4).

The spatial filter parameters in the proposed SDWT-2DWT approach are the wavelet mother function type and its level of decomposition. These two parameters were set using the same logic in the last subsection by comparing the classification accuracy OA with different types of wavelets for the second level of decomposition for each dataset. Then, we use the first two functions having the highest classification accuracies with more levels of decomposition to fix only one function with the best level of decomposition having the highest obtained classification accuracy. Table 7 lists the classification accuracy OA with different types of wavelets for the first level of decomposition on the Indian Pines, Pavia University, and Salinas datasets.

Table 7. Classification accuracies with respect to the 2D-DWT mother function on all datasets for the SDWT-2DWT_SVM.

Wavelet	Indian Pines	Pavia University	Salinas	Wavelet	Indian Pines	Pavia University	Salinas
haar	81.67	95.43	90.70	bior3.3	78.08	93.68	90.12
db2	81.01	95.17	90.57	bior3.5	79.04	94.42	89.54
db3	79.93	95.39	90.14	bior3.7	78.83	94.34	89.92
db4	80.58	94.96	90.15	bior4.4	80.08	95.15	90.58
db5	80.06	94.79	90.53	bior5.5	80.20	95.17	89.87
db6	80.17	95.28	90.20	coif1	80.75	95.51	90.66
db7	80.31	95.10	90.19	coif2	81.38	95.34	89.92
db8	80.19	95.15	90.24	coif3	80.97	95.25	90.19
db9	80.27	94.97	90.07	coif4	80.52	95.13	90.16
bior1.3	79.61	94.89	90.18	coif5	80.92	95.25	90.15
bior1.5	80.11	94.28	89.70	sym2	80.90	95.20	90.61
bior2.2	81.82	95.21	89.21	sym3	80.35	95.32	90.18
bior2.4	79.70	95.06	90.44	sym4	80.98	95.33	89.99
bior2.6	80.09	95.06	90.12	sym5	81.19	95.23	90.18
bior3.1	76.79	89.53	88.35	sym6	80.46	94.98	90.08

As can be seen in Table 7, for the Indian Pines dataset, ‘haar’ and ‘bior2.2’ functions have the highest classification accuracies (81.67% OA and 81.82% OA, respectively). For Pavia University dataset, ‘haar’ and ‘coif1’ functions have the highest classification accuracies (95.43% OA and 95.51%OA, respectively). For the Salinas dataset, the highest accuracies correspond to ‘haar’ with an OA of 90.70% and ‘coif1’ with an OA of 90.66%. For each dataset, the aforementioned functions corresponding to the first two highest accuracies are further decomposed to fix only one function to its best level of decomposition.

Table 8 lists the obtained classification results using the best two functions for the 2D-DWT for each dataset in the proposed SDWT-2DWT_SVM approach.

Table 8. Classification accuracies with respect to 2D-DWT level of decomposition on all datasets for the SDWT-2DWT_SVM approach.

Decomposition Level	Indian Pines		Pavia University		Salinas	
	Haar	Bior2.2	Haar	Coif1	Haar	Coif1
1	77.27	76.88	90.39	90.63	88.86	89.08
2	81.67	81.82	95.43	95.51	90.70	90.66
3	84.91	82.95	97.46	97.34	92.99	92.57
4	85.68	85.33	98.31	98.03	93.91	94.35
5	85.66		98.27	98.15	95.41	94.85
6	85.54		98.29	98.28	95.24	
7	85.26		98.33		95.90	
8			98.44			

As shown in Table 8, for the three datasets, the highest classification accuracy was obtained using ‘haar’ function. Specifically, the highest accuracy was obtained at the fifth level of decomposition for the Indian Pines dataset (85.66% OA), at the eighth level of decomposition for the Pavia University dataset, and at the seventh level of decomposition for the Salinas dataset (95.90% OA). Thus, for the SDWT-2DWT approach, we adopted the following spatial parameters (dataset, 2D-DWT function, and level of decomposition):

- (Indian Pines dataset, 2D-DWT, ‘haar’, 5);
- (Pavia University dataset, 2D-DWT, ‘haar’, 8);
- (Salinas dataset, 2D-DWT, ‘haar’, 7).

For the spatial filtering threshold, we used the default global threshold (universal threshold) developed by Donoho and Johnstone, as it provides easy, fast, and automatic thresholding [46].

5.1.2. The SDWT-WF_SVM Framework Parameter Estimation

The spectral filter parameters in this approach are the wavelet mother function type and its level of decomposition. These two parameters were set by comparing the classification accuracy OA with different types of wavelets for the first level of decomposition for each dataset as listed in Table 9 for Indian Pines, Pavia University, and Salinas datasets. For each dataset, two functions with the highest accuracies were selected and further decomposed to fix the best function with its level of decomposition, as listed in Table 10.

As can be seen in Table 9, for the Indian Pines dataset, ‘bior2.4’ and ‘sym3’ functions have the highest classification accuracies (90.70% OA and 90.52% OA, respectively). For the Pavia University dataset, ‘coif2’, and ‘sym4’ functions have the highest classification accuracies (95.33% OA and 94.91% OA, respectively). For the Salinas dataset, the highest accuracies correspond to ‘db4’ with an OA of 96.78% and ‘coif2’ with an OA of 96.65%. For each dataset, the aforementioned functions corresponding to the first two highest accuracies are further decomposed to fix only one function with its best level of decomposition. Table 10 lists the obtained classification results using the best two functions for each dataset in the proposed SDWT-WF_SVM approach.

Table 9. Classification accuracies with respect to the DWT mother function on all datasets for the SDWT-WF_SVM approach.

Wavelet	Indian Pines	Pavia University	Salinas	Wavelet	Indian Pines	Pavia University	Salinas
haar	90.02	-	-	bior3.3	89.04	93.61	95.76
db2	89.03	94.50	96.57	bior3.5	88.74	93.91	95.86
db3	90.05	94.48	95.83	bior3.7	89.97	93.82	96.04
db4	89.44	94.66	96.78	bior4.4	90.38	94.27	96.06
db5	90.24	94.75	96.62	bior5.5	90.23	94.45	96.44
db6	89.53	94.64	95.69	coif1	90.05	94.52	96.58
db7	90.08	94.26	96.26	coif2	89.65	95.33	96.65
db8	90.04	93.89	95.97	coif3	89.59	94.18	96.15
db9	90.22	94.19	95.81	coif4	89.41	93.92	96.43
bior1.3	-	-	-	coif5	89.80	94.25	95.56
bior1.5	90.41	-	-	sym2	89.28	94.48	96.40
bior2.2	89.55	94.15	96.30	sym3	90.52	94.40	96.17
bior2.4	90.70	94.69	95.66	sym4	90.47	94.91	95.97
bior2.6	89.38	94.60	95.96	sym5	89.47	94.38	96.49
bior3.1	87.61	93.43	95.09	sym6	90.33	94.59	96.29

Table 10. Classification accuracies with respect to the DWT level of decomposition on all datasets for the SDWT-WF_SVM approach.

Decomposition Level	Indian Pines		Pavia University		Salinas	
	BIOR2.4	Sym3	Coif2	Sym4	Db4	Coif2
1	84.72	84.87	90.31	91.04	93.09	93.33
2	87.38	88.65	95.72	95.61	94.85	94.48
3	90.26	90.20	97.34	97.03	96.78	96.45
4	91.01	90.68			97.53	97.18
5		92.37				

As depicted from Table 10, the highest classification accuracy (92.37% OA) was obtained using ‘sym3’ function at the fifth level of decomposition for the Indian Pines dataset. For the Pavia University dataset, the highest accuracy (97.34% OA) was obtained using ‘coif2’ function at the third level of decomposition. For the Salinas dataset, the highest accuracy (97.53% OA) was obtained using ‘db4’ function at its fourth level of decomposition. Thus, for the SDWT-WF_SVM approach, we adopted the following spectral parameters (dataset, SDWT, wavelet function, level of decomposition):

- (Indian Pines dataset, SDWT, ‘sym3’, 5);
- (Pavia University dataset, SDWT, ‘coif2’, 3);
- (Salinas dataset, SDWT, ‘db4’, 4).

Wiener filter patch size represents the spatial parameter in the SDWT-WF_SVM approach. This parameter was set by comparing the classification accuracy OA with varying Wiener filter patch sizes where the opted patch size corresponds to the best classification OA for each dataset. Figure 9 depicts the variation of the classification accuracy OA with varying filter patch sizes, ranging from 3×3 to 63×63 pixels, on the Indian Pines, Pavia University, and Salinas datasets.

As can be seen in Figure 9, classification accuracy (OA) increased with a larger patch size value for the three datasets. Thus, accuracies tend to become stable at around a 41×41 patch size for the Indian Pines dataset, at around 27×27 patch size for the Pavia University dataset, and at around a 47×47 patch size for the Salinas dataset. Therefore, these patch sizes corresponding to each dataset were selected for our experiments.

Figure 9. Classification accuracy (overall accuracy (OA)%) with varying Wiener filter patch size in SDWT-WF_SVM on the Indian Pines, Pavia University, and Salinas datasets.

5.1.3. The SDWT-2DCT_SVM Framework Parameter Estimation

The spectral filter parameters in the proposed SDWT-2DCT_SVM approach are the wavelet mother function type and its level of decomposition. These two parameters were set by comparing the classification accuracy OA with different types of wavelets for the first level of decomposition for each dataset as listed in Table 11 for the Indian Pines, Pavia University, and Salinas datasets. Then, for each dataset, two functions with the highest accuracies were selected and further decomposed to fix the best function with the level of decomposition as listed in Table 12.

As can be seen in Table 11, for the Indian Pines dataset, 'db7' and 'db9' functions have the highest classification accuracies (85.64% OA and 85.47% OA, respectively). For Pavia University dataset, 'bior1.3' and 'coif2' functions have the highest classification accuracies (98.36% OA and 98.13% OA, respectively). For the Salinas dataset, the highest accuracies correspond to 'bior3.7' with an OA of 98.70% and 'bior4.4' with an OA of 98.67%. For each dataset, the aforementioned functions corresponding to the first two highest accuracies are further decomposed to fix only one function with its best level of decomposition. Table 12 lists the obtained classification results using the best two functions for each dataset in the proposed SDWT-2DCT_SVM approach.

As shown in Table 12, the highest classification accuracy (85.86% OA) was obtained using 'db7' function at the second level of decomposition for the Indian Pines dataset. For the Pavia University dataset, the highest accuracy (98.51% OA) was obtained using 'bior1.3' function at the fourth level of decomposition. For the Salinas dataset, the highest accuracy (98.65% OA) was obtained using 'bior3.7' function at its second level of decomposition. Thus, for the SDWT-2DCT_SVM approach, we opted the following spectral parameters (dataset, SDWT, wavelet function, level of decomposition):

- (Indian Pines dataset, SDWT, 'db7', 2);
- (Pavia University dataset, SDWT, 'bior1.3', 4);
- (Salinas dataset, SDWT, 'bior3.7', 2).

The spatial filter parameter in the SDWT-2DCT_SVM approach was the threshold for 2D-DCT. This parameter was set by comparing the classification accuracy OA with varying threshold values where the opted threshold corresponds to the best classification OA for each dataset. Figure 10 depicts the variation of the classification accuracy OA with varying threshold values, ranging from 50 to 1000 with the step of 50, on the Indian Pines, Pavia University, and Salinas datasets.

Table 11. Classification accuracies with respect to the DWT mother function on all datasets for the SDWT-2DCT approach.

Wavelet	Indian Pines	Pavia University	Salinas	Wavelet	Indian Pines	Pavia University	Salinas
haar	84.33	98.28	98.28	bior3.3	82.38	95.79	98.48
db2	84.46	98.19	98.44	bior3.5	81.33	96.94	98.67
db3	84.57	98.04	98.31	bior3.7	83.35	95.34	98.70
db4	83.05	98.35	98.20	bior4.4	84.63	97.60	98.67
db5	82.58	96.99	98.41	bior5.5	84.53	97.60	98.64
db6	85.37	98.12	98.56	coif1	83.95	98.13	98.46
db7	85.64	97.70	98.02	coif2	84.83	98.36	98.50
db8	84.40	97.87	98.38	coif3	84.41	97.50	98.10
db9	85.47	97.46	98.41	coif4	84.59	97.96	98.16
bior1.3	84.96	98.62	97.90	coif5	84.35	97.53	98.31
bior1.5	83.68	98.29	98.27	sym2	84.51	98.06	98.36
bior2.2	84.20	98.01	98.54	sym3	84.21	97.90	98.65
bior2.4	84.45	97.00	98.49	sym4	83.78	97.62	98.47
bior2.6	83.94	97.70	98.58	sym5	84.32	97.75	98.12
bior3.1	78.95	96.37	98.58	sym6	84.75	97.56	98.29

Table 12. Classification accuracies with respect to the DWT level of decomposition on all datasets for the spectral DWT multilevel decomposition with 2D-DCT filter (SDWT-2DCT_SVM) approach.

Decomposition Level	Indian Pines		Pavia University		Salinas	
	Db7	Db8	Bior1.5	Coif2	Bior3.7	Bior4.4
1	83.16	81.97	96.17	87.53	98.26	98.15
2	85.86	85.42	97.86	96.37	98.65	98.52
3	85.42	84.83	98.49	98.40	98.65	98.64
4			98.51			98.59

Figure 10. Classification accuracy (OA %) with varying 2D-DCT thresholds in SDWT-2DCT_SVM on the Indian Pines, Pavia University, and Salinas datasets.

As can be seen in Figure 10, the OA increased with a higher threshold value for all datasets and tended to be stable around the threshold of 1400 for Indian Pines, while this degree of stability was obtained at around a threshold of 1050 for the Pavia University dataset and around 850 for the Salinas dataset. Therefore, these threshold values corresponding to each dataset were selected for our experiments.

5.2. Classification of Indian Pines Dataset

In this subsection, we compare the classification effectiveness of the proposed frameworks SDWT-2DWT_SVM, SDWT-WF_SVM, and SDWT-2DCT_SVM with the other classification methods on the Indian Pines dataset. For this dataset, the parameter tunings for the three proposed approaches are recapitulated in Table 13. The number of PCs obtained using HySime criterion was 18 and therefore was used in PCA_SVM method.

Table 13. The parameter tuning for the proposed approach in the Indian Pines dataset.

Approach	SDWT (Function, Level)	2D-DWT (Function, Level)	Wiener Filter Patch Size	2D-DCT Threshold
SDWT-2DWT_SVM	('db9', 3)	('haar', 5)	/	/
SDWT-WF_SVM	('sym4', 4)	/	41	/
SDWT-2DCT_SVM	('db7', 2)	/	/	1400

The first experiment assessed the validity of our proposed approaches relative to the other methods. We randomly chose 100 samples for each class from the reference data as training samples, and the rest were used as test samples as given in Table 1. The results reported in Table 14 represent the averages of the individual classification accuracies (%), κ statistic, standard deviation, AA, OA, and computational time in seconds. Classification maps for all the approaches are shown in Figure 11. From Table 14 and Figure 11, several observations can be made.

The approaches incorporating only the spectral information, including SVM (70.40% OA) and PCA_SVM (64.40% OA) attain lower classification results when compared with the other spectral-spatial techniques. This can be confirmed in Figure 11, where it is clear that the classification maps resulting from these pixel-wise techniques are degraded by salt-and-pepper noise.

The methods considering both the spectral and spatial information, including 3D_SVM, 3DG_SVM, EPF, IFRF, BM4D_SVM, PARAFAC_SVM, and our three proposed approaches are most accurate. Specifically, the two proposed approaches combining two different filters and taking advantage of both of them, including SDWT-2DCT_SVM and SDWT-WF_SVM, achieved high classification results. The proposed SDWT-2DCT_SVM approach achieved the highest values for the three classification criteria (91.47% κ , 92.59% OA, and 95.46% AA), followed by the proposed SDWT-WF_SVM with the second-best performance (90.76% κ , 91.97% OA, and 94.89% AA). Furthermore, the proposed SDWT-2DCT_SVM and SDWT-WF_SVM approaches achieve 2.27%–28.19% and 1.65%–27.57% advantages in OA, respectively, over the other state-of-the-art methods. The classification maps produced from both the SDWT-2DCT_SVM and SDWT-WF_SVM approaches are noiseless and the closest to the ground truth map, as shown in Figure 11. These results demonstrate that the classification performance was improved by exploring the DWT multilevel decomposition in the spectral domain when spatially filtering the details coefficients using another different filter.

The proposed SDWT-2DWT_SVM approach using DWT in both the spectral and spatial filtering steps achieved worse classification results than the 3DG_SVM method and better results than the 3D_SVM (85.46% OA, 90.32% OA, and 85.04% OA, respectively). While the edge-preserving filtering (EPF)-based approach achieved the fourth-best classification result with an OA of 86.61%, the second EPF-based approach, which is the IFRF with an OA of 74.30%, achieved worse classification results than the two denoising-based approaches, namely the including BM4D_SVM and PARAFAC_SVM approaches (83.08% OA and 81.62% OA, respectively).

For the computational cost, Table 14 shows that the shortest classification time was 1.86 s achieved with the IFRF, which is a feature extraction-based method. Nevertheless, it had low classification accuracy (74.30% OA). Denoising based methods are time-consuming with low classification accuracies where CDCT-BM4D_SVM achieves an OA of 83.08% in 351.37 s, and the PARAFAC_SVM achieves an OA of 81.62% in 297.39 s.

Table 14. Classification results (%) for the Indian Pines dataset with standard deviation (in parentheses).

Class	SVM	PCA_SVM	3D_SVM	3DG_SVM	EPF	IFRF	BM4D_SVM	PARAFAC_SVM	SDWT-2DWT_SVM	SDWT-WF_SVM	SDWT-2DCT_SVM
1	88.70	78.26	98.70	98.70	100	71.70	91.74	93.48	94.35	97.83	95.22
2	72.71	56.44	76.42	87.31	85.71	74.25	83.25	77.64	82.48	90.56	92.25
3	70.11	59.40	81.59	90.92	90.79	49.74	80.78	73.90	87.97	89.52	93.81
4	80.95	75.26	98.32	99.64	65.07	63.15	91.68	96.50	90.95	97.23	95.91
5	92.69	88.98	96.89	97.10	96.01	80.51	95.48	92.87	94.33	96.16	96.45
6	95.22	91.70	97.22	99.56	99.46	89.60	96.60	97.08	98.02	98.87	99.17
7	92.14	78.57	95.00	97.14	100	00	87.86	87.86	94.29	92.14	93.57
8	96.75	95.79	100	100	100	100	98.99	99.68	98.65	99.95	99.39
9	78.00	66.00	100	94.00	100	00	89.00	94.00	95.00	97.00	97.00
10	72.99	55.09	80.91	89.91	71.34	61.48	82.26	78.41	86.40	91.89	91.09
11	56.24	46.82	75.23	83.72	91.69	94.19	68.74	70.28	72.27	85.82	84.52
12	72.72	51.89	91.03	93.41	66.39	42.11	83.87	80.34	83.59	87.77	94.71
13	98.76	96.86	99.14	99.71	100	81.76	99.14	99.33	99.14	99.52	99.33
14	85.95	86.47	95.72	96.21	99.34	97.89	92.46	91.79	95.59	98.52	99.06
15	71.47	60.24	95.10	99.93	78.47	79.09	88.43	92.69	92.48	98.08	98.08
16	97.39	97.61	98.91	94.83	93.67	97.50	96.74	99.35	98.48	97.39	97.83
κ	70.40 (0.79)	59.63 (1.24)	82.84 (0.78)	88.87 (1.18)	84.68 (1.96)	71.06 (1.66)	80.66 (0.64)	78.95 (1.03)	83.36 (0.69)	90.76 0.73	91.47 (0.74)
OA	73.97 (0.71)	64.40 (1.13)	85.04 (0.70)	90.32 (1.04)	86.61 (1.72)	74.30 (1.46)	83.08 (0.56)	81.62 (0.92)	85.46 (0.62)	91.97 0.64	92.59 (0.65)
AA	82.67 (0.97)	74.09 (1.41)	92.51 (0.72)	94.41 (0.70)	89.87 (1.60)	67.69 (1.80)	89.19 (0.87)	89.08 (1.00)	91.50 (0.69)	94.89 0.93	95.46 (0.55)
Time (s)	3.10	42.72	53.41	210.16	6.80	1.86	351.37	297.39	40.04	47.74	32.03

The bold values indicate critical values.

Figure 11. Classification maps of different methods on the Indian Pines dataset.

The proposed SDWT_2DCT_SVM and SDWT_WF_SVM approaches achieved the two highest classification accuracies within a short execution time (less than one minute). For example, SDWT-2DCT_SVM achieved an OA of 92.59% in 32.03 s, while SDWT-WF_SVM achieved an OA of 91.97% in 47.74 s. The third highest classification accuracy, achieved by the 3DG_SVM (90.32% OA) method within 210.16 s, represents the longest execution time among all the compared methods.

In the second experiment, the performance of the proposed approaches was evaluated against state-of-the-art approaches under different training conditions, in which 2%, 4%, 6%, 8%, and 10% labeled samples of each class are randomly selected as training samples, while the rest are used as test samples. Each experiment was performed 20 times, and the average results in terms of the OA values for each method are plotted in Figure 12.

Figure 12 illustrates that the two proposed approaches SDWT-2DCT_SVM and SDWT-WF_SVM can significantly improve classification results when varying the training sample size, as the size of training samples changes from 2% to 10%, these two proposed approaches always achieved the first and the second highest OA. Meanwhile, with 2% and 4% as a training sample, the proposed CDCT-WF_SVM approach outperforms the SDWT-2DCT_SVM approach that takes back the best performance when using more than 6% training samples per class. In this scenario, our proposed SDWT-2DWT_SVM outperforms the 3DG_SVM method when using only 2% training sample per class. Meanwhile, using 4% and above, the 3DG_SVM method is more accurate than the proposed SDWT-2DWT_SVM approach.

Figure 12. Classification accuracy (OA, %) achieved by the compared methods using different training sample sizes on the Indian Pine dataset.

5.3. Classification of Pavia University Dataset

Regarding the Pavia University dataset experiments, the parameter tunings of the three proposed approaches are recapitulated in Table 15. In the PCA_SVM method, the number of PCs obtained using HySime criterion on this dataset was 45.

Table 15. The parameter tuning for the proposed approaches in the Pavia University dataset.

Approach	SDWT (Function, Level)	2D-DWT (Function, Level)	Wiener Filter Patch Size	2D-DCT Threshold
SDWT-2DWT_SVM	('bior1.3', 4)	('haar', 8)	/	/
SDWT-WF_SVM	('coif2', 3)	/	27	/
SDWT-2DCT_SVM	('bior1.3', 4)	/	/	1050

In the first experiment, the effectiveness of our proposed frameworks against other methods was evaluated. We randomly chose 300 samples for each class from the reference data as training samples, and the rest were used as test samples, as summarized in Table 2. The quantitative classification results, including the individual classification accuracies (%), κ statistic, standard deviation, OA, AA, and computational time in seconds are reported in Table 16. The classification maps of the compared approaches on the Pavia University dataset are shown in Figure 13.

As can be seen in Table 16, the highest classification accuracies are always obtained by the spectral–spatial classification techniques, especially the structural filtering based approaches. The proposed method SDWT-2DCT_SVM achieved the highest values for the three classification criteria (98.51% κ , 98.90% OA, and 98.94% AA), followed by the proposed SDWT-2DWT_SVM (97.95% κ , 98.49% OA, and 98.37% AA). The proposed SDWT-2DCT_SVM and SDWT-2DWT_SVM approaches achieve 0.83%–18.41% and 0.42%–18% advantages in OA, respectively, over the other state-of-the-art methods. However, for this dataset, 3DG_SVM method (98.07% OA) achieved better classification results than the proposed SDWT-2DCT_SVM (97.29% OA) and the 3D_SVM method (96.95% OA). The BM4D_SVM, a denoising-based method, achieved classification results comparable with the EPF method (87.23% OA and 86.81% OA, respectively). Meanwhile, PARAFAC_SVM and IFRF were the least accurate among the spectral–spatial classifiers with OAs of 83.99% and 83.14%, respectively.

As can be visually seen from Figure 13, the classification maps produced from methods combining spectral and spatial information are smooth. In particular, wavelet-based approaches have smoother classification maps than the other approaches. The classification maps obtained from our proposed SDWT-2DCT_SVM and SDWT-2DWT_SVM approaches closely match the ground-truth map.

Although wavelet-based approaches have the highest classification accuracies, the processing time of our proposed frameworks is faster than 3DG_SVM and 3D_SVM methods. For example, the proposed SDWT-2DCT_SVM achieve an OA of 98.90% in 426.28 s, while 3DG_SVM achieves an OA of 98.07% in 2142.40 s, which is five times slower than our proposed SDWT-2DCT_SVM approach. The shortest processing time was achieved by the IFRF method within 18.14 s; however, this method delivered the lowest classification accuracy among the spectral–spatial approaches with an OA of 83.14%. In addition, denoising-based approaches are time-consuming with low classification accuracies. For example, the BM4D_SVM achieved an OA of 87.23% in 1535.13 s and the PARAFAC_SVM achieved an OA of 83.99% in 1275.06 s.

In the second experiment, we evaluated the performance of our proposed approaches on the Pavia University dataset against state-of-the-art methods with different training conditions, in which 2%, 4%, 6%, 8%, and 10% labeled samples of each class were randomly selected as training samples, and the rest were used for testing. The average results in terms of OA values are plotted in Figure 14 for each compared method.

Figure 14 illustrates that all wavelet-based approaches outperform the other compared methods when varying the training sample size. Specifically, the proposed SDWT-2DCT_SVM and SDWT-2DWT_SVM approaches always significantly improved the classification accuracy significantly with a different number of training samples, and as the size of training samples changed from 2% to 10%, these two proposed approaches always achieved the two highest OA values, followed by 3DG_SVM and the proposed SDWT-WF_SVM methods. Additionally, BM4D_SVM provided higher classification accuracy than the IFRF and EPF methods even when varying the training sample size.

Table 16. Classification results (%) for the Pavia University dataset with standard deviation (in parentheses).

Class	SVM	PCA_SVM	3D_SVM	3DG_SVM	EPF	IFRF	BM4D_SVM	PARAFAC_SVM	SDWT_2DWT_SVM	SDWT_WF_SVM	SDWT_2DCT_SVM
1	71.71	72.13	96.89	98.88	98.68	76.18	80.07	77.90	96.42	96.00	97.64
2	82.91	82.79	97.64	98.91	97.20	97.90	89.49	85.37	98.97	98.66	99.43
3	78.89	78.23	88.15	90.08	92.51	52.39	82.61	70.81	98.28	95.29	98.69
4	91.89	91.00	99.01	99.42	67.81	84.99	94.15	93.85	97.67	97.25	97.91
5	99.70	99.70	100	100	99.89	98.85	99.66	99.78	99.77	99.67	99.88
6	84.39	77.94	96.68	99.16	64.31	92.74	89.76	84.83	99.10	98.44	99.29
7	76.96	77.57	98.91	99.53	77.25	64.22	85.59	77.89	99.15	99.41	99.65
8	69.44	71.07	94.64	96.08	84.40	53.24	75.88	74.30	96.07	89.89	98.04
9	99.91	99.86	99.63	99.72	96.17	43.63	99.88	99.92	99.95	99.94	99.94
κ	75.24 (1.43)	74.25 (2.16)	95.88 (0.30)	97.82 (0.12)	82.79 (1.03)	77.63 (0.25)	83.00 (1.40)	80.91 (0.47)	97.95 (0.31)	96.33 (0.54)	98.51 (0.30)
OA	81.18 (1.15)	80.49 (1.75)	96.95 (0.22)	98.07 (0.09)	86.81 (1.68)	83.14 (0.19)	87.23 (1.10)	83.99 (0.40)	98.49 (0.23)	97.29 (0.40)	98.90 (0.22)
AA	83.98 (0.69)	83.37 (1.68)	96.84 (0.14)	97.98 (0.12)	86.47 (1.01)	73.79 (0.17)	88.57 (0.79)	84.96 (0.40)	98.37 (0.21)	97.17 (0.36)	98.94 (0.14)
Time (s)	80.05	78.68	485.9	2142.40	52.84	18.14	1535.13	1275.06	432.03	349.64	426.28

The bold values indicate critical values.

Figure 13. Classification maps of different methods on the Pavia University dataset.

Figure 14. Classification accuracy (OA, %) achieved by the compared methods using different training sample sizes on the Pavia University dataset.

5.4. Classification of Salinas Dataset

Parameter tuning the three proposed approaches for Salinas dataset is recapitulated in Table 17. For PCA_SVM method, the number of PCs obtained using HySime criterion for this dataset was 20.

Table 17. The parameter tuning for the proposed approaches in the Salinas dataset.

Approach	SDWT (Function, Level)	2D-DWT (Function, Level)	Wiener Filter Patch Size	2D-DCT Threshold
SDWT-2DWT_SVM	('sym4', 4)	('haar', 7)	/	/
SDWT-WF_SVM	('db4', 4)	/	47	/
SDWT-2DCT_SVM	('bior3.7', 2)	/	/	850

In the first experiment, to compare our proposed approaches and other methods, we randomly chose 100 training samples for each class from the reference data, and the rest were used as test samples, as summarized in Table 3. Table 18 reports the quantitative classification results, including the individual classification accuracies (%), κ statistic, standard deviation, AA, OA, and computational time in seconds. The classification maps for all the compared approaches on the Salinas image are illustrated in Figure 15.

From Table 18, we can see that our three proposed approaches provide the highest three classification accuracies in terms of all criteria (AA, OA, and kappa) with stable performance for the proposed SDWT-2DCT_SVM (0.21 as the lowest standard deviation value for OA). Hence, the proposed SDWT-2DCT_SVM approach achieved the highest values for the three classification criteria (98.91% κ , 99.02% OA, and 99.19% AA), followed by the proposed SDWT-WF_SVM with the second-best performance (97.46% κ , 97.72% OA, and 98.61% AA). The third-best performance was achieved by our proposed SDWT-2DWT_SVM (96.13% κ , 96.71% OA, and 97.97% AA). Furthermore, the proposed approaches achieve 2.61%–17.4%, 1.31%–16.1%, and 0.3%–15.09% OA advantage over the other state-of-the-art methods in the SDWT-2DCT_SVM, the SDWT-WF_SVM, and the SDWT-2DWT_SVM approaches, respectively.

In addition, the classification maps produced from the three proposed approaches are smooth and close to the ground truth map, as shown in Figure 15. Specifically, the classification map resulting from the proposed SDWT-2DWT_SVM is noiseless and very similar to the ground truth map.

Table 18. Classification results (%) for the Salinas dataset with standard deviation (in parentheses).

Class	SVM	PCA_SVM	3D_SVM	3DG_SVM	EPF	IFRF	BM4D_SVM	PARAFAC_SVM	SDWT-2DWT_SVM	SDWT-WF_SVM	SDWT-2DCT_SVM
1	99.13	99.36	99.19	99.70	100	98.99	99.25	93.48	99.72	99.81	99.54
2	99.51	99.52	99.35	99.68	99.97	100	99.63	77.64	99.64	99.69	99.82
3	99.58	99.09	97.48	98.95	96.07	99.84	99.77	73.90	99.50	99.47	99.82
4	99.34	99.35	99.36	99.43	98.43	90.05	99.30	96.50	99.23	99.63	99.30
5	98.74	97.78	98.98	99.46	99.80	99.98	98.55	92.87	98.06	98.30	98.94
6	99.68	99.67	99.86	99.95	99.99	100	99.67	97.08	99.68	99.62	99.59
7	99.57	99.56	99.66	99.67	99.93	99.79	99.55	87.86	99.58	99.60	99.76
8	67.98	72.40	86.27	88.67	85.40	99.53	78.00	99.68	91.05	93.41	98.54
9	99.00	97.50	98.43	99.18	98.73	99.98	99.47	94.00	98.96	99.86	99.56
10	95.01	94.51	93.79	95.60	93.60	99.72	96.36	78.41	96.09	98.01	97.03
11	98.90	98.75	99.68	99.93	97.58	99.02	98.97	70.28	98.94	98.34	99.20
12	99.82	99.70	99.97	99.87	99.67	98.82	99.84	80.34	99.95	99.91	99.91
13	99.58	99.55	99.02	98.97	99.93	97.70	99.72	99.33	99.75	99.46	99.96
14	97.76	97.92	98.01	98.19	97.89	97.58	98.43	91.79	97.95	96.96	98.79
15	70.04	70.51	75.92	85.52	77.68	83.06	79.72	92.69	90.80	97.05	98.50
16	98.86	98.88	98.83	98.70	99.71	100	98.48	99.35	98.71	98.59	98.79
κ	87.04 (0.62)	87.86 (1.07)	92.05 (0.41)	94.44 (0.89)	92.13 (0.90)	96.01 (0.27)	91.04 (0.69)	78.95 (1.03)	96.13 (0.61)	97.46 (0.35)	98.91 (0.23)
OA	88.36 (0.57)	89.10 (0.97)	92.88 (0.37)	95.02 (0.79)	92.95 (0.81)	96.41 (0.25)	91.96 (0.62)	81.62 (0.92)	96.71 (0.55)	97.72 (0.32)	99.02 (0.21)
AA	95.16 (0.20)	95.25 (0.41)	96.49 (0.15)	97.59 (0.36)	96.52 (0.31)	97.75 (0.29)	96.54 (0.21)	89.08 (1.00)	97.97 (0.21)	98.61 (0.16)	99.19 (0.08)
Time (s)	99.13	13.18	183.5	1142.52	19.29	4.77	1854.07	297.39	228.56	273.15	169.79

The bold values indicate critical values.

Figure 15. Classification maps of different methods on the Salinas dataset.

In this scenario, the IFRF method with an OA of 96.41% outperforms the 3DG_SVM (95.02% OA), with a smoother classification map. The EPF, 3D_SVM, and BM4D_SVM methods provided close classification results (92.95% OA, 92.88% OA, and 91.96% OA, respectively), resulting in similar classification maps as shown in Figure 15. The PARAFAC_SVM approach delivered the lowest classification accuracy (81.62% OA) with a noisy produced classification map.

As for the computational time, Table 18 shows that feature extraction-based methods achieved the shortest times, using 4.77 s for the IFRF method and 13.18 s for the PCA_SVM method. In general, classifiers incorporating spatial information are slower than the pixel-wise methods. Our proposed approaches, including SDWT-2DCT_SVM, SDWT-WF_SVM, and SDWT-2DWT_SVM, had the fourth, the fifth, and the sixth shortest execution times, among all the tested spectral-spatial methods. Specifically, the highest classification accuracy (99.09% OA) was obtained by our proposed SDWT-2DCT_SVM approach using 169.79 s and was less than three minutes. The 3DG_SVM and BM4D_SVM methods were time-consuming, using 1142.52 and 1854.07 s, respectively. Spectral DWT is performed on each pixel, and the spatial filters performed on each band in the three proposed approaches, which means that they can be parallelized. Thus, the computational time could be greatly improved.

The robustness of the proposed approaches regarding the state-of-the-art methods with different training conditions was evaluated in the second experiment, in which 2%, 4%, 6%, 8%, and 10% labeled samples of each class were randomly selected as training samples. The rest were used as testing samples. The average results of the OA values of 20 trials are plotted in Figure 16.

Figure 16. Classification accuracy (OA, %) achieved by the compared methods using different training sample sizes on the Salinas dataset.

Figure 16 shows that the three proposed approaches can improve the classification accuracy significantly with different training sample sizes, as the number of training samples varies from 2% to 10%. Our proposed approaches (SDWT-2DCT_SVM and SDWT-WF_SVM) always achieved the first- and the second-highest OA. However, the IFRF method outperforms our proposed SDWT-2DWT_SVM approach when using a smaller number of training samples (2%–6%). When the training sample size was greater than 8%, the proposed SDWT-2DWT_SVM approach tends to be more accurate than the IFRF method.

In general, several observations can be made from the experiments reported in this section on the three widely used hyperspectral datasets:

- The proposed approaches outperform the other considered methods; especially, our proposed SDWT-2DCT_SVM framework in terms of all classification criteria (AA, OA, and κ). It delivers higher accuracy on the three datasets with smoother classification maps than the other compared methods.
- The proposed approaches deliver higher classification accuracies under the constraint of the small size of training samples.
- The proposed frameworks are computationally efficient with a reasonable tradeoff between accuracy and computational time. Subsequently, they will be quite useful for applications requiring a fast response.
- The proposed methods can deal with different spatial resolutions (20 m, 3.7 m, and 1.7 m). Particularly, for the Indian Pines scene at a spatial resolution of 20 m, our proposed SDWT-2DCT_SVM and SDWT-WF_SVM approaches achieved the first- and the second-highest classification accuracies. Our proposed frameworks are effective for the classification of low spatial resolution images.

Indeed, the potential strength of our proposed strategies consists of the following advantages:

- The spectral and spatial dimensions are filtered separately as the noise variance is more drastic in the spectral dimension than in the spatial dimension.
- The performed spectral filters can potentially separate the valuable information from the details using the wavelet multi-resolution analysis.
- The spatial filter is performed in the transform domain, where the separation of the significant information from the detail is easy.
- The spatial filter in the transform domain is performed only on the detail coefficients and does not alter the significant information resulting from the spectral filter.

These characteristics made this filtering process very effective on the three tested datasets as the highest classification accuracies were always achieved by our proposed approaches.

6. Conclusions

In this paper, three effective classification approaches based on spectral–spatial preprocessing using DWT multilevel decomposition and spatial filtering are proposed. In these approaches, we perform spectral DWT multilevel decomposition on the original hyperspectral image to separate the approximation coefficients that represent the most significant information from the detail coefficients. Then, for each level of decomposition, only the details coefficients are spatially filtered in the transform domain to extract the rest of the meaningful spatial information and further enhance the spectral filtered part. After the spatial filtering step using three different filters, namely 2D-DWT, 2D-DCT, and AWF, corresponding to the three proposed frameworks, the DWT reconstruction was carried out on both the approximate and the spatially filtered detail coefficients. Finally, the output-preprocessed image is fed into a simple linear SVM classifier.

Different performance indicators were used to assess our experiments, including the classification accuracy, execution time, the classification maps, and the tests with different training sample sizes. Our experimental results, conducted with three hyperspectral images, confirm that the proposed approaches yield good classification results that were comparable to or better than those obtained by other state-of-the-art classifiers. In particular, the proposed SDWT-2DCT_SVM approach exhibited a significantly higher performance and outperformed all the comparable methods in terms of classification accuracy, even with small training sample sizes, and within a reasonable execution time on all the three datasets. Furthermore, in the three proposed approaches, spectral DWT multilevel decomposition is performed on each pixel separately, and the spatial filters are performed on each band, so they could be parallelized easily for significantly reduced computational times.

Author Contributions: All authors have made great contributions to the work. Conceptualization, R.B. and K.B.; Software, R.B. and K.B.; validation, H.W., R.B. and K.B.; formal analysis, R.B.; supervision, H.W., Writing—original draft, R.B.; and Writing—review and editing, H.W. and K.B.

Funding: This research received no external funding.

Acknowledgments: The authors are grateful for the valuable comments and propositions from the reviewer of this paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Tong, Q.; Xue, Y.; Zhang, L. Progress in Hyperspectral Remote Sensing Science and Technology in China Over the Past Three Decades. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 70–91. [\[CrossRef\]](#)
2. Adão, T.; Hruška, J.; Pádua, L.; Bessa, J.; Peres, E.; Morais, R.; Sousa, J.J. Hyperspectral Imaging: A Review on UAV-Based Sensors, Data Processing and Applications for Agriculture and Forestry. *Remote Sens.* **2017**, *9*, 1110. [\[CrossRef\]](#)
3. Yokoya, N.; Chan, J.C.-W.; Segl, K. Potential of Resolution-Enhanced Hyperspectral Data for Mineral Mapping Using Simulated EnMAP and Sentinel-2 Images. *Remote Sens.* **2016**, *8*, 172. [\[CrossRef\]](#)
4. He, J.; He, Y.; Zhang, C. Determination and Visualization of Peimine and Peiminine Content in *Fritillaria thunbergii* Bulbi Treated by Sulfur Fumigation Using Hyperspectral Imaging with Chemometrics. *Molecules* **2017**, *22*, 1402. [\[CrossRef\]](#)
5. Ghamisi, P.; Plaza, J.; Chen, Y.; Li, J.; Plaza, A.J. Advanced spectral classifiers for hyperspectral images: A review. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 8–32. [\[CrossRef\]](#)
6. Melgani, F.; Bruzzone, L. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Trans. Geosci. Remote Sens.* **2004**, *42*, 1778–1790. [\[CrossRef\]](#)
7. Ham, J.; Chen, Y.; Crawford, M.M.; Ghosh, J. Investigation of the random forest framework for classification of hyperspectral data. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 492. [\[CrossRef\]](#)
8. Ratle, F.; Camps-Valls, G.; Weston, J. Semisupervised neural networks for efficient hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 2271–2282. [\[CrossRef\]](#)
9. Zhong, Y.; Zhang, L. An adaptive artificial immune network for supervised classification of multi-/hyperspectral remote sensing imagery. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 894–909. [\[CrossRef\]](#)
10. Li, J.; Bioucas-Dias, J.M.; Plaza, A. Spectral–spatial classification of hyperspectral data using loopy belief propagation and active learning. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 844–856. [\[CrossRef\]](#)
11. Di, W.; Crawford, M.M. View generation for multiview maximum disagreement based active learning for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 1942–1954. [\[CrossRef\]](#)
12. Iordache, M.-D.; Bioucas-Dias, J.M.; Plaza, A. Total variation regularization in sparse hyperspectral unmixing. In Proceedings of the 2011 3rd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), Lisbon, Portugal, 6–9 June 2011; pp. 1–4.
13. Chen, Y.; Nasrabadi, N.M.; Tran, T.D. Hyperspectral Image Classification via Kernel Sparse Representation. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 217–231. [\[CrossRef\]](#)
14. Castrodad, A.; Xing, Z.; Greer, J.B.; Bosch, E.; Carin, L.; Sapiro, G. Learning Discriminative Sparse Representations for Modeling, Source Separation, and Mapping of Hyperspectral Imagery. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 4263. [\[CrossRef\]](#)
15. Wang, Y.; Duan, H. Classification of Hyperspectral Images by SVM Using a Composite Kernel by Employing Spectral, Spatial and Hierarchical Structure Information. *Remote Sens.* **2018**, *10*, 441. [\[CrossRef\]](#)
16. Camps-Valls, G.; Tuia, D.; Bruzzone, L.; Benediktsson, J.A. Advances in hyperspectral image classification: Earth monitoring with statistical learning methods. *IEEE Signal Process. Mag.* **2014**, *31*, 45–54. [\[CrossRef\]](#)
17. Plaza, A.; Benediktsson, J.A.; Boardman, J.W.; Brazile, J.; Bruzzone, L.; Camps-Valls, G.; Chanussot, J.; Fauvel, M.; Gamba, P.; Gualtieri, A.; et al. Recent advances in techniques for hyperspectral image processing. *Remote Sens. Environ.* **2009**, *113*, S110–S122. [\[CrossRef\]](#)
18. Casalino, G.; Gillis, N. Sequential dimensionality reduction for extracting localized features. *Pattern Recognit.* **2017**, *63*, 15–29. [\[CrossRef\]](#)
19. Quesada-Barriuso, P.; Arguello, F.; Heras, D.B. Spectral–spatial classification of hyperspectral images using wavelets and extended morphological profiles. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 1177–1185. [\[CrossRef\]](#)

20. Jia, S.; Wu, K.; Zhu, J.; Jia, X. Spectral-Spatial Gabor Surface Feature Fusion Approach for Hyperspectral Imagery Classification. *IEEE Trans. Geosci. Remote Sens.* **2018**, 1–13. [[CrossRef](#)]
21. Tang, Y.Y.; Lu, Y.; Yuan, H. Hyperspectral image classification based on three-dimensional scattering wavelet transform. *IEEE Trans. Geosci. Remote Sens.* **2015**, 53, 2467–2480. [[CrossRef](#)]
22. Fauvel, M.; Benediktsson, J.A.; Chanussot, J.; Sveinsson, J.R. Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles. *IEEE Trans. Geosci. Remote Sens.* **2008**, 46, 3804–3814. [[CrossRef](#)]
23. Ghamisi, P.; Benediktsson, J.A.; Cavallaro, G.; Plaza, A. Automatic Framework for Spectral–Spatial Classification Based on Supervised Feature Extraction and Morphological Attribute Profiles. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, 7, 2147. [[CrossRef](#)]
24. Sun, S.; Zhong, P.; Xiao, H.; Wang, R. An MRF model-based active learning framework for the spectral-spatial classification of hyperspectral imagery. *IEEE J. Sel. Top. Signal Process.* **2015**, 9, 1074–1088. [[CrossRef](#)]
25. Pan, B.; Shi, Z.; Xu, X. MugNet: Deep learning for hyperspectral image classification using limited samples. *ISPRS J. Photogramm. Remote Sens.* **2017**, 145, 108–119. [[CrossRef](#)]
26. Liu, B.; Yu, X.; Yu, A.; Zhang, P.; Wan, G. Spectral-spatial classification of hyperspectral imagery based on recurrent neural networks. *Remote Sens. Lett.* **2018**, 9, 1118–1127. [[CrossRef](#)]
27. Song, B.; Li, J.; Dalla Mura, M.; Li, P.; Plaza, A.; Bioucas-Dias, J.M.; Benediktsson, J.A.; Chanussot, J. Remotely sensed image classification using sparse representations of morphological attribute profiles. *IEEE Trans. Geosci. Remote Sens.* **2014**, 52, 5122. [[CrossRef](#)]
28. Ni, D.; Ma, H. Hyperspectral image classification via sparse code histogram. *IEEE Geosci. Remote Sens. Lett.* **2015**, 12, 1843–1847.
29. Wang, Y.; Song, H.; Zhang, Y. Spectral-spatial classification of hyperspectral images using joint bilateral filter and graph cut based model. *Remote Sens.* **2016**, 8, 748. [[CrossRef](#)]
30. Fang, L.; Li, S.; Kang, X.; Benediktsson, J.A. Spectral–spatial classification of hyperspectral images with a superpixel-based discriminative sparse model. *IEEE Trans. Geosci. Remote Sens.* **2015**, 53, 4186–4201. [[CrossRef](#)]
31. He, L.; Li, J.; Liu, C.; Li, S. Recent advances on spectral–spatial hyperspectral image classification: An overview and new guidelines. *IEEE Trans. Geosci. Remote Sens.* **2018**, 56, 1579–1597. [[CrossRef](#)]
32. Cao, X.; Ji, B.; Ji, Y.; Wang, L.; Jiao, L. Hyperspectral image classification based on filtering: A comparative study. *J. Appl. Remote Sens.* **2017**, 11, 35007. [[CrossRef](#)]
33. Liu, K.-H.; Lin, Y.-Y.; Chen, C.-S. Linear spectral mixture analysis via multiple-kernel learning for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2015**, 53, 2254–2269. [[CrossRef](#)]
34. Zhou, Y.; Peng, J.; Chen, C.L.P. Extreme learning machine with composite kernels for hyperspectral image classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, 8, 2351–2360. [[CrossRef](#)]
35. Bazine, R.; Wu, H.; Boukhechba, K. Spatial Filtering in DCT Domain-Based Frameworks for Hyperspectral Imagery Classification. *Remote Sens.* **2019**, 11, 1405. [[CrossRef](#)]
36. Oktem, R.; Ponomarenko, N.N. Image filtering based on discrete cosine transform. *Telecommun. Radio Eng.* **2007**, 66. [[CrossRef](#)]
37. Guo, X.; Huang, X.; Zhang, L. Three-Dimensional Wavelet Texture Feature Extraction and Classification for Multi/Hyperspectral Imagery. *IEEE Geosci. Remote Sens. Lett.* **2014**, 11, 2183. [[CrossRef](#)]
38. Kordi Ghasrodashti, E.; Helfroush, M.S.; Danyali, H. A wavelet-based classification of hyperspectral images using Schroedinger eigenmaps. *Int. J. Remote Sens.* **2017**, 38, 3608–3634. [[CrossRef](#)]
39. Kordi Ghasrodashti, E.; Helfroush, M.S.; Danyali, H. Spectral-spatial classification of hyperspectral images using wavelet transform and hidden Markov random fields. *Geocarto Int.* **2018**, 33, 771–790. [[CrossRef](#)]
40. Qian, Y.; Ye, M.; Zhou, J. Hyperspectral image classification based on structured sparse logistic regression and three-dimensional wavelet texture features. *IEEE Trans. Geosci. Remote Sens.* **2013**, 51, 2276–2291. [[CrossRef](#)]
41. Cao, X.; Xu, L.; Meng, D.; Zhao, Q.; Xu, Z. Integration of 3-dimensional discrete wavelet transform and Markov random field for hyperspectral image classification. *Neurocomputing* **2017**, 226, 90–100. [[CrossRef](#)]
42. Zhou, X.; Prasad, S.; Crawford, M.M. Wavelet-domain multiview active learning for spatial-spectral hyperspectral image classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, 9, 4047–4059. [[CrossRef](#)]
43. Shi, C.; Pun, C.-M. 3D multi-resolution wavelet convolutional neural networks for hyperspectral image classification. *Inf. Sci. (N.Y.)* **2017**, 420, 49–65. [[CrossRef](#)]
44. Othman, H.; Qian, S.-E. Noise reduction of hyperspectral imagery using hybrid spatial-spectral derivative-domain wavelet shrinkage. *IEEE Trans. Geosci. Remote Sens.* **2006**, 44, 397–408. [[CrossRef](#)]

45. Mallat, S.G. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **1989**, *11*, 674. [[CrossRef](#)]
46. Donoho, D.L. Denoising by softthresholding. *IEEE Trans. Inf. Theory* **1995**, *41*, 613–627. [[CrossRef](#)]
47. Prasad, S.; Bruce, L.M. Limitations of principal components analysis for hyperspectral target recognition. *IEEE Geosci. Remote Sens. Lett.* **2008**, *5*, 625–629. [[CrossRef](#)]
48. Kang, X.; Li, S.; Benediktsson, J.A. Spectral-Spatial Hyperspectral Image Classification with Edge-Preserving Filtering. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 2666–2677. [[CrossRef](#)]
49. Kang, X.; Li, S.; Benediktsson, J.A. Feature extraction of hyperspectral images with image fusion and recursive filtering. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 3742–3752. [[CrossRef](#)]
50. Maggioni, M.; Foi, A. Nonlocal transform-domain denoising of volumetric data with groupwise adaptive variance estimation. In Proceedings of the IS&T/SPIE Electronic Imaging 2012, Burlingame, CA, USA, 22–26 January 2012; Volume 8296, p. 82960O.
51. Liu, X.; Bourennane, S.; Fossati, C. Denoising of hyperspectral images using the PARAFAC model and statistical performance analysis. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 3717–3724. [[CrossRef](#)]
52. Chang, C.-C.; Lin, C.-J. {LIBSVM}: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.* **2011**, *2*, 27:1–27:27. [[CrossRef](#)]
53. Bioucas-Dias, J.M.; Nascimento, J.M.P. Estimation of signal subspace on hyperspectral data. In Proceedings of the SPIE Remote Sensing, Bruges, Belgium, 19–22 September 2005; Volume 5982, p. 59820L.
54. Story, M.; Congalton, R.G. Accuracy assessment: A user's perspective. *Photogramm. Eng. Remote Sens.* **1986**, *52*, 397–399.

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).