Article

Spatial Filtering in DCT Domain-Based Frameworks 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

check for
updates

Abstract: In this article, we propose two effective frameworks for hyperspectral imagery classification based on spatial filtering in Discrete Cosine Transform (DCT) domain. In the proposed approaches, spectral DCT is performed on the hyperspectral image to obtain a spectral profile representation, where the most significant information in the transform domain is concentrated in a few low-frequency components. The high-frequency components that generally represent noisy data are further processed using a spatial filter to extract the remaining useful information. For the spatial filtering step, both two-dimensional DCT (2D-DCT) and two-dimensional adaptive Wiener filter (2D-AWF) are explored. After performing the spatial filter, an inverse spectral DCT is applied on all transformed bands including the filtered bands to obtain the final preprocessed hyperspectral data, which is subsequently fed into a linear Support Vector Machine (SVM) classifier. Experimental results using three hyperspectral datasets show that the proposed framework Cascade Spectral DCT Spatial Wiener Filter (CDCT-WF_SVM) outperforms several state-of-the-art methods in terms of classification accuracy, the sensitivity regarding different sizes of the training samples, and computational time.

Keywords: Spectral-spatial hyperspectral classification; discrete cosine transform; structural filtering; SVM; wiener filter; transform domain filtering

1. Introduction

Hyperspectral imagery is collected by hyperspectral remote sensors at hundreds of narrow spectral bands. It contains rich discriminative spectral and spatial characteristics regarding material surfaces [1]. This attribute makes hyperspectral imagery an interesting source of information for a wide variety of applications in areas such as agriculture, environmental planning, surveillance, target detection, medicine [2–4], etc. Most of these hyperspectral applications build upon classification tasks. The classification of hyperspectral images is often performed using supervised classification methods, in which the training of the classification model requires the availability of labeled samples.

However, supervised classification has issues related to the imbalance between the high dimensionality and the limited availability of the acquired training samples, leading to the Hughes phenomenon. Moreover, noises from the environment and optical sensors can further reduce the classification effectiveness. Thus, to address these problems, intensive work has been devoted to providing accurate classifiers for hyperspectral images including, SVM [5], random forests [6], neural networks [7,8], sparse representation [9,10], and active learning [11,12] methods. Notably, the SVM classifier is efficient [5], since it requires relatively few training samples to obtain high classification accuracy and robustness regarding the high spectral dimensionality [5]. Although these pixel-wise classifiers can fully use the spectral information in the hyperspectral imagery, the spatial dimension is not takenin consideration. Consequently, the resulting classification maps are corrupted with salt-and-pepper noise [13].

In recent years, the integration of spatial to spectral information in a classifier model has improved classification accuracy, eliminating salt-and-pepper noise in classification maps [14,15]. Subsequently, spectral-spatial classification methods have been proposed to improve the accuracy further. The proposed methods can be roughly divided into three basic paradigms in terms of the fusion stage of the spectral and the spatial information [16], including preprocessing-based classification [17,18], post-processing-based classification [19,20], and integrated classification [21,22]. If we use more than one basic paradigm, we obtain a hybrid classification [23,24]. For these different adopted paradigms, there are typical methodologies including structural filtering-based approaches [25,26], random field-based approaches [27,28], morphological profile (MP)-based approaches [29,30], deep learning-based approaches [31–33], sparse representation-based approaches (SRC) [34,35], and segmentation-based approaches such as hierarchical segmentation [36], graph cut [37], and superpixel [38] approaches.

The structural filtering methodology is one of the most studied methods for preprocessing hyperspectral imagery [16]. Compared with other spatial methodologies, spatial filtering is simple and easy to implement [39]. This advantage makes it very suitable for practical applications. Much research is currently underway using structural filtering to obtain contextual features. The simplest way is to extract the spatial information based on the moment criteria [40,41]. Performing local harmonic analysis is another research direction for the structural filtering. The related work includesspectral-spatial wavelet features proposed in [42], 3D wavelet feature extraction and classification [26], spectral-spatial Gabor features developed in [43,44]. Recently, the trend in structural filtering has shifted towards methods that extract features using adaptive structures [45], such as adaptive multidimensional wiener filtering [46]. Even though these methods perform effectively in some scenarios, they still do not make full use of spatial information and only exploit this information in the preprocessing feature extraction phase.

Based on the success of the structural filtering methodology for hyperspectral imagery classification, and aiming to exploit the spectral and spatial information fully, in this work, we propose two effectivespectral-spatial hyperspectral classification frameworks based on performing the spatial filtering process in the DCT transform domain. We perform a spectral DCT on the original dataset to separate the most significant information concentrated in a few low-frequency components from the noisy data embedded in the high-frequency components. Then, spatial filtering is performed on the separated noisy data in the frequency domain to extract the remaining meaningful information. In the spatial filtering step, both 2D-DCT and 2D-AWF are explored given their effectiveness in noise removal [47]. Indeed, the variation of the noise variance in the spectral dimension is, in general, more drastic than that in the spatial dimensions [48]. Thus, the proposed approach is based on exploring DCT energy compaction property in the spectral domain to preserve the valuable part of the spectral signature presented by few low-frequency coefficients where further spatial filtering is only performed on the high-frequency coefficients that represent details and noise. This methodology makes the filtering process very effective. Furthermore, as we fully exploit the spectral and spatial information, the SVM classifier is adopted in our framework for its robustness to the Hughes phenomenon, in addition to its good generalization capacity [5].

The main contributions of this paper are threefold. The first contribution is the application of the spatial filter in the transform domain on the noisy data rather than discarding this part of information as usually adopted by the feature extraction-based approaches [49,50]. Thus, the proposed framework makes full use of the filtered spectral and spatial information of the hyperspectral image. The second contribution 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. Thus, we provide an effective framework for hyperspectral image classification that involves the tuning of only a couple of parameters regarding the spectral and the spatial filters respectively. The third contribution is the low computational time compared with the other spectral-spatial classification methods. Indeed, the efficient exploitation of the filtering process improves the classification performance in less computational time.

The rest of this paper is structured as follows. The mathematical formulation of DCT and Wiener filter are briefly introduced in Section 2. The proposed approaches are described in Section 3. Datasets and the evaluation process are presented in Section 4. Section 5 gives the experimental setup and reports the results with the related analysis. Finally, Section 6 outlines a summary and draws conclusions.

2. Materials and Methods

2.1. Discrete Cosine Transform (DCT)

The DCT termed the cosine basis function which represents a summation of cosine functions at different increasing frequencies [51]. Using these cosine basis functions, DCT compacts the energy of a signal in a few low-frequency DCT coefficients [50]. Thus, DCT is widely used in image and signal processing, particularly for data compression, due to its high "energy compaction" property [52,53].

From the viewpoint of signal processing, considering each pixel X as a discrete signal in N dimension space where $x = [x_0, x_1, x_2, \ldots, x_{N-1}]$, and N represents the number of the spectral bands. The DCT coefficients of this discrete signal X are defined as [52]:

$$\begin{cases} d = [d_0,\ d_1, \ldots, d_{N-1}]^T \\ d_0 = \frac{\sqrt{2}}{N} \sum\limits_{n=0}^{N-1} x_n \\ d_u = \frac{2}{N} \sum\limits_{n=0}^{N-1} x_n \cos \frac{(2n+1)u\pi}{2N}; u = 1, 2, \ldots . N-1 \end{cases} \tag{1}$$

where d_u represents the uth DCT coefficient in N dimension space and each vector d corresponds to the original pixel X. Our approach is based on transforming the original spectral feature space to a transform feature space by performing the spectral DCT (SDCT) on each pixel spectral curve.

The Inverse DCT is defined as:

$$X(u) = \sum_{n=0}^{N-1} c[u] d_n \cos \left(\frac{u2\pi n}{N} \right) \tag{2}$$

where $X(u)$ is the uth IDCT reconstruction value, $u = 0, 1, \ldots, N - 1$, d_n is the nth DCT coefficient, $c[u] = 1$ for $u = 0$ and $c[u] = 2$ for $u = 1, 2, \ldots, N - 1$.

The DCT algorithm is very effective owing to its symmetry and simplicity. It is a reliable replacement for a Fast Fourier Transform (FFT), as it considers the real components of the image data. Generally, FFT is used for general spectral analysis applications where DCT is frequently used in lossy data compression and denoising applications, given its ability to concentrate the energy of the signal in a small set of DCT coefficients. Moreover, using only "real-valued" cosine functions makes DCT computationally simpler than FFT, which is a complex algorithm using magnitude and phase [54].

2.2. 2D-Discrete Cosine Transform (2D-DCT)

The 2D-DCT is obtained by performing the DCT on the first dimension and then the second one. The 2D-DCT of an $f[m, n]$, which is N by N image pixels, is expressed by Equation (3):

$$F[u, v] = \frac{1}{N^2} \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} f[m, n] \cos[(2m+1)u\pi/2N] \cos[(2n+1)v\pi/2N] \tag{3}$$

where $F[u, v]$ is the $(u,v)^{th}$ 2D-DCT coefficient, $f[m, n]$ is (m,n) value of the matrix f and (u,v) are the discrete frequency variables $(0,0), \ldots, (N - 1, N - 1)$.

The 2D basis functions are the multiplication result of the matrix row of 1D-basis functions with the same function matrix column. The basis function frequencies are increasing with regard to both the vertical and horizontal directions [51]. The 2D Inverse DCT is expressed by Equation (4):

$$f[m,n] = \sum_{u=0}^{N-1}\sum_{v=0}^{N-1} c[u]c[v]F[u,v]\cos[(2u+1)m\pi/2N]\cos[(2v+1)n\pi/2N] \tag{4}$$

where $f[m,n]$ is the $(m,n)^{th}$ IDCT reconstruction value, $F[u,v]$ is the $(u,v)^{th}$ 2D-DCT coefficients, c[λ] = 1 for λ = 0 and c[λ] = 2 for λ = 1, 2, ... , $N-1$.

Due to its energy concentration ability, DCT provides optimal decorrelation for images [55]. Thus, in our approach, after performing DCT to obtain a frequency profile where the meaningful information is compacted in few low-frequency components is retained, and further spatial filtering is executed on the high-frequency components by exploring the effectiveness of both 2D-DCT and 2D-AWF filtering in the transform domain.

2.3. Spatial Adaptive Wiener Filter (2D-AWF)

Hyperspectral imagery can be affected by two kinds of noise which are signal-independent noise and signal-dependent noise, many references [56,57] argue that only signal-independent noise needs be reduced, as it is the dominant component. This signal-independent noise is referred to as Additive White Gaussian Noise (AWGN) [58]. Therefore, in the spatial filtering step of our approach, we have also explored the most widely known spatial filter for AWGN removal, which is the 2D-AWF.

Indeed, the Wiener filter was proposed by Norbert Wiener during the 1940s; it is based on the minimization of the mean square error (MSE) between the output signal and the estimated signal [59]. The 2D-AWF filter applies the idea of the Wiener filter in the spatial domain. It estimates the image noise and filters the image based on this noise [60].

Considering an image $x(i,j)$ corrupted with AWGN $n(i,j)$ expressed as follows:

$$y(i,j) = x(i,j) + n(i,j) \tag{5}$$

where $x(i,j)$ is the original image and $y(i,j)$ denotes the noisy image.

If we assume that the noise is stationary of zero mean, uncorrelated with the original image. The original image $x(i,j)$ can be modeled by:

$$x(i,j) = m_x + \sigma_x w(i,j) \tag{6}$$

where m_x is the local mean and σ_x is the standard deviation, and $w(i,j)$ representsthe zero mean white noise of unit variance. The spatial Wiener filter that minimizes the (MSE) between the original image $x(i,j)$ and the enhanced image $\hat{x}(i,j)$ is expressed as follows [61]:

$$\hat{x}(i,j) = m_x + \frac{\sigma_x^2}{\sigma_x^2 + \sigma_n^2}[y(i,j) - m_x] \tag{7}$$

In Equation (7), m_x and σ_x are estimated from the observed noisy image and updated for each pixel as proposed in the Lee filter [62]:

$$\hat{m}_x(i,j) = \frac{1}{(2m+1)(2n+1)}\sum_{k=i-m}^{i+m}\sum_{l=j-n}^{j+n} y(k,l) \tag{8}$$

$$\hat{\sigma}_y^2(i,j) = \frac{1}{(2m+1)(2n+1)}\sum_{k=i-m}^{i+m}\sum_{l=j-n}^{j+n}[y(k,l) - \hat{m}_x(i,j)]^2 \tag{9}$$

In our approach, the filter size setting $(2m + 1)(2n + 1)$ is fixed regarding the best-obtained classification accuracy.

$$\hat{\sigma}_x^2(i, j) = \max\{0, \hat{\sigma}_y^2(i, j) - \sigma_n^2\} \tag{10}$$

After substituting $\hat{m}_x(i, j)$ and $\hat{\sigma}_x(i, j)$ into Equation (7), it becomes

$$\hat{x}(i, j) = \hat{m}_x(i, j) + \frac{\hat{\sigma}_x^2(i, j)}{\hat{\sigma}_x^2(i, j) + \sigma_n^2}[y(i, j) - \hat{m}_x(i, j)] \tag{11}$$

In our approach, 2D-AWF is explored as a spatial filter in the DCT domain to extract the useful information embedded in the high-frequency components resulting from the spectral filtering using SDCT.

3. The Proposed Approach

A general outline description of the proposed classification frameworks based on spectral DCT combined with a spatial filter (2D-DCT or 2D-AWF) and SVM classifier is provided in Figure 1.
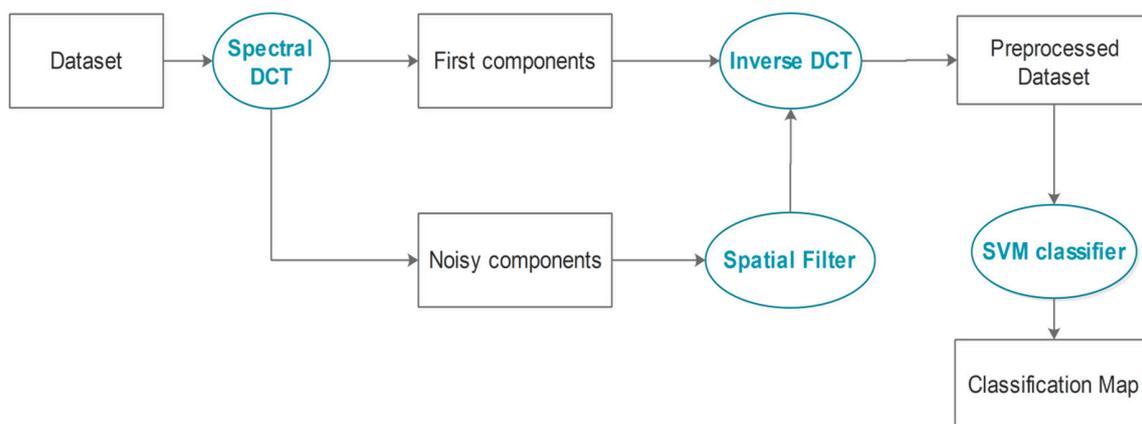

Figure 1. General flowchart of the spatial filtering in DCT domain-based classification framework for hyperspectral imagery.

As shown in Figure 1, in these approaches, we perform a spectral DCT on the original dataset to separate the most significant information concentrated in few low-frequency components from the noisy data embedded in the high-frequency components. Spatial filtering is performed in the transform domain on the high DCT spectral components to extract the remaining of meaningful information. An inverse spectral DCT is performed on all components including the filtered ones to obtain the final preprocessed hyperspectral dataset. Finally, the resulted data is introduced into the linear SVM classifier. Figure 2 provides a detailed flowchart of the proposed approaches.

As depicted in Figure 2, the proposed frameworks consist of the following steps:

Step 1: Consider an original hyperspectral dataset cube, with M rows, L columns, and P bands.

Step 2: Flatten the dataset cube to obtain a matrix E(N,P) with N = (M × L) rows and P columns.

Step 3: Perform a spectral DCT on matrix E to obtain two matrices: a matrix F(N,K) for the K first retained low-frequency components and matrix R(N,P-K) for the remaining high-frequency components (K is the cut-off channel count).

Step 4: Unflatten the matrix R(N,P-K) to obtain the data cube Rc(M,L,P-K) with M rows, L columns, and P-k bands.

Step 5: Perform spatial filtering on each matrix Rm(M, L) in the cube Rc using 2D-DCT or 2D-AWF as follows:

-    The 2D-DCT filtering step consists of performing a global 2D-DCT filter on each matrix Rm(M,L) where the high-frequency components are discarded using an empirical estimated hard threshold.

Then, an inverse DCT is applied to the matrix Rm(M, L). Figure 3 illustrates this spatial filtering step.

- 2D-AWF approach filtering step consists on dividing the matrix Rm(M, L) into blocks or patches of a specified size and processing them using a local 2D-AWF.

Step 6: Flatten the filtered data cube Rc to get back to the matrix R.

Step 7: Stack the two matrices F and R in order to obtain the matrix D, which has the same size as the matrix E, in step 2.

Step 8: Perform an Inverse DCT on the matrix D to obtain the preprocessed matrix E.

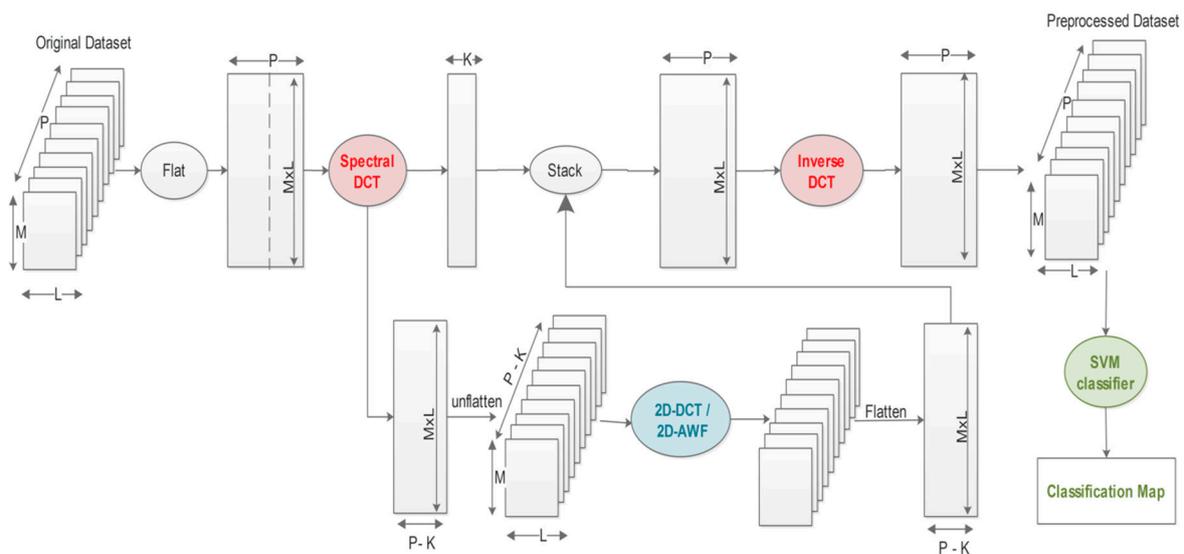Step 9: Unflatten matrix E to obtain the final matrix with same size as the original one (M × L × P) .

Figure 2. Detailed flowchart of the proposed spatial filtering in DCT domain-based frameworks for hyperspectral imagery classification.
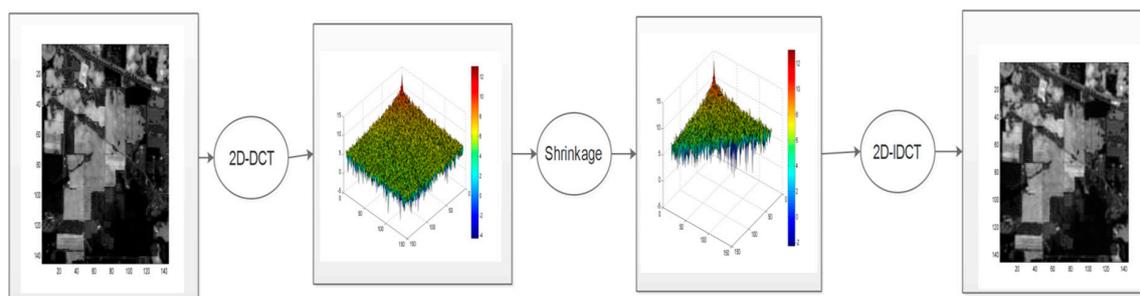

Figure 3. Flowchart of global 2D-DCT filtering step.

4. Data and Evaluation Process

4.1. Data

Three hyperspectral scenes with variant spatial resolutionsdownloaded from http://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes were used for the experiments including the Indian Pines, Salinas, and Pavia University datasets.

The first dataset concerns an Indian Pines scene. This scene was captured over northwest Indiana's Indian Pines area by the Airborne Visible Infrared Imaging Spectrometer (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. Regarding the spectral similarity between the classes and its low spatial resolution, this scene represents challenging classification. 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
Total		10,249	1294	8955

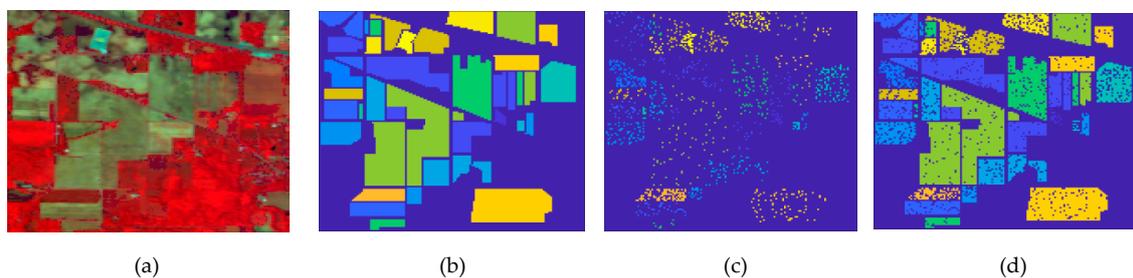Figure 4 shows the original band and the ground truth data of the Indian Pines dataset.

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

Figure 4. The ground truth related information of Indian Pines image: (**a**) false-color composite, (**b**) ground truth map, (**c**) training map, (**d**) test map.

The second test dataset was the Salinas scene. This dataset was captured the area over Salinas Valley in California by AVIRIS (Airborne Visible Infrared Imaging Spectrometer) 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 2 lists these classes with the related number of samples, training, and testing points in this dataset. Figure 5 illustrates the Salinas image and its ground truth categorization maps.

The third 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 as listed in Table 3 with the number of samples that belong to each region, the number of training and testing points.

Table 2. Classes of the Salinas scene with the related number of test and training samples.

Class	Type	Samples	Training	Testing
1	Brocoli greenweeds1	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
Total		54,129	1600	52,529

(a)   (b)   (c)   (d)
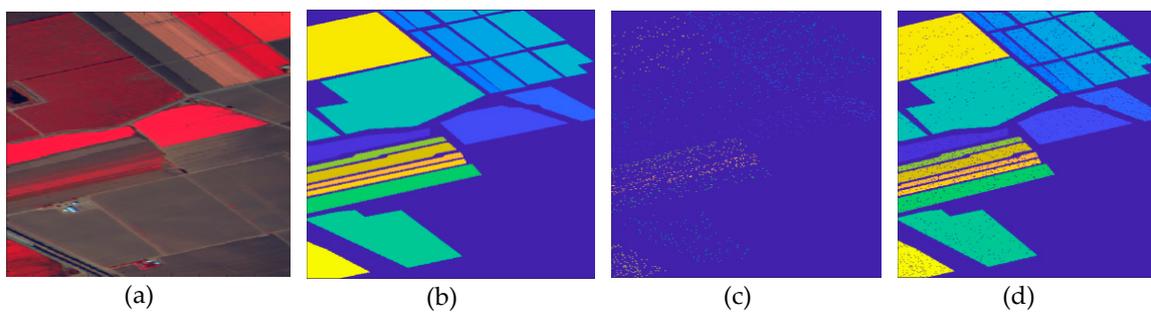
Figure 5. The ground truth related information of Salinas image: (**a**) false-color composite, (**b**) ground truth map, (**c**) training map, (**d**) test map.

Table 3. Classes of the Pavia University scene with the related number 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
Total		42,776	2700	40,076

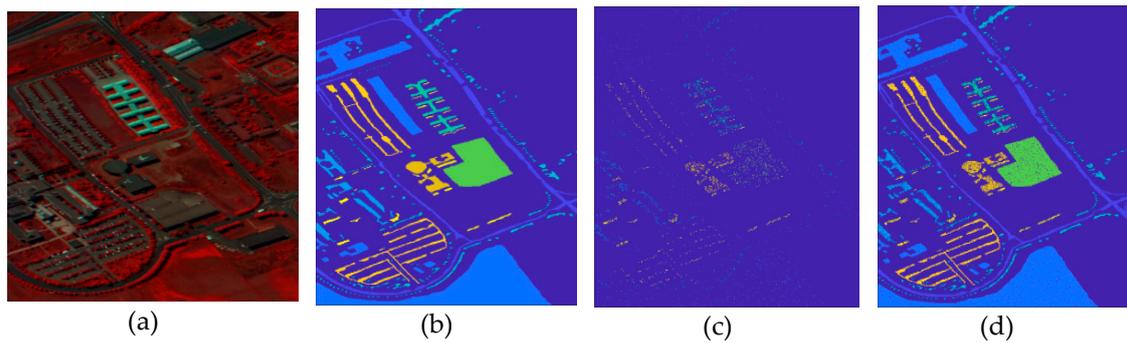Figure 6 illustrates the Pavia University image and its related ground truth maps.

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

Figure 6. The ground truth related information of Pavia University image: (**a**) false-color composite, (**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 CDCT-WF_SVM and CDCT-2DCT_SVM frameworks. These included two pixel-wise classifiers represented by the standard SVM [5], and PCA_SVM [63]. Besides, as nowadays three-dimensional Wavelet filter is widely explored in the structural filtering methodology, we opted for comparison two recent spectral-spatial classification methods including 3D-DWT-based approach (3D_SVM) [62], 3D-DWT with Graph Cut-based approach (3DG_ SVM) [62]. The bilateral filter, which is considered to be an effective filter for removing noises and small details while preserving large-scale structures, is also included in our comparison, represented by the Edge-Preserving Filtering-based approach (EPF) [64] and the Image Fusion and Recursive Filtering-based approach (IFRF) [65]. Moreover, two well-known denoising methods, Block-Matching 4-D Filtering (BM4D_SVM) [66] and the Parallel Factor Analysis-based approach (PARAFAC_SVM) [67], were included for comparison.

The parameter settings selected in the compared techniques were the default settings used in the corresponding research works, and the related source codes were provided by the respective authors. Specifically, for a fair comparison, linear SVM was adopted as a pixel-wise classifier for all the compared methods. The SVM classifier was implemented using the Matlab R2014a libsvm library [68]. 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 [69].

Furthermore, to emphasize the efficiency of our filtering approach, other framework configurations based on performing spectral and/or spatial filtering were implemented for comparison purposes, including performing only the spectral DCT filtering (DCT_SVM), performing only the spatial filtering (2DCT_SVM), performing serial spectral and spatial filtering on all components (SDCT-2DCT_SVM), and serial spatial and spectral filtering (S2DCT-DCT_SVM). Table 4 summarizes all the compared approaches with their abbreviations.

The compared methods are assessed numerically using five criteria: Average Accuracy (AA), Overall Accuracy (OA), statistical kappa coefficient (κ) [70], 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. The experiments were repeated 20 times, and the average of classification accuracies (kappa, OA, AA) was retained for comparison.

To verify the classification robustness of our proposed frameworks in case of 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 are 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 are 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 R2014a ®.

Table 4. Summary of the compared methods with their abbreviations.

Abbreviations	Methods
SVM	Support Vector Machine
PCA_SVM	Principal Component Analysis followed by SVM
DCT_SVM	Spectral one-dimensional Discrete Cosine Transform followed by SVM
2DCT_SVM	Two- dimensional DCT followed by SVM
CDCT-2DCT_SVM	Cascade spectral DCT spatial 2D-DCT followed by SVM
CDCT-WF_SVM	Cascade spectral DCT spatial Wiener filter followed by SVM
SDCT-2DCT_SVM	Serial spectral DCT spatial 2D-DCT followed by SVM
S2DCT-DCT_SVM	Serial spatial 2D-DCT spectral DCT followed by SVM
3D_SVM	Three-dimensional Wavelet followed by SVM
3DG_ SVM	Three-dimensional Wavelet with Graph Cut followed by SVM
EPF	Edge-Preserving Filtering
IFRF	Image Fusion and Recursive Filtering
BM4D_SVM	Block-Matching 4-D Filtering followed by SVM
PARAFAC_SVM	Parallel Factor Analysis followed by SVM

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 parameters' estimation for our proposed approaches are described and the experimental results analyzed to evaluate the performance of our proposed approaches relative to other closely related hyperspectral classification methods.

5.1. Parameter Settings

Two key parameters related to the spectral and spatial filters must be independently fixed for each of our proposed approaches. The estimation of these two parameters is explained in the following two subsections for the CDCT-2DCT_SVM and CDCT-WF_SVM frameworks.

5.1.1. CDCT-2DCT_SVM Framework Parameter Estimation

The count of the retained coefficients after performing the spectral DCT represents the spectral filter parameter. This parameter is set by comparing the classification accuracy OA with varying numbers of DCT coefficients where the count of the retained coefficients corresponds to the most accurate classification OA for each dataset. Figure 7 illustrates the classification accuracies in terms of OA for the CDCT-2DCT_SVM framework with a different number of DCT coefficients on the Indian Pines, Salinas, and Pavia University datasets.

As shown in Figure 7, it is clear that the classification accuracy OA is the the highest when incorporating up to ten first coefficients for the three datasets. Classification accuracy decreases due to the influence of noise present in the high-frequency components when selecting more than 40, 70, and 20 coefficients for the Indian Pines, Salinas, and Pavia University datasets, respectively. In our experiments, therefore, only ten DCT coefficients were retained for the three datasets. For each dataset, ten DCT coefficient were also set in the other methods used for comparison, based on performing a spectral DCT including DCT_SVM, SDCT-2DCT_SVM, and S2DCT-DCT_SVM.

The spatial filter parameter in the CSDCT-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 8 depicts the variation of the classification accuracy OA with varying threshold values, ranging from 50 to 1000 with the step of 50, on Indian Pines, Salinas, and Pavia University datasets.
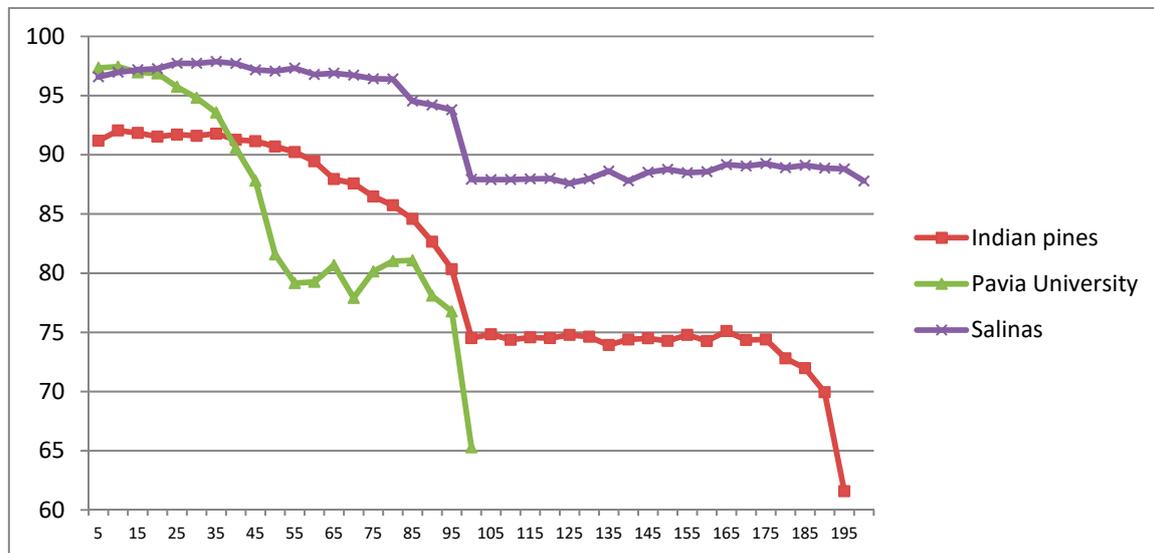

Figure 7. Overall Accuracy (OA%) with respect to the number of the first retained DCT coefficients in CDCT-2DCT_SVM on the Indian Pines, Salinas, and Pavia University datasets.
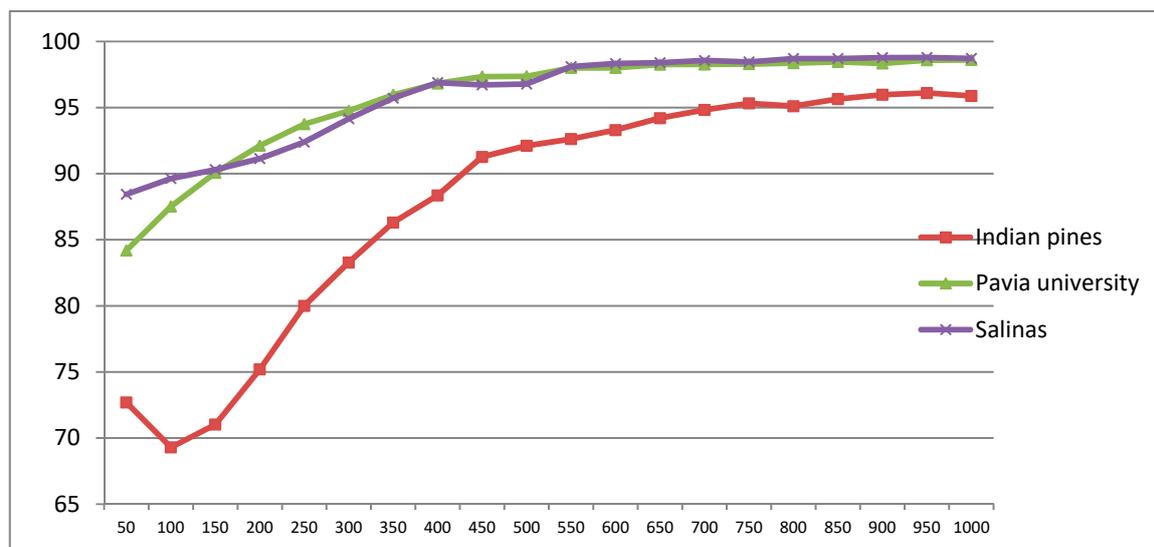

Figure 8. Overall Accuracy (OA%) with respect to the different selected 2D-DCT thresholds in CDCT-2DCT_SVM on Indian Pines, Salinas, and Pavia University datasets.

As can be seen in Figure 8, the OA increased with a higher threshold value for all datasets and tend to be stable around the threshold of 500 for Indian Pines, while this degree of stability was obtained at around a threshold of 450 for Salinas and Pavia University datasets. Therefore, a maximum threshold of 500was selected for the three datasets in our experiments. This threshold value is also selected in the other compared methods based on using 2D-DCT, including 2DCT_SVM, S2DCT-DCT_SVM, and SDCT-2DCT_SVM. The parameter settings of the CDCT-WF_SVM framework are explained in the next subsection.

5.1.2. CDCT-WF_SVM Framework Parameter Estimation

The spectral filter parameter in this approach is the count of the retained coefficients after performing the spectral DCT, estimated in the same way as in the first approach. Thus, this parameter was set by comparing the classification accuracy OA with varying number of DCT coefficients where the count of the retained coefficients corresponds to the most accurate classification OA for each dataset. Figure 9 illustrates the classification accuracies in terms of OA for the CDCT-WF_SVM framework with a different number of DCT coefficients on Indian Pines, Salinas, and Pavia University datasets.
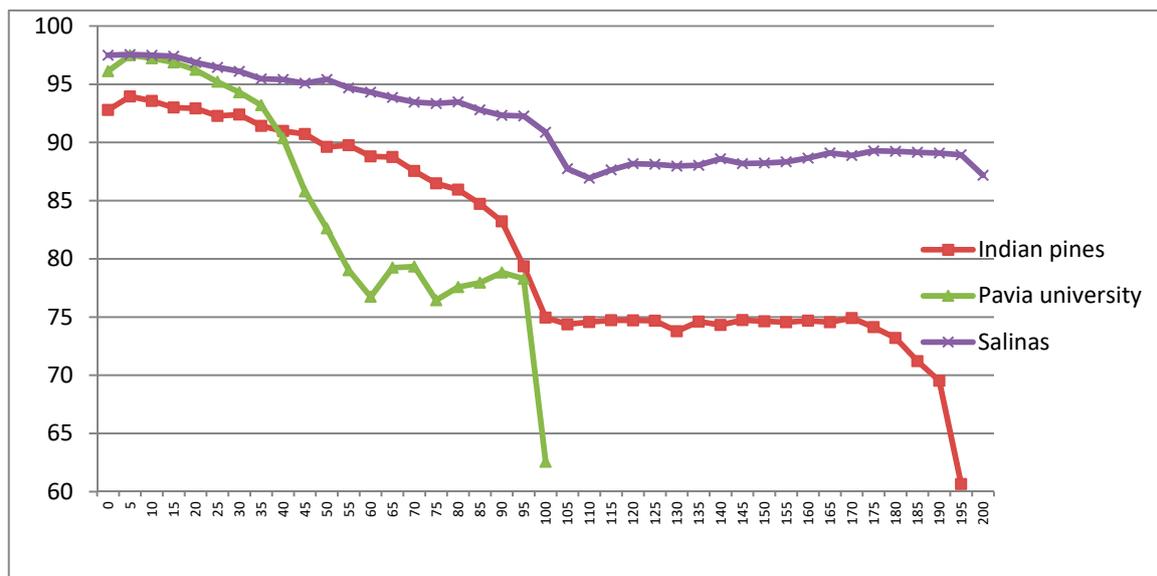

Figure 9. Overall Accuracy (OA%) with respect to the number of the first retained DCT coefficients in CDCT-WF_SVM on Indian Pines, Salinas, and Pavia University datasets.

As shown in Figure 9, for the CDCT-WF_SVMapproach, the classification accuracy OA was maximal when incorporating the five first coefficients for the three datasets. Also, the OA decreased due to the influence of noise present in the high-frequency components when selecting more than 15 coefficients for the three datasets. In our experiments, therefore, only five DCT coefficients were retained for the three dataset.

Wiener filter patch size represents the spatial parameter in the CSDCT-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 10 depicts the variation of the classification accuracy OA with varying filter patch sizes, ranging from 3×3 to 63×63 pixels, on Indian Pines, Salinas, and Pavia University datasets.

As can be seen in Figure 10, classification accuracy (OA) increased with a larger patch size value for the three datasets. Thus, accuracies tend to become stable at around a 39×39 patch size for the Indian Pines dataset, at around 43×43 patch size for Salina dataset, and a 31×31 patch size for Pavia University dataset. Therefore, these patch sizes were selected for our experiments corresponding to each dataset.

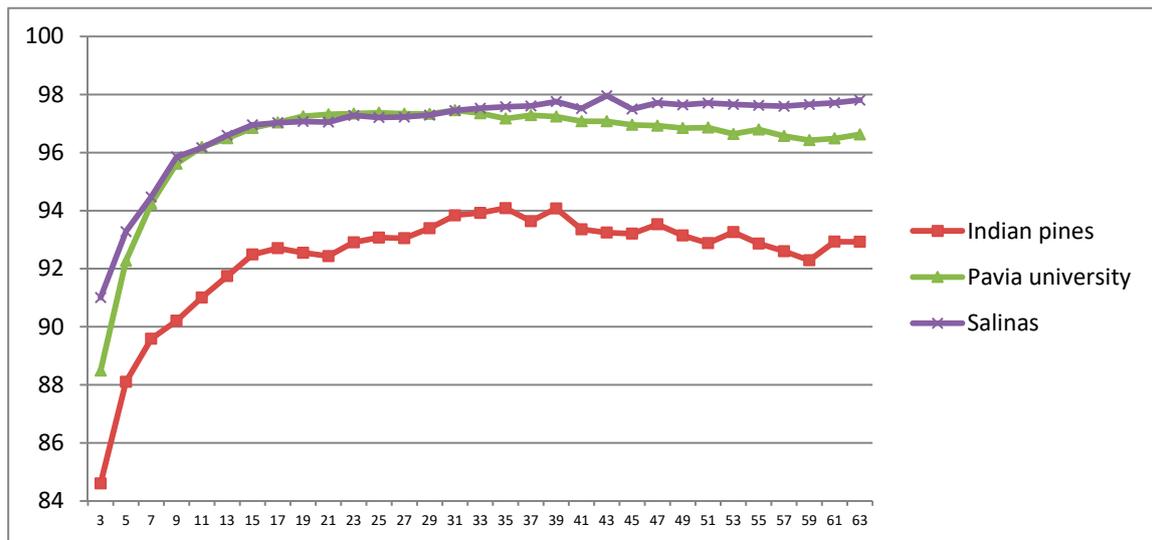The summary of the important parameters involved in our experiments on different datasets is given in Table 5.

Figure 10. Overall Accuracy (OA%)with respect to the Wiener filter patch size in CDCT-WF_SVM on Indian Pines, Salinas, and Pavia University datasets.

Table 5. Summary of the important parameters on different datasets.

Methods	Parameters	Indian Pines	Salinas	Pavia University
CDCT-2DCT_SVM	- The count of the retained Spectral DCT coefficients - 2D-DCT threshold	10 500	10 500	10 500
CDCT-WF_SVM	- The count of the retained Spectral DCT coefficients - Wiener filter patch size	5 39×39	5 43×43	5 31×31
PCA_SVM	The count of PCs	18	20	45
All methods	The training samples size per class	100	100	300

5.2. Classification of Indian Pines Dataset

In this subsection, we compare the classification effectiveness of the proposed frameworks CDCT-2DCT_SVM and CDCT-WF_SVM with the other classification methods on the Indian Pines dataset. The first experiment assessed the validity of our proposed approaches relative to the other methods. The results reported in Table 6 represent the averages of the individual classification accuracies (%), κ statistic, standard deviation, AA, OA, and computational time in seconds. Classification maps for all the compared approaches are shown in Figure 11.

Table 6. Classification results (%)on the Indian Pines dataset with standard deviation (in bracket).

| Class	SVM	PCA_SVM	DCT_SVM	2DCT_SVM	SDCT-2DCT_SVM	S2DCT-DCT_SVM	3D_SVM	3DG_SVM	EPF	IFRF	BM4D_SVM	PARAFAC_SVM	CDCT-WF_SVM	CDCT-2DCT_SVM
1	88.70	78.26	86.96	90.00	94.35	84.78	98.68	**98.70**	100.00	71.70	91.74	93.48	**98.70**	96.96
2	72.71	56.44	71.50	63.83	77.51	57.58	76.42	81.31	85.71	74.25	83.25	77.64	**92.12**	90.16
3	70.11	59.40	66.55	68.16	81.90	54.77	81.59	90.92	90.79	49.74	80.78	73.90	94.44	**95.00**
4	80.95	75.26	86.57	90.80	97.15	80.00	98.32	**99.64**	65.07	63.15	91.68	96.50	97.88	98.83
5	92.69	88.98	86.74	92.66	88.12	81.67	96.89	97.10	96.01	80.51	95.48	92.87	**98.09**	97.42
6	95.22	91.70	87.54	95.25	93.40	84.44	97.22	**99.56**	99.46	89.60	96.60	97.08	99.29	99.24
7	92.14	78.57	89.29	97.86	95.00	88.57	95.00	97.14	100.00	0.00	87.86	87.86	96.43	92.14
8	96.75	95.79	96.35	98.84	98.57	95.08	100.00	100.00	100.00	100.00	98.99	99.68	**100.00**	99.79
9	78.00	66.00	80.00	84.00	89.00	72.00	100.00	97.00	100.00	0.00	89.00	94.00	**100.00**	99.00
10	72.99	55.09	68.14	60.56	71.36	48.86	80.91	89.91	71.34	61.48	82.26	78.41	**90.92**	90.01
11	56.24	46.82	48.91	46.37	53.93	34.00	75.23	83.72	91.69	**94.19**	68.74	70.28	90.64	82.70
12	72.72	51.89	70.65	59.47	77.87	44.99	91.03	**96.41**	66.39	42.11	83.87	80.34	94.02	95.72
13	98.76	96.86	97.05	99.52	99.05	97.33	99.14	99.71	100.00	81.76	99.14	99.33	99.71	99.43
14	85.95	86.47	83.58	90.64	88.88	78.64	95.72	96.21	99.24	99.24	92.46	91.79	98.89	**99.30**
15	71.47	60.24	68.32	86.71	92.90	71.57	95.10	**99.93**	78.47	79.09	88.43	92.69	98.15	98.60
16	97.39	97.61	97.17	**100.00**	99.78	99.57	98.91	92.83	93.67	97.50	96.74	99.35	96.96	98.48
κ	70.40 (0.79)	59.63 (1.24)	65.83 (1.32)	65.27 (0.74)	73.07 (0.87)	53.34 (1.62)	83.84 (0.78)	88.87 (1.18)	84.68 (1.96)	71.06 (1.66)	80.66 (0.64)	78.95 (1.03)	**93.44** (1.03)	90.81 (1.05)
OA	73.97 (0.71)	64.40 (1.13)	69.82 (1.19)	69.33 (0.71)	76.27 (0.80)	58.46 (1.63)	86.12 (0.70)	90.32 (1.04)	85.61 (1.72)	74.30 (1.46)	83.08 (0.56)	81.62 (0.92)	**94.31** (1.05)	92.01 (0.92)
AA	82.67 (0.97)	74.09 (1.41)	80.33 (1.59)	82.79 (0.89)	87.42 (1.13)	73.37 (1.36)	92.51 (0.72)	95.01 (0.70)	89.87 (1.60)	67.69 (1.80)	89.19 (0.87)	89.08 (1.00)	**96.64** (0.56)	95.80 (0.43)
Time(s)	3.10	42.72	43.37	4.12	15.95	53.24	53.41	210.16	6.80	1.48	351.37	297.39	13.30	13.91

The bold values indicate critical values.
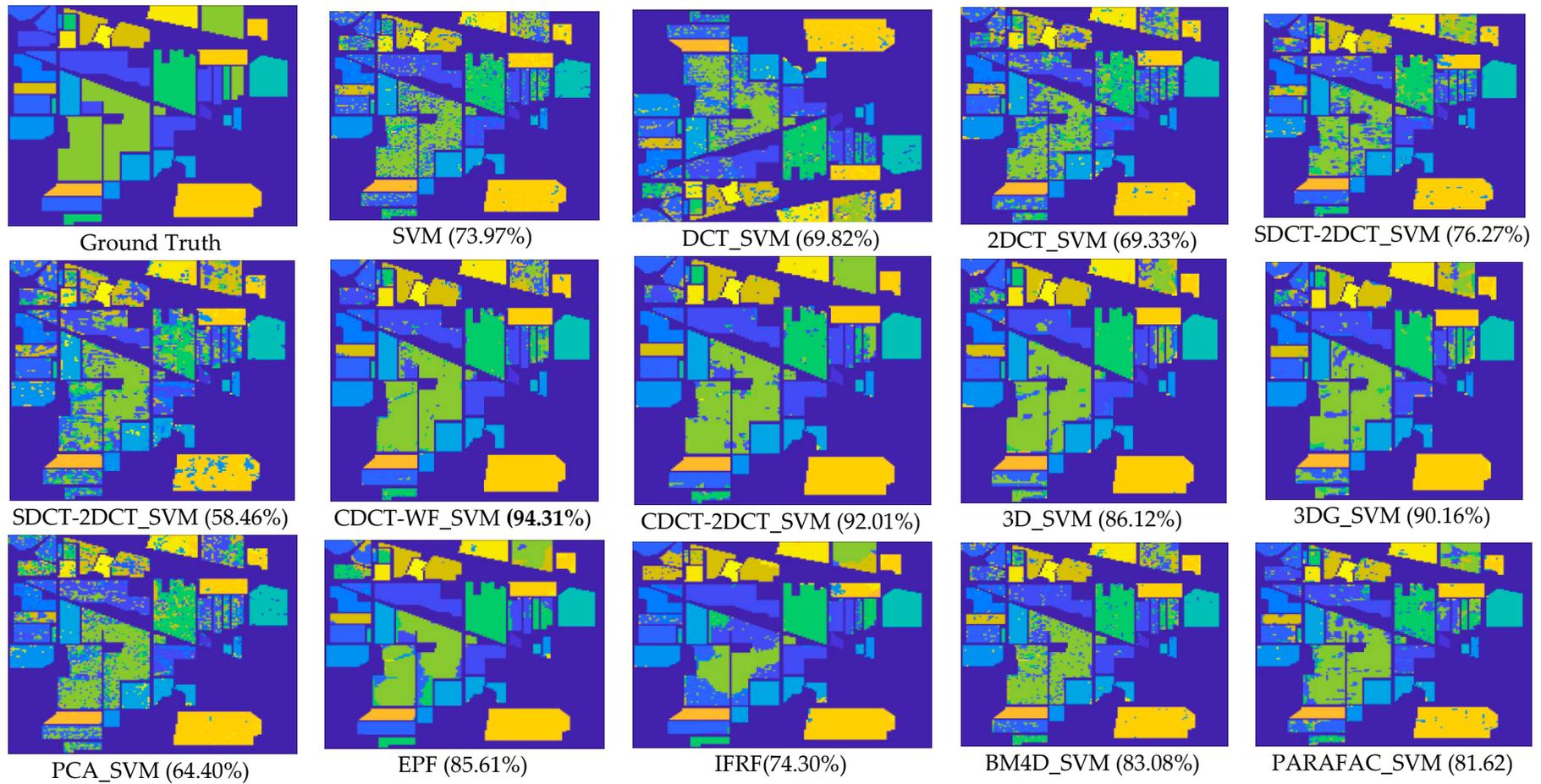
Figure 11. Classification maps created from the results of different methods on the Indian Pines dataset.

Figure 11 and Table 6 show that CDCT-WF_SVM achieved the highest values for the three criteria (AA, OA, and κ), followed by CDCT_2DCT_SVM, which had the second best performance (94.31% OA and 92.01% OA, respectively). The classification maps produced with both the CDCT-WF_SVM and CDCT-2DCT_SVM approaches are smooth and close to the ground truth map, as shown in Figure 11. Several observations can also be made:

- The approaches exploring only the spectral information, including SVM, PCA_SVM, and SDCT_SVM, attain poorer results when compared with the other techniques, which can be confirmed in Figure 11, where it is clear that the classification maps resulted from these techniques are degraded by salt-and-pepper noise. Nevertheless, the classification accuracy of SDCT_SVM (73.07% OA) is higher than PCA_SVM (64.40% OA) owing to the effectiveness ofDCT energy compaction that concentrates the energy of the spectral signature in a few low-frequency components where noisy data is embedded in the high-frequency components; instead of selecting the first PCs in PCA which cannot guarantee that image quality decreases for PCs with lower variances. The 2D-DCT_SVM approach only filters spatial information and therefore yielded less accurate results.

- The methods considering both the spectral and spatial information, including 3DG_SVM, 3D_SVM, EPF, IFRF, BM4D_SVM, and PARAFAC_SVM, are most accurate. The 3DG_SVM, and_3DSV methods delivered the third and fourth highest accuracies (90.32% OA and 86.12% OA, respectively) given the efficiency of wavelet features in structural filtering. Nevertheless, by considering hyperspectral image data as a 3D cube where the spatial and spectral features must be treated as a whole in 3D-DWT approaches, we implicitly assume that the noise variance is the same in the three dimensions and ignores dissimilarity between the spatial and the spectral dimensions where the degree of irregularity is higher in the spectral dimension than in the spatial dimensions. The proposed CDCT-WF_SVM and CDCT-2DCT_SVM approaches overcome this issue by filtering the spectral dimension and the spatial dimensions separately to achieve higher performance than wavelet-based filtering approaches. Moreover, the highest accuracy (94.31% OA) was achieved by CDCT-WF_SVM, which combines two different filters and take advantage of both of them. Additionally, it can be seen in Figure 11 that 3D_SVM, 3DG_SVM, EPF, and IFRF achieved smoother classification maps than the two well-known denoising-based approaches, BM4D_SVM and PARAFAC_SVM.

- Serial filtering-based approaches including SDCT-2DCT_SVM and S2DCT-DCT_SVM obtained by performing spectral and spatial filters on the same information cannot be effective in improving the classification accuracy, since performing a second filter on the already filtered information will alter the useful information rather than discarding more noise. In contrast, the proposed CDCT-WF_SVM and CDCT-2DCT_SVM perform a spatial filter on the noisy part from the spectral filter.

- Regarding the computational cost, Table 6 shows that the shortest classification time was 1.48 s and was achieved with the IFRF method, which had low classification accuracy (74.30% OA). The other spectral-spatial-based methods including EPF (85.61%OA), 3D_SVM (86.12%OA), 3DG_VM ((90.32%OA) achieved higher classification accuracies, but they are time-consuming using 85.61s, 86.12s, and 210.16s, respectively. Similarly, denoising-based techniques are also time-consuming with 83.08s for BM4D_SVM and 81.62 for PARAFAC_SVM. However, the proposed CDCT-WF_SVM and CDCT-2DCT_SVM achieved the two first highest classification accuracies within a short execution time. For example, CDCT-WF_SVM achieved an OA of 94.31% in 13.30s where CDCT-2DCT_SVM achieves an OA of 92.01% in 13.91s. Moreover, in the proposed approaches, the SDCT is performed on each pixel, and the spatial filter is performed on each band. Hence, our approaches are easily parallelized, which could further reduce the computational time.

The performance of the classifiers is greatly affected by the number of training samples. Therefore, in our second experiment, the performance of the proposed frameworks 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.
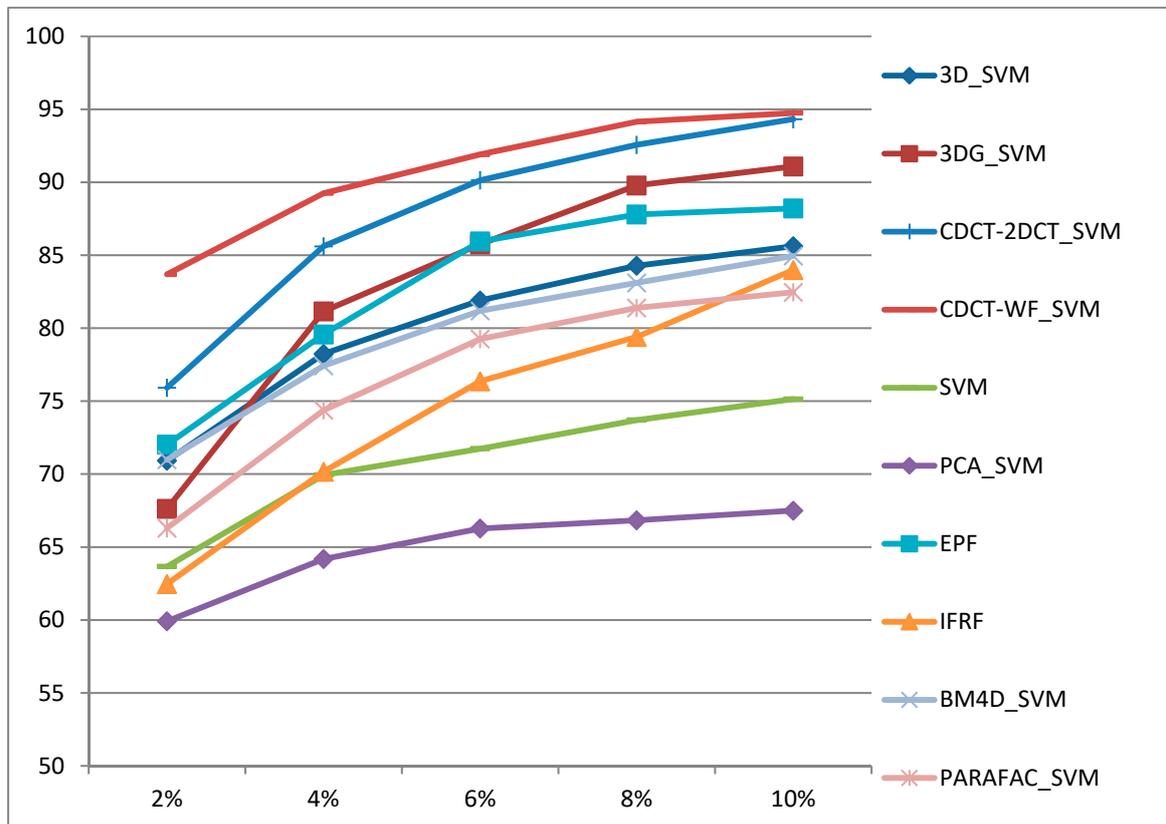

Figure 12. Overall Accuracy (OA%) of the compared methods with different training samples proportions on Indian Pine dataset.

Figure 12 illustrates that the proposed frameworks can significantly improve classification results when varying the training samples size, as the size of training samples changes from 2% to 10%, our proposed approaches (CDCT-WF_SVM and CDCT-2DCT_SVM) always achieved the first and the second highest OA. Their advantages are apparent from the small size of the training samples. The proposed CDCT-WF_SVM has an 11.67–23.78% gain in accuracy over the other tested compared methods with a 2% training sample per class. This indicates that a large number of unlabeled samples can be discerned with a small labeled samples size, which further demonstrates the robustness of the proposed approaches.

5.3. Classification of SALINAS Dataset

In the first experiment, we compared our proposed approaches and other methods on the Salinas dataset. Table 7 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 Salinas scene are illustrated in Figure 13.

Table 7. Classification results (%) on the Salinas dataset with standard deviation (in brackets).

| Class	SVM	PCA_SVM	DCT_SVM	2DCT_SVM	SDCT-2DCT_SVM	S2DCT-DCT_SVM	3D_SVM	3DG_SVM	EPF	IFRF	BM4D_SVM	PARAFAC_SVM	CDCT-WF_SVM	CDCT-2DCT_SVM
1	99.13	99.36	97.53	99.64	97.02	96.89	99.19	99.70	**100.00**	98.99	99.25	97.33	99.52	99.50
2	99.51	99.52	98.83	99.13	96.64	97.43	99.35	99.68	99.97	**100.00**	99.63	96.05	99.78	99.59
3	99.58	99.09	99.47	94.88	91.92	90.36	97.48	98.95	96.07	**99.84**	99.77	96.08	98.97	99.39
4	99.34	99.35	99.18	98.98	98.56	97.98	99.36	99.43	98.43	90.05	99.30	96.02	99.23	**99.51**
5	98.74	97.78	98.03	96.57	94.84	94.17	98.98	99.46	99.80	**99.98**	98.55	86.14	98.36	98.25
6	99.68	99.67	99.68	99.71	99.74	99.74	99.86	99.95	99.99	**100.00**	99.67	99.44	99.62	99.65
7	99.57	99.56	99.45	99.65	99.02	97.31	99.66	99.67	**99.93**	99.79	99.55	98.58	99.65	99.64
8	67.98	72.40	74.40	70.87	63.49	59.75	86.27	88.67	85.40	**99.53**	78.00	77.13	94.85	91.46
9	99.00	97.50	98.09	98.60	96.47	96.86	98.43	99.18	98.73	**99.98**	99.47	88.52	99.46	99.41
10	95.01	94.51	93.17	93.54	90.19	89.74	93.79	95.60	93.60	**99.72**	96.36	73.28	97.76	95.15
11	98.90	98.75	97.23	97.22	94.29	94.63	99.68	**99.93**	97.58	99.02	98.97	85.23	98.80	99.52
12	99.82	99.70	99.57	97.62	95.64	89.59	**99.97**	99.87	99.67	98.82	99.84	82.19	99.93	99.93
13	99.58	99.55	98.13	99.17	98.03	95.69	99.02	98.97	99.83	97.70	99.72	89.34	99.46	**99.90**
14	97.76	97.92	95.59	97.70	98.43	96.55	98.01	98.19	97.89	97.58	98.43	93.89	98.12	**99.04**
15	70.04	70.51	69.71	70.55	65.84	64.22	75.92	85.52	77.68	83.06	79.72	81.56	**96.19**	93.62
16	98.86	98.88	98.51	98.71	97.87	96.07	98.83	98.70	99.71	**100.00**	98.48	95.72	98.87	99.21
κ	87.04 (0.62)	87.86 (1.07)	87.96 (0.86)	87.20 (0.73)	83.52 (0.53)	81.86 (0.95)	92.05 (0.41)	94.44 (0.89)	92.13 (0.90)	96.28 (0.27)	91.04 (0.69)	85.49 (0.47)	**97.62** (0.33)	96.30 (0.80)
OA	88.36 (0.57)	89.10 (0.97)	89.20 (0.78)	88.51 (0.65)	85.18 (0.48)	83.69 (0.86)	92.88 (0.37)	95.02 (0.79)	92.95 (0.81)	96.46 (0.25)	91.96 (0.62)	86.94 (0.73)	**97.86** (0.30)	96.68 (0.72)
AA	95.16 (0.20)	95.25 (0.41)	94.78 (0.34)	94.53 (0.36)	92.37 (0.23)	91.06 (0.45)	96.49 (0.15)	97.59 (0.36)	96.52 (0.31)	97.65 (0.29)	96.54 (0.21)	89.78 (0.39)	**98.66** (0.16)	98.30 (0.32)
Time(s)	7.24	13.18	64.46	12.81	90.42	102.00	183.59	1142.52	19.29	4.77	1854.07	1634.3	61.22	70.56

The bold values indicate critical values.
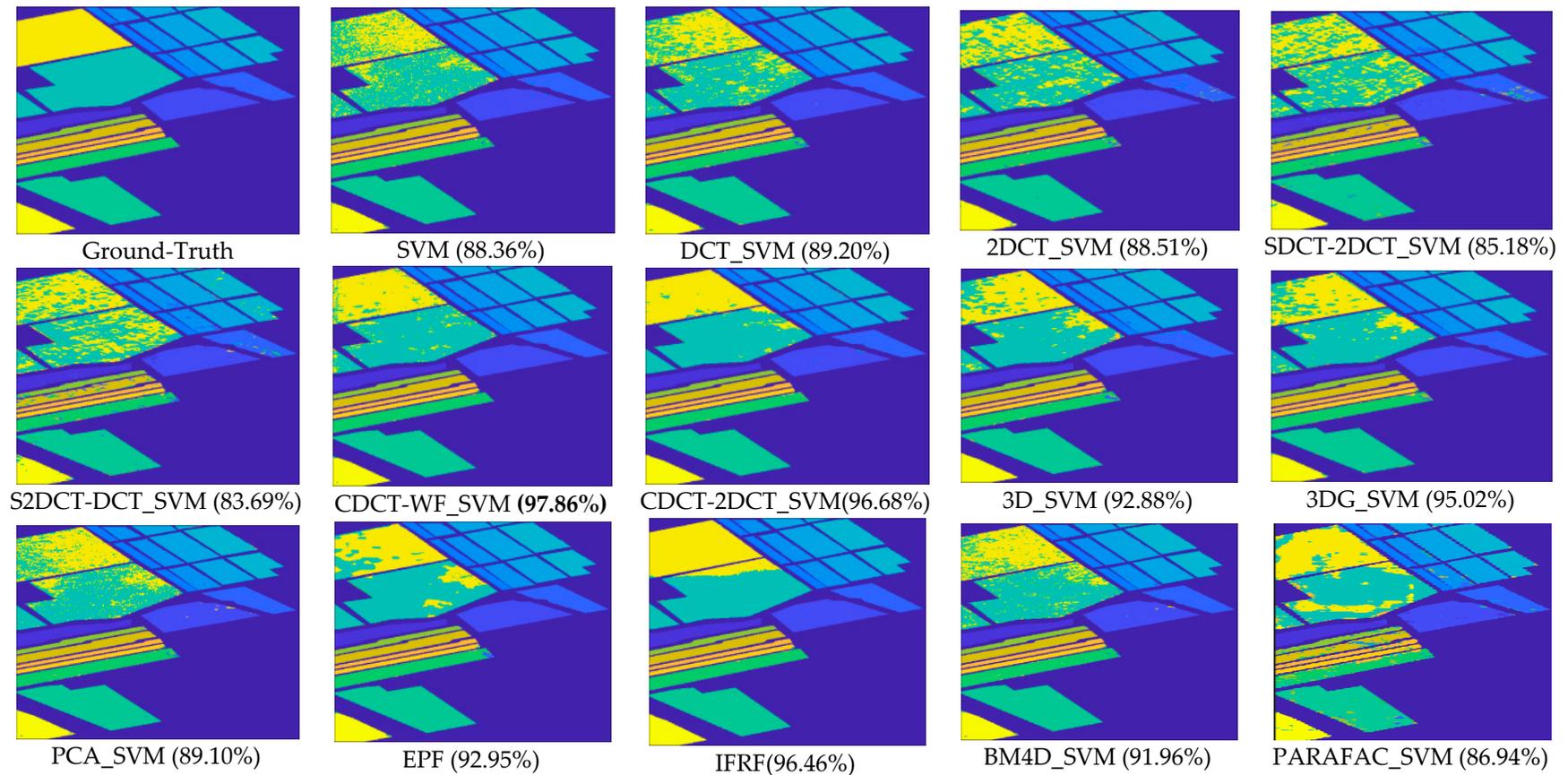
Figure 13. Classification maps created from the results of different methods on the Salinas dataset.

From Table 7, we can see that our proposed CDCT-WF_SVM approach provided the highest classification accuracy in terms of all the three criteria (AA, OA, and kappa), with stable performance (lower standard deviation). Hence, it achieved the highest classification accuracy in terms of OA (97.86%). Our proposed CDCT-2DCT_SVM method provided the second highest classification with an OA of 96.68%. In addition, the classification maps produced from both the CDCT-WF_SVM and CDCT-2DCT_SVM approaches are smooth and close to the ground truth map, as shown in Figure 13. From Table 7, it is clear that IFRF (96.46%OA) produced more accurate classification results than 3DG_SVM (83.34%OA) in this scenario, with much smoother classification map. The EPF, 3D_SVM and BM4D_SVM methods achieved comparable classification OA values (92.95%, 92.88% and 91.96%, respectively), resulting in similar classification maps, as shown in Figure 13. The PARAFAC_SVM method delivers lower classification accuracy (86.94%OA) than the pixel-wise classifiers including DCT_SVM, PCA_SVM, 2DCT_SVM, and SVM. Also, the two approaches based on serial filtering, SDCT-2DCT_SVM and S2DCT-DCT_SVM yielded low classification OA values (88.51%, and 83.69%, respectively).This is likely because the second filter altered the meaningful information that had been filtered by the first filter already.

Table 7 shows that classifiers with spatial information cost more time than their pixel-wise counterparts. Meanwhile, the IFRF, which is a spectral-spatial feature extraction-based method, achieves the shortest time of 4.77 s. Our proposed CDCT-WF_SVM approach achieved the third least execution time with 61.22 s among the spectral-spatial classifiers after IFRF and EPF (19.02 s) methods. From Table 7, we can also conclude that wavelet-based and denoising approaches are time-consuming. For example, 3DG_SVM requires 1142.52 s to achieve an OA of 95.02% where our proposed CDCT-WF_SVM approach requires only 65.22 s to achieve an OA of 97.86%, and the proposed CDCT-2DCT_SVM requires 70.56 s to provide an OA of 96.68%.

In the second experiment, the robustness of the proposed approaches regarding the state-of-the-art methods with different training conditions was evaluated, in which 2%, 4%, 6%, 8%, and 10% labeled samples of each class were randomly selected as training samples, while the rest are used as testing samples. Thus, for stability, each experiment was executed 20 times, and the averages result in terms of the OA values are plotted in Figure 14.

Figure 14 shows that the proposed frameworks can always improve the classification accuracy with a different number of training samples, as the number of training samples varies from 2% to 10%, our proposed approaches (CDCT-2DCT_SVM and CDCT-WF_SVM) always achieved the first and the second highest OA. Additionally, it should be noted that the IFRF method had the third-highest accuracy when varying the number of training samples and outperforms wavelet-based approaches (3DG_SVM and 3D_SVM).
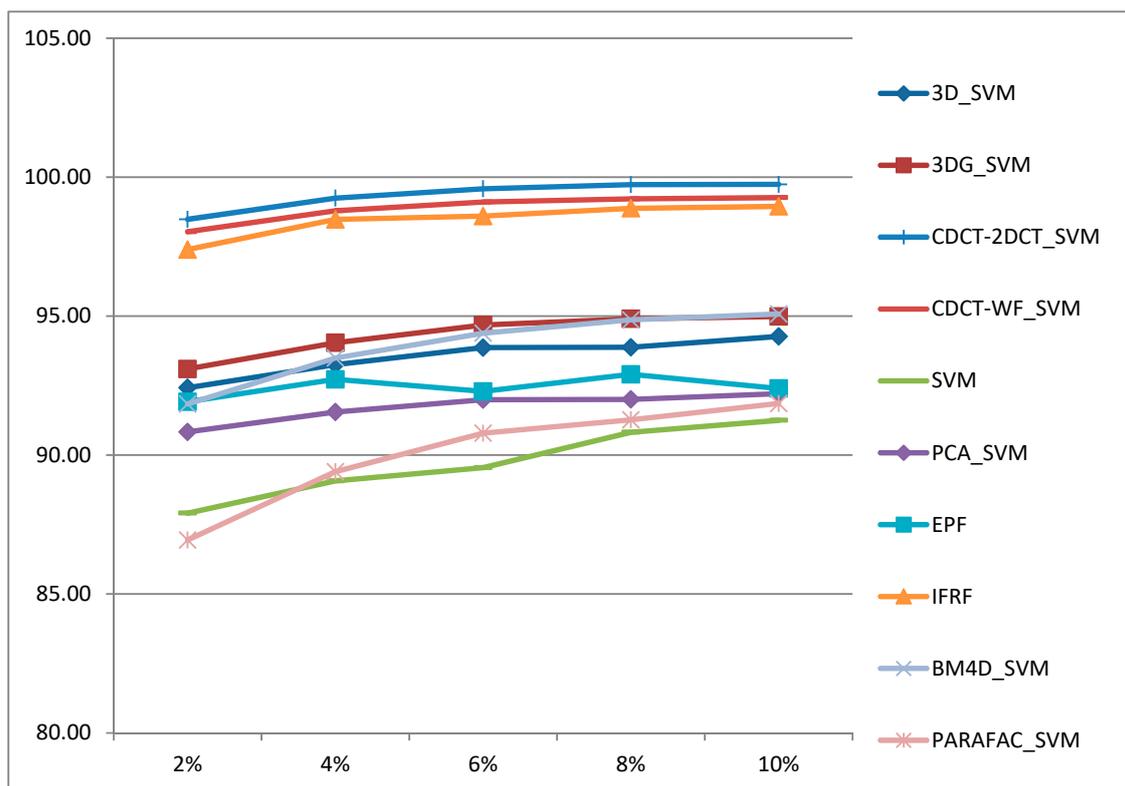
Figure 14. Overall Accuracy (OA%) of all methods with different training samples proportions on Salinas dataset.

5.4. Classification of Pavia University Dataset

In the first experiment, we compare the classification effectiveness of the proposed frameworks with the other classification methods on Pavia University dataset. Table 8 reports the quantitative classification results. Details of the classification maps of the compared approaches are shown in Figure 15.

As can be seen in Table 8, the highest classification accuracies are always obtained by the spectral-spatial classification techniques, especially the structural filtering-based approaches. The proposed method CDCT-WF_SVM achieved the highest accuracy. For example, the OA obtained by this approach was around 98.71%, which is 17.53% higher than OA obtained by the pixel-wise SVM classifier.

In this scene, 3DG_SVM provides competitive performance, which is 98.39% in terms of OA followed by our proposed CDCT-2DCT_SVM approach with an OA of 97.40% and the 3D_SVM method with an OA of 96.95%. Although the structural filtering approaches have comparable accuracy, the processing time of our proposed frameworks is faster than wavelet-based methods. For example, the proposed CDCT-WF_SVM achieve an OA of 98.71% in 124.93 s, where 3DG_SVM achieved an OA of 98.39% in 2142.40 s, which is 17 times slower than our proposed CDCT-WF_SVM approach. Another observation is that BM4D_SVM and EPF methods achieve comparable classification OA (87.23% and 86.81%, respectively). Meanwhile, PARAFAC_SVM and IFRF were the least accurate among the spectral-spatial classifiers, with an OA of 83.99% and 83.14%, respectively. The pixel-wise classifiers including SVM, PCA_SVM, and DCT_SVM methods produced the least accurate classification results. For example, the OA values were 81,18%, 80,49%, and 78.51%, respectively. Similar to the experiments with other datasets, serial filtering-based approaches yielded the least accurate classification results.

From Figure 15, we can see visually that combining spectral and spatial information produce less noisy and smoother results. In particular, structural filtering-based approaches have smoother classification maps than the other methods. The results obtained from our proposed CDCT-WF_SVM approach closely match the ground-truth map.

Table 8. Results (%) on the Pavia University dataset with standard deviation (in brackets).

Class	SVM	PCA_SVM	DCT_SVM	2DCT_SVM	SDCT-2DCT_SVM	S2DCT-DCT_SVM	3D_SVM	3DG_SVM	EPF	IFRF	BM4D_SVM	PARAFAC_SVM	CDCT-WF_SVM	CDCT-2DCT_SVM
1	71.71	72.13	74.59	58.67	54.54	65.52	96.89	**98.88**	98.68	76.18	80.07	77.90	97.96	96.21
2	82.91	82.79	79.16	65.90	60.75	67.14	97.64	98.91	97.20	97.90	89.49	85.37	**99.66**	98.02
3	78.89	78.23	81.89	61.86	54.57	62.01	88.15	90.08	92.51	52.39	82.61	70.81	**98.87**	97.23
4	91.89	91.00	86.28	87.64	83.49	80.34	99.01	**99.42**	67.81	84.99	94.15	93.85	95.92	97.38
5	99.70	99.70	99.75	99.89	99.92	99.88	100.00	**100.00**	99.89	98.85	99.66	99.78	99.78	99.77
6	84.39	77.94	68.45	63.63	56.13	55.24	96.68	99.16	64.31	92.74	89.76	84.83	**99.39**	98.16
7	76.96	77.57	84.98	60.96	57.36	71.27	98.91	**99.53**	77.25	64.22	85.59	77.89	99.49	98.63
8	69.44	71.07	75.54	53.59	47.35	59.56	94.64	96.08	84.40	53.24	75.88	74.30	**97.62**	93.75
9	99.91	99.86	99.94	99.72	99.77	99.86	99.63	99.72	96.17	43.63	99.88	99.92	**99.88**	99.91
κ	75.24 (1.43)	74.25 (2.16)	71.74 (1.15)	56.52 (1.95)	50.32 (1.39)	57.58 (1.36)	95.88 (0.30)	97.82 (0.12)	82.79 (1.03)	77.63 (0.25)	83.00 (1.40)	80.91 (0.47)	**97.87** (0.26)	96.48 (0.42)
OA	81.18 (1.15)	80.49 (1.75)	78.51 (1.14)	66.07 (1.56)	60.95 (2.00)	67.01 (1.82)	96.95 (0.22)	98.39 (0.09)	86.81 (1.68)	83.14 (0.19)	87.23 (1.10)	83.99 (0.40)	**98.71** (0.19)	97.40 (0.32)
AA	83.98 (0.69)	83.37 (1.68)	83.40 (1.73)	72.43 (1.47)	68.21 (1.64)	73.42 (1.52)	96.84 (0.14)	97.98 (0.12)	86.47 (1.01)	73.79 (0.17)	88.57 (0.79)	84.96 (0.40)	**98.73** (0.19)	97.67 (0.18)
Time(s)	80.05	78.68	186.10	116.79	252.69	216.92	485.92	2142.40	52.84	18.14	1535.13	1275.06	124.93	142.74

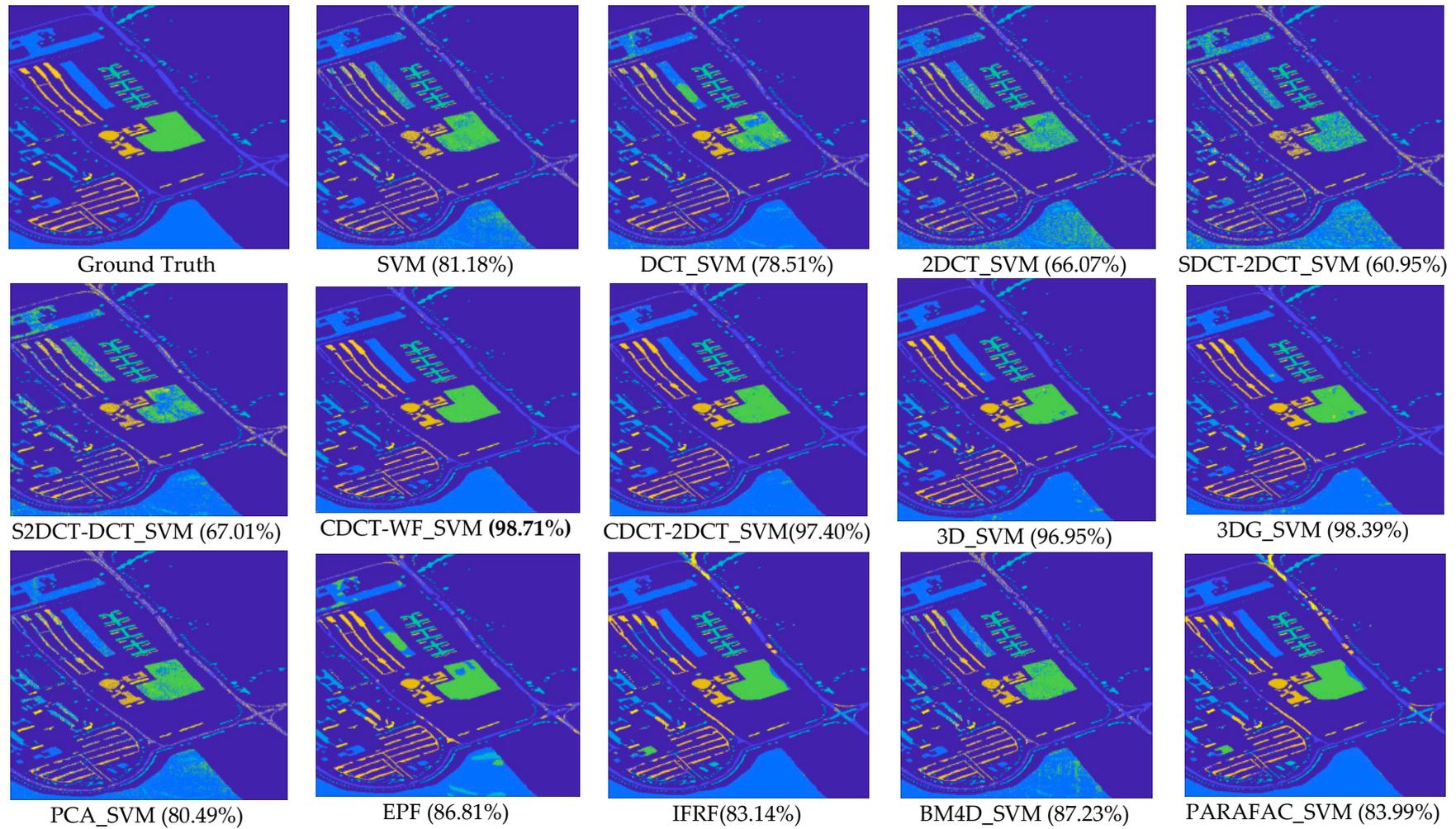The bold values indicate critical values.

Figure 15. Classification maps created from the results of different methods on thePavia University dataset.

In the second experiment, we evaluated the performance of the proposed frameworks on Pavia University image against state-of-the-art methods with different training conditions, in which 2%, 4%, 6%, 8%, and 10% labeled samples of each class are randomly selected as training samples, and the rest are used for testing. The averages results in terms of OA values for each compared method are plotted in Figure 16.
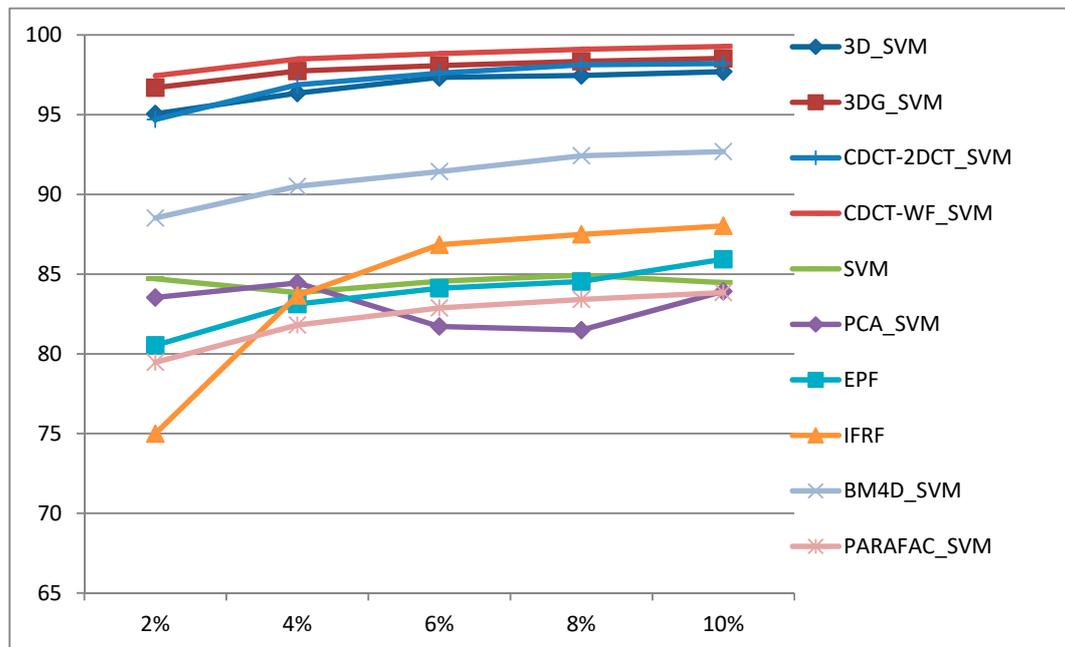

Figure 16. Overall Accuracy (OA%) of all methods with different training samples proportions on the Pavia University dataset.

Figure 16 illustrates that the proposed frameworks can always improve the classification accuracy significantly with a different number of training samples, as the size of training samples changes from 2% to 10%, the proposed approach (CDCT-WF_SVM) always achieved the first highest OA followed by 3DG_SVM, CDCT-2DCT_SVM, and 3D_SVM, which have a comparable performance when varying the size of training samples. Meanwhile, it should be noted that BM4D_SVM provided higher accuracy than EPF and IFRF methods when varying the training samples size.

5.5. Classification in Noisy Scenario

In the noisy scenario experiment, we evaluated the robustness of our proposed approaches on the Indian Pines dataset with all 224 spectral bands without removing the 24 noisy bands as done in the former experiment. This noisy Indian pines dataset is publicly available at http://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes. As in the experiment on the clean Indian Pines dataset discussed in Section 5.2, 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 classification accuracy OA of all the state-of-the-art methods with the clean Indian Pines dataset (with only 200 bands) and the noisy Indian Pines dataset (224 bands including the 24 noisy bands) are compared in Figure 17.
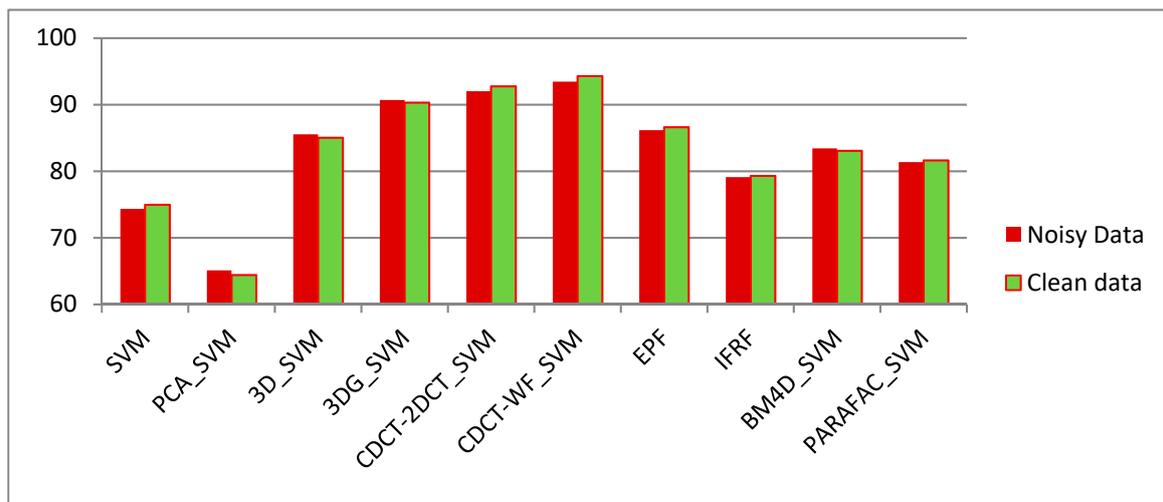
Figure 17. Overall Accuracy (%) of all competing methods on clean and noisy Indian Pines dataset.

From Figure 17, although the classification accuracy of all the compared methods on the noisy scenario was lower than on the clean dataset, nevertheless, the proposed CDCT-WF_SVM and CDCT-2DCT_SVM still provided a higher OA than the other methods in this noisy scenario, which further shows the robustness of our proposed approaches in the noisy scenario relative to the other tested methods. Another observation is that denoising-based approaches including BM4D_SVM and PARAFAC_SVM methods, achieved comparable classification OA on both clean and noisy scenarios. Meanwhile, PCA_SVM was the least accurate among all compared methods. This is because PCA is a global transformation and neither maintains the local spectral signatures nor considers noise. Consequently, it might not preserve all the required useful information for providing effective and reliable classification [71].

6. Summary and Conclusions

In this paper, we present two effective classification frameworks based on cascaded spectral and spatial filtering where the spatial filter is performed in the DCT domain. In our proposed frameworks, we perform a spectral DCT to separate the most significant information presented in a few low-frequency components from the noisy data embedded in high-frequency components. Spatial filtering exploiting both 2D-DCT and WF are performed on the high DCT frequency components to extract the rest of useful spatial information. After applying an inverse DCT on all components, including the filtered components, the linear SVM classifier was applied to the preprocessed data. Various performance metrics were used to support our experimental studies including the classification accuracy, computational time, the classification maps, and the sensitivity tests for different sizes of the training samples. Experimental results on three real hyperspectral datasets show that our approaches outperform the state-of-the-art methods investigated in this paper. The most significant results of this work are listed below:

- The proposed approaches outperform the other considered methods; in particular, our proposed CDCT-WF_SVM method, which delivers higher accuracy on the three datasets with a smoother classification map than the other tested techniques.
- The proposed approaches deliver higher classification accuracies regardless of the size of training samples, and they are robust to noise.
- A major advantage of the proposed frameworks is that they are computationally efficient along with a reasonable tradeoff between accuracy and computational time. Thus, they will be quite useful for applications such as flood monitoring and risk management, which require a fast response. Moreover, as the DCT is performed on each pixel, and both WF with 2D-DCT are

- performed on each band, our approaches are easily parallelized, which could further reduce the computational time.
- The results obtained illustrate that the proposed approaches can deal with different spatial resolutions (20 m, 3.7 m, and 1.7 m). In particular, for Indian pines dataset with 20m spatial resolution, our proposed CDCT-WF_SVM and CDCT-2DCT_SVM approaches achieve first and the second highest accuracy. Thus, our proposed approaches can effectively deal with low spatial resolution images.
- For the three datasets, structural-based filtering methods have stable performance, including our proposed frameworks and wavelet-based methods. However, the IFRF method cannot provide stable performance by providing the third-highest accuracy on Salinas dataset and low accuracy on the two other datasets.
- The proposed approaches require the selection of only two parameters to achieve high classification accuracy in low computational time, which allows their potential use in practical applications.

Author Contributions: All authors have made great contributions to the work. Conceptualization, R.B. and K.B.; Software, R.B. and K.B.; validation, W.H., 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.

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, A.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. Melgani, F.; Bruzzone, L. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Trans. Geosci. Remote Sens.* **2004**, *42*, 1778–1790. [CrossRef]
6. 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]
7. 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]
8. 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]
9. 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]
10. 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*, 4263–4281. [CrossRef]
11. 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]
12. 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]

13. 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]

14. 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]

15. 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]

16. 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]

17. Plaza, A.; Martinez, P.; Plaza, J.; Perez, R. Dimensionality reduction and classification of hyperspectral image data using sequences of extended morphological transformations. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 466. [CrossRef]

18. Benediktsson, J.A.; Palmason, J.A.; Sveinsson, J.R. Classification of Hyperspectral Data from Urban Areas Based on Extended Morphological Profiles. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 480. [CrossRef]

19. Xia, J.; Chanussot, J.; Du, P.; He, X. Spectral-spatial classification for hyperspectral data using rotation forests with local feature extraction and Markov random fields. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 2532–2546. [CrossRef]

20. Khodadadzadeh, M.; Li, J.; Plaza, A.; Ghassemian, H.; Bioucas-Dias, J.M.; Li, X. Spectral–Spatial Classification of Hyperspectral Data Using Local and Global Probabilities for Mixed Pixel Characterization. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 6298. [CrossRef]

21. Peng, J.; Zhou, Y.; Chen, C.L.P. Region-kernel-based support vector machines for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 4810–4824. [CrossRef]

22. Chen, Y.; Nasrabadi, N.M.; Tran, T.D. Hyperspectral Image Classification Using Dictionary-Based Sparse Representation. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 3973. [CrossRef]

23. Lu, Z.; He, J. Spectral-spatial hyperspectral image classification with adaptive mean filter and jump regression detection. *Electron. Lett.* **2015**, *51*, 1658–1660. [CrossRef]

24. Golipour, M.; Ghassemian, H.; Mirzapour, F. Integrating Hierarchical Segmentation Maps with MRF Prior for Classification of Hyperspectral Images in a Bayesian Framework. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 805. [CrossRef]

25. Oktem, R.; Ponomarenko, N.N. Image filtering based on discrete cosine transform. *Telecommun. Radio Eng.* **2007**, *66*. [CrossRef]

26. 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.

27. 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]

28. Li, W.; Prasad, S.; Fowler, J.E. Hyperspectral image classification using Gaussian mixture models and Markov random fields. *IEEE Geosci. Remote Sens. Lett.* **2014**, *11*, 153–157. [CrossRef]

29. 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]

30. Gu, Y.; Liu, T.; Jia, X.; Benediktsson, J.A.; Chanussot, J. Nonlinear multiple kernel learning with multiple-structure-element extended morphological profiles for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 3235–3247. [CrossRef]

31. Pan, B.; Shi, Z.; Xu, X. R-VCANet: a new deep-learning-based hyperspectral image classification method. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 1975–1986. [CrossRef]

32. Pan, B.; Shi, Z.; Xu, X. MugNet: deep learning for hyperspectral image classification using limited samples. *ISPRS J. Photogramm. Remote Sens.* **2017**. [CrossRef]

33. Pan, B.; Shi, Z.; Xu, X. Hierarchical guidance filtering-based ensemble classification for hyperspectral images. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 4177–4189. [CrossRef]

34. Ni, D.; Ma, H. Hyperspectral image classification via sparse code histogram. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 1843–1847.

35. Du, P.; Xue, Z.; Li, J.; Plaza, A. Learning discriminative sparse representations for hyperspectral image classification. *IEEE J. Sel. Top. Signal Process.* **2015**, *9*, 1089–1104. [CrossRef]

36. Song, H.; Wang, Y. A spectral-spatial classification of hyperspectral images based on the algebraic multigrid method and hierarchical segmentation algorithm. *Remote Sens.* **2016**, *8*, 296. [CrossRef]

37. 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]

38. 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]

39. 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]

40. 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]

41. 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]

42. 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]

43. Rajadell, O.; Garcia-Sevilla, P.; Pla, F. Spectral–Spatial Pixel Characterization Using Gabor Filters for Hyperspectral Image Classification. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 860. [CrossRef]

44. He, L.; Li, J.; Plaza, A.; Li, Y. Discriminative low-rank Gabor filtering for spectral-spatial hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 1381–1395. [CrossRef]

45. Phillips, R.D.; Blinn, C.E.; Watson, L.T.; Wynne, R.H. An Adaptive Noise-Filtering Algorithm for AVIRIS Data With Implications for Classification Accuracy. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 3168–3179. [CrossRef]

46. Bourennane, S.; Fossati, C.; Cailly, A. Improvement of classification for hyperspectral images based on tensor modeling. *IEEE Geosci. Remote Sens. Lett.* **2010**, *7*, 801–805. [CrossRef]

47. Wang, J.; Wu, Z.; Jeon, G.; Jeong, J. An efficient spatial deblocking of images with DCT compression. *Digit. Signal Process. A Rev. J.* **2015**, *42*, 80–88. [CrossRef]

48. 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]

49. Boukhechba, K.; Wu, H.; Bazine, R. DCT-Based Preprocessing Approach for ICA in Hyperspectral Data Analysis. *Sensors* **2018**, *18*, 1138. [CrossRef]

50. Jing, L.; Yi, L. Hyperspectral remote sensing images terrain classification in DCT SRDA subspace. *J. China Univ. Posts Telecommun.* **2015**, *22*, 65–71. [CrossRef]

51. Pennebaker, W.B.; Mitchell, J.L. *JPEG: Still Image Data Compression Standard*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 1992.

52. Ahmed, N.; Natarajan, T.; Rao, K.R. Discrete Cosine Transform. *Comput. IEEE Trans.* **1974**, *C-23*, 90–93. [CrossRef]

53. Fevralev, D.V.; Ponomarenko, N.N.; Lukin, V.V.; Abramov, S.K.; Egiazarian, K.O.; Astola, J.T. Efficiency analysis of DCT-based filters for color image database. In *Image Processing: Algorithms and Systems IX*; IS&T—The Society for Imaging Science and Technology: San Francisco, CA, USA, 2011; Volume 7870, p. 78700R, doi:10.1117/12.871944.

54. Roy, A.B.; Dey, D.; Banerjee, D.; Mohanty, B. Comparison of FFT, DCT, DWT, WHT Compression Techniques on Electrocardiogram and Photoplethysmography Signals. *IJCA Spec. Issue Int. Conf. Comput. Commun. Sens. Netw.* **2013**, *CCSN2012*, 6–11.

55. Clarke, R.J. *Transform Coding of Images. Astrophysics*; Academic Press, Inc.: Orlando, FL, USA, 1985.

56. Gao, L.R.; Zhang, B.; Zhang, X.; Zhang, W.J.; Tong, Q.X. A New Operational Method for Estimating Noise in Hyperspectral Images. *IEEE Geosci. Remote Sens. Lett.* **2008**, *5*, 83–87. [CrossRef]

57. Chen, G.; Qian, S.-E. Denoising of hyperspectral imagery using principal component analysis and wavelet shrinkage. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 973–980. [CrossRef]

58. Acito, N.; Diani, M.; Corsini, G. Signal-Dependent Noise Modeling and Model Parameter Estimation in Hyperspectral Images. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 2957–2971. [CrossRef]

59. Wiener, N. *The Interpolation, Extrapolation and Smoothing of Stationary Time Series*; MIT Press: New York, NY, USA, 1949.

60. Lim, J.S. *Two-dimensional Signal and Image Processing*; Prentice-Hall, Inc.: Upper Saddle River, NJ, USA, 1990; ISBN 0-13-935322-4.

61. Lee, J.-S. Digital image enhancement and noise filtering by use of local statistics. *IEEE Trans. Pattern Anal. Mach. Intell.* **1980**, *PAMI-2*, 165–168. [CrossRef]

62. 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* **2016**. [CrossRef]

63. Prasad, S.; Bruce, L.M. Limitations of principal components analysis for hyperspectral target recognition. *IEEE Geosci. Remote Sens. Lett.* **2008**, *5*, 625–629. [CrossRef]

64. 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]

65. 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]

66. Maggioni, M.; Foi, A. Nonlocal transform-domain denoising of volumetric data with groupwise adaptive variance estimation. *Comput. Imaging X* **2012**, *8296*, 82960O.

67. 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]

68. 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]

69. Bioucas-Dias, J.M.; Nascimento, J.M.P. Estimation of signal subspace on hyperspectral data. *Proc. SPIE* **2005**, *5982*, 59820L.

70. Story, M.; Congalton, R.G. Accuracy assessment: A user's perspective. *Photogramm. Eng. Remote Sens.* **1986**, *52*, 397–399.

71. Kaewpijit, S.; Le Moigne, J.; El-Ghazawi, T. Automatic Reduction of Hyperspectral Imagery using Wavelet Spectral Analysis. *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 863–871. [CrossRef]