

Effective Denoising and Classification of Hyperspectral Images using Curvelet Transform and Singular Spectrum Analysis

Tong Qiao, Jinchang Ren, Zheng Wang, Jaime Zabalza, Meijun Sun, Huimin Zhao, Shutao Li, Jón Atli Benediktsson, Qingyun Dai, and Stephen Marshall

Abstract—Hyperspectral imaging (HSI) classification has become a popular research topic in recent years, and effective feature extraction is an important step before the classification task. Traditionally, spectral feature extraction techniques are applied to the HSI data cube directly. This paper presents a novel algorithm for HSI feature extraction by exploiting the curvelet transformed domain via a relatively new spectral feature processing technique – singular spectrum analysis (SSA). Although the wavelet transform has been widely applied for HSI data analysis, the curvelet transform is employed in this paper since it is able to separate image geometric details and background noise effectively. Using the support vector machine (SVM) classifier, experimental results have shown that features extracted by SSA on curvelet coefficients have better performance in terms of classification accuracies over features extracted on wavelet coefficients. Since the proposed approach mainly relies on SSA for feature extraction on the spectral dimension, it actually belongs to the spectral feature extraction category. Therefore, the proposed method has also been compared with some state-of-the-art spectral feature extraction techniques to show its efficacy. In addition, it has been proven that the proposed method is able to remove the undesirable artefacts introduced during the data acquisition process as well. By adding an extra spatial post-processing step to the classified map achieved using the proposed approach, we have shown that the classification performance is comparable with several recent spectral-spatial classification methods.

Index Terms—Hyperspectral imaging (HSI), the curvelet transform, singular spectrum analysis (SSA), classification, support vector machine (SVM), spatial post-processing.

Manuscript received on October xx, 2015. This work is partially funded by the following projects: National Natural Science Foundation of China (61202165, 61572351), the Science and Technology Major Project of Education Department of Guangdong Province (2014KZDXM060), the Natural Science Foundation of Guangdong Province (2016A030311013), International Scientific and Technological Cooperation Projects of Education Department of Guangdong Province (2015KJHZ021).

T. Qiao, J. Ren, J. Zabalza and S. Marshall are Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow, G1 1XW, UK. (Corresponding author email: jinchang.ren@strath.ac.uk).

Z. Wang and M. Sun are respectively with School of Computer Software and School of Computers, Tianjin University, Tianjin, 300072, China.

H. Zhao and Q. Dai are with School of Electronic and Information, Guangdong Polytechnic Normal University, Guangzhou, 510665, China.

S. Li is with the College of Electrical and Information Engineering, Hunan University, Changsha, 410082, China.

J. A. Benediktsson is with the Faculty of Electrical and Computer Engineering, University of Iceland, 101, Reykjavik, Iceland.

I. INTRODUCTION

HYPERSPECTRAL imaging (HSI) sensors, which capture hundreds of continuous bands in a broad spectral range covering visible, near-infrared, and beyond, play an important role in many research areas, including applications in food quality control and analysis [1], pharmaceuticals [2], computer-based forensics and security [3], as well as planet surface investigation such as Mars [4]. In addition to these applications, one of the most active areas of HSI is remote sensing, where researchers develop diverse algorithms based on it, e.g. target detection for military surveillance [5], data compression for faster transmission [6, 7], surface and data classification for land-cover analysis [8–12]. However, due to the characteristics of HSI, data redundancy is inevitable. Furthermore, for remote sensing HSI, noise could be involved during the process of data acquisition and transmission. Therefore, effective feature extraction and denoising of HSI data is necessary for remote sensing applications, in particular for supervised classification problems as discussed in this paper.

Feature extraction could be considered as a linear or non-linear data transformation. Several unsupervised and supervised feature extraction methods have been proposed by researchers, including the most widely used principal component analysis (PCA) [8, 12], along with other approaches such as independent component analysis (ICA) [13], linear discriminant analysis (LDA) [14], minimum noise fraction (MNF) [15], non-negative matrix factorisation (NMF) [16] and singular value decomposition (SVD) [17]. The above mentioned approaches not only extract features, but also they could reduce the dimensionality of HSI data. Other alternatives exist for supervised feature extraction without reducing the data dimensionality, such as spectral curve fitting of hyperspectral image bands [18]. Additionally, an impressive technique, called singular spectrum analysis (SSA), has recently demonstrated its ability in effective feature extraction for HSI data. SSA is a great improvement in terms of potential classification accuracies compared with another popular technique, empirical mode decomposition (EMD) [9–11]. Rather than reducing the dimensionality, SSA ‘smooths’ the spectral profile of HSI data so that effective features can be extracted and classification accuracies can be improved accordingly.

It is assumed that directly applying denoising techniques on the original image could probably remove fine features and noise at the same time, which is undesirable for following applications. To avoid this problem, the multi-scale wavelet transform based approaches are widely used for image denoising, where the transform decomposes the image into a set of wavelet coefficients at different decomposition levels, and noise in the low-energy channels of the transformed domain can be removed. A famous algorithm called soft thresholding in the wavelet domain was proposed by Donoho in [19]. By removing small coefficients under a certain threshold and shrinking large coefficients, most unwanted noise can be discarded easily. For HSI data, Othman and Qian proposed a hybrid spatial-spectral derivative-domain wavelet shrinkage approach based on the soft thresholding [20]. This algorithm works in the spectral derivative domain, in which the noise level is elevated and the signal regularity is dissimilar in the spatial and spectral domains of HSI data. Recently, Chen and Qian combined feature extraction and denoising, leading to a more effective denoising method for HSI data using PCA and wavelet shrinkage [21]. In this approach, first PCA transform is performed on the HSI data and then a 2D bivariate wavelet thresholding method is used to remove noise for low-energy PCA channels.

In this paper, we are also aiming at combining ideas of feature extraction and denoising together for improving classification accuracies of remote sensing hyperspectral images. This work is inspired by [22] where the curvelet transform is applied to HSI data and the representation of noise free signals in the curvelet domain is predicted using multiple linear regression (MLR), a regression that has been previously used for noise analysis [23–24]. Although the wavelet transform has been widely applied for image denoising, many studies have concluded that the wavelet transform cannot provide a good representation of an anisotropic singularity, such as curves or edges in the image [25–27]. For this reason, soft-thresholding directional curvelet coefficients that match image edges could achieve better noise reduction effect than the coefficients obtained in the wavelet domain. It has been proven that the curvelet transform could represent piecewise linear contours on multiple scales through a few significant coefficients, leading to a better separation between geometric details and background noise [28]. Therefore, the curvelet transform is a good candidate for image denoising and enhancement. After the curvelet transform, the coefficients of two adjacent bands still maintain the correlation similarity of the original HSI data [29], which means spectral processing techniques could then be applied in the curvelet domain. Feature extraction by MLR in [22] is achieved by estimating the noise free band by using all pixels in adjacent bands at the same time, while more accurate spectral feature extraction is proposed to be achieved by applying the above mentioned SSA to each pixel. The denoising performance in [22] is then assessed by adding noise to the original HSI data and comparing the mean squared errors (MSE) and the mean structural similarity (MSSIM) between the original data and denoised data, where the classification performance on the

denoised image is unknown. We therefore adopt the curvelet transform instead of the wavelet transform in our approach and combine SSA to extract effective spectral features in the curvelet domain in order to improve classification accuracies for remote sensing hyperspectral images.

The paper is organized as follows: the theory of the curvelet transform and SSA is introduced in Section II. The proposed denoising and feature extraction approach applied on HSI data is given in Section III. Section IV introduces the experimental setup, including data sets, the classifier and tuning of optimal parameters. Results and analysis are presented in Section V. The paper ends with a conclusion section.

II. BACKGROUND PRINCIPLES

In this section, the theory of the curvelet transform and SSA [9, 26, 30–33], which will be used in following sections, is reviewed.

A. The Curvelet Transform

Similar to the wavelet transform, the curvelet transform could also provide multi-scale analysis on images. It was first proposed by Candes and Donoho in 1999, and their version is also called the first-generation curvelet transform [34]. However, this transform is rather complicated and needs at least four steps to complete, including a sub-band decomposition, smooth partitioning, renormalization and ridgelet analysis [34]. A few years later, a simpler and faster version of the curvelet transform was developed, which is the second-generation transform also called the fast discrete curvelet transform (FDCT) [31]. Two forms of FDCT with the same computational complexity were proposed by Candes et al., based on the unequally-spaced fast Fourier transforms (USFFT) and the wrapping of specially selected Fourier samples, respectively. The curvelet transform used here is the one based on USFFT.

Let us assume there is a pair of smooth, non-negative and real-valued windows called the radial window $W(r)$ and the angular window $V(t)$, such that W is supported on $r \in (1/2, 2)$ and V is supported on $t \in [-1, 1]$. These windows will always satisfy the admissibility conditions:

$$\sum_{j=-\infty}^{\infty} W^2(2^j r) = 1, \quad r \in (3/4, 3/2); \quad (1)$$

$$\sum_{l=-\infty}^{\infty} V^2(t-l) = 1, \quad t \in (-1/2, 1/2). \quad (2)$$

For each scale j , the frequency window U_j is defined in the Fourier domain by

$$U_j(r, \theta) = 2^{-3j/4} W(2^{-j} r) V\left(\frac{2^{\lfloor j/2 \rfloor} \theta}{2\pi}\right), \quad (3)$$

where $\lfloor j/2 \rfloor$ is the integer part of $j/2$. Consequently the support of U_j is defined by the support of W and V , which will be a polar wedge.

A mother curvelet waveform $\varphi_j(x)$ is defined related to $U_j(\omega)$ ($U_j(r, \theta)$ will be abbreviated as $U_j(\omega)$), where the Fourier transform of φ_j is equal to U_j . Thus, all curvelets of scale 2^{-j} can be acquired by rotations and translations of φ_j , where the equispaced sequence of rotation angles are $\theta_l = 2\pi \cdot 2^{-\lfloor j/2 \rfloor} \cdot l$, with $l = 0, 1, \dots$ such that $0 \leq \theta_l < 2\pi$, and the sequence of translation parameters $k = (k_1, k_2) \in \mathbb{Z}^2$. Then, we can have the curvelets at scale 2^{-j} , orientation θ_l and position $x_k^{(j,l)} = R_{\theta_l}^{-1}(k_1 \cdot 2^{-j}, k_2 \cdot 2^{-j/2})$, defined as

$$\varphi_{j,l,k}(x) = \varphi_j(R_{\theta_l}(x - x_k^{(j,l)})), \quad (4)$$

where R_θ is the rotation by θ radians and R_θ^{-1} is its inverse,

$$R_\theta = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}. \quad (5)$$

A curvelet coefficient of an element $f \in \mathfrak{R}^2$ is the inner product of f and a curvelet $\varphi_{j,l,k}$,

$$c(j, l, k) := \langle f, \varphi_{j,l,k} \rangle = \int_{\mathbb{R}^2} f(x) \overline{\varphi_{j,l,k}(x)} dx. \quad (6)$$

Our discussion of the digital curvelet transforms will always be in the frequency domain. Therefore, the curvelet coefficient could be re-expressed in the frequency domain using Plancherel's theorem,

$$\begin{aligned} c(j, l, k) &:= \frac{1}{(2\pi)^2} \int \hat{f}(\omega) \overline{\hat{\varphi}_{j,l,k}(\omega)} d\omega \\ &= \frac{1}{(2\pi)^2} \int \hat{f}(\omega) U_j(R_{\theta_l} \omega) e^{i \langle x_k^{(j,l)}, \omega \rangle} d\omega \end{aligned} \quad (7)$$

In practice, instead of using the polar window defined in (3), it is more common to use Cartesian equivalents. The Cartesian frequency window is shown as follows:

$$\tilde{U}_j(\omega) := \psi_j(\omega_1) V_j(\omega), \quad (8)$$

where we define $\psi(\omega_1) = \sqrt{\phi(\omega_1/2)^2 - \phi(\omega_1)^2}$ as a bandpass profile in $\psi_j(\omega_1) = \psi(2^{-j} \omega_1)$ and $V_j(\omega) = V(2^{\lfloor j/2 \rfloor} \omega_2 / \omega_1)$ with V still obeying (2). By introducing the set of equispaced slopes $\tan \theta_l := l \cdot 2^{-\lfloor j/2 \rfloor}$ with $l = -2^{\lfloor j/2 \rfloor}, \dots, 2^{\lfloor j/2 \rfloor} - 1$, for each $\theta_l \in [-\pi/4, \pi/4]$, the Cartesian window can be rewritten as:

$$\tilde{U}_{j,l}(\omega) := \psi_j(\omega_1) V_j(S_{\theta_l} \omega), \quad (9)$$

where S_θ is the shear matrix defined as below:

$$S_\theta := \begin{pmatrix} 1 & 0 \\ -\tan \theta & 1 \end{pmatrix}. \quad (10)$$

Similarly, curvelets should be re-defined in Cartesian form:

$$\tilde{\varphi}_{j,l,k}(x) = 2^{3j/4} \tilde{\varphi}_j(S_{\theta_l}^T(x - S_{\theta_l}^{-T} b)), \quad (11)$$

with $b := (k_1 \cdot 2^{-j}, k_2 \cdot 2^{-j/2})$, where the superscript T represents the transpose of the matrix. Accordingly, the curvelet coefficient will be

$$c(j, l, k) = \int \hat{f}(S_{\theta_l} \omega) \tilde{U}_j(\omega) e^{i \langle b, \omega \rangle} d\omega. \quad (12)$$

Assume the input Cartesian arrays for an n by n image are in the form of $f[t_1, t_2]$, $0 \leq t_1, t_2 < n$, a four-step implementation of FDCT via USFFT can be obtained as follows:

1) *2D FFT*: Apply two-dimensional (2D) fast Fourier transform (FFT) to input arrays and obtain Fourier samples as:

$$\hat{f}[n_1, n_2] = \sum_{t_1, t_2=0}^{n-1} f[t_1, t_2] e^{-i2\pi(n_1 t_1 + n_2 t_2)/n}, \quad (13)$$

with $-n/2 \leq n_1, n_2 < n/2$.

2) *Interpolation*: For each pair of scale and angle (j, l) , Fourier samples $\hat{f}[n_1, n_2]$ are interpolated to get new values, $\hat{f}[n_1, n_2 - n_1 \tan \theta_l]$, for $n_1, n_2 \in P_j$. P_j is defined as a set below:

$$P_j = \{(n_1, n_2) : n_{10} \leq n_1 < n_{10} + L_{1,j}, n_{20} \leq n_2 < n_{20} + L_{2,j}\}, \quad (14)$$

where $L_{1,j}$ and $L_{2,j}$ are the length and width of a rectangle and (n_{10}, n_{20}) is the index of the pixel at the bottom-left of the rectangle.

3) *Multiplication*: Multiply the interpolated samples with the frequency window \tilde{U}_j and obtain

$$\tilde{f}_{j,l}[n_1, n_2] = \hat{f}[n_1, n_2 - n_1 \tan \theta_l] \tilde{U}_j[n_1, n_2]. \quad (15)$$

4) *Inverse 2D FFT*: The last step is to apply the inverse 2D FFT to $\tilde{f}_{j,l}$ in order to get the discrete curvelet coefficients:

$$\begin{aligned} c^D(j, l, k) &= \sum_{n_1, n_2 \in P_j} \hat{f}[n_1, n_2 - n_1 \tan \theta_l] \\ &\quad \tilde{U}_j[n_1, n_2] e^{i2\pi(k_1 n_1 / L_{1,j} + k_2 n_2 / L_{2,j})} \end{aligned} \quad (16)$$

With all four steps, the discrete curvelet transform via USFFT requires $O(n^2 \log n)$ flops for computation and $O(n^2)$ for storage.

B. Singular Spectrum Analysis

As a well-established approach, SSA has been applied for time series analysis and forecasting and is widely used in different areas, including mathematics, economics and even biomedical engineering [35]. SSA shares the same theoretical foundations as PCA. Both of them are able to decompose the original time series into a linear combination of a new orthogonal basis, which includes eigenvectors generated from the diagonalization of the data correlation matrix [36]. The main capability of SSA is that it can decompose the original series into some interpretable components, such as the trend, oscillations and unstructured noise [9]. The SSA algorithm consists of two stages including the decomposition and the reconstruction, and it is briefly explained as follows.

1) *Decomposition*: Assume that \mathbf{X} is a non-zero one-dimensional (1D) series vector with length N , i.e., $\mathbf{X} = (x_1, x_2, \dots, x_N)$. Given a window size L ($1 < L < N$), the original series \mathbf{X} is mapped to K lagged vectors, $\mathbf{X}_i = (x_i, x_{i+1}, \dots, x_{i+L-1})^T$ for $i = 1, 2, \dots, K$, where $K = N - L + 1$. The window size should be chosen properly depending on the application. Then the trajectory matrix is formed as:

$$\mathbf{T} = \begin{pmatrix} \mathbf{X}_1 & \mathbf{X}_2 & \cdots & \mathbf{X}_K \end{pmatrix} = \begin{pmatrix} x_1 & x_2 & \cdots & x_K \\ x_2 & x_3 & \cdots & x_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_L & x_{L+1} & \cdots & x_N \end{pmatrix}. \quad (17)$$

It should be noted that the matrix \mathbf{T} in (17) is a Hankel matrix of size $L \times K$, which means its entries along the anti-diagonals are equal.

The next step is to compute the SVD of the trajectory matrix \mathbf{T} . First, eigenvalues of $\mathbf{T}\mathbf{T}^T$ are calculated and sorted in a descending order, i.e., $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_L \geq 0$. Let the corresponding eigenvectors be $(\mathbf{U}_1, \mathbf{U}_2, \dots, \mathbf{U}_L)$, and the resulting trajectory matrix after the SVD is shown in (18),

$$\mathbf{T} = \mathbf{T}_1 + \mathbf{T}_2 + \cdots + \mathbf{T}_d, \quad (18)$$

where d is the rank of \mathbf{T} , $\mathbf{T}_i = \sqrt{\lambda_i} \mathbf{U}_i \mathbf{V}_i^T$ ($i = 1, 2, \dots, d$) is called the elementary matrix with rank 1, and $\mathbf{V}_i = \mathbf{T}^T \mathbf{U}_i / \sqrt{\lambda_i}$ are often referred to the principal components (PCs) of the matrix \mathbf{T} . Generally, the contribution of the elementary matrix \mathbf{T}_i to the trajectory matrix \mathbf{T} is determined by the ratio of each eigenvalue and the sum of all eigenvalues, as shown in (19),

$$\eta_i = \lambda_i / \sum_{i=1}^d \lambda_i. \quad (19)$$

2) *Reconstruction*: The first step of reconstruction is grouping, where the set of indices $\{1, 2, \dots, d\}$ is divided into m disjoint subsets $\{I_1, I_2, \dots, I_m\}$. Each I_j ($j = 1, 2, \dots, m$) contains one or several elementary matrices \mathbf{T}_i which are summed within each group. Then, the resultant trajectory matrices can be computed for each group and consequently (18) is expanded as follows:

$$\begin{aligned} \mathbf{T} &= \overbrace{\mathbf{T}_1 + \mathbf{T}_2 + \cdots + \mathbf{T}_d}^{SVD} \\ &= \overbrace{\mathbf{T}_{I_1} + \mathbf{T}_{I_2} + \cdots + \mathbf{T}_{I_m}}^{Grouping}. \end{aligned} \quad (20)$$

The last step is diagonal averaging, which transforms each grouped matrix \mathbf{T}_{I_j} into a new series with length N . This process is also known as Hankelisation of the matrix \mathbf{T}_{I_j} . Assume $\mathbf{Y}_1 = (y_1, y_2, \dots, y_N)$ is the transformed 1D series of \mathbf{T}_{I_1} , elements in \mathbf{Y}_1 can be calculated using (21) by averaging the corresponding diagonals of \mathbf{T}_{I_1} ,

$$y_k = \begin{cases} \frac{1}{k} \sum_{j=1}^k y_{j, k-j+1}^*, & 1 \leq k < L^* \\ \frac{1}{L^*} \sum_{j=1}^{L^*} y_{j, k-j+1}^*, & L^* \leq k \leq K^* \\ \frac{1}{N-k+1} \sum_{j=k-K^*+1}^{N-K^*+1} y_{j, k-j+1}^*, & K^* < k \leq N \end{cases}, \quad (21)$$

where $L^* = \min(L, K)$, $K^* = \max(L, K)$, $y_{j, k-j+1}^*$ refers to the elements in \mathbf{T}_{I_1} , $y_{j, k-j+1}^* = y_{j, k-j+1}$ if $L < K$ and $y_{j, k-j+1}^* = y_{k-j+1, j}$ if $L \geq K$. Then, the initial series $\mathbf{X} = (x_1, x_2, \dots, x_N)$ is decomposed into m series, like shown in (22),

$$\mathbf{X} = \mathbf{Y}_1 + \mathbf{Y}_2 + \cdots + \mathbf{Y}_m. \quad (22)$$

Therefore, the original series vector could be reconstructed by only using the first or the first a few groups generated from its eigenvalues and the rest could be discarded as noise. For example, if the window size L is set as ten, reconstruction using the first three to five components should usually be enough to achieve a good denoising performance. However, there is no general rule for the grouping. Like the window size L , one can choose the eigenvalue grouping (EVG) value depending on the application. If all decomposed components are involved in the EVG, the reconstructed series will be the same as the original series.

III. THE PROPOSED APPROACH: APPLYING SSA IN THE CURVELET DOMAIN

As briefly mentioned in Section I, the proposed denoising and feature extraction approach will be applied on hyperspectral images and the flowchart is shown in Fig. 1 for reference. Hyperspectral sensors have a relatively high spectral resolution and can generate hundreds of observation channels [37]. The obtained three-dimensional (3D) HSI data cube can be regarded as a stack of 2D images of the same scene corresponding to different wavelengths [22], and the correlation between each two adjacent bands is fairly high. The first step of the proposed method is to perform the curvelet transform on each band of the hyperspectral image and a few image stacks are generated after the transform as shown in Fig. 1. By applying the curvelet transform on each band, the band correlation property can be preserved so that SSA is able to exploit the spectral signature. The next step is to apply SSA in the spectral dimension for each detail image stack for feature extraction and denoising. After that, the denoised detail images at each band are gathered according to their original location, followed by the inverse curvelet transform to get the denoised hyperspectral image.

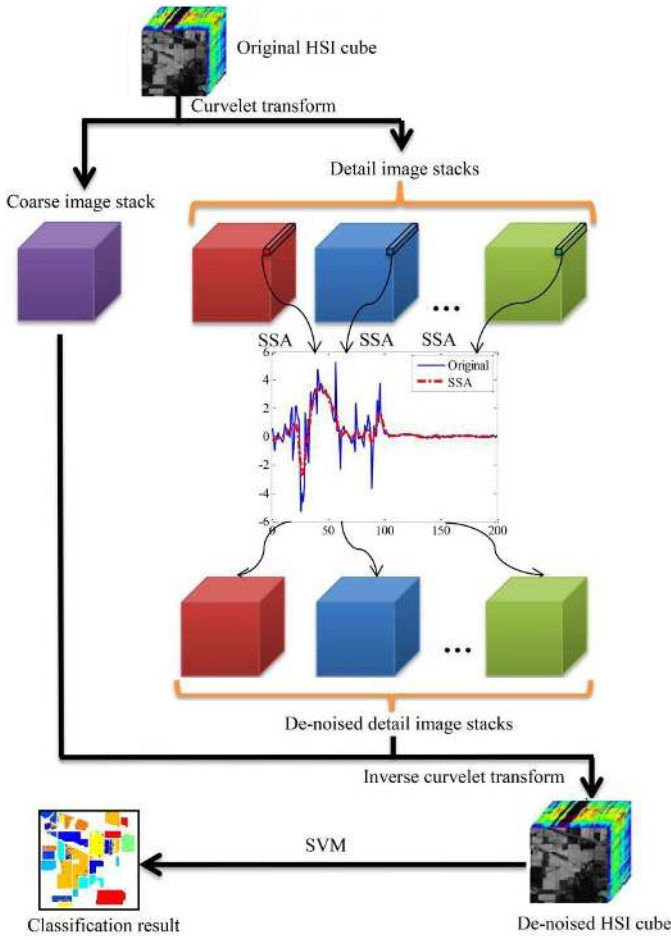


Fig. 1. Flowchart of the proposed methodology.

In this paper, a hyperspectral image is denoted as $\mathbf{I}(p, q) = (I_1(p, q), I_2(p, q), \dots, I_B(p, q)) \in \mathfrak{R}^B$, where

$p \in [1, P]$, $q \in [1, Q]$ and $P \times Q$ is the spatial dimension of the hyperspectral image and B is the number of bands. \mathfrak{R} represents the set of real numbers with the pixel intensity $I_b(p, q)$ at all sensor channels with $b \in [1, B]$. The 2D curvelet transform via USFFT employed in this paper is completed using the toolbox CurveLab (version 2.1.2) [38]. Similar to the wavelet transform, the curvelet transform can also decompose the image into a coarse image and several detail images. Just like most image processing algorithms, the curvelet transform requires the processed image to be a square whose dimension is a power of two. If the size of the original image is not a power of two, pixels with a value of zero are padded to the next larger power of two.

Given the zero-padded HSI data set, $\mathbf{I}_{pad}(m, n) = (I_{pad_1}(m, n), I_{pad_2}(m, n), \dots, I_{pad_B}(m, n))$ where $m, n \in [1, N]$, consisting of B bands where each band has N^2 pixels, the curvelet transform is performed on a band image I_{pad_b} to obtain the curvelet coefficients corresponding to that band c_b^D :

$$c_b^D = CT(I_{pad_b}), \quad (23)$$

where CT stands for the discrete curvelet transform operation explained in Section II. The decomposition results of the transform can be regarded as a superposition in the form as follows:

$$c_b^D = C_{b,J} + \sum_{j=1}^{J-1} W_{b,j}, \quad (24)$$

where $C_{b,J}$ is the coarse version of the original band image with low frequency contents and $W_{b,j}$ stands for the detail band image at scale 2^{-j} containing high frequency contents. With an N by N image, the default number of decomposition scales is $J = \log_2(N) - 3$ as set in the CurveLab toolbox. Take a band image with size of 128 by 128 for instance, the number of decomposition scales J will be four. According to the settings in the toolbox, there are eight orientations of the curvelet in the second and third scales, starting from the top-left wedge and increasing in a clockwise fashion, but for the coarsest and finest scales, there is only one direction. Thus, (24) is rewritten as below,

$$c_b^D = C_{b,J,1} + W_{b,1,1} + \sum_{j=2}^{J-1} \sum_{l=1}^8 W_{b,j,l}, \quad (25)$$

where $W_{b,1,1}$ represents the finest scale with one orientation.

The schematic diagram of four decomposition scales corresponding to a 128 by 128 image is shown in Fig. 2. Then, after applying the curvelet transform to all bands, there will be a coarse image stack and 17 detail image stacks.

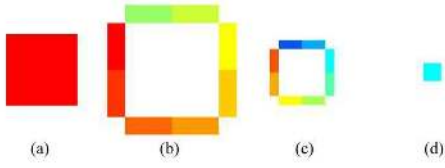


Fig. 2. Illustration of a 128 by 128 image with 4 decomposition scales, where (a) is the first (finest) scale with 1 orientation, (b) is the second scale with 8 orientations, (c) is the third scale with 8 orientations, and (d) is the fourth (coarsest) scale with 1 orientation.

As suggested by the commonly adopted wavelet denoising rule, noise is removed by thresholding only the wavelet coefficients of the detail sub-bands, while keeping the low frequency coefficients unchanged [39]. Therefore, in our approach, the coarse image stack stays unaltered as well. For each detail image stack, SSA is applied to the spectral dimension for smoothing the spectral profile, followed by the inverse curvelet transform to get the denoised HSI data cube.

Given a detail image stack at scale 2^{-j} and orientation l , which is denoted as $W_{b,j,l}(x, y)$ ($b \in [1, B]$, $x \in [1, X]$, $y \in [1, Y]$) containing $X \times Y$ coefficients in each band, the spectral series vector for one pixel can be constructed as $\mathbf{W}_{j,l}(x, y) = (W_{1,j,l}(x, y), W_{2,j,l}(x, y), \dots, W_{B,j,l}(x, y))$. Then, the smoothing process on the spectral dimension by SSA could be achieved, denoted as below,

$$\mathbf{Y}_{j,l}(x, y) = SSA_{L, EVG}(\mathbf{W}_{j,l}(x, y)), \quad (26)$$

in which SSA is the operation of the singular spectrum analysis listed in (18)-(22), L , EVG are SSA parameters, and $\mathbf{Y}_{j,l}(x, y) = (Y_{1,j,l}(x, y), Y_{2,j,l}(x, y), \dots, Y_{B,j,l}(x, y))$ is the smoothed spectral feature in the curvelet domain. After collecting all smoothed spectra for each detail stack, the inverse discrete curvelet transform is performed on the smoothed curvelet coefficients band by band using the same toolbox CurveLab and the denoised HSI data cube is achieved, shown as (27),

$$I_d_b = ICT\left(C_{b,J,l} + Y_{b,1,l} + \sum_{j=2}^{J-1} \sum_{l=1}^8 Y_{b,j,l}\right), \quad (27)$$

where ICT stands for the inverse curvelet transform operation. The last step is to crop the denoised HSI cube to its original size. The pseudo code of the proposed curvelet and SSA approach is summarized in Algorithm 1.

Algorithm 1 The Curvelet and SSA algorithm

- 1: **procedure** CURVELETSSA(I , L , EVG) \triangleright
 $P \times Q \times B$ HSI data set I , SSA window size L
and grouping parameter EVG
- 2: **if** $P \neq Q$ & P, Q are not power of two **then**
- 3: zero-pad each band of I to the next larger
power of two to get

- 4: a new data cube I_pad with size of
 - 5: $N \times N \times B$
 - 6: **else**
 - 7: $I_pad \leftarrow I$
 - 8: **end if**
 - 9: **for** $b \leftarrow 1, B$ **do**
 - 10: apply curvelet transform to I_pad_b to get
 - 11: coarse $C_{b,J}$ and detail $W_{b,j,l}$ images
 - 12: **end for**
 - 13: **for all** detail image stacks **do**
 - 14: **for** $x \leftarrow 1, X$ **do**
 - 15: **for** $y \leftarrow 1, Y$ **do**
 - 16: $\mathbf{W}_{j,l}(x, y) \leftarrow ((W_{1,j,l}(x, y), \dots,$
 - 17: $W_{B,j,l}(x, y))$
 - 18: $\mathbf{Y}_{j,l}(x, y) \leftarrow SSA(\mathbf{W}_{j,l}(x, y), L, EVG)$
 - 19: **end for**
 - 20: **end for**
 - 21: **end for**
 - 22: substitute original detail curvelet coefficients
 - 23: with SSA-processed detail coefficients
 - 24: **for** $b \leftarrow 1, B$ **do**
 - 25: apply inverse curvelet transform to $C_{b,J}$ and
 - 26: $\mathbf{Y}_{b,j,l}$ to get denoised band image I_d_b
 - 27: **end for**
 - 28: crop I_d_b to the original size $P \times Q \times B$
 - 29: **end procedure**
-

IV. DATA SETS AND EXPERIMENTAL SETUP

In this section, the experimental data sets will be introduced as well as the choice of the classification model and parameters.

Three publicly available and widely used HSI remote sensing data sets are employed to evaluate the performance of the proposed method in this paper, including the airborne visible/infrared imaging spectrometer (AVIRIS) natural scenes Indian Pines [40] and Salinas Valley [41], as well as the reflective optics system imaging spectrometer (ROSIS) urban scene Pavia University [42].

The Indian Pines data set, was collected in the Indian Pines test site in North-western Indiana, USA in June 12, 1992. The scene consists of two-thirds agriculture, and one-third forest or other natural perennial vegetation [43]. Acquired in band interleaved (BIL) format, it has 145 by 145 pixels with a spatial resolution of 18 m, and 220 continuous spectral channels ranging from 400 to 2500 nm covering the complete

VIS-NIR-SWIR spectrum. The nominal spectral resolution is 10 nm and the radiometric resolution is 16 bits [44]. Due to atmospheric water absorption, bands 104-108, 150-163 and 220 do not contain useful information and are consequently removed to prevent from decreasing the classification accuracy, resulting in a reduced data set with 200 spectral bands. There are 16 classes in the original ground truth map. However, some classes have insufficient samples for training the classification model. As suggested by other researchers [9, 37, 45], 7 out of 16 classes are discarded for more consistent results, leaving 9 classes considered in experiments. One band image and the reduced ground truth map of Indian Pines are shown in Fig. 3.

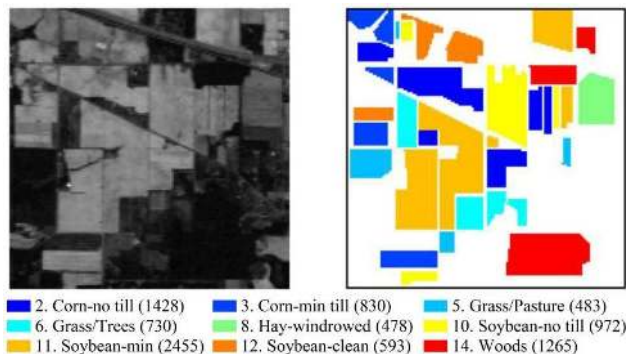


Fig. 3. Indian Pines data set: band 150 out of 200 bands (left) and the reduced 9-class ground truth map (right) with number of samples shown.

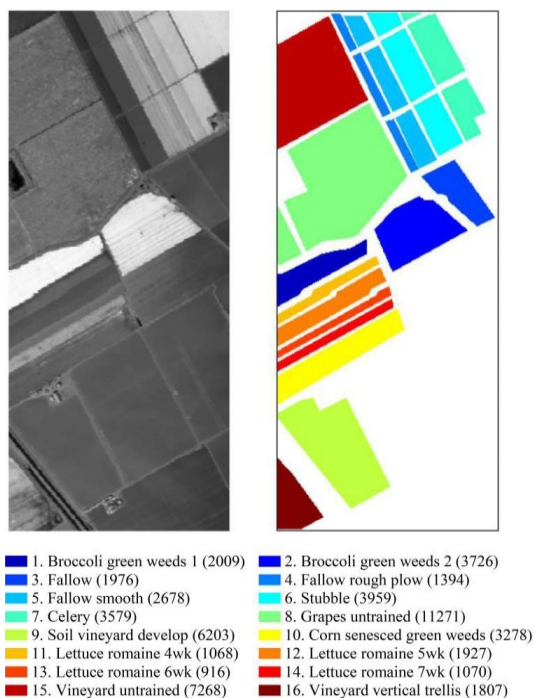


Fig. 4. Salinas Valley data set: band 50 out of 204 bands (left) and the 16 class-ground truth map (right) with number of samples shown.

The second data set was collected over Salinas Valley, CA at low altitude in October 9, 1998, resulting in a high spatial resolution of 3.7 m per pixel [46]. The full scene, which includes vegetables, bare soils and vineyard fields, comprises 512 lines by 217 samples with 224 spectral bands. The spectral

and radiometric resolutions of this data set are the same as for Indian Pines. After removing water absorption bands (108-112, 154-167 and 224), the obtained image then has 204 channels covering 400 to 2500 nm [44]. The ground truth map of Salinas Valley contains 16 classes as shown in Fig. 4.

The Pavia University data set was collected during a flight campaign over the Pavia district in north Italy, with a spatial resolution of 1.3 m per pixel [47]. There are originally 115 bands with a spectral coverage ranging from 430 to 860 nm. However, 12 channels have been removed due to noise, leaving 103 bands with 610×340 pixels per band. Nine classes of interest are provided in the ground truth map, including urban, soil and vegetation features, as shown in Fig. 5.

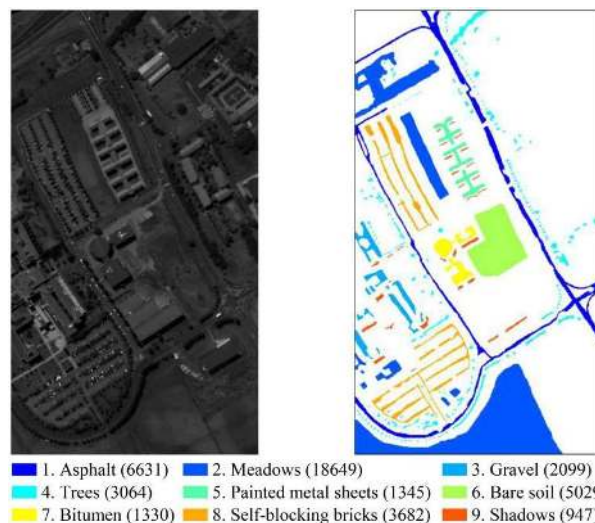


Fig. 5. Pavia University data set: band 70 out of 103 bands (left) and the 9 class-ground truth map (right) with number of samples shown.

The spatial area of all data sets is not a square, hence, before applying the curvelet transform, Indian Pines, Salinas Valley and Pavia University are zero-padded to the size of $256 \times 256 \times 200$, $512 \times 512 \times 204$ and $1024 \times 1024 \times 103$, respectively. Then, when the denoising and feature extraction process is finished, reconstructed images are cropped to their original sizes for the following performance evaluation.

In the context of supervised classification, a variety of methods have been developed for HSI data classification problems, including the famous artificial neural networks [48–50], multinomial logistic regression [51], as well as the widely used support vector machine (SVM) which shows its outstanding performance in many papers [1, 8, 37, 52]. The properties of SVM make it an effective tool for HSI data classification problems which are influenced by the Hughes phenomenon [53]. The Hughes phenomenon is also known as the ‘curse of dimensionality’, where the high number of spectral bands and low number of labelled training sample result in an decrease in the generalisation capability of the classifier [37]. In this paper, a publicly available SVM library called LIBSVM [54] is adopted and the Gaussian radial basis function (RBF) kernel is chosen as it has been shown to outperform other kernel functions for HSI data, such as linear and polynomial kernel

functions, in terms of classification accuracy [37]. Optimal values of the regularization parameter C and the width parameter γ in the Gaussian RBF kernel are achieved using ten-fold cross validation on training samples, where values of both parameters are tested within the exponentially increasing sequence $\{2^{-10}, 2^{-9}, \dots, 2^{10}\}$.

V. RESULTS AND ANALYSIS

In this section, the classification results based on SVM for three hyperspectral data sets are revealed. The classification performance is evaluated by the overall accuracy (OA) and the average accuracy (AA), where OA refers to the percentage of all pixels that are correctly labelled and AA stands for the average percentage of correctly labelled pixels for each class.

Since the proposed methodology is inspired by feature extraction approach based on SSA [9] and the denoising approach in [22] which uses MLR applied in the curvelet domain, it is compared with them in subsection A below. Besides, the curvelet transform is similar to the wavelet transform as a multi-scale geometric analysis (MGA). Therefore, SSA and MLR are also applied in the wavelet domain for comparison in this subsection. For the wavelet transform-based approaches, the hyperspectral data are also zero-padded and the number of decomposition scales is set the same as the curvelet transform, where the Cohen-Daubechies-Feauveau (CDF) 9/7 wavelet is adopted. For convenience, the proposed approach is denoted as CT-SSA and other inspirational approaches are named as SSA, CT-MLR (MLR applied in the curvelet domain), WT-MLR (MLR applied in the wavelet domain) and WT-SSA (SSA applied in the wavelet domain). The computational cost required for these methods is also discussed in this part.

Then, the CT-SSA method as well as its extended version CT-SSA-PCA, are further compared with some state-of-the-art spectral feature extraction techniques in subsection B, including PCA, LDA, NMF and ensemble EMD (EEMD) [55], to show the efficacy of the proposed methodology.

The proposed approach only takes into account the spectral information of the HSI data, while ignoring the important spatial information. For this reason, in subsection C, a simple yet powerful post-processing technique is applied to the classified ground truth map and it is compared with several recent spectral-spatial classification methods listed in [56].

A. Comparison with Inspirational Approaches

The SSA based feature extraction approach mentioned in [9] has two parameters: window size L and grouping parameter EVG . It is noticed in [9] that good results can be achieved with only the first component used for EVG , as long as a proper window size L is chosen. Another parameter in the proposed approach is the decomposition level in the curvelet transform. As mentioned previously, the default number of decomposition levels in the CurveLab toolbox is $J = \log_2(N) - 3$ and it is advised that the number of decomposition levels should not be

higher than the default setting. Therefore, different values have been tested to look for optimal parameters. Firstly, the default number of decomposition levels is used in the curvelet transform and the window size L is tested from 2 to 5 as plotted in Fig. 6. Similarly, experiments have been carried out ten times, and the average OA as well as the standard deviation are shown in the plots. By trial and error, optimal value for L is set as 5 for both Indian Pines and Salinas Valley. For Pavia University, L is set as 4. The same settings of SSA are used for CT-SSA and WT-SSA as well. With the chosen optimal values for window size L , the number of decomposition levels is tested from 2 to the default setting. Results plotted in Fig. 6 indicate that the default setting of decomposition levels should always be used in the curvelet transform.

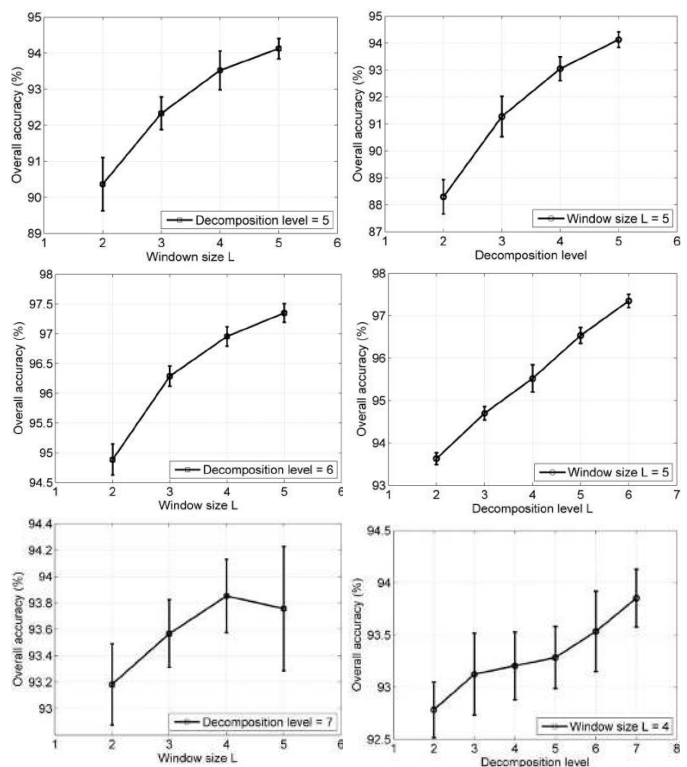


Fig. 6. Sensitivity analysis of window size (left) and number of decomposition levels (right) for Indiana Pines (top), Salinas Valley (middle) and Pavia University (bottom).

For the Indian Pines data set, in each class out of the 9 classes, 10% of pixels are randomly extracted into the training set with the rest allocated in the testing set. As Salinas Valley is a high resolution HSI data set, spectra with this scene are more separable than those with Indian Pines. Therefore, a lower percentage of samples in each class (5%) is used for training the classification model of SVM to give a more interpretable result. A recent publication proposed to use a fixed number (200) from each class in the Pavia University data set as training samples [56], which account for 4% of the whole labelled pixels. Therefore, 4% of pixels are also randomly chosen from Pavia University for training SVM in the following experiment. Figs. 7-9 have shown the classification maps as well as overall

accuracies for three data sets. Compared with those inspirational methods, the CT-SSA approach always presents better denoising and feature extraction performance in terms of highest classification accuracies.

Additionally, the capability of the proposed CT-SSA method in removing noise is compared. One noisy band is taken from each data set and the same band after being processed by CT-SSA is compared visually in Fig. 10 – 12. It can be observed that the CT-SSA method works for both high-noise (Fig. 10 and Fig. 11) and low-noise (Fig. 12) cases, where the noise in the corrupted band is effectively suppressed and local details of the original image can be kept simultaneously.

In order to avoid errors and to get a more consistent result, ten repeated experiments are carried out based on randomly chosen training and testing samples, and average classification accuracies are calculated, with numerical results shown in Table I – III for three experimental data sets. It can be noticed that compared with the original image, OA using the CT-SSA method is increased with a percentage of 10.02%, 4.44% and 0.89% respectively for three data sets, achieving an impressive improvement over other methods. For both Indian Pines and Salinas Valley, in most cases, the proposed method presents the

highest classification accuracy, especially in class 15 of Salinas Valley, the accuracy is improved by over 20% compared with the raw image. Although it was found that the classification of hyperspectral urban data is a challenging problem without combining the spatial and spectral information together [47], both AA and OA are slightly improved for Pavia University using the proposed spectral processing method CT-SSA. In comparison with CT-MLR, WT-MLR yields slightly worse results in both AA and OA. Same situations occur for WT-SSA and CT-SSA, where the latter presents slightly higher AA and OA than the former. It proves that the curvelet transform does have advantages over the wavelet transform for extracting geometric details in some ways, but the performance of the curvelet transform is largely dependent on the data set. Despite the fact that CT-SSA is inspired by CT-MLR, SSA is applied to the detail curvelet coefficients for each pixel while MLR is applied to each detail image stack all at once. This along with the embedding process and SVD decomposition in SSA is the main reason why CT-SSA overperforms MLR in potentially maximised the opportunity in reducing redundancy and noise within the hyperspectral image data for improved accuracy of classification.

TABLE I
MEAN CLASS-BY-CLASS, AVERAGE AND OVERALL ACCURACIES (%) OF TEN REPEATED EXPERIMENTS ON TESTING SAMPLES OF THE ORIGINAL INDIAN PINES DATA SET AND SSA, CT-MLR, WT-MLR, WT-SSA AND CT-SSA PROCESSED DATA SETS WITH 10% OF DATA USED FOR TRAINING, FOLLOWED BY THE STANDARD DEVIATION

Class	Method					
	Original	SSA	CT-MLR	WT-MLR	WT-SSA	CT-SSA
2	75.44 ± 2.09	80.02 ± 2.38	86.76 ± 1.35	86.15 ± 2.20	87.61 ± 1.97	89.58 ± 3.15
3	69.92 ± 5.11	77.01 ± 5.34	88.26 ± 3.16	89.64 ± 2.84	93.19 ± 1.42	92.38 ± 2.43
5	90.46 ± 3.41	91.31 ± 1.34	93.33 ± 2.10	92.60 ± 2.08	93.17 ± 2.06	94.76 ± 1.94
6	98.11 ± 0.86	97.79 ± 0.67	98.71 ± 0.95	98.54 ± 0.88	98.89 ± 0.95	98.95 ± 0.45
8	99.26 ± 0.42	99.63 ± 0.23	99.61 ± 0.35	99.77 ± 0.36	99.54 ± 0.48	99.84 ± 0.11
10	72.35 ± 2.59	80.87 ± 4.74	86.22 ± 1.52	85.49 ± 2.56	90.20 ± 2.33	89.45 ± 2.29
11	85.10 ± 2.32	87.52 ± 1.73	92.01 ± 1.52	91.65 ± 1.65	94.59 ± 2.02	93.96 ± 0.95
12	75.21 ± 3.96	82.42 ± 5.23	85.99 ± 2.16	87.83 ± 2.80	91.92 ± 2.00	90.83 ± 2.73
14	98.27 ± 0.52	98.37 ± 0.62	99.17 ± 0.35	99.07 ± 0.57	98.93 ± 1.04	99.54 ± 0.26
AA	84.90 ± 1.18	88.35 ± 1.01	92.34 ± 0.41	92.30 ± 0.55	94.23 ± 0.29	94.48 ± 0.32
OA	84.11 ± 0.87	87.54 ± 0.83	91.93 ± 0.50	91.76 ± 0.44	93.87 ± 0.38	94.13 ± 0.28

TABLE II
MEAN CLASS-BY-CLASS, AVERAGE AND OVERALL ACCURACIES (%) OF TEN REPEATED EXPERIMENTS ON TESTING SAMPLES OF THE ORIGINAL SALINAS VALLEY DATA SET AND SSA, CT-MLR, WT-MLR, WT-SSA AND CT-SSA PROCESSED DATA SETS WITH 5% OF DATA USED FOR TRAINING, FOLLOWED BY THE STANDARD DEVIATION

Class	Method					
	Original	SSA	CT-MLR	WT-MLR	WT-SSA	CT-SSA
1	99.02 ± 0.50	99.07 ± 0.61	99.48 ± 0.44	99.40 ± 0.41	99.32 ± 0.37	99.35 ± 0.33
2	99.85 ± 0.06	99.71 ± 0.20	99.68 ± 0.20	99.72 ± 0.20	99.68 ± 0.24	99.69 ± 0.27
3	99.07 ± 0.41	99.37 ± 0.49	99.39 ± 0.33	99.67 ± 0.08	99.65 ± 0.19	99.59 ± 0.22
4	99.21 ± 0.34	99.44 ± 0.34	99.32 ± 0.28	99.30 ± 0.41	99.09 ± 0.69	99.28 ± 0.33
5	98.79 ± 0.42	98.52 ± 0.75	99.07 ± 0.66	98.84 ± 0.60	98.98 ± 0.55	99.04 ± 0.32

6	99.80 ± 0.08	99.74 ± 0.17	99.80 ± 0.09	99.98 ± 0.14	99.83 ± 0.16	99.81 ± 0.08
7	99.62 ± 0.18	99.48 ± 0.23	99.52 ± 0.17	99.51 ± 0.23	99.59 ± 0.18	99.62 ± 0.15
8	88.05 ± 1.47	87.90 ± 1.20	91.28 ± 0.82	90.62 ± 0.77	92.26 ± 0.74	94.60 ± 0.75
9	99.65 ± 0.22	99.72 ± 0.14	99.60 ± 0.42	99.54 ± 0.21	99.86 ± 0.09	99.88 ± 0.09
10	94.98 ± 1.20	95.19 ± 0.80	95.36 ± 1.30	96.01 ± 1.11	97.35 ± 1.17	97.19 ± 1.02
11	97.63 ± 1.01	97.02 ± 2.10	97.84 ± 1.42	97.62 ± 2.44	97.78 ± 2.58	98.58 ± 0.30
12	99.53 ± 0.28	99.64 ± 0.19	99.78 ± 0.15	99.50 ± 0.80	99.79 ± 0.20	99.92 ± 0.06
13	98.09 ± 1.11	98.45 ± 1.40	98.43 ± 0.86	98.84 ± 0.72	98.52 ± 0.93	98.54 ± 0.95
14	94.56 ± 2.12	95.70 ± 1.99	96.86 ± 1.45	95.89 ± 2.14	94.83 ± 3.80	96.46 ± 1.62
15	71.73 ± 3.21	76.16 ± 1.85	82.70 ± 2.88	83.17 ± 1.26	87.12 ± 1.73	92.49 ± 0.86
16	98.34 ± 0.28	98.13 ± 0.51	98.72 ± 0.32	98.48 ± 1.18	98.50 ± 0.97	98.71 ± 0.26
AA	96.12 ± 0.11	96.45 ± 0.22	97.30 ± 0.28	97.25 ± 0.31	97.63 ± 0.44	98.30 ± 0.13
OA	92.91 ± 0.12	93.49 ± 0.25	95.18 ± 0.33	95.11 ± 0.21	96.08 ± 0.34	97.35 ± 0.16

TABLE III

MEAN CLASS-BY-CLASS, AVERAGE AND OVERALL ACCURACIES (%) OF TEN REPEATED EXPERIMENTS ON TESTING SAMPLES OF THE ORIGINAL PAVIA UNIVERSITY DATA SET AND SSA, CT-MLR, WT-MLR, WT-SSA AND CT-SSA PROCESSED DATA SETS WITH 4% OF DATA USED FOR TRAINING, FOLLOWED BY THE STANDARD DEVIATION

Class	Method					
	Original	SSA	CT-MLR	WT-MLR	WT-SSA	CT-SSA
1	93.05 ± 0.69	92.50 ± 1.11	93.27 ± 0.99	92.27 ± 0.66	92.81 ± 0.66	94.17 ± 0.91
2	97.63 ± 0.38	97.47 ± 0.54	97.74 ± 0.54	97.65 ± 0.39	98.10 ± 0.27	97.92 ± 0.21
3	76.45 ± 1.65	76.21 ± 2.02	79.05 ± 2.79	77.51 ± 2.78	80.48 ± 0.84	79.63 ± 1.59
4	93.15 ± 1.63	92.88 ± 1.42	93.40 ± 1.41	92.69 ± 1.99	92.43 ± 2.26	93.97 ± 1.66
5	99.08 ± 0.35	99.03 ± 0.29	99.11 ± 0.41	99.15 ± 0.32	99.01 ± 0.24	99.10 ± 0.31
6	86.19 ± 1.91	85.20 ± 1.56	86.99 ± 1.02	86.46 ± 1.53	88.10 ± 1.16	88.11 ± 1.67
7	83.07 ± 3.70	85.45 ± 2.14	83.55 ± 2.09	85.47 ± 2.99	86.59 ± 2.94	85.62 ± 1.72
8	87.43 ± 2.52	87.75 ± 1.44	87.66 ± 2.40	88.04 ± 1.97	88.02 ± 0.58	88.08 ± 1.90
9	99.84 ± 0.12	99.77 ± 0.23	99.85 ± 0.16	99.85 ± 0.13	99.76 ± 0.12	99.84 ± 0.11
AA	90.59 ± 0.49	90.69 ± 0.41	91.18 ± 0.36	91.01 ± 0.54	91.70 ± 0.58	91.82 ± 0.32
OA	92.96 ± 0.32	92.77 ± 0.37	93.34 ± 0.20	93.05 ± 0.30	93.67 ± 0.21	93.85 ± 0.28

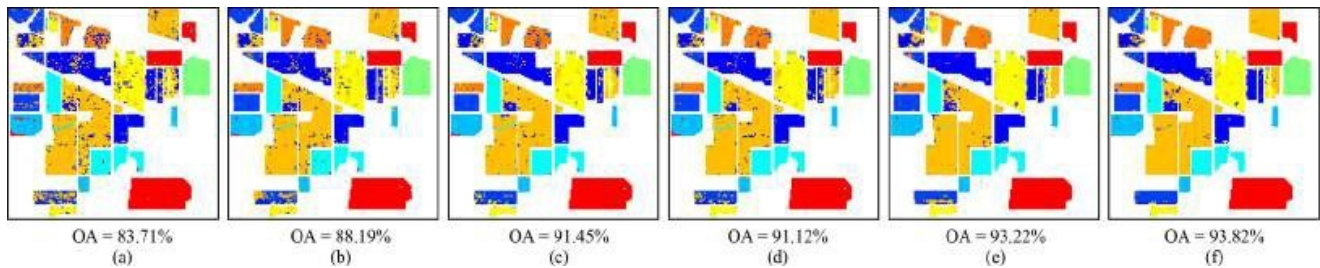


Fig. 7. Classification results (overall accuracy in percentage) of Indian Pines obtained on (a) original data, (b) SSA processed data, (c) CT-MLR processed data, (d) WT-MLR processed data, (e) WT-SSA processed data, and (f) CT-SSA processed data.

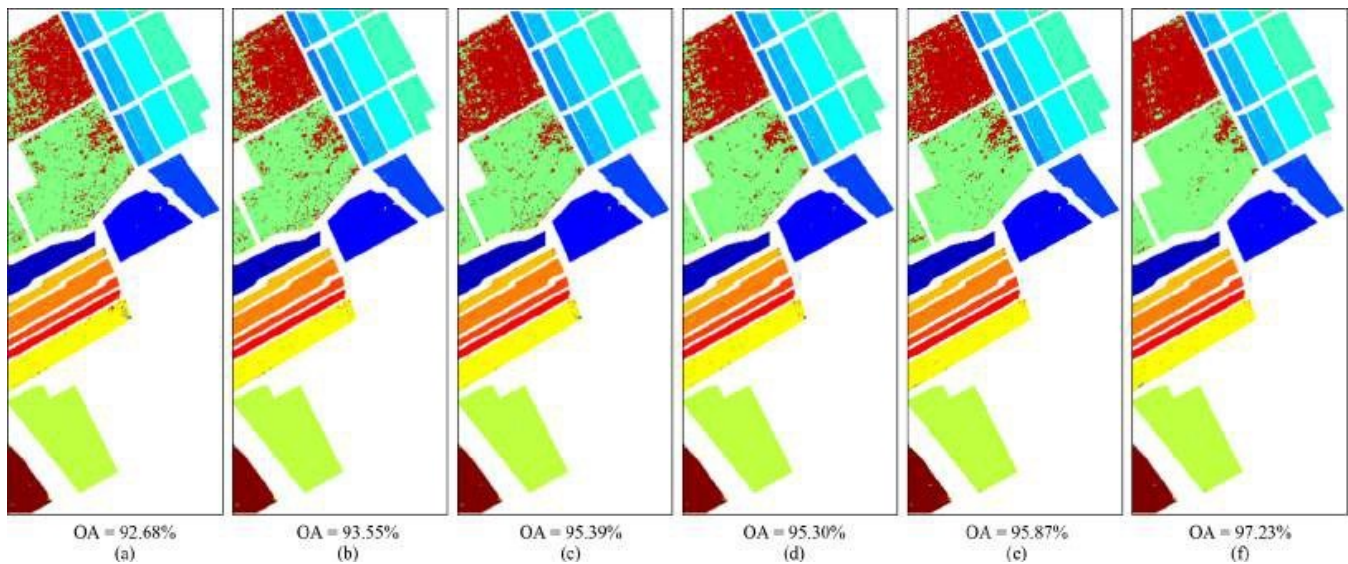


Fig. 8. Classification results (overall accuracy in percentage) of Salinas Valley obtained on (a) original data, (b) SSA processed data, (c) CT-MLR processed data, (d) WT-MLR processed data, (e) WT-SSA processed data, and (f) CT-SSA processed data.

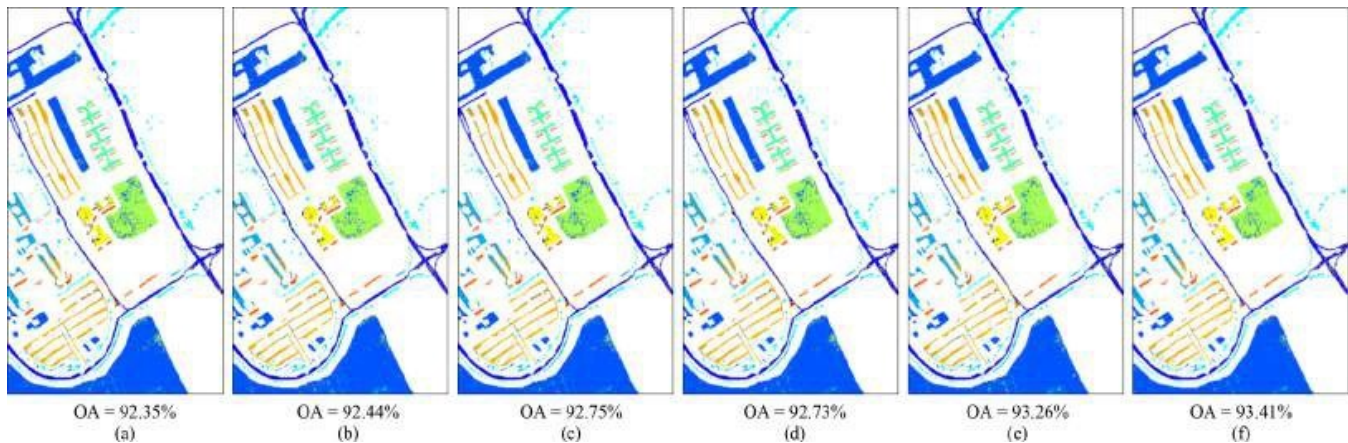


Fig. 9. Classification results (overall accuracy in percentage) of Pavia University obtained on (a) original data, (b) SSA processed data, (c) CT-MLR processed data, (d) WT-MLR processed data, (e) WT-SSA processed data, and (f) CT-SSA processed data.

For an $N \times N$ block, the computational complexity for the curvelet transform is given as $O(N^2 \log N)$ [31], while the computation of the wavelet transform requires $O(3/2(m_H + m_L)(1 - 1/4^J)N^2)$ flops for a block with the same size, where J is the decomposition level and m_H, m_L are the lengths of the filter [57]. With the CDF 9/7 filter used in the experiments, the lengths of the filter are both 9, i.e. $m_H = m_L = 9$. Therefore, when J is higher than 2, the computational complexity for the wavelet transform can be approximated as $O(27N^2)$, which is much higher than that of the curvelet transform. Besides, it should be noted that the processing time for MLR and SSA is largely dependent on the number of detail coefficient stacks and the number of total pixels in those stacks. However, it is not a problem for SSA, as fast implementation of SSA [10] has been proposed without

degrading the feature extraction performance. For the Indian Pines data set, the computing time of the proposed CT-SSA method (using the original SSA implementation) is about 5 min on a personal computer with an Intel Core i5-2400 CPU at 3.10 GHz using Matlab 2014a (Mathworks). Hence, it is reasonable to believe that with the fully optimised CurveLab toolbox and the fast implementation of SSA, the time cost of the proposed method should be much lower.

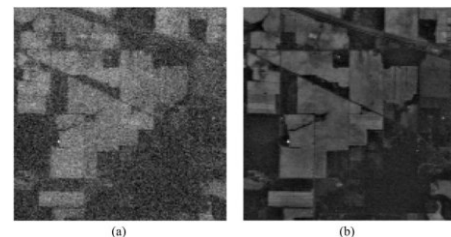


Fig. 10. Band 199 of Indian Pines in (a) original data and (b) CT-SSA processed data.

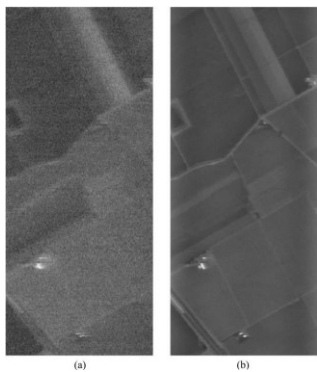


Fig. 11. Band 2 of Salinas Valley in (a) original data and (b) CT-SSA processed data.

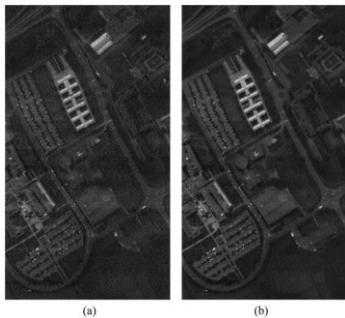


Fig. 12. Band 1 of Pavia University in (a) original data, (b) CT-SSA processed data.

B. Comparison with Other State-of-the-Art Spectral Processing Techniques

Although the curvelet transform is applied in the spatial domain of the hyperspectral images, the dominating part of the proposed denoising and feature extraction method, SSA, is applied in the spectral domain. As a matter of fact, the proposed CT-SSA approach actually belongs to the 1D spectral processing technique for hyperspectral images. Therefore, a few state-of-the-art 1D spectral feature extraction and dimensionality reduction techniques, including PCA, LDA, NMF are compared with the proposed approach for further performance assessments. Linear discriminant analysis is a standard supervised dimensionality reduction technique for pattern recognition, where the highest degree of class separability is obtained by maximizing the Raleigh quotient, i.e. the ratio of the between-class scatter matrix to the within-class scatter matrix [14]. Different from PCA which could only extract holistic features from the original HSI data, NMF is a parts-based learning algorithm, decomposing a non-negative matrix into two non-negative matrices which are more intuitive and interpretable. One of the decomposed matrices is formed with a set of basis vectors, which can be utilised to project the original data into the lower dimensional subspace [16]. The NMF algorithm is realized by the NMF Matlab toolbox [58]. Additionally, as an upgraded version of EMD, EEMD is also included for comparison as it overcomes the drawback of EMD, and therefore it outperforms EMD [55]. Being the fundamental part of the Hilbert-Huang transform (HHT), EMD is a non-linear and non-stationary signal decomposition method that

decomposes the signal into a finite number of intrinsic mode functions (IMFs) [59]. Each IMF represents a zero-mean frequency-amplitude modulation component that is often related to a specific physical process so that EMD is not biased with any predetermined basis. Although EMD has shown its efficacy in time series decomposition (e.g. speech recognition), results reported in [60] indicated that spectral feature extraction using EMD could potentially reduce the following classification accuracy. Consequently a more robust and data-driven technique, EEMD was proposed to alleviate the addressed problem. It is achieved by sifting an ensemble of white noise-added signal and similarly it generates a series of IMFs, where the sum of them represents the processed signal. In this paper, the fast EEMD Matlab toolbox is adopted [59]. The performance of PCA on the proposed approach for further feature extraction is also included for comparison.

The number of resulting features using PCA on both original data and CT-SSA processed data is tested within 5 – 50 in a step of 5 [11], while the dimensions of extracted features are tested from 2 to 10 for both LDA and NMF. Best results with highest classification accuracy are chosen for comparison. With regard to EEMD, it is suggested that the input white noise level ϵ_0 should be in the range of 0.1 – 0.4 and the number of ensemble N_E should be the order of 100 [61]. Therefore, we keep these parameters the same as [55], where $\epsilon_0 = 0.2$, $N_E = 200$ and the number of IMFs is 7. Experiments are also repeated ten times and comparison results are given in Table IV.

TABLE IV
MEAN OVERALL ACCURACIES (%) OF TEN REPEATED EXPERIMENTS ON THE ORIGINAL DATA SETS, SSA, CT-MLR, WT-MLR, WT-SSA, CT-SSA, CT-SSA-PCA AND SOME STATE-OF-THE-ART SPECTRAL FEATURE EXTRACTION TECHNIQUES PROCESSED DATA SETS, WITH DIMENSIONALITY OF FEATURES SHOWN IN PARENTHESES

Method	Data set		
	Indian Pines (10%)	Salinas Valley (5%)	Pavia University (4%)
Original	84.11 (200)	92.91 (204)	92.96 (103)
SSA	87.54 (200)	93.49 (204)	92.77 (103)
CT-MLR	91.93 (200)	95.18 (204)	93.34 (103)
WT-MLR	91.76 (200)	95.11 (204)	93.05 (103)
WT-SSA	93.87 (200)	96.08 (204)	93.67 (103)
CT-SSA	94.13 (200)	97.35 (204)	93.85 (103)
CT-SSA-PCA	95.75 (50)	98.33 (50)	91.78 (10)
PCA [12]	82.65 (50)	92.78 (40)	92.93 (50)
LDA [14]	87.03 (9)	91.62 (10)	92.06 (10)
NMF [58]	77.43 (10)	92.64 (10)	91.71 (10)
EEMD [59]	83.78 (200)	93.20 (204)	93.22 (103)

It can be seen that those dimensionality reduction techniques only give comparable results compared with original data set, even though they could speed up the classification process. As given in [9], with 10% training rate of Indian Pines, the highest classification OA achieved by EMD is 75.49%. Although EEMD has made an improvement of over 8% on this data set, it is not as good as the proposed CT-SSA approach. For Indian

Pines and Salinas Valley, adding an extra PCA step leads to further improvement in classification accuracy while reducing the classification time with less number of features. Even though the extra step of PCA has limited performance on the Pavia University data set, the proposed CT-SSA method still outperforms other spectral feature extraction techniques.

C. Comparison with Recent Spectral-spatial Classification Methods

For the majority of supervised classifiers, such as neural networks, decision trees, SVM and etc., the spectral feature-based classification chain is adopted. However, they are not able to incorporate spatial dependencies presented in the original scene into the classification process [62]. Several recent publications have demonstrated that the integration of spectral and spatial information could be beneficial to hyperspectral image classification problems, where some spectral-spatial classification methods, including the extended morphological profile (EMP) [47], the logistic regression and multi-level logistic (LMLL) [63], the edge-preserving filter (EPF) [64], image fusion and recursive filter (IFRF) [65] and the intrinsic image decomposition (IID) [56], have been proved to be superior to the spectral feature-based classification method. Therefore, a post-processing step is added to the proposed CT-SSA approach to increase the spatial consistency in the classification result, where a $T \times T$ spatial window is applied around each central pixel in the classification map and the final classified label is decided in favour of the class which appears most in the window. The window shape could be designed to conform to different scenes, but to make it simple, here the square window is used.

As can be noticed in ground truth maps in Fig. 3-5, Indian Pines has a smaller spatial area for some classes compared with Salinas Valley and Pavia University. Therefore, the spatial post-processing (SPP) window T is chosen as 5 for Indian Pines, 9 for Salinas Valley and 9 for Pavia University, respectively. Results using EMP, LMLL, EPF, IFRF and IID are cited directly from [56]. Since only 9 classes in Indian Pines are involved in our experiments, we have eliminated the other 7 classes from their results and recalculated AA and OA accordingly. For Salinas Valley in [56], 2% of pixels in each class have been used for training. Therefore, we have reperformed the experiment under the same training rate with other parameters unchanged. Numerical results are shown in Tables V–VII, where the proposed approach with spatial post-processing (CT-SSA-SPP) gives comparable accuracies with other spectral-spatial classification methods. For the proposed CT-SSA approach, it is found that the best overall accuracy values achieved are 94.13, 97.35 and 93.85 for the three datasets Indiana Pines, Salinas Valley and Pavia University, where the corresponding training ratios are 10%, 5% and 4%, respectively. With the introduced spatial post-processing and under the same or even less training ratios, the overall accuracy values for the three datasets have been dramatically improved to 98.40, 99.39 and 99.19 from the new

CCT-SSA-SPP approach. This has clearly demonstrated the efficacy of the SPP procedure in the proposed approach.

TABLE V
COMPARISON OF CLASS-BY-CLASS, AVERAGE AND OVERALL CLASSIFICATION ACCURACIES (%) OF THE CT-SSA-SPP APPROACH AND RECENT SPECTRAL-SPATIAL METHODS FOR THE INDIAN PINES DATA SET (10% TRAINING RATE)

Class	Method					
	EMP [47]	LMLL [63]	EPF [64]	IFRF [65]	IID [56]	CT-SSA- SPP (T=5)
2	86.63	90.64	95.03	95.41	96.90	96.18
3	89.74	87.23	95.68	96.07	97.20	99.63
5	93.23	96.39	98.39	98.07	97.88	97.08
6	97.50	99.90	98.90	98.62	99.40	100
8	100	100	99.49	100	100	100
10	86.18	93.49	82.09	95.26	95.49	95.45
11	93.82	89.75	95.93	98.42	98.03	99.38
12	87.09	98.60	92.81	96.72	97.23	97.32
14	99.59	92.45	99.11	99.75	99.76	99.96
AA	92.64	94.27	95.27	97.59	97.99	98.33
OA	92.45	92.57	95.07	97.56	97.89	98.40

TABLE VI
COMPARISON OF CLASS-BY-CLASS, AVERAGE AND OVERALL CLASSIFICATION ACCURACIES (%) OF THE CT-SSA-SPP APPROACH AND RECENT SPECTRAL-SPATIAL METHODS FOR THE SALINAS VALLEY DATA SET (2% TRAINING RATE)

Class	Method					
	EMP [47]	LMLL [63]	EPF [64]	IFRF [65]	IID [56]	CT-SSA- SPP (T=9)
1	99.93	99.83	100	99.99	100	100
2	99.60	99.96	100	100	99.97	100
3	96.75	99.80	98.36	99.89	99.85	100
4	98.25	99.22	97.17	97.32	97.80	99.93
5	99.14	98.75	99.96	99.94	99.96	99.73
6	99.95	99.66	99.98	100	99.98	100
7	99.87	99.55	99.98	99.70	99.90	100
8	94.79	85.97	91.78	99.34	99.07	98.15
9	99.41	99.69	99.65	99.99	99.86	99.92
10	95.10	94.68	93.88	99.81	98.80	99.69
11	96.39	98.47	98.80	99.26	99.98	100
12	98.79	100	99.99	99.90	99.99	99.99
13	98.87	98.48	100	99.28	100	100
14	97.06	98.60	99.19	97.96	97.51	97.15
15	86.67	78.78	85.50	97.07	99.52	99.04
16	97.31	99.35	99.99	99.95	97.92	99.99
AA	97.37	96.92	97.76	99.34	99.38	99.60
OA	96.11	93.55	95.59	99.27	99.45	99.39

TABLE VII
COMPARISON OF CLASS-BY-CLASS, AVERAGE AND OVERALL CLASSIFICATION ACCURACIES (%) OF THE CT-SSA-SPP APPROACH AND RECENT SPECTRAL-SPATIAL METHODS FOR THE PAVIA UNIVERSITY DATA SET (4% TRAINING RATE)

Class	Method					
	EMP [47]	LMLL [63]	EPF [64]	IFRF [65]	IID [56]	CT-SSA- SPP (T=9)
1	98.61	94.36	98.07	97.25	99.62	100
2	98.84	97.79	98.38	99.68	99.88	100
3	95.09	88.10	98.39	95.53	98.92	98.57
4	96.64	98.05	98.45	96.52	97.54	95.91
5	98.00	99.83	99.51	99.90	99.89	100
6	88.33	99.74	95.35	98.72	99.91	99.97

7	93.51	98.14	97.57	96.69	99.59	100
8	97.48	93.77	93.72	94.40	98.66	97.29
9	99.91	99.92	98.23	92.54	94.84	90.37
AA	96.27	96.63	97.63	96.80	98.76	98.01
OA	96.82	96.81	97.98	98.07	99.41	99.19

VI. CONCLUSIONS

In this paper, two powerful tools, SSA and the curvelet transform, are combined for HSI feature extraction. By applying SSA in the curvelet domain, noise can be smoothed from the decomposed signals, resulting in more effective feature extraction. Inspired by applying MLR in the curvelet domain and another method of solely applying SSA on HSI data, the proposed method takes advantage of those approaches. Using SVM on three publicly available data sets, i.e. Indian Pines, Salinas Valley and Pavia University for performance assessment, the proposed CT-SSA method significantly improves OA by as much as 10% compared with the result of the original Indian Pines data set. In addition, noise could be effectively removed by checking the band image visually. Compared with the inspirational approaches and some state-of-the-art spectral feature extraction techniques, the proposed method always stands out with the highest classification accuracies. Furthermore, it is noticed that with the spatial post-processing step, another 4% improvement in accuracy could be added to the final result of Indian Pines, while for Salinas Valley and Pavia University, classification results can even be improved to nearly 100%. It is worth noting that by adding the simple spatial post-processing step, the performance of the proposed denoising and feature extraction method is either comparable with in terms of accuracies or even higher than some recent spectral-spatial classification methods. The choice of the spatial post-processing techniques will be a topic of our future work.

Although in this paper the use of SSA is proposed for spectral processing in the curvelet domain, other filtering techniques could also give a relatively good result, where 2D filtering in the spatial domain in combination with SSA will be further investigated as well.

ACKNOWLEDGMENT

Authors would thank Prof. D. Landgrebe, Prof. L. Johnson and Prof. P. Gamba for providing experimental data sets online. We also wish to thank the Associated Editor and anonymous reviewers for their constructive suggestions to further improve the clarity and quality of the paper.

REFERENCES

- [1] T. Kelman, J. Ren, and S. Marshall, "Effective classification of Chinese tea samples in hyperspectral imaging," *Artif. Intell. Res.*, vol. 2, no. 4, pp. 87–96, Oct. 2013.
- [2] Y. Roggo, A. Edmond, P. Chalus, and M. Ulmschneider, "Infrared hyperspectral imaging for qualitative analysis of pharmaceutical solid forms," *Anal. Chim. Acta.*, vol. 535, no. 1-2, pp. 79–87, Apr. 2005.
- [3] K. Gill, J. Ren, S. Marshall, S. Karthick, and J. Gilchrist, "Quality-assured fingerprint image enhancement and extraction using hyperspectral imaging," in *ICDP 2011*, London, UK, 2011.
- [4] A. J. Brown, W. M. Calvin, P. C. McGuire, and S. L. Murchie, "Compact reconnaissance imaging spectrometer for Mars (CRISM) south polar mapping: first Mars year of observations," *J. Geophys. Res. - Planet.*, vol. 115, no. E2, Feb. 2010.
- [5] C. Zhao, X. Li, J. Ren, and S. Marshall, "Improved sparse representation using adaptive spatial support for effective target detection in hyperspectral imagery," *Int. J. Remote Sens.*, vol. 34, no. 24, pp. 8669–8684, Oct. 2013.
- [6] T. Qiao, J. Ren, M. Sun, J. Zheng, and S. Marshall, "Effective compression of hyperspectral imagery using an improved 3D DCT approach for land-cover analysis in remote-sensing applications," *Int. J. Remote Sens.*, vol. 35, no. 20, pp. 7316–7337, Oct. 2014.
- [7] X. Li, J. Ren, C. Zhao, T. Qiao, and S. Marshall, "Novel multivariate vector quantization for effective compression of hyperspectral imagery," *Opt. Commun.*, vol. 332, pp. 192–200, Dec. 2014.
- [8] J. Zabalza, J. Ren, M. Yang, Y. Zhang, J. Wang, S. Marshall, and J. Han, "Novel folded-pca for improved feature extraction and data reduction with hyperspectral imaging and SAR in remote sensing," *ISPRS J. Photogramm Remote Sens.*, vol. 93, no. 7, pp. 112–122, Jul. 2014.
- [9] J. Zabalza, J. Ren, W. Zheng, S. Marshall, and J. Wang, "Singular spectrum analysis for effective feature extraction in hyperspectral imaging," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 11, pp. 1886–1890, Nov. 2014.
- [10] J. Zabalza, J. Ren, Z. Wang, H. Zhao, J. Wang, and S. Marshall, "Fast implementation of singular spectrum analysis for effective feature extraction in hyperspectral imaging," *IEEE J. Sel. Topics Appl. Earth Observ. and Remote Sens.*, vol. 8, no. 6, pp. 2845–2853, Jun. 2015.
- [11] J. Zabalza, J. Ren, J. Zheng, J. Han, H. Zhao, S. Li, and S. Marshall, "Novel two-dimensional singular spectrum analysis for effective feature extraction and data classification in hyperspectral imaging," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 8, pp. 4418–4433, Aug. 2015.
- [12] J. Ren, J. Zabalza, S. Marshall, and J. Zheng, "Effective feature extraction and data reduction in remote sensing using hyperspectral imaging," *IEEE Signal Process. Mag.*, vol. 31, no. 4, pp. 149–154, Jul. 2014.
- [13] J. Wang and C.-I. Chang, "Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 6, pp. 1586–1600, Jun. 2006.
- [14] W. Liao, A. Pizurica, P. Scheunders, W. Philips, and Y. Pi, "Semisupervised local discriminant analysis for feature extraction in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 184–198, Jan. 2013.
- [15] U. Amato, R. M. Cavalli, A. Palombo, S. Pignatti, and F. Santini, "Experimental approach to the selection of the components in the minimum noise fraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 1, pp. 153–160, Jan. 2009.
- [16] X. Huang and L. Zhang, "An adaptive mean-shift analysis approach for object extraction and classification from urban hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 12, pp. 4173–85, 2008.
- [17] X. Geng, L. Ji, Y. Zhao, and F. Wang, "A small target detection method for the hyperspectral image based on higher order singular value decomposition (HOSVD)," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 6, pp. 1305–1308, Nov. 2013.
- [18] A. J. Brown, "Spectral curve fitting for automatic hyperspectral data analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 6, pp. 1601–1608, Jun. 2006.
- [19] D. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inf. Theory*, vol. 41, no. 3, pp. 613–627, May 1995.
- [20] H. Othman and S.-E. Qian, "Noise reduction of hyperspectral imagery using hybrid spatial-spectral derivative- domain wavelet shrinkage," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 2, pp. 397–408, Feb. 2006.
- [21] G. Chen and S.-E. Qian, "Denoising of hyperspectral imagery using principal component analysis and wavelet shrinkage," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 3, pp. 973–980, Mar. 2011.
- [22] D. Xu, L. Sun, J. Luo, and Z. Liu, "Analysis and denoising of hyperspectral remote sensing image in the curvelet domain," *Math. Probl. Eng.*, vol. 2013, 2013.

- [23] R.E. Roger and J.F. Arnold, "Reliably estimating the noise in AVIRIS hyperspectral images," *International Journal of Remote Sensing*, vol. 17, no. 10, pp. 1951-1962, 1996.
- [24] L. Gao, Q. Du, B. Zhang, W. Yang, and Y. Wu, "A comparative study on linear regression-based noise estimation for hyperspectral imagery," *IEEE JSTARS*, vol. 6, no. 2, pp. 488-498, 2013.
- [25] I. Selenick, R. Baraniuk, and N. Kingsbury, "The dual-tree complex wavelet transform," *IEEE Signal Process. Mag.*, vol. 22, no. 6, pp. 123-151, Nov. 2005.
- [26] J.-L. Starck, E. Candes, and D. Donoho, "The curvelet transform for image denoising," *IEEE Trans. Image Process.*, vol. 11, no. 6, pp. 670-684, Jun. 2002.
- [27] M. Do and M. Vetterli, "The contourlet transform: an efficient directional multiresolution image representation," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2091-2106, Dec. 2005.
- [28] F. Nencini, A. Garzelli, S. Baronti, and L. Alparone, "Remote sensing image fusion using the curvelet transform," *Inform. Fusion*, vol. 8, no. 2, pp. 143-156, 2007.
- [29] L. Sun and J. Luo, "Junk band recovery for hyperspectral image based on curvelet transform," *J. Cent. South Univ. T.*, vol. 18, pp. 816-822, 2011.
- [30] J.-L. Starck, F. Murtagh, E. Candes, and D. Donoho, "Gray and color image contrast enhancement by the curvelet transform," *IEEE Trans. Image Process.*, vol. 12, no. 6, pp. 706-717, Jun. 2003.
- [31] E. Candes, L. Demanet, D. Donoho, and L. Ying, "Fast discrete curvelet transforms," *Multiscale Model. Simul.*, vol. 5, no. 3, pp. 861-899, 2006.
- [32] N. Golyandina and A. Zhigljavsky, *Singular Spectrum Analysis for Time Series*. Berlin, Germany: Springer-Verlag, 2013.
- [33] N. Golyandina, V. Nekrutkin, and A. A. Zhigljavsky, *Analysis of Time Series Structure: SSA and Related Techniques*. London, UK: CRC Press, 2001.
- [34] E. Candes and D. Donoho, "Curvelets - a surprisingly effective nonadaptive representation for objects with edges," in *Curve and Surface Fitting*. Nashville, TN: Vanderbilt University Press, 1999, pp. 105-120.
- [35] S. Kouchaki, S. Sanei, E. Arbon, and D.-J. Dijk, "Tensor based singular spectrum analysis for automatic scoring of sleep eeg," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 23, no. 1, pp. 1-9, Jan. 2015.
- [36] W. Pereira, S. Bridal, A. Coron, and P. Laugier, "Singular spectrum analysis applied to backscattered ultrasound signals from in vitro human cancellous bone specimens," *IEEE Trans. Ultrason., Ferroelect., Freq. Control*, vol. 51, no. 3, pp. 302-312, Mar. 2004.
- [37] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778-1790, Aug. 2004.
- [38] E. Candes, L. Demanet, D. Donoho, and L. Ying. CurveLab Toolbox (v 2.1.2). Available: <http://www.curvelet.org/software.html>
- [39] S. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Trans. Image Process.*, vol. 9, no. 9, pp. 1532-1546, Sep. 2000.
- [40] D. Landgrebe. AVIRIS image Indian Pines. [Online]. Available: <https://engineering.purdue.edu/~biehl/MultiSpec>
- [41] L. Johnson. AVIRIS image Salinas Valley. [Online]. Available: <http://www.ehu.es/ccwintco/index.php>
- [42] P. Gamba. ROSIS image Pavia University. [Online]. Available: <http://www.ehu.es/ccwintco/index.php>
- [43] G. Camps-Valls and L. Bruzzone, Eds., *Kernel Methods for Remote Sensing Data Analysis*. Chichester, UK: Wiley, 2009.
- [44] A. Chakrabarty, O. Choudhury, P. Sarkar, A. Paul, and D. Sarkar, "Hyperspectral image classification incorporating bacterial foraging-optimized spectral weighting," *Artif. Intell. Res.*, vol. 1, no. 1, pp. 63-83, Sep. 2012.
- [45] R. Archibald and G. Fann, "Feature selection and classification of hyperspectral images with support vector machines," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 4, pp. 674-677, Oct. 2007.
- [46] A. Plaza, P. Martinez, J. Plaza, and R. Perez, "Dimensionality reduction and classification of hyperspectral image data using sequences of extended morphological transformations," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 466-479, Mar. 2005.
- [47] J. A. Benediktsson, M. Pesaresi, and K. Amason, "Classification and feature extraction for remote sensing images from urban areas based on morphological transformations," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 9, pp. 1940-1949, Sep. 2003.
- [48] D. L. Civco, "Artificial neural networks for land-cover classification and mapping," *Int. J. Geophys. Inf. Syst.*, vol. 7, no. 2, pp. 173-186, 1993.
- [49] H. Bischof and A. Leonardis, "Finding optimal neural networks for land use classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 1, pp. 337-341, Jan. 1998.
- [50] L. Bruzzone and D. F. Prieto, "A technique for the selection of kernel-function parameters in rbf neural networks for classification of remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 2, pp. 1179-1184, Mar. 1999.
- [51] P. Zhong and R. Wang, "Modeling and classifying hyperspectral imagery by crfs with sparse higher order potentials," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 2, pp. 688-705, Feb. 2011.
- [52] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 6, pp. 1351-1362, Jun. 2005.
- [53] G. P. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. Inf. Theory*, vol. 14, no. 1, pp. 55-63, Jan. 1968.
- [54] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines (version 3.20)", <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>
- [55] Z. He, Y. Shen, Q. Wang, and Y. Wang, "Optimized ensemble emd-based spectral features for hyperspectral image classification," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 5, pp. 1041-1056, May 2014.
- [56] X. Kang, S. Li, L. Fang, and J. A. Benediktsson, "Intrinsic image decomposition for feature extraction of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 2241-2253, Apr. 2015.
- [57] A. Ignat, "Combining features for texture analysis," in *Computer Analysis of Images and Patterns*. Springer, 2015, pp. 220-229.
- [58] Y. Li and A. Ngom, "The non-negative matrix factorization toolbox for biological data mining," *Source Code Biol. Med.*, vol. 8, no. 1, pp. 1-15, Apr. 2013.
- [59] Y.-H. Wang, C.-H. Yeh, H.-W. V. Young, K. Hu, and M.-T. Lo, "On the computational complexity of the empirical mode decomposition algorithm," *Phys. A*, vol. 400, pp. 159-167, Apr. 2014.
- [60] B. Demir and S. Ertürk, "Empirical mode decomposition of hyperspectral images for support vector machine classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4071-4084, Nov. 2010.
- [61] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: a noise-assisted data analysis method," *Adv. Adapt. Data Anal.*, vol. 1, no. 1, pp. 1-41, Jan. 2009.
- [62] M. Rojas, I. Döpfido, A. Plaza, and P. Gamba, "Comparison of support vector machine-based processing chains for hyperspectral image classification," in *Satellite Data Compression, Communications, and Processing VI*. International Society for Optics and Photonics, 2010.
- [63] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Hyperspectral image segmentation using a new bayesian approach with active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3947-3960, Oct. 2011.
- [64] X. Kang, S. Li, and J. A. Benediktsson, "Spectral-spatial hyperspectral image classification with edge-preserving filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2666-2677, May 2014.
- [65] —, "Feature extraction of hyperspectral images with image fusion and recursive filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3742-3752, Jun. 2014.



Tong Qiao received the B.Eng. degree (1st class) and M.Sc. degree (distinction) both in Electrical and Electronic Engineering from University of Manchester and University of Strathclyde in 2011 and 2012, respectively. In 2016, she was awarded a Ph.D. in Hyperspectral Imaging from the Centre for excellence in Signal and Image Processing (CeSIP), University of Strathclyde. Her research interests include hyperspectral imaging based applications and feature extraction for remote sensing image classification.



Jinchang Ren (M'07) received his B. E. degree in computer software, M.Eng. in image processing, D.Eng. in computer vision, all from Northwestern Polytechnical University, Xi'an, China. He was also awarded a Ph.D. in Electronic Imaging and Media Communication from Bradford University, Bradford, U.K.

Currently he is a Senior Lecturer with the Centre for excellence for Signal and Image Processing (CeSIP), also Deputy Director of the Strathclyde Hyperspectral Imaging Centre, University of Strathclyde, Glasgow, U.K. His research interests focus mainly on visual computing and multimedia signal processing, especially on semantic content extraction for video analysis and understanding and more recently hyperspectral imaging. He has published over 150 peer reviewed journal and conferences papers, and acts as an Associate Editor for two international journals including Multidimensional Systems and Signal Processing and Int. J. of Pattern Recognition and Artificial Intelligence.



Zheng Wang received the Ph.D. degree in Computer Science from Tianjin University (TJU), Tianjin, China, in 2009. He is now an Associate Professor in School of Computer Software, TJU.

During 2007 to 2008, he was a visiting scholar in INRIA institute, France. His current research interests include video analysis, hyperspectral imaging, and computer graphics.



Jaime Zabalza received the MEng in Industrial Engineering from the Universitat Jaume I (UJI), Spain, in 2006, the MAS in Electrical Technology at the Universitat Politècnica de Valencia, Spain, in 2010. From the University of Strathclyde, U.K., he was awarded the MSc in Electronic and Electrical Engineering with distinction in 2012 and Ph.D. in hyperspectral imaging in 2015.

From 2006 to 2011, he joined the UJI as a R&D engineer and worked in the Energy Technological Institute, Spain, in multidisciplinary fields comprising power electronics, automation and computer science. His interests are related to SAR and remote sensing, hyperspectral imaging and DSP, including signal processing in a wide range of applications.



Meijun Sun received the Ph.D. degree in Computer Science from Tianjin University (TJU), Tianjin, China, in 2009. She is now an Associate Professor in School of Computer Science and Technology, TJU. She was a visiting scholar of INRIA Institute, France, from 2007 to 2008. Her current research interests include computer graphics, hyperspectral imaging, and image processing.



Huimin Zhao, was born in Shaanxi, China, in 1966. He received the B.Sc. and the M.Sc. degree in signal processing in 1992 and 1997 from Northwestern Polytechnical University, Xian, China, respectively. He received the Ph.D. degree in electrical engineering from the Sun Yat-sen University in 2001. At present, he is a Professor with the Guangdong Polytechnic Normal

University. His research interests include image, video and information security technology.



Shutao Li (M'07-SM'15) received the B.S., M.S., and Ph.D. degrees from Hunan University, Changsha, China, in 1995, 1997, and 2001, respectively, all in electrical engineering.

In 2001, he joined the College of Electrical and Information Engineering, Hunan University. In 2013, he was granted the National Science Fund for Distinguished Young Scholars in China. He is currently a Full Professor with the College of Electrical and Information Engineering,

Hunan University. He is also a Chang-Jiang Scholar Professor appointed by the Ministry of Education of China. He has authored/coauthored more than 180 refereed papers. His professional interests include compressive sensing, sparse representation, image processing, and pattern recognition.

He is an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, a member of the Editorial Board of the Information Fusion and the Sensing and Imaging. He was a recipient of two 2nd-Grade National Awards at the Science and Technology Progress of China in 2004 and 2006.



Jón Atli Benediktsson (S'84-M'90-SM'99-F'04) received the Cand.Sci. degree in electrical engineering from the University of Iceland, Reykjavik, Iceland, in 1984 and the M.S.E.E. and Ph.D. degrees in electrical and computer engineering from Purdue University, West Lafayette, IN, USA, in 1987 and 1990, respectively.

He is currently the Rector and a Professor of electrical and computer engineering with the University of Iceland. He is a Cofounder of the biomedical start-up company Oxymap. His research interests are remote sensing, biomedical analysis of signals, pattern recognition, image processing, and signal processing, and he has published extensively in those fields.

Prof. Benediktsson was the 2011-2012 President of the IEEE Geoscience and Remote Sensing Society (GRSS), and he has been on the GRSS AdCom since 2000. He is a fellow of SPIE. He is a member of the Association of Chartered Engineers in Iceland (VFI), Societas Scintiarum Islandica, and Tau Beta Pi. He was an Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (TGRS) from 2003 to 2008, and he has served as an Associate Editor of TGRS since 1999, the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS since 2003, and IEEE ACCESS since 2013. He is on the International Editorial Board of the International Journal of Image and Data Fusion, and he was the Chairman of the Steering Committee of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (J-STARS) in 2007-2010. He received the Stevan J. Kristof Award from Purdue University in 1991 as an outstanding graduate student in remote sensing. He was the recipient of the Icelandic Research Councils Outstanding Young Researcher Award in 1997, he was granted the IEEE Third Millennium Medal in 2000, he was a corecipient of the University of Iceland's Technology Innovation Award in 2004, he received the yearly research award from the Engineering Research Institute of the University of Iceland in 2006, and he received the Outstanding Service Award from the IEEE Geoscience and Remote Sensing Society in 2007. He was the corecipient of the 2012 IEEE TGRS Paper Award. He received the 2013 IEEE/VFI Electrical Engineer of the Year Award, and a corecipient of the IEEE GRSS Highest Impact Paper Award.



Qingyun Dai received her Ph.D. degree in South China University of Technology. She is a Full Professor in the School of Information Engineering, the Director of Intelligent Signal Processing and Big Data Research Center and the Director of Science and Technology Research Office of the Guangdong University of Technology. Her research interests mainly include wavelets, image processing, image retrieval, pattern

recognition, manufacturing engineering systems, RFID, cloud computing and big data, etc.



Stephen Marshall received the BSc degree in electrical and electronic engineering and the PhD degree in image processing respectively from the University of Nottingham and the University of Strathclyde, U.K. With over 150 papers published, his research activities focus in nonlinear image processing and hyperspectral imaging. Currently he is a Professor with the Department of Electronic and Electrical Engineering in Strathclyde, and a Fellow of the IET.