PCA-domain Fused Singular Spectral Analysis for fast and Noise-Robust Spectral-Spatial Feature Mining in Hyperspectral Classification

Yijun Yan, Jinchang Ren, Qiaoyuan Liu, Huimin Zhao, Haijiang Sun, Jaime Zabalza

Abstract—The principal component analysis (PCA) and 2-D singular spectral analysis (2DSSA) are widely used for spectral domain and spatial domain feature extraction in hyperspectral images (HSI). However, PCA itself suffers from low efficacy if no spatial information is combined, whilst 2DSSA can extract the spatial information yet has a high computing complexity. As a result, we propose in this paper a PCA domain 2DSSA approach for spectral-spatial feature mining in HSI. Specifically, PCA and its variation, folded-PCA are utilized to fuse with the 2DSSA, as folded-PCA can extract both global and local spectral features. By applying 2DSSA only on a small number of PCA components, the overall computational complexity has been significantly reduced whilst preserving the discrimination ability of the features. In addition, with the effective fusion of spectral and spatial features, the proposed approach can work well on the uncorrected dataset without removing the noisy and water absorption bands, even under a small number of training samples. Experiments on two publicly available datasets have fully demonstrated the superiority of the proposed approach, in comparison to several state-of-the-art HSI classification methods and deep-learning models.

Index Terms—Hyperspectral image (HSI); spectral-spatial feature mining; principal component analysis (PCA); singular spectrum analysis (SSA).

I. INTRODUCTION

With rich spectral and spatial information in a 3-D hypercube, HSI can well characterize the material and objects based on their physical, e.g. moisture and temperature, and chemical properties. As a result, different HSI processing tasks, including data classification [1], spectral unmixing [2], and image restoration [3], have been explored to tackle various challenges in remote sensing.

An HSI is usually composed of 2-D scenes in hundreds of contiguous wavelengths, in which each pixel has a 1-D spectral signature [4]. Aside from spectral and spatial information, HSI data contains redundant content and noise due to environmental noise, sensor limitations and atmospheric impacts. As a result, even sophisticated classifiers like support vector machine (SVM) and deep learning (DL) models have limited classification accuracy. Herein, the bottleneck is how to derive the most representative features from the HSI data, i.e. spectral and spatial feature mining especially of the uncorrected dataset.

Considering the high redundancy in contiguous spectral bands, spectral feature extraction and dimensionality reduction has been popularly used in some early studies. Although PCA is most widely used for unsupervised dimension reduction and spectral feature extraction, it often fails to extract the useful local spectral information. To tackle this issue, several variations have been explored, such as a correlation based segmented PCA (SPCA) [6], where the spectral bands are segmented into groups for group based PCA followed by feature concatenation. In Tsai et al [7], a spectrally segmented PCA was proposed and shown better performance than PCA and SPCA for mapping of the plant species. Similar to the SPCA, FPCA was also developed to extract both the local and global structures in the spectral domain [8]. However, the main difference is that FPCA reallocates the spectrum of each pixel into a matrix form, based on which, a partial covariance matrix can be directly determined and accumulated for subsequent Eigenvalue decomposition and data projection. In this case, it can be more efficient and effective than PCA and SPCA. More recently, Uddin et al [9], proposed a Segmented-FPCA approach, which was superior to PCA, FPCA and SPCA. However, due to noise caused intra-class variations and high inter-class similarity, those methods still suffer from lack of robustness and limited discriminability.

Recently, a new techniques, named 1D-SSA [10], was developed for more effectively exploiting the spectral features. It can extract the trend from the original signal as well as the oscillations and noise components. By only taking the main trend and selected oscillations as features whilst abandoning the noisy components, the classification accuracy can be much improved. In an extended 2DSSA [4], spatial features can be effectively extracted for significantly improved classification accuracy. However, 1D-SSA and 2DSSA needs to be applied either to every pixel or every spectral band of the HSI, thus it is very time-consuming. To reduce the overall computational complexity whilst maintaining the classification accuracy, fast implementation of 1D-SSA and 2DSSA were also developed [11], though the overall reduction of computational cost is still very limited. Recently, a 1.5D-SSA [12] has been proposed for near real-time HSI analysis yet with a much compromised classification accuracy.

When applying the DL-based approaches to HSI, some models prominent in computer vision are modified for data

Y. Yan and J. Ren are with the National Subsea Centre, Robert Gordon University, Aberdeen, U.K. (J Ren (✉), email: jinchang.ren@ieee.org).

Q. Liu and H. Sun are with Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Chuangchun, China.

H. Zhao is with School of Computer Sciences, Guangdong Polytechnic Normal University, Guangzhou, China.

J. Zabalza are with Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow, UK.

classification [13]. Nonconvexity is also applied into DL models for improved interpretability in complicated real-world situations. As a result, many well-performing nonconvex DL models were investigated for HSI classification [3], often via extraction of spectral and/or spatial features, where the results can be fairly good in classification of HSI. However, they may suffer from either a very high computational cost or lack of sufficient training data. This is why classical machine learning models, such as SVM, is still widely used, in combination with an effective feature extractor, which may achieve comparable performance as deep learning models in classification of HSI for land cover mapping [14].

These challenges motivate us to propose a new framework of applying the 2DSSA on the PCA domain (PCA+2DSSA, FPCA+2DSSA), resulting in improved classification accuracy yet with significantly reduced computational complexity. By fusion of FPCA and PCA with 2DSSA, we further propose Fusion+2DSSA, for more improved data storage efficiency, classification accuracy and computation cost. The main contributions are summarized below:

1) We proposed a new framework of PCA domain 2DSSA for spectral-spatial feature extraction in HSI, where the computation cost can be significantly reduced whilst improving the classification accuracy.

2) In the proposed framework, three different schemes i.e., PCA+2DSSA, FPCA+2DSSA, and Fusion+2DSSA, are introduced to balance the efficiency and efficacy to satisfy various practical needs, with parameters adaptively determined for ease of implementation.

3) The superiority of our approach has been validated in two corrected HSI datasets and two uncorrected HSI datasets when benchmarked with traditional feature extraction methods and deep learning models.

II. PROPOSED APPROACH

Fig. 1 shows the workflow of the proposed method, which is composed of three main steps, i.e., spectral feature extraction and dimension reduction in HSI, 2DSSA based PCA domain spatial feature extraction, and feature fusion, followed by data classification using SVM as detailed below.

A. PCA based spectral feature mining in HSI

Given an HSI hypercube $D \in \mathfrak{R}^{D_x \times D_y \times D_\lambda}$, the spectral vector of a given pixel can be denoted as $x_n = [x_{n1}, x_{n2}, \dots, x_{nD_\lambda}]^T$,

where $n \in [1, N]$, and $N = D_x D_y$ is the total number of pixels. The mean-adjusted vector I_n of x_n will be used to calculate the covariance matrices of PCA.

$$C_{PCA} = \frac{1}{N} \sum_{n=1}^{N} I_n I_n^T \qquad (1)$$

Let $A_n \in \mathfrak{R}^{H \times W}$ be the converted matrix where H is the number of band group and W is the band number in each band group and $HW = D_\lambda$, the covariance matrices of FPCA can be obtained by [8]

$$C_{FPCA} = \frac{1}{N} \sum_{n=1}^{N} A_n^T A_n \qquad (2)$$

For a covariance matrix, the Eigen problem can be solved by decomposing C into the multiplication of three matrices as

$$C = \Lambda D \Lambda^T \qquad (3)$$

where D is the diagonal matrix composed by the Eigenvalues of C, and Λ denotes the orthonormal matrix composed by the corresponding Eigenvectors $[v_1, v_2, \dots, v_{D_\lambda}]$. To reduce the dimension of spectral features, top Eigenvectors corresponding bigger Eigenvalues are selected. For PCA, we take the first q_{PCA} components as the spectral features of x_n as follows.

$$x_n(PCA) = \Lambda^T I_n \in \mathfrak{R}^{1 \times q_{PCA}} \qquad (4)$$

For FPCA, we take the first $\hat{q}$ components for each band group, and the spectral features of x_n can be derived as

$$x_n(FPCA) = \Lambda^T A_n \in \mathfrak{R}^{H \times \hat{q}} \qquad (5)$$

where the total number of components in FPCA will be $q_{FPCA} = H\hat{q}$. For convenience, the spectral feature of D can be represented as $D(PCA) \in \mathfrak{R}^{D_x \times D_y \times q_{PCA}}$ and $D(FPCA) \in \mathfrak{R}^{D_x \times D_y \times q_{FPCA}}$.

B. PCA domain spatial feature extraction with 2DSSA

After spectral feature mining, the original HSI hypercube D is represented by PCA features $D(PCA) \in \mathfrak{R}^{D_x \times D_y \times q_{PCA}}$ and FPCA features $D(FPCA) \in \mathfrak{R}^{D_x \times D_y \times q_{FPCA}}$. Note that each of the PCA/FPCA component is actually of the same size as the original spectral band, i.e. $D_x \times D_y$, to which the 2DSSA [4] is applied to extract the spectral-spatial features. First, a squared window $L \in \mathfrak{R}^{L_x \times L_y}$, where, $L_x \in [1, D_x]$ and $L_y \in [1, D_y]$, is used to construct a trajectory matrix $T \in \mathfrak{R}^{S \times K}$ of featured image (embedding step) where $S = L_x \times L_y$, $K = (D_x - L_x + 1)(D_y - L_y + 1)$. Often, we have $L_x = L_y$ for simplicity.

For the derived trajectory matrix T, the singular value decomposition (SVD) is applied to extract the Eigenvalues $e_1 \geq e_2 \geq \dots \geq e_S$ and Eigenvectors $U \in \mathfrak{R}^{S \times S}$. As a result, T
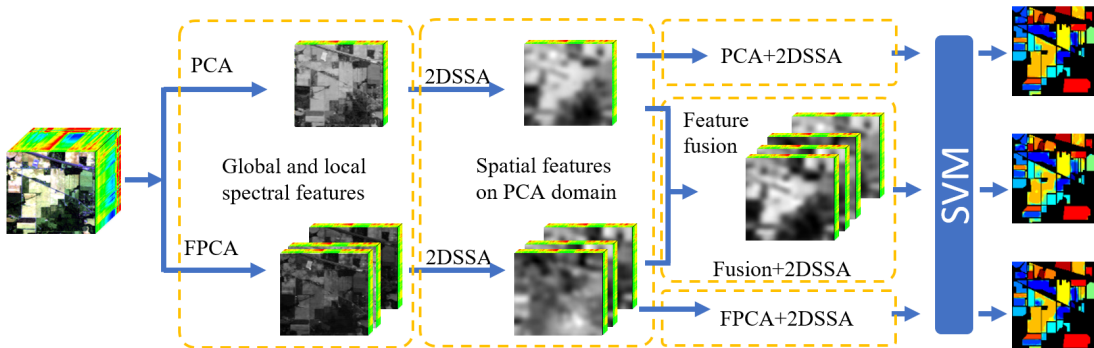

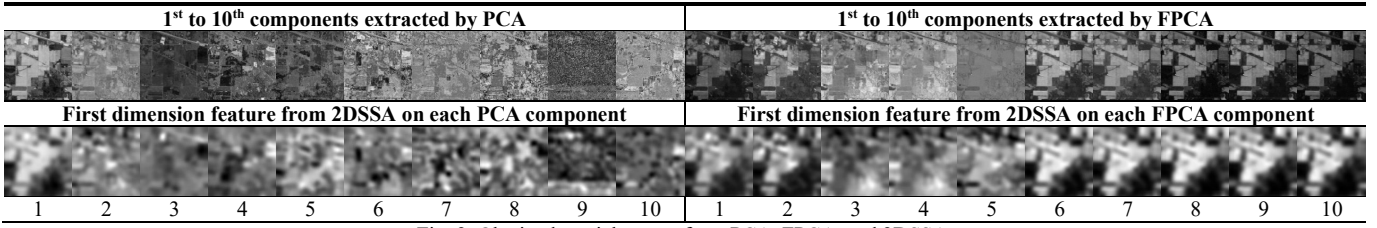Fig. 1. The workflow of our proposed PCA domain 2DSSA schemes.

1st to 10th components extracted by PCA	1st to 10th components extracted by FPCA

First dimension feature from 2DSSA on each PCA component	First dimension feature from 2DSSA on each FPCA component

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |

Fig. 2. Obtained spatial scenes from PCA, FPCA, and 2DSSA.

is decomposed in $T = T_1 + T_2 + \cdots + T_S$ components. After that, the grouping and diagonal averaging step are applied to invert the embedding step and obtain the reconstructed image Z. Accordingly, each featured image in $D(PCA)$ and $D(FPCA)$ can be represented by

$$D(\cdot)' = Z_1 + Z_2 + \cdots + Z_M = \sum_{m=1}^{M} Z_m \qquad (6)$$

where M is the number of selected Eigenvalues in the SVD. When $M = S$, the reconstructed image is equal to the original image. Herein, we denote $D(PCA + 2DSSA)$ and $D(FPCA + 2DSSA)$ as the PCA-based spectral-spatial features and FPCA-based spectral-spatial features, respectively.

For consistency, the same configuration of 2DSSA in [4] is adopted, where $L=10$ and only the first Eigenvalue component, $M=1$, i.e. the trend, is used. Although varying parameters may affect the final classification performance for different datasets, the overall difference from different configurations is estimated to be less than 1%. Therefore, the parameters L and M are set to 10 and 1 in all the experiments for simplicity.

C. Feature fusion

Applying the 2DSSA on the PCA/FPCA domains can reduce the computation cost compared to using 2DSSA on each band. On the other hand, as demonstrated in Fig. 2, the discrimination ability of features extracted from the Indian Pines dataset by PCA+2DSSA, and FPCA+2DSSA can be enhanced. In PCA, we choose $q_{PCA} = 10$. For FPCA, we have $H=10$, $W=20$ and $q_{FPCA} = 10$, i.e. the 200 bands are grouped into 10 groups, and only one component is extracted from each group.

As seen in Fig. 2, low-order PCA components have smooth spatial features while high-order PCA components are noisy. 2DSSA applied to PCA components makes these noisy components helpful in the derived trend signal. This has shown the added-value of 2DSSA on the PCA domain as the extracted spatial-domain trend signal can effectively suppress the noise and enhance the discrimination ability of the spectral-spatial features. On the other hand, PCA can extract the global spectral structure using a small number of low-order principal components, whilst FPCA can preserve local spectral features. As seen, PCA and FPCA features are quite supplementary to each other, which has motivated our fused solution below.

As the FPCA components are extracted from locally grouped spectral bands, they appear to be significantly smoother than those from PCA. This actually shows that FPCA is more robust to spectral noise, hence it has the potential to achieve noise-robust feature extraction and data classification in HSI, especially from the uncorrected dataset without removing the noisy and water absorption bands. On the other hand, the features extracted from FPCA seem to be more redundant, due

possibly to inappropriate grouping of bands. In addition, when applying 2DSSA to FPCA components, the effect of spatial smoothing is not as strong as those on the PCA components. This actually indicates potential limitations of FPCA+2DSSA hence the need for fusion with PCA+2DSSA.

For an HSI, the obtained spectral-spatial features $D(PCA + 2DSSA)$ and $D(FPCA + 2DSSA)$, can be separately used for classification of the HSI. Meanwhile, they can also be fused to form a combined feature vector, denoted as

$$\begin{aligned} D(Fusion + 2DSSA) = \{ &D(PCA + 2DSSA), \\ &D(FPCA + 2DSSA)\} \in \Re^{D_x \times D_y \times (q_{PCA} + q_{FPCA})} \end{aligned} \qquad (7)$$

The combined feature has a dimension of $(q_{PCA} + q_{FPCA})$, which can be much smaller than D_λ, though the spatial dimension remains the same. Note that q_{PCA} and q_{FPCA} here are adaptively decided as follows. For FPCA in Fusion+2DSSA, we divide each spectrum into 10 groups and select the first component of each group to form 10 combined components, i.e., $q_{FPCA} = 10$. For PCA, the q_{PCA} is decided based on the accumulated variance of the PCA components no less than a threshold of the total variance, and this threshold is empirically determined as 99.98% as it can help to produce particular good results for all the datasets. Accordingly, the q_{PCA} values for the Indian Pines and Salinas are adaptively determined as 90, and 20, respectively. To this end, the total number of combined features after the feature fusion for the Indian Pines and Salinas is 100 and 30, respectively. The detail experimental results and efficacy of the PCA+2DSSA, FPCA+2DSSA and Fusion+2DSSA schemes are presented in Section III.

III. EXPERIMENTS

A. Data description

In our experiments, two publicly available HSI datasets are used for performance evaluation. The first is Indian Pines, which is collected by the AVIRIS in 1992 in the USA. This dataset is labelled in 16 land cover classes and contains 145×145 pixels in 220 spectral bands. The second is Salinas, also collected by AVIRIS of the Salinas Valley in California, the USA, it has 512×217 pixels in 224 spectral bands labelled in 16 classes. After removing 20 noisy and water absorption bands, both HSI datasets will become corrected datasets.

B. Experimental Setup

The optimal numbers of principal components (PCs) for PCA, FPCA, PCA+2DSSA and FPCA+2DSSA are determined within [10, 100] at a step of 10 by maximizing the KP (%). To validate the efficacy of the extracted features, a standard Support Vector Machine (SVM) classifier [15] is employed for data classification. Consequently, the radical base function

TABLE I
CLASSIFICATION PERFORMANCE FOR INDIAN PINES DATASET WITH 5 PIXELS PER CLASS FOR TRAINING.

		RAW	PCA	FPCA	1DSSA	2DSSA	CCJSR	SuperPCA	JSRC	PCA+2DSSA	FPCA+2DSSA	Fusion+2DSSA
ND		200	20	20	200	200	200	30	200	40	40	100
Time(s)		0.14	0.27	0.74	12.01	8.78	38.05	11.11	68.37	3.31	3.06	7.28
Corrected	AA	60.40±2.58	53.77±2.36	65.08±2.31	65.99±2.07	73.43±2.21	70.37±1.61	*83.66±1.57*	77.21±1.71	83.28±1.93	77.38±2.15	**85.12±1.49**
	OA	46.48±4.07	41.41±3.02	51.50±3.46	53.03±4.16	59.64±4.50	56.65±4.43	71.77±2.87	64.01±2.99	*72.46±3.68*	62.99±3.65	**75.13±2.78**
	KP×100	40.36±4.18	34.83±2.94	45.84±3.66	47.47±4.47	54.96±4.75	51.50±4.39	68.29±3.04	59.70±3.19	*69.10±3.97*	58.80±3.89	**72.06±3.00**
Uncorrected	AA	56.62±1.42	53.77±2.36	64.58±1.71	62.71±2.71	75.31±1.96	70.62±1.23	83.99±1.84	77.18±1.82	*84.01±1.77*	77.40±2.01	**85.15±1.50**
	OA	43.14±3.36	41.41±3.02	51.47±3.95	48.61±4.23	61.53±3.12	56.82±4.37	*73.00±2.48*	62.96±3.10	72.56±3.09	63.41±4.15	**75.14±2.76**
	KP×100	36.63±3.27	34.83±2.94	45.74±4.11	42.67±4.48	56.97±3.26	51.69±4.32	*69.63±2.71*	58.66±3.22	69.28±3.32	59.24±4.48	**72.07±2.98**

TABLE II
CLASSIFICATION PERFORMANCE FOR SALINAS DATASET WITH 5 PIXELS PER CLASS FOR TRAINING.

		RAW	PCA	FPCA	1DSSA	2DSSA	CCJSR	SuperPCA	JSRC	PCA+2DSSA	FPCA+2DSSA	Fusion+2DSSA
ND		200	10	30	200	200	200	30	200	20	50	30
Time(s)		0.75	1,89	2.55	84.84	47.06	140.78	27.64	945.52	5.25	2.63	10.17
Corrected	AA	89.23±0.93	90.77±1.23	89.82±1.10	88.77±0.80	91.44±1.31	86.74±2.93	94.68±2.09	90.20±1,63	*96.13±0.68*	94.21±1.02	**96.46±0.52**
	OA	82.18±2.30	83.75±3.62	82.45±2.87	81.69±1.92	86.52±2.92	80.61±2.44	92.60±3.48	84.34±1.89	*93.34±0.95*	91.53±2.20	**93.86±0.75**
	KP×100	80.26±2.51	81.98±3.99	80.53±3.18	79.72±2.08	85.08±3.18	78.52±2.69	91.80±3.83	82.65±2.08	*92.61±1.05*	90.59±2.43	**93.19±0.83**
Uncorrected	AA	89.23±0.93	90.71±1.21	89.43±1.05	88.79±0.82	91.46±1.31	86.97±2.79	94.51±1.90	90.16±1.61	*96.13±0.69*	93.07±0.99	**96.46±0.50**
	OA	82.19±2.31	83.71±3.62	82.66±3.62	81.70±1.91	86.52±2.94	80.84±2.27	91.61±2.25	84.32±1.87	*93.44±1.01*	89.72±1.84	**94.01±0.76**
	KP×100	80.26±2.52	81.94±3.99	80.74±3.99	79.73±2.07	85.08±3.20	78.77±2.51	90.71±2.49	82.62±2.05	*92.73±1.11*	88.59±2.04	**93.35±0.84**

(RBF) is used as the kernel for the SVM, where the cost (c) and the gamma (γ) are optimized through a grid search [8]. The overall accuracy (OA), average accuracy (AA) and Kappa coefficient (KP) are used for quantitative evaluation. Each experiment was repeated 10 times, where training and testing samples are randomly selected without overlap. Average results are taken for statistical significance analysis and comparison.

C. Experimental results

The quantitative comparison between our proposed method and other benchmarking techniques on two HSI datasets is shown in Table I and Table II. ND is the number of feature dimensions. Time is the running time of each method. The best results and the second-best results are highlighted in bold and italic shading, respectively. The optimal selection of principal component number in PCA+2DSSA, FPCA+2DSSA, PCA and FPCA is decided after massive experiments. It can be seen that Fusion+2DSSA always leads to a higher accuracy, thanks to the strong fusion of PCA, FPCA and, making full use of local/global-spectral, and spatial information while suppressing data noise. PCA+2DSSA and FPCA+2DSSA consistently produce better results than 2DSSA, and this is because PCA and FPCA reduce the redundant information in the spectral domain making 2DSSA more effective. In contrast, absence of spatial information causes PCA, FPCA, and 1D-SSA to generate low accuracy in benchmarking approaches. In addition, PCA and 1DSSA produce worse performance than raw data in Indian Pines and Salinas, respectively. All these adverse factors reflect the importance of combining spatial and spectral features for HSI classification. Last but not the least, applying 2DSSA on PCA domain makes the computation cost much lower which reflects on the running time. It can be seen that three proposed schemes can produce faster and better classification results than 2DSSA. Among our three schemes, FPCA+2DSSA has the fastest running speed, Fusion+2DSSA has the best classification performance and PCA+2DSSA is a balanced

solution. Compared with other benchmarking methods such as CCJSR [16], SuperPCA [17] and JSRC[18], our methods are more effective and efficient.

D. Comparison with deep learning methods

To further validate the efficacy of our proposed method, we also do the comparison against another 4 deep learning models [18-21] using 200 training pixels per class (Table III). To be more specific, after removing classes with fewer than 200 pixels, only nine classes are used in the Indian Pines dataset. The experimental results shown that our proposed frameworks, Fusion+2DSSA and PCA+2DSSA, can consistently yield the best and second-best OA on both datasets. In this way, the effectiveness of our approaches is validated.

TABLE III
THE OA OF OUR PROPOSED METHODS COMPARED WITH DEEP-LEARNING METHODS USING 200 TRAINING SAMPLES PER CLASS

Datasets	[18]	[19]	[20]	[21]	PCA+2DSSA	FPCA+2DSSA	Fusion+2DSSA
Indian Pines	95.81	98.43	96.76	98.99	*99.04*	97.75	**99,36**
Salinas	96.07	98.33	97.42	99.58	*99.64*	98.51	**99.77**

E. Computational complexity

The suggested spectral-spatial fusion approach improves the efficiency of the standard 2DSSA by integrating PCA and FPCA to minimize dimensionality in the spectral domain. In this subsection, we briefly analyze the computational complexity and memory requirement of each implementation stage in Table IV and Table V. As seen the saving factor referring to 2DSSA in Table IV, applying 2DSSA on PCA domain decreases the 2DSSA band repetition process, which turns to lower computation burden. Fusion+2DSSA has slightly higher complexity than the other two, because of the fusion of both PCA and FPCA. As we only apply on the principal components, this has significantly reduced the computational cost from conventional. As presented in Table V, due to the fusion of the spectral and spatial processing, our proposed three

TABLE IV
COMPUTATIONAL COMPLEXITY IN THE DIFFERENT STAGES AND SAVING FACTORS REFERRING TO 2DSSA

Stage	PCA	FPCA	2DSSA	PCA+2DSSA	FPCA+2DSSA	Fusion+2DSSA
Covariance matrix	$ND_\lambda{}^2$	$ND_\lambda W$	$ND_\lambda{}^2$		$ND_\lambda W$	$ND_\lambda(D_\lambda + W)$
Eigen problem	$D_\lambda{}^3$	W^3	$D_\lambda{}^3$		W^3	$D_\lambda{}^3 + W^3$
Data projection	$ND_\lambda q_{PCA}$	NWq_{FPCA}	$ND_\lambda q_{PCA}$		NWq_{FPCA}	$ND_\lambda(q_{PCA} + q_{FPCA})$
Embed.	N/A	N/A	N/A	N/A	N/A	N/A
SVD	N/A	N/A	$(L^2K + L^3) \times D_\lambda$	$(L^2K + L^3) \times q_{PCA}$	$(L^2K + L^3) \times q_{FPCA}$	$(L^2K + L^3) \times (q_{PCA} + q_{FPCA})$
Grouping	N/A	N/A	$2LKM \times D_\lambda$	$2LKM \times q_{PCA}$	$2LKM \times q_{FPCA}$	$2LKM \times (q_{PCA} + q_{FPCA})$
D.Av.	N/A	N/A	$D_xD_y \times D_\lambda$	$D_xD_y \times q_{PCA}$	$D_xD_y \times q_{FPCA}$	$D_xD_y \times (q_{PCA} + q_{FPCA})$
Saving factor				$\approx 3D_\lambda/q_{PCA}$	$\approx 3D_\lambda/q_{FPCA}$	$\approx 3D_\lambda/(q_{PCA} + q_{FPCA})$

frameworks need more memory than the 2DSSA and PCA/FPCA alone. However, overall the memory requirement is modest, which is very close to the size of the hypercube. For Indian Pines and Salinas datasets, the memory requirements are only up to about 25M and 102M bytes respectively, a very small portion of the computer RAM at 32G or even more.

Therefore, it validates the computational efficiency of the proposed method. Detailed computational cost in terms of MACs, running time and memory need on two HSI datasets can be found in the supplementary material (Table S1-S3).

TABLE V
MEMORY REQUIREMENT OF DIFFERENT METHODS USING PCA/FPCA AND 2D SSA (L=10, M=1, 20 PCs).

	PCA	FPCA	2DSSA	PCA+2DSSA	FPCA+2DSSA	Fusion+2DSSA
Data matrix size	ND_λ	D_λ	ND_λ	$ND_\lambda + Nq_{PCA}$	$ND_\lambda + Nq_{FPCA}$	$N(q_{PCA} + q_{FPCA})$
Covariance matrix size	$D_\lambda{}^2$	W^2	L^2	$D_\lambda{}^2 + L^2$	$W^2 + L^2$	$D_\lambda{}^2 + W^2 + L^2$
Projection matrix size	$D_\lambda q_{PCA}$	Wq_{FPCA}/H	N/A	$D_\lambda q_{PCA}$	Wq_{FPCA}/H	$D_\lambda q_{PCA} + Wq_{FPCA}/H$

IV. CONCLUSION

In this letter, a novel PCA domain 2DSSA framework is proposed, where three schemes are introduced for noise robust spectral-spatial feature extraction. By applying PCA/FPCA in the PCA domain, the computational cost of band-wise 2DSSA can be significantly reduced whilst preserving the dominant spectral information for more effective data classification in HSI. Experiments on two publicly available datasets have fully validated both the efficiency and efficacy of the proposed framework. Among our proposed schemes, FPCA+2DSSA has the lowest computation cost, yet Fusion+2DSSA can produce consistently the best classification accuracy on the corrected and uncorrected datasets when benchmarked with several state-of-the-art approaches. Besides, PCA+2DSSA has relatively a good balance between the computation cost and the classification accuracy.

With the advantages of low computational cost, high classification accuracy and robustness to noise, the proposed methods have many potential application scenarios in hyperspectral remote sensing. As the future work, superpixel segmentation and band selection will be focused for improved spatial feature extraction and dimension reduction.

REFERENCES

[1] D. Hong et al., "Graph convolutional networks for hyperspectral image classification," IEEE Trans. on Geos. and Rem. Sens., 2020.

[2] D. Hong et al., "An augmented linear mixing model to address spectral variability for hyperspectral unmixing," IEEE Transactions on Image Processing, vol. 28, no. 4, pp. 1923-1938, 2018.

[3] D. Hong et al., "Interpretable hyperspectral artificial intelligence: When nonconvex modeling meets hyperspectral remote sensing," IEEE Geoscience and Remote Sensing Magazine, vol. 9, no. 2, pp. 52-87, 2021.

[4] J. Zabalza et al., "Novel 2D singular spectrum analysis for effective feature extraction and data classification in hyperspectral imaging," IEEE Trans. on Geos. and Rem. Sens., vol. 53, no. 8, pp. 4418-4433, 2015.

[5] X. Jia et al., "Segmented principal components transformation for efficient hyperspectral remote-sensing image display and classification," IEEE Trans. on Geos. and Rem. Sens., vol. 37, no. 1, pp. 538-542, 1999.

[6] F. Tsai et al., "Spectrally segmented principal component analysis of hyperspectral imagery for mapping invasive plant species," International Journal of Remote Sensing, vol. 28, no. 5, pp. 1023-1039, 2007.

[7] J. Zabalza et al., "Novel folded-PCA for improved feature extraction and data reduction with hyperspectral imaging and SAR in remote sensing," ISPRS Journal of Photog. and Rem. Sens., vol. 93, pp. 112-122, 2014.

[8] M. P. Uddin et al., "Effective feature extraction through segmentation-based folded-PCA for hyperspectral image classification," International Journal of Remote Sensing, vol. 40, no. 18, pp. 7190-7220, 2019.

[9] J. Zabalza et al., "Singular spectrum analysis for effective feature extraction in hyperspectral imaging," IEEE Geoscience and Remote Sensing Letters, vol. 11, no. 11, pp. 1886-1890, 2014.

[10] J. Zabalza, et al., "Fast implementation of two-dimensional singular spectrum analysis for effective data classification in hyperspectral imaging," J. of the Franklin Institute, vol. 355, no. 4, pp. 1733-1751, 2018.

[11] F. Hang et al., "A Novel Spectral-spatial Singular Spectrum Analysis Technique for Near Real-time In-Situ Feature Extraction in Hyperspectral Imaging," IEEE JSTARS, 2020.

[12] B. Rasti et al., "Feature extraction for hyperspectral imagery: The evolution from shallow to deep: Overview and toolbox," IEEE Geoscience and Remote Sensing Magazine, vol. 8, no. 4, pp. 60-88, 2020.

[13] A. Vali et al., "Deep Learning for Land Use and Land Cover Classification based on Hyperspectral and Multispectral Earth Observation Data: A Review," Remote Sensing, vol. 12, no. 15, p. 2495, 2020.

[14] C.-C. Chang et al., "LIBSVM: A library for support vector machines," ACM Trans. on Intel. syst. and tech. (TIST), vol. 2, no. 3, p. 27, 2011.

[15] B. Tu, et al., "Hyperspectral image classification via fusing correlation coefficient and joint sparse representation," IEEE Geoscience and Remote Sensing Letters, vol. 15, no. 3, pp. 340-344, 2018.

[16] J. Jiang et al., "SuperPCA: A superpixelwise PCA approach for unsupervised feature extraction of hyperspectral imagery," IEEE Trans. on Geos. and Rem. Sens., vol. 56, no. 8, pp. 4581-4593, 2018.

[17] T. Qiao et al., "Joint bilateral filtering and spectral similarity-based sparse representation," Pattern Recognition, vol. 77, pp. 316-328, 2018.

[18] P. Zhou et al., "Learning compact and discriminative stacked autoencoder for hyperspectral image classification," IEEE Transactions on Geoscience and Remote Sensing, vol. 57, no. 7, pp. 4823-4833, 2019.

[19] Y. Kong et al., "Spectral–spatial feature extraction for HSI classification based on supervised hypergraph and sample expanded CNN," IEEE JSTARS, vol. 11, no. 11, pp. 4128-4140, 2018.

[20] S. Mei et al., "Learning sensor-specific spatial-spectral features of hyperspectral images via convolutional neural networks," IEEE Trans. on Geos. and Rem. Sens., vol. 55, no. 8, pp. 4520-4533, 2017.

[21] G. Cheng et al., "Exploring hierarchical convolutional features for hyperspectral image classification," IEEE TGRS, vol. 56, no. 11, pp. 6712-6722, 2018.