

PCA-Domain Fused Singular Spectral Analysis for Fast and Noise-Robust Spectral–Spatial Feature Mining in Hyperspectral Classification

Yijun Yan¹, Member, IEEE, Jinchang Ren, Senior Member, IEEE, Qiaoyuan Liu², Huimin Zhao³, Haijiang Sun⁴, and Jaime Zabalza⁵, Member, IEEE

Abstract—The principal component analysis (PCA) and 2-D singular spectral analysis (2DSSA) are widely used for spectral and spatial-domain feature extraction in hyperspectral images (HSIs). However, PCA itself suffers from low efficacy if no spatial information is combined, while 2DSSA can extract the spatial information yet has a high computing complexity. As a result, we propose in this letter a PCA domain 2DSSA approach for spectral–spatial feature mining in HSI. Specifically, PCA and its variation, folded PCA (FPCA) are fused with the 2DSSA, as FPCA can extract both global and local spectral features. By applying 2DSSA only on a small number of PCA components, the overall computational cost can be significantly reduced while preserving the discrimination ability of the features. In addition, with the effective fusion of spectral and spatial features, our 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 validated the superiority of the proposed approach, in comparison to several state-of-the-art methods and deep learning models.

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

I. INTRODUCTION

WITH rich spectral and spatial information in a 3-D hypercube, hyperspectral image (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], change detection [2], and image denoising [3], have been explored to tackle various challenges in remote sensing.

An HSI is composed of 2-D scenes in hundreds of contiguous wavelengths, where each pixel has a 1-D spectral profile [4]. Aside from spectral and spatial information, HSI data contains redundant content and noise due to environmental effect, sensor limitations, and atmospheric impacts. Thus, even sophisticated classifiers, e.g., support vector machine (SVM) and deep learning (DL), have limited classification accuracy. Here, 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 have been popularly used in some early studies. Although principal component analysis (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) [5], where the spectral bands are segmented into groups for group-based PCA followed by feature concatenation. In [6], a spectrally SPCA was proposed and showed better performance than PCA and SPCA for mapping of the plant species. Similar to the SPCA, folded PCA (FPCA) was also developed to extract both the local and global structures in the spectral domain [7]. 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.* [8] proposed a segmented-FPCA approach, which was superior to PCA, FPCA, and SPCA. However, due to noise caused intraclass variations and high interclass similarity, those methods still suffer from lack of robustness and limited discriminability.

Recently, a new technique, named 1-D singular spectrum analysis (SSA) [9], 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 trend and selected oscillations as features while abandoning the noisy components, the classification accuracy can be much improved. In an extended 2-D SSA (2DSSA) [4], spatial features can be effectively extracted for significantly improved classification accuracy. However, both 1-D-SSA and 2DSSA need to be applied either to each pixel or each spectral band of the HSI, and thus, it is very time-

Manuscript received 6 July 2021; revised 22 September 2021; accepted 14 October 2021. Date of publication 19 October 2021; date of current version 20 June 2023. This work was supported in part by the International Cooperation Project of CIOFMP with RGU under Grant Y9U933T190, the National Natural Science Foundation of China under 62072122, and Key Discipline Improvement Project of Guangdong Province (2022ZDJ015). (Corresponding authors: Jinchang Ren; Huimin Zhao.)

Yijun Yan is with the National Subsea Centre (NSC), Robert Gordon University (RGU), Aberdeen AB21 0BH, U.K.

Jinchang Ren is with the NSC, RGU, Aberdeen AB21 0BH, U.K., and also with the School of Computer Sciences, Guangdong Polytechnic Normal University (GPNU), Guangzhou 510665, China (e-mail: jinchang.ren@ieee.org).

Qiaoyuan Liu and Haijiang Sun are with the Changchun Institute of Optics, Fine Mechanics and Physics (CIOFMP), Chinese Academy of Sciences, Changchun 130033, China.

Huimin Zhao is with the School of Computer Sciences, GPNU, Guangzhou 510665, China (e-mail: zhaohuimin@gpnu.edu.cn).

Jaime Zabalza is with the Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow G1 1XQ, U.K.

This article has supplementary material provided by the authors and color versions of one or more figures available at <https://doi.org/10.1109/LGRS.2021.3121565>.

Digital Object Identifier 10.1109/LGRS.2021.3121565

1558-0571 © 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.

See <https://www.ieee.org/publications/rights/index.html> for more information.

consuming. To reduce the overall computational complexity while maintaining the classification accuracy, fast implementation of 1-D-SSA and 2DSSA was also developed [10], though the overall reduction of computational cost is still very limited. Recently, a tensor-SSA [11] was proposed for much improved classification accuracy.

When applying the DL-based approaches to HSI, some models prominent in computer vision are adapted for data classification [12]. Nonconvexity is also applied to DL models for improved interpretability in complicated real-world situations. When applying denoising for HSI classification [3], extraction of spectral and/or spatial features is focused, where the results can be fairly good in HSI classification. 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, are still widely used. By combining with an effective feature extractor, SVM may achieve comparable performance as DL models [13].

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 as follows.

- 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 while 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 DL 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 next.

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

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

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]

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

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

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

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 to bigger eigenvalues are selected. For PCA, we take the first q_{PCA} components as the spectral features of x_n as follows:

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

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

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

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

B. PCA Domain Spatial Feature Extraction With 2DSSA

After spectral feature mining, the original HSI hypercube D is represented by PCA features $D(\text{PCA}) \in \mathfrak{R}^{D_x \times D_y \times q_{\text{PCA}}}$ and FPCA features $D(\text{FPCA}) \in \mathfrak{R}^{D_x \times D_y \times q_{\text{FPCA}}}$. Note that each of the PCA/FPCA components 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 $S = L_x \times L_y$ and $K = (D_x - L_x + 1)(D_y - L_y + 1)$. 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 is decomposed in $T = T_1 + T_2 + \dots + T_S$ 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 $D(\text{PCA})$ and $D(\text{FPCA})$ can be represented by

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

where M is the number of selected Eigenvalues in the SVD. When $M = S$, the reconstructed image is equal to the original image. Here, we denote $D(\text{PCA} + 2\text{DSSA})$ and $D(\text{FPCA} + 2\text{DSSA})$ 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 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 band-wise operations. In addition, as shown in Fig. 2, the discrimination ability

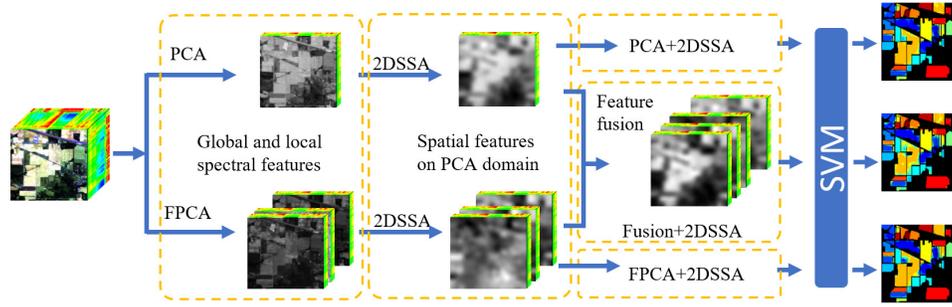


Fig. 1. Workflow of our proposed PCA domain 2DSSA schemes.

1 st to 10 th components extracted by PCA										1 st to 10 th components extracted by FPCA									
1										1									
2										2									
3										3									
4										4									
5										5									
6										6									
7										7									
8										8									
9										9									
10										10									
First dimension feature from 2DSSA on each PCA component										First dimension feature from 2DSSA on each FPCA component									
1										1									
2										2									
3										3									
4										4									
5										5									
6										6									
7										7									
8										8									
9										9									
10										10									

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

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 put into ten groups, and only one component is utilized from each group.

As shown in Fig. 2, low-order principal components (PCs) have smoothed the spatial features, while high-order PCs are quite noisy. Applying the 2DSSA to PCs can make these noisy components usable again 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 PCs, while 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 FPCA components are extracted from locally grouped spectral bands, they appear to be much smoother than those from PCA. This actually shows that FPCA is more robust to spectral noise, and 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, possibly due 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 and, hence, the need for fusion with PCA + 2DSSA.

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

$$\begin{aligned}
 & D(\text{Fusion} + 2\text{DSSA}) \\
 &= \{D(\text{PCA} + 2\text{DSSA})D(\text{FPCA} + 2\text{DSSA})\} \\
 &\in \mathfrak{R}^{D_x \times D_y \times (q_{PCA} + q_{FPCA})}. \tag{7}
 \end{aligned}$$

The combined feature has a dimension of $(q_{PCA} + q_{FPCA})$, which can be much smaller than D_x , 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 ten groups and select the first component of each group to form ten combined components, i.e., $q_{FPCA} = 10$. For PCA, 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 particularly 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 detailed experimental results and efficacy analysis 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 Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) in 1992 in the USA. This dataset is labeled in 16 land cover classes and contains 145×145 pixels in 220 spectral bands. The second is Salinas, also collected by AVIRIS over the Salinas Valley in CA, USA, and it has 512×217 pixels in 224 spectral bands labeled in 16 classes. After removing 20 noisy and water absorption bands, the corrected datasets of the two images are produced.

B. Experimental Setup

The optimal numbers of PCs for PCA, FPCA, PCA + 2DSSA, and FPCA + 2DSSA are determined within $[10, 100]$ at a step of 10 by maximizing the Kappa coefficient (KP) (%). To validate the efficacy of the extracted features, a standard SVM classifier [14] is employed for data classification. Consequently, the radial base function (RBF) is used as the kernel for the SVM, where the cost (c) and the gamma (γ) are optimized through a grid search [7]. The overall accuracy (OA), average accuracy (AA), and KP are used for quantitative evaluation. Each experiment was repeated ten times, where

TABLE I
CLASSIFICATION PERFORMANCE FOR INDIAN PINES DATASET WITH FIVE 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±2.6	53.77±2.4	65.08±2.3	65.99±2.1	73.43±2.2	70.37±1.6	<i>83.66±1.5</i>	77.21±1.7	83.28±1.9	77.38±2.1	85.12±1.5
	OA	46.48±4.1	41.41±3.0	51.50±3.5	53.03±4.2	59.64±4.5	56.65±4.4	71.77±2.9	64.01±3.0	<i>72.46±3.7</i>	62.99±3.6	75.13±2.8
	KP%	40.36±4.2	34.83±3.0	45.84±3.7	47.47±4.5	54.96±4.7	51.50±4.4	68.29±3.0	59.70±3.2	<i>69.10±4.0</i>	58.80±3.9	72.06±3.0
Uncorrected	AA	56.62±1.4	53.77±2.4	64.58±1.7	62.71±2.7	75.31±1.9	70.62±1.2	83.99±1.8	77.18±1.8	<i>84.01±1.8</i>	77.40±2.0	85.15±1.5
	OA	43.14±3.4	41.41±3.0	51.47±4.0	48.61±4.2	61.53±3.1	56.82±4.4	<i>73.00±2.5</i>	62.96±3.1	72.56±3.1	63.41±4.1	75.14±2.7
	KP%	36.63±3.3	34.83±3.0	45.74±4.1	42.67±4.5	56.97±3.3	51.69±4.3	<i>69.63±2.7</i>	58.66±3.2	69.28±3.3	59.24±4.5	72.07±2.9

TABLE II
CLASSIFICATION PERFORMANCE FOR SALINAS DATASET WITH FIVE 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±0.9	90.77±1.2	89.82±1.1	88.77±0.8	91.44±1.3	86.74±2.9	94.68±2.1	90.20±1.6	<i>96.13±0.7</i>	94.21±1.0	96.46±0.5
	OA	82.18±2.3	83.75±3.6	82.45±2.9	81.69±1.9	86.52±2.9	80.61±2.4	92.60±3.5	84.34±1.9	<i>93.34±0.9</i>	91.53±2.2	93.86±0.7
	KP%	80.26±2.5	81.98±4.0	80.53±3.2	79.72±2.1	85.08±3.2	78.52±2.7	91.80±3.8	82.65±2.1	<i>92.61±1.0</i>	90.59±2.4	93.19±0.8
Uncorrected	AA	89.23±0.9	90.71±1.2	89.43±1.0	88.79±0.8	91.46±1.3	86.97±2.8	94.51±1.9	90.16±1.6	<i>96.13±0.7</i>	93.07±1.0	96.46±0.5
	OA	82.19±2.3	83.71±3.6	82.66±3.6	81.70±1.9	86.52±2.9	80.84±2.3	91.61±2.2	84.32±1.9	<i>93.44±1.0</i>	89.72±1.8	94.01±0.7
	KP%	80.26±2.5	81.94±4.0	80.74±4.0	79.73±2.1	85.08±3.2	78.77±2.5	90.71±2.5	82.62±2.0	<i>92.73±1.1</i>	88.59±2.0	93.35±0.8

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 Tables I and 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 the PC number in PCA + 2DSSA, FPCA + 2DSSA, PCA, and FPCA is decided after massive experiments. As seen, Fusion + 2DSSA always leads to a higher accuracy, due to the strong fusion of PCA and 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, the absence of spatial information causes PCA, FPCA, and 1-D-SSA to generate low accuracy in benchmarking approaches. In addition, PCA and 1-D-SSA 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 least, applying 2DSSA on PCA domain makes the computation cost much lower, which reflects on the running time. It can be seen that the 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 correlation coefficient and joint sparse representation (CCJSR) [15], SuperPCA [16], and joint bilateral filtering and spectral similarity-based sparse representation (JSRC) [17], our methods are more effective and efficient.

D. Comparison With DL Methods

To further validate the efficacy of our proposed method, we also do the comparison against another four DL mod-

TABLE III
OA OF OUR PROPOSED METHODS COMPARED WITH DL METHODS USING 200 TRAINING SAMPLES PER CLASS

Datasets	[18]	[19]	[20]	[21]	PCA+	FPCA+	Fusion+
					2DSSA	2DSSA	2DSSA
Indian Pines	95.81	98.43	96.76	98.99	<i>99.04</i>	97.75	99.36
Salinas	96.07	98.33	97.42	99.58	<i>99.64</i>	98.51	99.77

els [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 show 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.

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 section, we briefly analyze the computational complexity and memory requirement of each implementation stage in Tables IV and 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 PCs, this has significantly reduced the computational cost from conventional. As shown in Table V (D.M, C.M, and P.M represent the size of input data matrix, covariance matrix, and projection matrix, respectively), our proposed three frameworks need slightly more memory than the 2DSSA and PCA/FPCA alone due to the fusion of the spectral and spatial processing. However, the overall 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 25 and 102 MB, respectively, a very small portion of the computer RAM at 32 G or even more.

This has validated the computational efficiency of the proposed method. The detailed comparison of

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_λ^2	$ND_\lambda W$		ND_λ^2	$ND_\lambda W$	$ND_\lambda(D_\lambda + W)$
Eigen problem	D_λ^3	W^3		D_λ^3	W^3	$D_\lambda^3 + W^3$
Data projection	$ND_\lambda q_{PCA}$	$NW q_{FPCA}$		$ND_\lambda q_{PCA}$	$NW q_{FPCA}$	$ND_\lambda(q_{PCA} + q_{FPCA})$
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	$D_x D_y \times D_\lambda$	$D_x D_y \times q_{PCA}$	$D_x D_y \times q_{FPCA}$	$D_x D_y \times (q_{PCA} + q_{FPCA})$
Saving factor				$\approx 3D_\lambda/q_{PCA}$	$\approx 3D_\lambda/q_{FPCA}$	$\approx 3D_\lambda/(q_{PCA} + q_{FPCA})$

TABLE V

MEMORY REQUIREMENT OF DIFFERENT METHODS IN DIFFERENT STAGES USING PCA/FPCA AND 2DSSA ($L = 10$, $M = 1$, AND 20 PCS)

Stage	PCA	FPCA	2DSSA	PCA+2DSSA	FPCA+2DSSA	Fusion+2DSSA
D.M	ND_λ	D_λ	ND_λ	ND_λ $+ Nq_{PCA}$	ND_λ $+ Nq_{FPCA}$	$ND_\lambda + N$ $(q_{PCA} + q_{FPCA})$
C.M	D_λ^2	W^2	L^2	$D_\lambda^2 + L^2$	$W^2 + L^2$	$D_\lambda^2 + W^2 + L^2$
P.M	$D_\lambda q_{PCA}$	$W q_{FPCA} / H$	N/A	$D_\lambda q_{PCA}$	$W q_{FPCA} / H$	$D_\lambda q_{PCA}$ $+ W q_{FPCA} / H$

multiply-accumulates (MACs), running time, and memory requirements on the two HSI datasets can be found in the Supplementary Material (Tables S1–S3).

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 while 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, L. Gao, J. Yao, B. Zhang, A. Plaza, and J. Chanussot, "Graph convolutional networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5966–5978, Jul. 2021.
- [2] Y. Li *et al.*, "CBANet: An end-to-end cross-band 2-D attention network for hyperspectral change detection in remote sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, p. 5513011, 2023.
- [3] P. Ma, "Multiscale superpixelwise prophet model for noise-robust feature extraction in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, p. 5508912, 2023.
- [4] J. Zabalza *et al.*, "Novel 2D 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.
- [5] X. Jia and J. A. Richards, "Segmented principal components transformation for efficient hyperspectral remote-sensing image display and classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 1, pp. 538–542, Jan. 1999.
- [6] F. Tsai, E. K. Lin, and K. Yoshino, "Spectrally segmented principal component analysis of hyperspectral imagery for mapping invasive plant species," *Int. J. Remote Sens.*, vol. 28, no. 5, pp. 1023–1039, Mar. 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 J. Photogramm. Remote Sens.*, vol. 93, pp. 112–122, Jul. 2014.
- [8] M. P. Uddin, M. Al Mamun, and M. A. Hossain, "Effective feature extraction through segmentation-based folded-PCA for hyperspectral image classification," *Int. J. Remote Sens.*, vol. 40, no. 18, pp. 7190–7220, Sep. 2019.
- [9] J. Zabalza, J. Ren, Z. Wang, 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, C. Qing, P. Yuen, G. Sun, H. Zhao, and J. Ren, "Fast implementation of two-dimensional singular spectrum analysis for effective data classification in hyperspectral imaging," *J. Franklin Inst.*, vol. 355, no. 4, pp. 1733–1751, 2018.
- [11] H. Fu, G. Sun, A. Zhang, B. Shao, J. Ren, and X. Jia, "Tensor singular spectrum analysis for 3D feature extraction in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023.
- [12] B. Rasti *et al.*, "Feature extraction for hyperspectral imagery: The evolution from shallow to deep: Overview and toolbox," *IEEE Geosci. Remote Sens. Mag.*, vol. 8, no. 4, pp. 60–88, Dec. 2020.
- [13] A. Vali, S. Comai, and M. Matteucci, "Deep learning for land use and land cover classification based on hyperspectral and multispectral Earth observation data: A review," *Remote Sens.*, vol. 12, no. 15, p. 2495, Jul. 2020.
- [14] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, p. 27, 2011.
- [15] B. Tu, X. Zhang, X. Kang, G. Zhang, J. Wang, and J. Wu, "Hyperspectral image classification via fusing correlation coefficient and joint sparse representation," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 3, pp. 340–344, Mar. 2018.
- [16] J. Jiang, J. Ma, C. Chen, Z. Wang, Z. Cai, and L. Wang, "SuperPCA: A superpixelwise PCA approach for unsupervised feature extraction of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4581–4593, Aug. 2018.
- [17] T. Qiao *et al.*, "Joint bilateral filtering and spectral similarity-based sparse representation," *Pattern Recognit.*, vol. 77, pp. 316–328, Mar. 2018.
- [18] P. Zhou, J. Han, G. Cheng, and B. Zhang, "Learning compact and discriminative stacked autoencoder for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4823–4833, Jul. 2019.
- [19] Y. Kong, X. Wang, and Y. Cheng, "Spectral–spatial feature extraction for HSI classification based on supervised hypergraph and sample expanded CNN," *IEEE J. Sel. Topics Appl. Earth Observat., Remote Sens.*, vol. 11, no. 11, pp. 4128–4140, Nov. 2018.
- [20] S. Mei, J. Ji, J. Hou, X. Li, and Q. Du, "Learning sensor-specific spatial-spectral features of hyperspectral images via convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 8, pp. 4520–4533, Aug. 2017.
- [21] G. Cheng, Z. Li, J. Han, X. Yao, and L. Guo, "Exploring hierarchical convolutional features for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 11, pp. 6712–6722, Nov. 2018.