Functional Feature Extraction for Hyperspectral Image Classification With Adaptive Rational Function Approximation

Zhijing Ye, Tao Qian, Liming Zhang, *Member, IEEE*, Lei Dai, Hong Li, *Member, IEEE*, and Jón Atli Benediktsson, *Fellow, IEEE*

***Abstract*— A functional feature extraction method based on rational function approximation for hyperspectral image (HSI) classification is proposed. In digital imagery, the spectral information of a pixel can be regarded as a 1-D signal. An HSI is composed of these 1-D signals arranged in a certain spatial structure. According to the functional characteristic of hyperspectral data, 1-D signals can be approximated by a linear combination of basis functions. Thus, a joint rational basis function system (JRBFS) based on class adaptivity is here first built for an HSI by adaptive Fourier decomposition (AFD). Second, the functional representations (FRs) and corresponding reconstructed spectral curves are obtained by decomposing the original spectral information in a JRBFS. Furthermore, the functional spectral–spatial features are extracted on the basis of FRs by an edge-preserving filtering method, FR-EPFs. Finally, the functional spectral–spatial features are used for HSI classification by SVM. Experimental results for five commonly used HSI data sets demonstrate the effectiveness and advantages of the proposed method FR-EPFs.**

***Index Terms*— Adaptive Fourier decomposition (FD) (AFD), functional spectral–spatial features, hyperspectral image (HSI) classification, rational orthogonal function system.**

Manuscript received May 18, 2020; revised October 12, 2020 and December 22, 2020; accepted January 7, 2021. Date of publication January 29, 2021; date of current version August 30, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 62001337 and Grant 61877021, in part by the Natural Science Foundation of Hubei Province under Grant 2020CFB136, in part by the Science and Technology Development Fund of Macao SAR FDCT under Grant 079/2016/A2 and Grant 0123/2018/A3, in part by the University of Macau Multi-year Research under Grant MYRG2018-00111-FST and Grant MYRG2018-00168-FST, and in part by the Fundamental Research Funds for the Central Universities under Grant WUT: 2020IB003. *(Corresponding author: Liming Zhang.)*

Zhijing Ye is with the School of Science, Wuhan University of Technology, Wuhan 430070, China (e-mail: xkinghust@163.com).

Tao Qian is with the Macao Center for Mathematical Sciences, Macau University of Science and Technology, Macau, China (e-mail: tqian@must.edu.mo).

Liming Zhang and Lei Dai are with the Faculty of Science and Technology, University of Macau, Macau, China (e-mail: lmzhang@um.edu.mo; yb87427@um.edu.mo).

Hong Li is with the School of Mathematics and Statistics, Huazhong University of Science and Technology, Wuhan 430074, China (e-mail: hongli@hust.edu.cn).

Jón Atli Benediktsson is with the Faculty of Electrical and Computer Engineering, University of Iceland, 107 Reykjavik, Iceland (e-mail: benedikt@hi.is).

Digital Object Identifier 10.1109/TGRS.2021.3052807

I. Introduction

With the rapid development of hyperspectral remote sensing technology, hyperspectral sensors can collect data in the electromagnetic spectrum from the visible to the near-infrared wavelength ranges, which can be processed to form hyperspectral images (HSIs) with the high spatial resolution and high spectral resolution. In HSIs, each pixel records the spectral information of dozens or even hundreds of continuous bands corresponding to some land cover. The rich spectral information makes accurately discriminating land covers of interest a possibility. Due to the advantage, HSIs have been widely used in geological exploration, precision agriculture, environmental monitoring, and so on [1]. However, the development of hyperspectral remote sensing technology also faces some problems, such as large data volume, high dimensionality, information redundancy, and processing efficiency to be improved [2].

HSI classification is an important research topic, which has attracted the attention of many researchers. HSI classification aims at assigning a unique class label to each pixel in HSIs. Because of the high dimensionality of the data, the limited number of available labeled samples, and the spatial variability of the spectral information [3]–[5], HSI classification is a very challenging task. Early classification methods for HSI data were mainly based on spectral feature matching, which relies heavily on the spectral library. In the past two decades, with extensive research on machine learning theory, many HSI classification methods based on machine learning have been proposed. With respect to whether spatial information is used or not, HSI classification methods are mainly divided into two categories: classification methods based on spectral information only (spectral classification methods) and classification methods for combining spatial and spectral information (spatial–spectral classification methods).

Classical spectral classification methods mainly identify the class of each pixel based on the statistical characteristics of the available spectral information. Considering how prior knowledge is used, classification methods are mainly divided into three categories: unsupervised learning, supervised learning, and semisupervised learning [6], in which supervised learning is probably the most commonly used approach. In this article, our focus is on supervised classification. Common supervised

classifiers include K-nearest neighbor, support vector machine (SVM), random forest, multiple logistic regression, deep learning [3], [7], and so on. Due to the high dimensionality, strong correlation and redundancy of the spectral information, and insufficient labeled samples, the direct classification of the spectral information by supervised classifiers is usually not successful in terms of accuracies. Therefore, in general, the spectral features are first extracted from the spectral information by many classical feature extraction methods, such as principal component analysis (PCA) [8], independent component analysis [9], discriminant analysis [10], nonparametric weighted feature extraction [11], orthogonal total variation component analysis (OTVCA) [12], and their variants [13]. Then, supervised classifiers are used for classifying the features after dimensionality reduction. Although these feature extraction methods have achieved good results in HSI classification, these methods usually treat the spectral information as discrete vectors and ignore their spectral arrangement structures and the law of continuous change with the wavelength [14]. In other words, if the spectral information is reordered, the new spectral information would not change the original classification results, but the land covers they represent actually would change substantially [15]. This is a disadvantage of machine learning methods that are separated from physical (imaging) models by only considering the data itself. Based on the continuous characteristics of spectral information, this article continues to explore a new functional feature extraction method especially suitable for HSIs and even higher dimensional ultraspectral images from a functional point of view.

Functional data learning is based on functional data analysis [16] and statistical machine learning theory. It focuses on machine learning problems with continuously varying data (i.e., functional data). Its salient feature is that the object it analyzes and processes is a function, not a vector. The spectral information of each pixel can be regarded as a sampling of the reflectance curve as a function of the wavelength. This remarkable feature of an HSI makes it have inherent functional characteristics, which provides a theoretical basis for exploring and studying functional feature extraction methods in HSIs. In recent years, a few researchers have carried out exploratory works in this direction. Li *et al.* [17] proposed a functional data fitting model based on the B-Spline basis function system and the functional principal component analysis (FPCA) method for HSI classification. Ye *et al.* [18] improved the functional data fitting models based on the B-Spline basis function system and proposed the functional data discriminant analysis (FDDA) method for HSI classification. Lv *et al.* [19] proposed a spatial–spectral classification method based on spatial FPCA with local mean filtering of HSIs. Zullo *et al.* [20] experimentally compared some classical machine learning methods with functional data learning methods, showing that functional data learning methods have significant advantages in the case of a small number of labeled samples and noise labels. Several existing methods have mainly been proposed based on the B-Spline basis function system and FPCA. Although the B-Spline basis function system shows a good approximation performance in nonperiodic data processing, it faces some problems, such as fixed basis functions, the high dimensionality of FRs, and slow convergence speed. Moreover, it needs enough basis functions to reach a preset error. Therefore, more effective basis function systems need to be further developed. The rational basis function system established by AFD has significant advantages in discrete signal and image decomposition, which has been studied in a series of work of Qian *et al.* [21]–[23]. However, the rational basis function system is a challenge in modeling and calculation [24], [25]. Here, it is considered to introduce it into the function space to establish new functional feature extraction methods for HSI classification.

In the past decade, research on spatial–spectral classification methods has attracted much attention, and research on such classification methods has gradually become a hot topic in the field of hyperspectral remote sensing [26]. The limited number of available labeled samples and the spatial variability of the spectral information in HSI classification make the accuracies obtained by spectral classification methods unsatisfactory. To overcome these problems, many scholars have tried to mine the spatial information of HSIs based on the assumption of homogeneous regions and spatial consistency in order to improve the classification performance. Heretofore, some spatial information extraction techniques have been used to implement spatial–spectral classification of HSIs, such as local mean filtering [27], extended morphological and attribute profiles [28], [29], invariant attribute profiles [30], superpixel segmentation [31], and neighboring sparse representation methods [32], [33]. Recently, the edge-preserving filtering method has also been used successfully for spatial information extraction, giving a very good performance [34]. However, direct use of edge-preserving features (EPFs) extracted by the edge-preserving filtering method in HSIs is not effective enough, and the dimensionality of the features is high [35]. Although the edge-preserving filtering method combined with some simple band selection and feature extraction methods can overcome the problems [36], extracting more discriminative spatial features is still a key task in HSI classification, especially within the framework of functional data learning.

Based on our previous work [18], in this article, we propose a new functional feature extraction method for HSI classification with adaptive rational function approximation. In this approach, in order to take full advantage of the high dimensionality, strong correlation, and redundancy of spectral information, a rational basis function system is at first used to fit the spectral information into a spectral curve. Note that the advantage of rational function approximation is that it can adaptively select the parameters of basis functions according to different spectral curves, in order to achieve the best approximation. Then, a JRBFS based on class adaptivity is built for different land covers by AFD. With the JRBFS, we use the idea of minimum mean squared error to establish a functional data fitting model in the function space. By solving the model, we can obtain the spectral curve corresponding to the spectral information and its FRs. Due to the JRBFS being an orthonormal system, the FRs can fully characterize the corresponding spectral curve. Next, the functional spectral–spatial feature extraction method FR-EPFs is built based on the

FRs by the edge-preserving filtering. Finally, SVM is used for the classification task of the functional spectral–spatial features. The proposed functional feature extraction method can not only effectively handle the problems of the high dimensionality, strong correlation, and redundancy of the spectral information but also mitigate the lack of available labeled samples and the spatial variability of the spectral information. The contributions of this article consist of the following.

1) A rational basis function system with adaptivity is built by AFD, which can produce a good fit for spectral curves from different classes with few basis functions. Compared with the two most commonly used basis function systems in functional data analysis, i.e., the Fourier basis function system and the B-spline basis function system, the rational basis function system has a faster convergence speed and gives a better approximation.
2) From a functional point of view, by introducing rational approximation into the functional data learning framework, a functional data fitting model based on the JRBFS is constructed. Compared with two classical functional data fitting models, the proposed model has significant advantages in terms of reconstruction and classification as can be seen by comparing classification accuracies of the spectral information before and after fitting.
3) The use of spatial information is a challenge in the framework of functional data learning. However, the FRs can be regarded as features of the spectral curves under the JRBFS. Thus, the functional spectral–spatial features are extracted by FR-EPFs. Compared with some popular spectral–spatial feature extraction methods, the proposed method FR-EPFs can demonstrate better classification performances.

The rest of this article is organized into five sections. Section II introduces the rational function systems built by AFD and certifies its effectiveness and advantages in spectral curve fitting. Section III presents the JRBFS based on class adaptivity and preliminarily demonstrates the superiority of the functional data fitting model based on the JRBFS in spectral classification. Section IV builds the spectral–spatial classification with the functional spectral–spatial features. Section V displays the experimental accuracies on five HSI data sets and the analysis of the proposed method in comparison with some popular methods. Finally, conclusions are drawn in Section VI.

II. Mathematical Foundation

With the functional characteristics of hyperspectral data, each pixel of the HSI records a complete and continuous spectral curve. However, due to the influence of noise, the spectral curve directly connected by the spectral vector is not very regular, even with varying degrees of jagged changes. To overcome the noise and reconstruct the spectral curve, the rational orthogonal function system based on weighted Blaschke products is built for approximating the intrinsic spectral curve [37].

A. Rational Function System

This section mainly studies the decomposition of 1-D spectral signals by AFD in the weighted Blaschke products. AFD is a new transformation of real-valued functions in classical Lebesgue L^2 spaces, especially L^2 space in the unit circle. AFD aims at expanding a signal by adaptively selecting a suitable Takenaka–Malmquist (TM) system [38], which is a rational orthogonal function system. At this point, AFD and classical Fourier decomposition (FD) share similar ideas and principles. However, different from the FD with a fixed basis (Fourier basis function system) in a Hilbert space, the rational function system (also known as the TM system) in AFD is adaptively selected based on the given spectral signal. In other words, different signals produce different TM systems. The Fourier basis function system is a special case of the TM system. On the other hand, AFD can be regarded as a further extension of FD. Compared with FD, AFD is more flexible and adaptable in signal decomposition. The mathematical principles of AFD are given in the following.

Assume that we have an HSI $\boldsymbol{X} \in \mathbb{R}^{N\times m}$ consisting of N samples $\{\boldsymbol{x}_i\}_{i=1}^{N}$, and the m elements in the spectral vector $\boldsymbol{x}_i = [\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{im}]$ are acquired in sequence from the corresponding spectral signal (curve) $x_i(t)$, where t denotes the wavelength. For convenience, the range of wavelengths t is transformed into $[0, 2\pi]$. Then, the jth element in $\boldsymbol{x}_i$ can be transformed into a point at the wavelength t_j on the signal $x_i(t)$, that is

$$\boldsymbol{x}_{ij} = x_i(t_j), \quad j = 1, 2, \ldots, m. \tag{1}$$

Utilizing (1), the raw discrete data are transformed into functional data. Next, we use the suitable basis function system to approximate the functional data.

In practice, any spectral signal $x(t) \in \{x_i(t)\}_{i=1}^{N}$ is energy-limited, namely, $||x|| < \infty$ ($x \in L^2$). According to FD, the following formula holds:

$$x(t) = \sum_{k=-\infty}^{+\infty} c_k e^{\mathbf{i}kt}, \quad \sum_{k=-\infty}^{+\infty} |c_k|^2 < \infty. \tag{2}$$

Here, $c_k = \langle x, e^{\mathbf{i}kt}\rangle = (1/2\pi)\int_0^{2\pi} x(t)e^{-\mathbf{i}kt}dt$, and $\mathbf{i}$ is the imaginary unit. Although the Fourier basis function system is a complete standardized orthogonal system and has some important mathematical properties, it faces the following problems in actual signal decomposition. First, the Fourier expansion usually converges slowly, for the entries $c_k e^{\mathbf{i}kt}$ in the expansion that build up the essential part of the total energy may arrive late. Thus, sufficient expansion entries $c_k e^{\mathbf{i}kt}$ are required to achieve the preset approximation effect. Second, the fixed basis system $\{e^{\pm\mathbf{i}kt}\}_{k=0}^{\infty}$ is selected. For a massive amount of spectral signals from different classes of land covers, some of the signals may get a good approximation in the fixed basis system, others may not. To avoid the problems, the AFD built in the Hardy space is introduced to achieve the decomposition of spectral signals.

Let us define the mathematical notation used. $\mathbb{C}$ is the complex plane. $D \triangleq \{z = re^{\mathbf{i}t} \in \mathbb{C} : 0 \le r < 1\}$ denotes an open unit disk centered on the origin in the complex plane $\mathbb{C}$. $\partial D \triangleq \{z = e^{\mathbf{i}t} \in \mathbb{C}\}$ denotes a unit circle around D. ∂D^+ is the set ∂D with the anticlockwise direction. The complex Hardy space $H^2(D)$ is the class of analytic functions f on

the unit disk D, which satisfies the following condition:

$$f(z)=\sum_{k=0}^{+\infty}c_k z^k,\ \sum_{k=0}^{+\infty}|c_k|^2<\infty. \tag{3}$$

In the following, we transform the study of the energy finite function into a function in the Hardy space $H^2(D)$. There are two main advantages: one is the deterministic of the function decomposition in the Hardy space, i.e., specific representations under the optimal rational basis function system; the other is applying the mature and effective complex analysis methods for the function study in the Hardy space, such as Cauchy's integral theorem. According to the Fourier expansion in (2), the spectral signal $x(t)$ can be transformed as a function $f(z)$ in the Hilbert space $L^2(\partial D)$ defined on ∂D, that is

$$x(t)=\sum_{k=-\infty}^{+\infty}c_k e^{\mathbf{i}kt}=\sum_{k=-\infty}^{+\infty}c_k z^k\triangleq f(z),\quad z=e^{\mathbf{i}t}. \tag{4}$$

Furthermore, the function $f(z)$ in (4) can be decomposed into the sum of two parts, i.e., a positive frequency part f^+ and a negative frequency part f^-

$$f(e^{\mathbf{i}t})=\sum_{k=0}^{+\infty}c_k e^{\mathbf{i}kt}+\sum_{k=-\infty}^{-1}c_k e^{\mathbf{i}kt}\triangleq f^+(e^{\mathbf{i}t})+f^-(e^{\mathbf{i}t}). \tag{5}$$

Due to the fact that $x(t)$ is a real-valued signal, the following holds: $c_{-k}=\bar{c}_k$. Here, $\bar{c}_k$ is the complex conjugate of c_k. Thus, (5) can be translated into

$$\begin{aligned}f(e^{\mathbf{i}t})&=\sum_{k=0}^{+\infty}c_k e^{\mathbf{i}kt}+\sum_{k=-\infty}^{0}c_k e^{\mathbf{i}kt}-c_0\\&=\sum_{k=0}^{+\infty}\left(c_k e^{\mathbf{i}kt}+c_{-k}e^{-\mathbf{i}kt}\right)-c_0\\&=\sum_{k=0}^{+\infty}\left(c_k e^{\mathbf{i}kt}+\bar{c}_k e^{-\mathbf{i}kt}\right)-c_0\\&=\sum_{k=0}^{+\infty}\left(c_k e^{\mathbf{i}kt}+\overline{c_k e^{\mathbf{i}kt}}\right)-c_0\\&=2\mathrm{Re}(\sum_{k=0}^{+\infty}\left(c_k e^{\mathbf{i}kt}\right)-c_0\\&=2\mathrm{Re}(f^+(e^{\mathbf{i}t}))-c_0\end{aligned} \tag{6}$$

where $\mathrm{Re}(\cdot)$ means taking the real part of a function. Based on (6), we can now discuss the decomposition of the positive frequency part f^+ instead of the original spectral signal f. Furthermore, $f^+(e^{\mathbf{i}t})$ is boundary values of the analytic function $f^+(z)\in H^2(D)$, that is

$$\lim_{r\to 1^-}f^+(re^{\mathbf{i}t})=\lim_{r\to 1^-}\sum_{k=0}^{+\infty}c_k r^k e^{\mathbf{i}kt}=\sum_{k=0}^{+\infty}c_k e^{\mathbf{i}kt}=f^+(e^{\mathbf{i}t}). \tag{7}$$

We learn that approximation to functions in the Hardy space $H^2(D)$ implies that also to those in the Hilbert space $L^2(\partial D)$. Thus, based on the transformations in (2)–(7), we study the decomposition of $f^+\in H^2(D)$ instead of the original spectral signal $f\in L^2(\partial D)$.

According to the spectral signal f, the analytic function f^+ can be obtained through the Hilbert transform $\mathcal{H}$ on the unit circle ∂D

$$\begin{aligned}f^+&=\frac{1}{2}(f+\mathbf{i}\mathcal{H}f+c_0)\\\mathcal{H}f(e^{\mathbf{i}t})&\triangleq \mathrm{P.V.}\frac{1}{2\pi}\int_0^{2\pi}\frac{f(e^{\mathbf{i}\tau})}{e^{\mathbf{i}t}-e^{\mathbf{i}\tau}}de^{\mathbf{i}\tau}.\end{aligned} \tag{8}$$

Note that the above integral is understood in the principal value (P.V.) sense. The Hilbert transform is very important in signal processing, where it derives the analytic representation f^+ of a real-valued signal f. Under the FD, f^+ in (8) is consistent with the one in (5), denoting the positive frequency part of f. This article aims at studying the decomposition of the spectral signal in $H^2(D)$. A TM system $\{B_k(z)\}_{k=1}^{+\infty}$ consists of rational functions in the Hardy space $H^2(D)$ that can approximate any functions $f^+(z)$ in the same Hardy space $H^2(D)$

$$f^+(z)=\sum_{k=1}^{+\infty}c_k B_k(z),\quad \sum_{k=-\infty}^{+\infty}|c_k|^2<\infty \tag{9}$$

where $c_k=\langle f^+,B_k\rangle$, and the TM system $\{B_k(z)\}_{k=1}^{+\infty}$ satisfies the following conditions [21]:

$$B_k(z)=\begin{cases}\dfrac{\sqrt{1-|a_1|^2}}{1-\bar{a}_1 z} & k=1\\ \dfrac{\sqrt{1-|a_k|^2}}{1-\bar{a}_k z}\displaystyle\prod_{l=1}^{k-1}\frac{z-a_l}{1-\bar{a}_l z} & k=2,3,\ldots\end{cases}\qquad a_l\in D. \tag{10}$$

Here, $e_{a_k}(z)=((1-|a_k|^2)^{1/2}/1-\bar{a}_k z)$ is the normalized Szegö kernel at a_k [39]. Although there is no orthogonality between the $\{e_{a_k}(z)\}_{k=1}^{+\infty}$, with the attachments $\prod_{l=1}^{k-1}(z-a_l/1-\bar{a}_l z)$ as Blaschke products, any TM system is orthonormal. In fact, by using the Cauchy formula, one can show the following properties.

1) Reproducing property

$$\begin{aligned}\langle F,e_{a_k}\rangle&=\sqrt{1-|a_k|^2}F(a_k)\quad \forall F\in H^2(\partial D^+)\\&\text{especially, }\langle e_{a_k},e_{a_k}\rangle=1.\end{aligned} \tag{11}$$

2) Normalization

$$\langle B_k,B_k\rangle=1,\quad k=1,2,3,\ldots \tag{12}$$

3) Orthogonality

$$\langle B_k,B_l\rangle=0\quad \forall k\neq l. \tag{13}$$

Thus, the TM system is an orthonormal system, which depends on the choices of the complex parameters $\{a_k\}_{k=1}^{+\infty}\subset D$. In particular, when $a_1=a_2=\cdots=a_k\cdots=0$, this TM system becomes the Fourier basis function system $\{z^k\}_{k=0}^{+\infty}$ $(z=e^{\mathbf{i}t})$. The AFD method aims at searching the optimal parameters $\{a_k\}_{k=1}^{+\infty}$ to construct the optimal TM system for different spectral signals. Thus, for a given HSI $\boldsymbol{X}=\{\boldsymbol{x}_i\}_{i=1}^N$, the functional data fitting model based on the TM system can be built by minimizing the sum of mean-squared errors (SMSE)

$$\min_{a_1,a_2,\ldots,a_k,\ldots}\mathrm{SMSE}=\sum_{i=1}^{N}\left\|x_i(t)-\sum_{k=1}^{+\infty}c_{ik}B_k(z)\right\|_2^2. \tag{14}$$

In the following, we take $N = 1$ as an example to formulate AFD for the single component case to build the optimal TM system.

B. Adaptive Fourier Decomposition

Because HSIs generally contain noise, we do not need to decompose the signal completely to obtain its infinite representation coefficients and rational basis functions. On the contrary, in HSI processing based on functional data analysis, we hope to use very few basis functions (i.e. very few parameters $\{a_k\}_{k=1}^K$) to approximate the original spectral curve with a very small error. Thus, the analytic signal f^+ derived from any spectral signal f in (8) can be represented as a linear combination of a finite TM system with a noise item, that is

$$f^+(z) = \sum_{k=1}^{K} c_k B_k(z) + \varepsilon(z) \tag{15}$$

where $\varepsilon(z)$ is the noise function. Combining with the complete expansion (9), the noise function $\varepsilon(z)$ can be denoted as

$$\varepsilon(z) = \sum_{k=K+1}^{+\infty} c_k B_k(z). \tag{16}$$

It is obvious that the noise is expected to have as little energy as possible in signal processing. Thus, the functional data fitting model of a sample can be built based on the TM system

$$\begin{aligned} \min_{a_1,a_2,\ldots,a_K} \text{MSE} &= ||f^+(z) - \sum_{k=1}^{K} c_k B_k(z)||_2^2 \\ &= \langle \varepsilon(z), \varepsilon(z) \rangle \\ &= \left\langle \sum_{k=K+1}^{+\infty} c_k B_k(z), \sum_{k=K+1}^{+\infty} c_k B_k(z) \right\rangle \\ &= \sum_{k=K+1}^{+\infty} |c_k|^2. \end{aligned} \tag{17}$$

Simultaneously, by utilizing the basis expansion in (9) and the properties in (13)–(15), We can derive the following energy relationship of the signal decomposition:

$$||f^+||_2^2 = \sum_{k=1}^{+\infty} |c_k|^2 = \sum_{k=1}^{K} |c_k|^2 + \sum_{k=K+1}^{+\infty} |c_k|^2. \tag{18}$$

Because the energy of the signal f^+ is finite, the optimization problem (17) is equivalent to the following model:

$$\begin{aligned} &\min_{a_1,a_2,\ldots,a_K} \left\| f^+(z) - \sum_{k=1}^{K} c_k B_k(z) \right\|_2^2 \\ &\quad \Leftrightarrow \min_{a_1,a_2,\ldots,a_K} \sum_{k=K+1}^{+\infty} |c_k|^2 \\ &\quad \Leftrightarrow \max_{a_1,a_2,\ldots,a_K} \sum_{k=1}^{K} |c_k|^2 = \sum_{k=1}^{K} |\langle f^+, B_k \rangle|^2. \end{aligned} \tag{19}$$

Thus, the minimization problem of the sum of mean-squared errors is transformed into a modulus maximization problem of the basis expansion coefficients. Next, we focus on searching the optimal parameters $\{a_k\}_{k=1}^K \subset D$ and make the energy of the signal f^+ maximize on these projection directions (the TM basis functions $\{B_k(z)\}_{k=1}^K$). In practice, the optimization process of the parameters is along with the process of achieving the Gram–Schmidt orthogonalization of $\{e_{a_k}(z)\}_{k=1}^K$. This is the core idea of AFD. Here, we use a step-by-step optimization strategy to find an optimal parameter K-tuple in turn.

Let $f_1 = f^+$, and then

$$f^+(z) = \langle f_1, e_{a_1} \rangle e_{a_1}(z) + r_1(z) \tag{20}$$

where $r_1(z)$ is the error term at the first step. According to the model (19) and the reproducing property (11), the first optimal parameter a_1 can be obtained by the maximal projection principle (MPP)

$$\begin{aligned} \hat{a}_1 &= \arg\max_{a_1 \in D} |\langle f_1, e_{a_1} \rangle|^2 \\ &= \arg\max_{a_1 \in D} (1 - |a_1|^2) |f_1(a_1)|^2. \end{aligned} \tag{21}$$

The first basis function $\hat{B}_1(z) = e_{\hat{a}_1}(z)$ can be determined by the parameter $\hat{a}_1$. Thus, the first error term $r_1(z)$ can be written as

$$r_1(z) = f_1(z) - \langle f_1, \hat{B}_1 \rangle \hat{B}_1(z) = f_1(z) - \frac{1 - |\hat{a}_1|^2}{1 - \bar{\hat{a}}_1 z} f_1(\hat{a}_1). \tag{22}$$

It is observed that $z = \hat{a}_1$ is a zero point of the error term $r_1(z)$, and $r_1(z)$ is orthogonal to $\hat{B}_1(z)$. The following transform at $\hat{a}_1$ is used for removing the zero point of $r_1(z)$:

$$f_2(z) = \frac{r_1(z)}{\tau_{\hat{a}_1}(z)} = \frac{f_1(z) - \langle f_1, \hat{B}_1 \rangle \hat{B}_1(z)}{\frac{z - \hat{a}_1}{1 - \bar{\hat{a}}_1 z}} \tag{23}$$

and achieving $f_2(z) \in H^2(D)$. The expansion (20) can be transformed into

$$f^+(z) = \langle f_1, \hat{B}_1 \rangle \hat{B}_1(z) + f_2(z) \frac{z - \hat{a}_1}{1 - \bar{\hat{a}}_1 z}. \tag{24}$$

Next, we continue to implement a decomposition step for $f_2(z)$ similar to $f_1(z)$ by searching the second optimal parameter a_2, that is

$$f_2(z) = \langle f_2, e_{a_2} \rangle e_{a_2}(z) + r_2(z) \tag{25}$$

which can also be obtained by solving the following optimization problem based on MPP [22]:

$$\begin{aligned} \hat{a}_2 &= \arg\max_{a_2 \in D} |\langle f_2, e_{a_2} \rangle|^2 \\ &= \arg\max_{a_2 \in D} (1 - |a_2|^2) |f_2(a_2)|^2. \end{aligned} \tag{26}$$

By substituting (25) and (26) into (24), the expansion (20) can be written as

$$f^+(z) = \langle f_1, \hat{B}_1 \rangle \hat{B}_1(z) + \langle f_2, e_{\hat{a}_2} \rangle \hat{B}_2(z) + r_2(z) \frac{z - \hat{a}_1}{1 - \bar{\hat{a}}_1 z} \tag{27}$$

where $\hat{B}_2(z) = e_{\hat{a}_2}(z)(z - \hat{a}_1 / 1 - \bar{\hat{a}}_1 z)$. Similar to $f_2(z)$, $f_3(z)$ can be defined by the Möbius transform at $\hat{a}_2$ for removing the zero point of the second error term $r_2(z)$, that is

$$f_3(z) = \frac{r_2(z)}{\tau_{\hat{a}_2}(z)} = \frac{f_2(z) - \langle f_2, e_{\hat{a}_2} \rangle e_{\hat{a}_2}(z)}{\frac{z - \hat{a}_2}{1 - \bar{\hat{a}}_2 z}} \tag{28}$$

and $f_3(z) \in H^2(D)$. After repeating the decomposition process, such as in (25) until the Kth step, the corresponding optimal parameters $\{\hat{a}_3, \hat{a}_4, \ldots, \hat{a}_K\}$ can be found in turn. Finally, the signal f^+ can be decomposed into

$$f^+(z) = \sum_{k=1}^{K} \langle f_k, e_{\hat{a}_k} \rangle \hat{B}_k(z) + f_{K+1}(z) \prod_{k=1}^{K} \frac{z - \hat{a}_k}{1 - \bar{\hat{a}}_k z} \quad (29)$$

where $\hat{B}_k(z)$ is the rational function $B_k(z)$ in the TM system (10) under taking the optimal parameters $\{\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_K\}$ for the given signal f^+, and

$$f_{k+1}(z) = \frac{r_k(z)}{\tau_{\hat{a}_k}(z)} = \frac{f_k(z) - \langle f_k, e_{\hat{a}_k} \rangle e_{\hat{a}_k}(z)}{\frac{z - \hat{a}_k}{1 - \bar{\hat{a}}_k z}}, \ k = 1, 2, \ldots, K. \quad (30)$$

The following conclusion can be proven:

$$\langle f_k, e_{\hat{a}_k} \rangle = \langle f^+, \hat{B}_k \rangle \triangleq \hat{c}_k, k = 1, 2, \ldots, K. \quad (31)$$

In the same way as in (15), the decomposition in (29) can be rewritten as

$$f^+(z) = \sum_{k=1}^{K} \hat{c}_k \hat{B}_k(z) + \hat{\varepsilon}(z). \quad (32)$$

Here, the residual is $\hat{\varepsilon}(z) = f_{K+1}(z) \prod_{k=1}^{K} (z - \hat{a}_k / 1 - \bar{\hat{a}}_k z)$. Due to the orthogonality of the TM system $\{B_k(z)\}_{k=1}^{K}$, the energy relation can be deduced based on (32)

$$||f^+||^2 = \sum_{k=1}^{K} |\hat{c}_k|^2 + ||\hat{\varepsilon}||^2 = \sum_{k=1}^{K} |\langle f^+, \hat{B}_k \rangle|^2 + ||f_{K+1}||^2. \quad (33)$$

Based on the MPP, in the manner of stepwise approximation, AFD can make the energy of the given signal f^+ be distributed in the first K terms as great as possible and, simultaneously, make the residual energy $||f_{K+1}||^2$ converge quickly to 0 along with the growth of K. The residual can be regard as the noise with very little energy. Thus, the signal f^+ can be approximated by a linear combination of the TM system based on AFD in the mean-squared error minimization sense, that is

$$f^+(z) \approx \sum_{k=1}^{K} \hat{c}_k \hat{B}_k(z). \quad (34)$$

Here, $\{\hat{c}_k\}_{k=1}^{K}$ is the FRs of f^+ under the TM system $\{\hat{B}_k(z)\}_{k=1}^{K}$. According to (6), the reconstructed spectral curve $x(t)$ can be obtained by taking the real part of f^+, i.e., $x(t) = 2\mathrm{Re}(f^+(z)) - c_0, z = e^{\mathbf{i}t}$. For clarification regarding the use of AFD, Algorithm 1 is shown.

To certify the effectiveness and advantages of the TM system by AFD in spectral curve fitting, we take randomly four samples from different classes in the Indian Pines data set that is used in Section V as examples and compare the TM system with the Fourier basis function system [16] and the B-spline basis function system [17] in terms of curve fitting. Fig. 1 shows the original spectral signals and the fitted spectral curves of these three basis function systems. Table I shows the fitting errors for these three basis function systems corresponding to Fig. 1. Although the B-spline basis function system can give a very good fit, it needs enough basis functions (202), depending on the B-spline order (4) and the band number (200). Note here that we have extended three nodes, respectively, outward at both ends of the bands. In order to achieve smaller fitting errors than the B-spline basis function system, the TM system only needs 45 basis functions. However, the Fourier basis function system gives a very bad fit with 45 basis functions. Compared with the Fourier and B-spline basis function systems, the TM system achieves a better spectral curve fit. To illustrate the fast convergence of the TM system obtained by AFD, Fig. 2 shows the fitting errors under different numbers of basis functions by taking the sample from Class 2 as an example. As the number of basis functions increases, the errors of both the TM and Fourier basis function systems decrease. However, the error of the TM system drops faster. This shows that the TM system has a faster convergence speed and gives a better approximation.

Algorithm 1 TM System by AFD

Input: Any spectral signal $\boldsymbol{x} \in \mathbb{R}^m$, the randomly initialized parameter set $\{a_k\}_{k=1}^{\mathcal{K}} \subset D$ and the decomposition step K ($\mathcal{K} \gg K$).

Output: The optimal parameters $\{\hat{a}_k\}_{k=1}^{K}$, the FRs $\{\hat{c}_k\}_{k=1}^{K}$, the fitted spectral curve $x(t)$, and the reconstructed error $||f_{K+1}||$.

1: Obtain the analytic form f^+ of the spectral signal $\boldsymbol{x}$ by (8).
2: Initialize $f_1 = f^+$, $z = e^{\mathbf{i}t}$, $t \in \{\frac{2j\pi}{m}\}_{j=0}^{m}$, and $x(t) = 0$.
3: **for** $k = 1 : K$ **do**
4: Compute $\hat{a}_k = \arg\max_{a \in \{a_k\}_{k=1}^{\mathcal{K}}} (1 - |a|^2)|f_k(a)|^2$.
5: Obtain the rational function $\hat{B}_k(z)$ by (10).
6: Compute the functional representation $\hat{c}_k = \langle f_k, \hat{B}_k \rangle$.
7: Obtain $f_{k+1}(z) = \frac{f_k(z) - \hat{c}_k \hat{B}_k(z)}{\frac{z - \hat{a}_k}{1 - \bar{\hat{a}}_k z}}$.
8: Compute $x(t) = x(t) + \hat{c}_k \hat{B}_k(z)$.
9: **end for**
10: Compute $x(t) = 2Re(x(t)) - \hat{c}_0$.
11: Compute the reconstructed error $||f_{K+1}||$.
12: **return** $\{\hat{a}_k\}_{k=1}^{K}$, $\{\hat{c}_k\}_{k=1}^{K}$, $x(t)$, and $||f_{K+1}||$.

TABLE I

FITTING ERRORS OF FOUR SAMPLES FROM FOUR CLASSES FOR THESE THREE METHODS (NUMBER OF BASIS FUNCTIONS ARE INDICATED IN PARENTHESIS) ON THE INDIAN PINES DATA SET

Methods	Fourier(45)	B-spline(202)	TM(45)
Class 2	2.00	0.53	**0.51**
Class 6	1.88	0.49	**0.38**
Class 11	1.95	0.59	**0.56**
Class 16	1.42	0.46	**0.43**

III. JOINT RATIONAL BASIS FUNCTION SYSTEM BASED ON CLASS ADAPTIVITY

According to the above results, the TM system is a very good tool for HSI functional data processing from a functional point of view. Next, we build a joint rational basis function system (JRBFS) based on class adaptivity for each HSI. Assume that the training samples $\{\boldsymbol{x}_j^l, y^l\}_{j=1}^{n_l} (l = 1, 2, \ldots, C)$ in an HSI are given, where $\boldsymbol{x}_j^l$ represents the jth sample of the

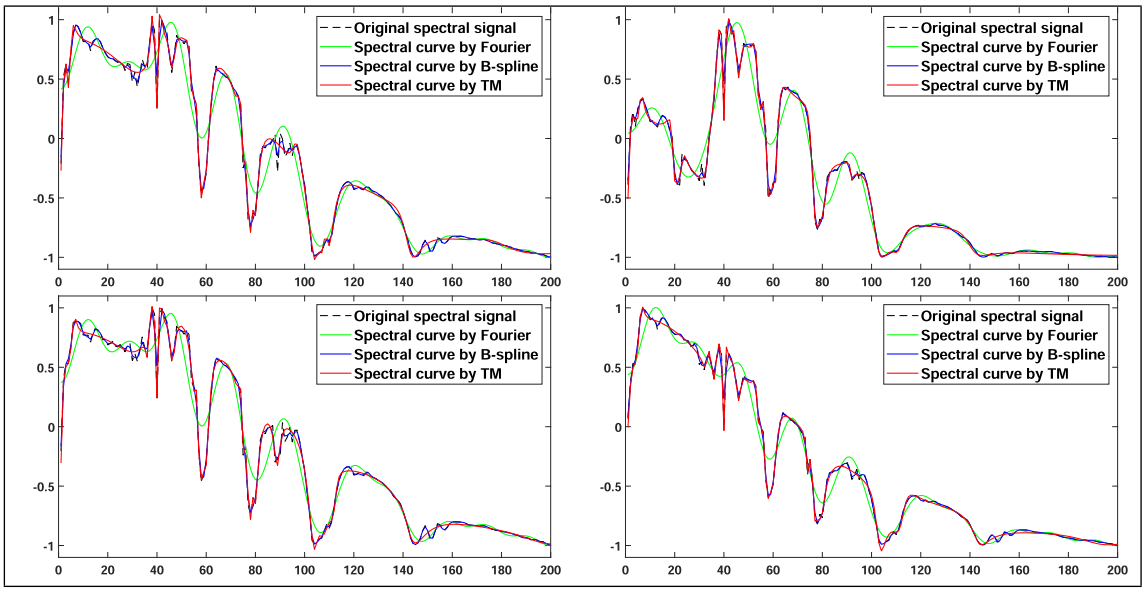


Fig. 1. Original spectral signals (black) and spectral curves of four samples from Class 2 (Top left), Class 6 (Top right), Class 11 (Bottom left), and Class 16 (Bottom right) after functional data fitting by the Fourier (green), B-spline (blue), and TM (red) basis function systems on the Indian Pines data set.

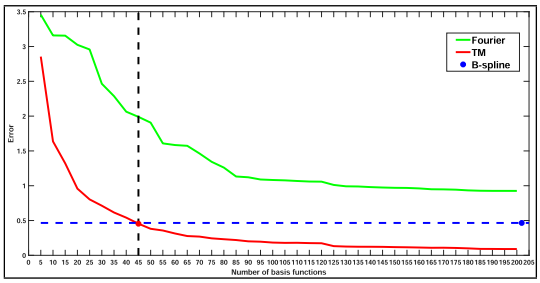


Fig. 2. Fitting errors for the Fourier and TM basis function systems under different numbers of basis functions, and the B-spline basis function system under a fixed number of basis functions (202) by taking the sample from Class 2 as an example.

lth class, y^l is the lth class label, n_l is the number of training samples from the lth class, and C is the class number.

To ensure that spectral curves from each class can get a good fit, we obtain a set of basis functions (or strictly a set of parameters $\{\hat{a}_k^l\}_{k=1}^{K_l}$) by AFD based on the mean of training samples $\{\boldsymbol{x}_j^l\}_{j=1}^{n_l}$ for each class. Due to the similarity of spectral curves from different land covers, the different sets of basis functions for all classes on an HSI may contain some common elements. Thus, to avoid repetition, we fuse these C sets of parameters $\{\hat{a}_k^l\}_{k=1}^{K_l}(l = 1, 2, \ldots, C)$ and form a new set of parameters $\{\tilde{a}_k\}_{k=1}^{K}$. Thereupon, a JRBFS $\{\tilde{B}_k(z)\}_{k=1}^{K}$ with the parameters $\{\tilde{a}_k\}_{k=1}^{K}$ can be constructed based on (10). It is obvious that the JRBFS is a new TM system. It is not a basis functions system for a particular class but a common basis function system for all classes on an HSI. Moreover, due to the fact that some common elements exist, the number of the joint rational basis functions is usually less than the sum of the number of the basis functions from each class, i.e., $K < \sum_{l=1}^{C} K_l$. Thus, for a given HSI $\boldsymbol{X} = \{\boldsymbol{x}_i\}_{i=1}^{N}$, the functional data fitting model can be built based on the JRBFS in the sense of the sum of mean-squared errors (SMSE)

$$\min_{c_{11},c_{12},\ldots,c_{\mathrm{NK}}} \mathrm{SMSE} = \sum_{i=1}^{N} \left\| x_i(t) - \sum_{k=1}^{K} c_{ik}\tilde{B}_k(z) \right\|_2^2. \quad (35)$$

Under the fixed orthogonal rational basis function system $\{\tilde{B}_k(z)\}_{k=1}^{K}$, the model can be solved easily by the projection method. The FRs are obtained as follows:

$$\tilde{c}_{ik} = \langle x_i(t), \tilde{B}_k(z)\rangle, z = e^{\mathbf{i}t}, \quad k = 1, 2, \ldots K. \quad (36)$$

Whereupon, any fitted spectral curve in an HSI can be represented as a continuous function under the JRBFS, that is

$$x_i(t) = \sum_{k=1}^{K} \tilde{c}_{ik}\tilde{B}_k(z), z = e^{\mathbf{i}t}, \quad i = 1, 2, \ldots N. \quad (37)$$

To further help understand the construction of the JRBFS based on class adaptivity and functional data fitting model, we summarize this in Algorithm 2.

To demonstrate the superiority of the functional data fitting model based on the JRBFS, we take the whole Indian Pines data set as an example and compare the proposed model with the regularized fitting model based on the Fourier basis function system (RFM_Fourier) [16] and the B-spline basis function system (RFM_B-spline) [17] in terms of classification accuracies. For convenience, we uniformly select six parameters for each class to construct the JRBFS, i.e., $K_l = 6$. As mentioned earlier, due to the fact that some common elements exist, $K < \sum_{l=1}^{C} K_l$. In fact, the number of the joint rational basis functions will change slightly with different training sets. In these experiments, the number of Fourier basis functions for RFM_Fourier is consistent with the proposed model, while the one for RFM_B-spline is still 202.

Unless particularly stated, we randomly select 10% of the labeled samples for each class as the training set for the HSI and the remaining 90% as the test set. The SVM with the

Algorithm 2 Functional Data Fitting Model Based on the JRBFS

Input: An HSI $\boldsymbol{X} = \{\boldsymbol{x}_i\}_{i=1}^{N} \in \mathbb{R}^{N\times m}$ with training samples $\{\boldsymbol{x}_j^l, y^l\}_{j=1}^{n_l}$ $(l = 1, 2, \ldots, C)$, the randomly initialized parameter set $\{a_k\}_{k=1}^{\mathcal{K}} \subset D$ and the decomposition step K_l for the l-th class $(\mathcal{K} \gg K_l)$.

Output: The fused parameter set $\{\tilde{a}_k\}_{k=1}^{K}$, the corresponding JRBFS $\{\tilde{B}_k(z)\}_{k=1}^{K}$, the FRs $\{\tilde{c}_{ik}\}_{k=1}^{K}(i = 1, 2, \ldots, N)$, and the fitted spectral curve $\{x_i(t)\}_{i=1}^{N}$.

1: **for** $l = 1 : C$ **do**
2: Compute the average spectral signal for the l-th class $\bar{\boldsymbol{x}}^l = \frac{1}{n_l}\sum_{j=1}^{n_l} \boldsymbol{x}_j^l$.
3: Compute $\{\hat{a}_k^l\}_{k=1}^{K_l}$ by **Algorithm** 1.
4: **end for**
5: Obtain the fused parameter set $\{\tilde{a}_k\}_{k=1}^{K} = UNIQUE\{\hat{a}_k^l, k = 1, 2, \ldots K_l, l = 1, 2, \ldots, C\}$.
6: Construct the JRBFS $\{\tilde{B}_k(z)\}_{k=1}^{K}$ under the fused parameter set $\{\tilde{a}_k\}_{k=1}^{K}$.
7: **for** $i = 1 : N$ **do**
8: **for** $k = 1 : K$ **do**
9: Compute $\tilde{c}_{ik} = \langle \boldsymbol{x}_i, \tilde{B}_k(z)\rangle$, where $z = e^{\mathbf{i}t}$ and $t \in \{\frac{2j\pi}{m}\}_{j=0}^{m}$.
10: **end for**
11: Obtain the fitted spectral curve $x_i(t) = \sum_{k=1}^{K} \tilde{c}_{ik}\tilde{B}_k(z)$.
12: **end for**
13: **return** $\{\tilde{a}_k\}_{k=1}^{K}$, $\{\tilde{B}_k(z)\}_{k=1}^{K}$, $\{\tilde{c}_{ik}\}_{k=1}^{K}(i = 1, 2, \ldots, N)$, and $\{x_i(t)\}_{i=1}^{N}$.

TABLE II
CLASSIFICATION RESULTS FOR THE RFM_FOURIER, RFM_B-SPLINE, AND PROPOSED MODELS ON THE INDIAN PINES DATA SET (OA AND AA ARE IN PERCENTAGE)

	Spect	RFM_Fourier		RFM_B-spline		Proposed model	
		RSpect	FRs	RSpect	FRs	RSpect	FRs
OA	80.68	81.64	81.65	82.29	81.52	82.92	**83.20**
AA	74.01	79.74	79.75	79.40	77.99	79.93	**80.80**
κ	0.779	0.790	0.791	0.798	0.789	0.805	**0.808**

TABLE III
NUMBER OF THE JOINT RATIONAL BASIS FUNCTIONS UNDER TEN RANDOM EXPERIMENTS ON THE INDIAN PINES DATA SET

Test	1	2	3	4	5	6	7	8	9	10
Number	43	40	44	44	44	44	43	45	43	45

Gaussian kernel is chosen as the classifier. The corresponding parameters are set optimally. The overall accuracy (OA [%]), average accuracy (AA [%]), and kappa statistic (κ) are quantitative indexes for evaluating the classification performance. These accuracies are computed by taking the mean of ten runs. The same applies hereinafter.

Table II shows the classification accuracies of the original spectral signals (Spect), the reconstructed spectral signals (RSpect), and the FRs by the RFM_Fourier, RFM_B-spline, and proposed model. It is easy to see that RSpect based on functional data fitting models obtained higher classification accuracies than Spect. This shows that the functional data fitting models demonstrate good fitting performances. Comparing RSpect based on the proposed model with the ones based on RFM_Fourier and RFM_B-spline, we find that the proposed model obtains the highest classification accuracies, followed by RFM_B-spline, but the worst performer is RFM_Fourier. This is consistent with the fitting performance shown in Fig. 1. This shows again that the proposed model demonstrates both good fitting and reconstruction. The FRs are the features extracted from the reconstructed spectral curves. After preliminary feature extraction, the FRs based on the proposed model achieve better classification accuracies than the ones obtained by RFM_Fourier and RFM_B-spline and are also more effective than the RSpect based on different models, including the proposed model. Note that the FRs based on RFM_Fourier outperform the ones based on RFM_B-spline, which is mainly because the dimension of the FRs based on RFM_B-spline (202) is too high. To observe the change of the dimension for the FRs based on the proposed model and RFM_Fourier, we show the number of the joint rational basis functions under ten random experiments in Table III. They are significantly lower than the number of the B-spline basis functions. In the case of the same number of basis functions, the proposed model based on the JRBFS can achieve a faster convergence than RFM_Fourier. The FRs based on the proposed model can also obtain more energy from the original spectral signals than the ones based on RFM_Fourier, as described by (33). Thereby, the proposed model can obtain better classification results. Furthermore, we give the OA in Fig. 3 when different numbers of rational basis functions are selected for each class, that is, K_l takes different values. As the number of rational basis functions increases, the classification accuracies for the FRs and RSpect increase, but, after taking the maximum value when $K_l = 6$, the accuracies gradually decrease. Moreover, the accuracy based on the RSpect decreases faster. The increase in the number of basis functions makes the fit better and better, even almost the same as the original spectral vectors. Thus, the accuracy becomes the same as the one obtained by Spect. This shows that too many basis functions lead to overfitting, which reduces the fitting performance.

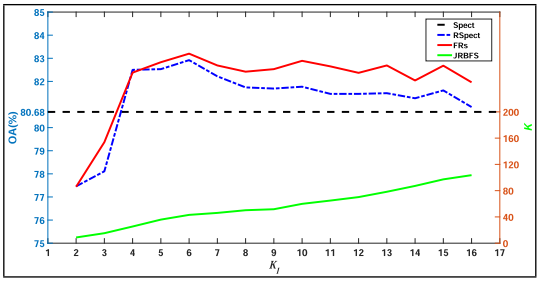


Fig. 3. OA for Spect, RSpect, and FRs based on the proposed model as a function of the number of rational basis functions for each class in left coordinate system (OA versus K_l). The cardinality K of the JRBFS as a function of the number of rational basis functions for each class in the right coordinate system (K versus K_l).

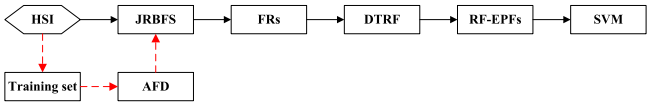


Fig. 4. Flowchart of the spectral–spatial classification scheme based on the functional feature extraction by using the adaptive rational function approximation and DTRF.

IV. Spectral–Spatial Classification With FR-EPFs

A. FR-EPFs

Large quantities of researches certify that the spatial structural information can improve the classification accuracies [40]. Edge-preserving features (EPFs) have been proven to effectively build spatial structural features, which were derived from the edge-preserving filtering by the domain transform recursive filter (DTRF) in HSIs [34]. In the above experiments, it is shown that the FRs are very effective functional features for discriminating the spectral curves. According to the MPP, the energy of the spectral information is mainly distributed in the first few components of FRs, which is similar to the contribution rate distribution of the principal components in PCA. Thus, this article further builds the functional spectral–spatial features based on the FRs by the edge-preserving filtering method, which is abbreviated as FR-EPFs. First, a single DTRF with two smoothing parameters δ_s and δ_r is applied for the first d components of FRs (i.e., $\mathrm{FR}_1, \mathrm{FR}_2, \ldots, \mathrm{FR}_d$) and stacked into a FR-EPF, that is

$$\begin{aligned}\text{FR-EPF}(\delta_s, \delta_r) = \{&\text{DTRF}(\mathrm{FR}_1, \delta_s, \delta_r), \text{DTRF}(\mathrm{FR}_2, \delta_s, \delta_r), \\ &\ldots, \text{DTRF}(\mathrm{FR}_d, \delta_s, \delta_r)\}. \quad (38)\end{aligned}$$

Second, FR-EPFs are built by stacking the multiple FR-EPF with τ different sets of smoothing parameters δ_s and δ_r

$$\begin{aligned}\text{FR-EPFs} = \big\{&\text{FR-EPF}\big(\delta_s^1, \delta_r^1\big), \text{FR-EPF}\big(\delta_s^2, \delta_r^2\big), \\ &\ldots, \text{FR-EPF}\big(\delta_s^\tau, \delta_r^\tau\big)\big\}. \quad (39)\end{aligned}$$

The resulting FR-EPFs fuse the spectral features and spatial structural features, which can improve the discrimination of samples in HSIs.

B. Spectral–Spatial Classification

To test the effectiveness of the proposed method, we design a spectral–spatial classification scheme based on the functional feature extraction by using the adaptive rational function approximation and DTRF. The flowchart of the proposed spectral–spatial classification scheme is shown in Fig. 4. First, in order to explore the function characteristic of the spectral vectors, a TM system is adaptively selected for each class by AFD, which makes the spectral vectors from each class be well fitted under its own function system. This is mainly due to the similarity of spectral vectors from the same class and the dissimilarity of ones from the different classes. Second, because spectral vectors from different classes may also exhibit the similarity in some continuous bands, this results in some same elements among different TM systems. Thus, a JRBFS is built by integrating TM systems from all classes. In contrast to the TM system suitable for some classes, the JRBFS is suitable for all classes from this HSI. The HSIs with different land covers have different JRBFSs. Third, based on the resulting FRs, the functional spectral–spatial features FR-EPFs are extracted by DTRF. Finally, the FR-EPFs consist of the inputs for the SVM classification.

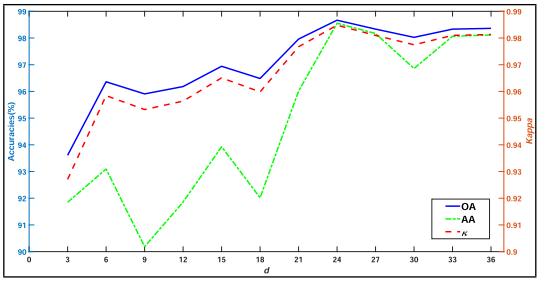


Fig. 5. OA, AA (left coordinate system, i.e., OA versus d and AA versus d), and κ (right coordinate system, i.e., κ versus d) as functions of d of FRs in the proposed method FR-EPFs.

V. Experimental Results and Analysis

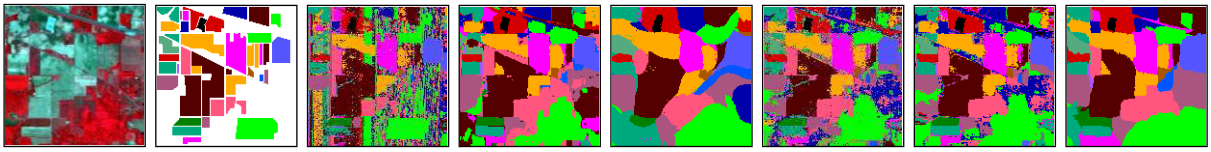
To prove the effectiveness and superiority of the proposed method FR-EPFs, experiments are first performed on three commonly used HSI data sets and two new HSI data sets, including the Indian Pines, Salinas, Pavia University,[1] and Houston scenes.[2] Furthermore, some state-of-the-art vector data and functional data classification methods are used for comparison, including SS-SVM [41], PCA-EPFs [36], OTVCA [13], FPCA [17]. and FDDA [18]. SS-SVM is a benchmarking spectral–spatial classification method that classifies the spectral–spatial features by stacking the spectral vectors and spatial vectors obtained by the local mean filtering method [41].

A. Experimental Design and Parameter Selection

For the following experiments, some experimental settings and parameter selections are shown in Section III. For SS-SVM [41], PCA-EPFs [36], OTVCA [13], FPCA [17], and FDDA [18], the parameters are set to the same as the corresponding publication. The SVM with the Gaussian kernel is used as a common classifier, whose penalty term c and width of the Gaussian kernel σ are tuned in the range $\{2^{-10}, 2^{-9}, \ldots, 2^{10}\}$. These optimal parameters are obtained by using tenfold cross-validation. In the proposed FR-EPFs method, the two most important parameters are considered as follows: K_l in the range from 2 to 16 and d in the range from 3 to 36 at intervals of 3. Besides, the smoothing parameters δ_s and δ_r in the FR-EPFs refer to the setting of PCA-EPFs [36]. The experiments are implemented using MATLAB R2018a in a PC with Intel Core i7-4770 CPU and 24-GB DDR3 RAM.

[1] http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes

[2] https://hyperspectral.ee.uh.edu/?page_id=1075



(a) (b) (c) (d) (e) (f) (g) (h)

Fig. 6. Whole classification maps for different methods applied on the Indian Pines data set. (a) False-color image. (b) Ground reference map. (c) SS-SVM. (d) PCA-EPFs. (e) OTVCA. (f) FPCA. (g) FDDA. (h) FR-EPFs.

B. Results and Analysis

1) Indian Pines Data Set: The experimental setup, as described above, was applied to the Indian Pines data set. Fig. 3 shows that, for these data, the FRs achieve the best pure spectral classification accuracies when the number of rational basis functions for each class is $K_l = 6$, which is adopted in the following experiments on the same data set. Meanwhile, $\tau = 3$, i.e., three sets of different smoothing parameters δ_s and δ_r are set as ($\delta_s^1 = 30$, $\delta_r^1 = 0.3$), ($\delta_s^2 = 115$, $\delta_r^2 = 0.6$), and ($\delta_s^3 = 200$, $\delta_r^3 = 0.9$). Fig. 5 shows the classification accuracies, including the OA, AA, and κ for different numbers d of FRs in the proposed method FR-EPFs. The OA and κ keep improving with d increased and then gradually become stable with a slight decline when $d > 24$. Although the AA also shows this trend, there are relatively large oscillations when $d < 24$. This is mainly due to serious misclassifications of two small classes in the Indian Pines data set, i.e., Class 7 with 28 samples and Class 9 with 20 samples. However, as d continues to increase, this phenomenon is significantly improved. Thus, in the next experiment, the number d of FRs is set as 24.

To further demonstrate the advantage of the proposed method FR-EPFs, we compare the approach with other state-of-the-art vector data and functional data classification methods in terms of OA, AA, and κ, including SS-SVM [41], PCA-EPFs [36], OTVCA [13], FPCA [17], and FDDA [18]. Table IV displays the classification performance of different methods for each class and all samples. The proposed method FR-EPFs outperforms the other methods in terms of classification accuracies. The PCA-EPFs also produces a competitive classification performance. When comparing FR-EPFs to PCA-EPFs, it illustrates that the functional features FRs from the functional data fitting model based on the JRBFS are more effective than the original spectral features obtained by band selection in PCA-EPFs. Furthermore, the classification accuracies of these two methods based on DTRF are higher than those of the other methods, which verifies the superiority of DTRF in terms of modeling the spatial structural information. Furthermore, in contrast to three vector data classification methods (SS-SVM, OTVCA, and PCA-EPFs), three functional data classification methods (FPCA, FDDA, and FR-EPFs) achieve correspondingly better classification results. This illustrates the effectiveness and advantages of using functional data learning methods in the field of HSI processing.

Meanwhile, Table IV also gives the computational time in seconds and the number of features for different classification methods in the Indian Pines scene. The proposed method requires very low computational cost, second only to PCA-EPFs. This is mainly because PCA-EPFs quickly realize the dimensionality reduction of HSIs by using the band selection technology and the PCA code in MATLAB, which greatly saves the computational cost. Compared with FPCA and FDDA, the functional data fitting model in FR-EPFs can produce a better fit with fewer basis functions. Fewer basis functions lead to less computational cost. OTVCA, as a state-of-the-art method, also has a low computational cost. The benchmarking method SS-SVM has the highest computational cost because the number of features is too high.

TABLE IV

CLASSIFICATION PERFORMANCE FOR THE STATE-OF-THE-ART METHODS AND PROPOSED METHOD ON THE INDIAN PINES DATA SET WITH 10% LABELED SAMPLES PER CLASS (NUMBER OF FEATURES IN BRACKETS)

Class	SS-SVM (400)	PCA-EPFs (5)	OTVCA (20)	FPCA (70)	FDDA (30)	FR-EPFs (72)
1	53.66	**100.00**	89.27	63.41	75.61	99.51
2	81.56	95.87	95.60	92.53	96.81	**97.89**
3	76.44	97.73	96.60	86.08	94.51	**98.31**
4	59.15	97.67	97.51	86.38	97.18	**98.03**
5	92.63	**99.76**	93.04	94.70	94.47	97.63
6	98.78	99.10	96.67	96.50	**100.00**	99.60
7	76.00	96.15	86.80	92.00	**100.00**	98.00
8	99.30	**100.00**	**100.00**	99.07	**100.00**	**100.00**
9	83.33	**100.00**	60.00	**100.00**	83.33	**100.00**
10	79.41	94.67	94.37	82.27	95.08	**98.62**
11	90.22	95.96	96.66	92.94	95.20	**98.99**
12	66.60	**97.14**	95.74	84.43	97.00	97.04
13	97.83	**100.00**	93.86	98.91	**100.00**	98.10
14	93.59	**100.00**	96.93	95.61	99.82	99.87
15	69.16	**99.71**	97.26	80.12	95.68	97.69
16	77.11	98.70	94.46	**100.00**	91.57	97.59
OA	85.38	97.41	96.09	91.20	96.59	**98.67**
AA	80.92	98.28	92.80	90.31	94.77	**98.55**
κ	0.832	0.970	0.955	0.900	0.961	**0.985**
Time(s)	2770.79	**58.79**	280.65	458.34	255.60	167.97

To visually reflect on the classification of these methods, we show their classification maps in the whole Indian Pines scene, including the background class, as shown in Fig. 6. Fig. 6(h) obtained by the proposed method demonstrates the best performance, which produces very smooth and accurate classification boundaries. Moreover, the dividing lines between different classes are also very distinct in this map, while that is opposite in Fig. 6(e) produced by OTVCA. The classification map of PCA-EPFs also looks good visually and outperforms the remaining three maps, i.e., Figs. 6(c), (f), and (g), in this respect. However, compared with Fig. 6(h), there is a small amount of salt and pepper noise in Fig. 6(d), and the classification boundaries are slightly damaged. In general, these two spectral–spatial classification methods based on EPFs achieve very good classification maps. This once again demonstrates the advantages of DTRF in extracting spatial structural features. Then, FDDA with local mean filtering also gives a

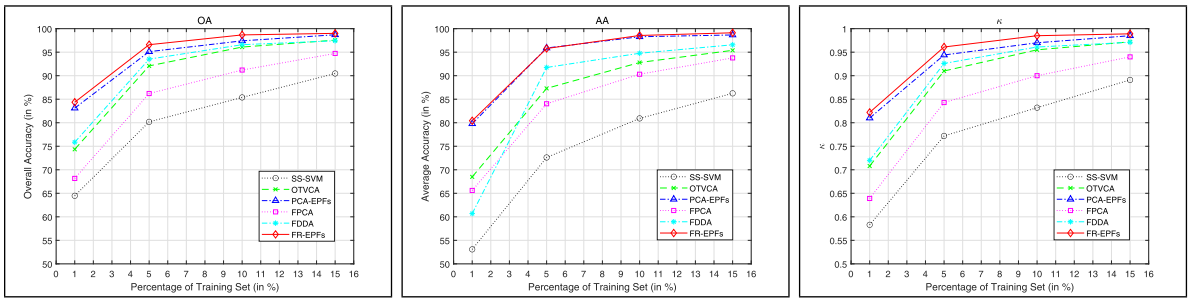


Fig. 7. Experimental results obtained by different methods when randomly selected 1%, 5%, 10%, and 15% of samples for each class as the training set. (a) OA. (b) AA. (c) κ.

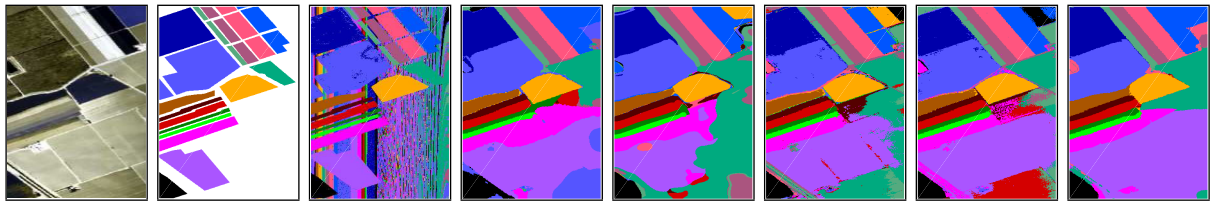


Fig. 8. Whole classification maps for different methods applied on the Salinas data set. (a) False-color image. (b) Ground reference map. (c) SS-SVM. (d) PCA-EPFs. (e) OTVCA. (f) FPCA. (g) FDDA. (h) FR-EPFs.

visually good classification map, but it has obvious salt and pepper noise. By contrast, the class boundaries of classification maps for SS-SVM and FPCA are heavily blurred, especially for SS-SVM.

To demonstrate the effectiveness and advantages of the proposed method in different situations, we discuss the influence of different training samples for different methods applied on the Indian Pines data set. We randomly select 1%, 5%, 10%, and 15% of the available samples for each class as the training set, respectively. The experimental results are displayed in Fig. 7. As the number of training samples increases, the classification accuracies of all methods are improved. The proposed method achieves the best classification results in these situations. In particular, when 15% of samples are selected as the training set, the proposed method exceeds 99% in terms of OA and AA.

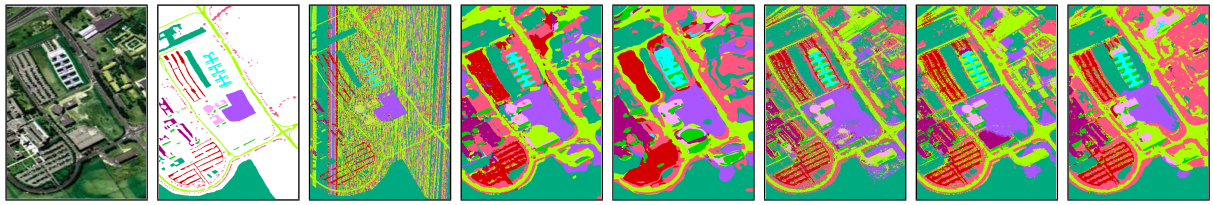
2) Salinas Data Set: This scene was acquired at a higher spatial resolution than the Indian Pines scene. The spectral information among different land covers is very different and can be easily distinguished. In this experiment, the two important parameters are set as $K_l = 7$ and $d = 27$. Table V shows the individual class accuracies, OA, AA, and κ for the different methods. All three functional data classification methods achieve very high accuracies, which again demonstrates that it is feasible and effective to implement the classification tasks of HSIs from a functional point of view. Moreover, the PCA-EPFs, OTVCA, FDDA, and FR-EPFs achieve near-perfect classification results, especially FR-EPFs that gives the best OA and κ. This further reflects the advantages of the combination of functional data processing and the EPFs for extracting the functional spectral–spatial features of the HSIs. Furthermore, this benchmarking method SS-SVM also achieves an OA of up to 97.65%.

TABLE V

CLASSIFICATION RESULTS FOR THE STATE-OF-THE-ART METHODS AND PROPOSED METHOD ON THE SALINAS DATA SET WITH 10% LABELED SAMPLES PER CLASS (NUMBER OF FEATURES IN BRACKETS)

Class	SS-SVM (408)	PCA-EPFs (5)	OTVCA (20)	FPCA (120)	FDDA (30)	FR-EPFs (81)
1	99.56	**100.00**	**100.00**	99.83	99.94	**100.00**
2	99.67	**100.00**	99.94	99.67	**100.00**	99.95
3	99.78	99.94	**100.00**	99.83	**100.00**	**100.00**
4	98.88	97.74	**99.84**	98.80	99.52	99.27
5	99.38	**99.92**	99.83	99.34	99.75	99.87
6	99.75	**100.00**	99.92	99.94	99.89	99.80
7	99.75	**100.00**	99.97	99.53	**100.00**	99.79
8	95.22	99.87	**99.96**	98.82	98.99	99.94
9	99.96	99.98	**100.00**	99.98	**100.00**	**100.00**
10	98.37	99.53	**99.97**	99.63	**99.97**	99.88
11	99.90	99.79	99.48	**100.00**	**100.00**	98.99
12	99.94	98.75	96.89	**100.00**	**100.00**	99.81
13	99.39	**100.00**	98.30	99.64	99.39	98.58
14	96.26	98.44	98.65	99.27	**99.79**	97.73
15	92.54	99.68	**99.89**	97.94	98.98	99.78
16	99.51	**100.00**	**100.00**	98.89	**100.00**	99.99
OA	97.65	99.75	99.77	99.26	99.60	**99.80**
AA	98.62	99.60	99.54	99.44	**99.76**	99.58
κ	0.974	0.997	0.997	0.992	0.996	**0.998**

Fig. 8 shows the whole classification maps for different methods applied on the Salinas data set. The classification map produced by FR-EPFs has both the best classification accuracies and visual performance, as shown in Fig. 8(h). The boundaries between different classes in this map are well maintained. This visually illustrates the advantages of the proposed method. Although PCA-EPFs, OTVCA, and FDDA achieve comparable classification accuracies, their classification maps are not as smooth as the map for the proposed method, and their boundaries are blurred to varying degrees. Fig. 8(f) seems even worse. Due to more misclassifications, the noise in the maps is obvious. The worst performance is seen in Fig. 8(c), generated by the benchmarking method SS-SVM, whose map is chaotic.



(a) (b) (c) (d) (e) (f) (g) (h)

Fig. 9. Whole classification maps for different methods applied on the Pavia University data set. (a) False-color image. (b) Ground reference map. (c) SS-SVM. (d) PCA-EPFs. (e) OTVCA. (f) FPCA. (g) FDDA. (h) FR-EPFs.

TABLE VI

CLASSIFICATION RESULTS FOR THE STATE-OF-THE-ART METHODS AND PROPOSED METHOD ON THE PAVIA UNIVERSITY DATA SET WITH 10% LABELED SAMPLES PER CLASS (NUMBER OF FEATURES IN BRACKETS)

Class	SS-SVM (206)	PCA-EPFs (6)	OTVCA (20)	FPCA (100)	FDDA (16)	FR-EPFs (72)
1	97.92	98.90	98.83	95.79	98.29	**99.64**
2	99.38	99.62	99.87	99.74	99.88	**99.92**
3	57.01	99.54	98.36	90.10	98.36	**99.76**
4	96.12	**98.97**	87.52	97.46	97.82	97.98
5	84.38	**99.97**	97.52	99.83	99.75	99.88
6	93.70	**100.00**	99.99	98.59	99.89	99.97
7	27.82	98.75	98.67	95.91	98.16	**99.25**
8	78.18	97.14	98.21	94.48	96.80	**99.03**
9	7.04	**99.36**	87.30	94.01	96.48	97.32
OA	89.61	99.26	98.23	97.64	99.02	**99.58**
AA	71.28	99.14	96.25	96.21	98.38	**99.20**
κ	0.861	0.990	0.977	0.969	0.987	**0.994**

3) Pavia University Data Set: Compared with the first two scenes, this scene was acquired with fewer bands. Thus, the number of rational basis functions for each class is set as $K_l = 5$. Another important parameter is set as $d = 24$. The classification results for different methods on this data set are shown in Table VI. The proposed method still achieves the best classification accuracies. The effectiveness and advantages of the proposed method are demonstrated again. At the same time, in this complex scene, EPFs show great advantages and can effectively extract more discriminative spectral–spatial features. This is again reflected in PCA-EPFs, which achieves competitive classification results. In general, three functional data classification methods achieve good classification results. The effectiveness of functional feature extraction is illustrated once again. However, SS-SVM does not perform very well for such complex images with irregularly distributed land covers.

The whole classification maps for different methods are shown in Fig. 9. Due to the complex spatial structures of this HSI, there are great differences between the different classification maps, especially for the background class. However, the classification areas in Fig. 9(d) and (h) seem smoother, and the class boundaries are distinct. In contrast, the classification of the background class seems a bit messy in Fig. 9(f) and (g). This is mainly due to the advantages of DTRF in preserving spatial structural features, which is more effective than the common local mean filtering approach used in FPCA and FDDA. In addition, the classification map produced by SS-SVM is heavily blurred, which also results from the local mean filtering. Although mean filtering can achieve a certain degree of denoising effect on the spectral information, it cannot protect the details in the spatial structures well.

TABLE VII

CLASSIFICATION RESULTS FOR THE STATE-OF-THE-ART METHODS AND PROPOSED METHOD ON THE HOUSTON 2013 AND 2018 DATA SETS

Dataset		SS-SVM	PCA-EPFs	OTVCA	FPCA	FDDA	FR-EPFs
Houston 2013	OA	79.49	72.72	81.15	82.82	83.84	**84.49**
	AA	82.99	78.31	83.43	85.61	**86.49**	86.28
	κ	0.779	0.704	0.797	0.814	0.826	**0.832**
Houston 2018	OA	89.79	95.98	95.96	91.30	92.43	**97.48**
	AA	87.30	**95.31**	93.68	88.69	90.11	94.85
	κ	0.867	0.948	0.947	0.887	0.901	**0.967**

4) Houston Data Sets: To further demonstrate the effectiveness and superiority of the proposed method FR-EPFs on newer HSI data sets, we test the classification performance of the proposed method on the Houston 2013 and 2018 data sets. These two data sets were acquired on June 23, 2012, and February 16, 2017, over the University of Houston campus and the neighboring urban area, respectively. The Houston 2013 HSI consists of 144 spectral bands in the 380–1050-nm region and 349 × 1905 pixels with 15 classes shown in Fig. 10(b). The Houston 2018 HSI covers the spectral range 380–1050 nm with 48 bands and contains 1202 × 4768 pixels and 20 land cover classes of interest shown in Fig. 11(b). The training and test samples have been given in advance for the Houston 2013 data set, as described in the 2013 IEEE GRSS Data Fusion Contest [42]. For the Houston 2018 data set, 10% labeled samples per class are still randomly selected as the training set. Table VII gives the classification accuracies of different methods on these two data sets. As can be seen from Table VII, the proposed method FR-EPFs almost obtains the highest classification accuracies, higher than other methods in terms of OA and κ. Figs. 10 and 11 also show the classification maps of different methods in the whole Houston 2013 and 2018 scenes. The experimental results for these two data sets once again demonstrate that the effectiveness of the proposed method, reflecting the advantages of the combination of functional feature extraction methods and edge-preserving filtering for modeling the spatial structure information.

C. Discussion

The proposed FR-EPFs method gives excellent classification results, and overall speaking outperforms the other methods in terms of OA, AA, and κ on the above five HSI data sets. This illustrates the advantages of the proposed method in processing high-dimensional continuous data that are mainly reflected as follows.

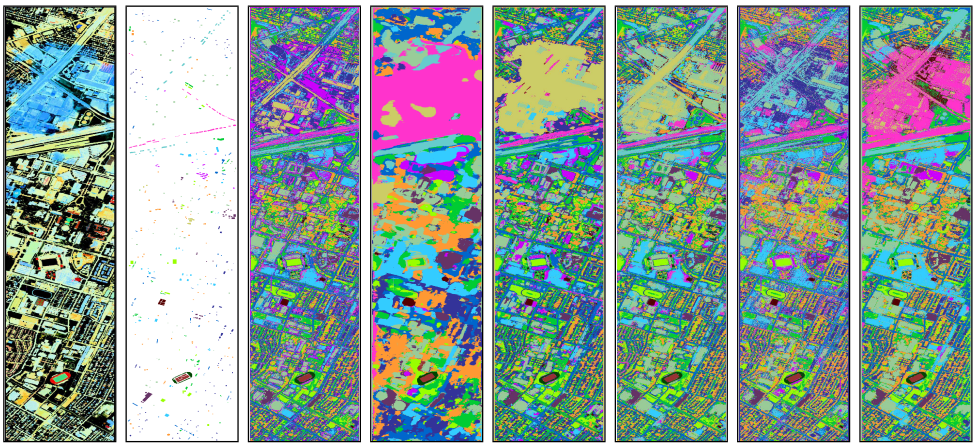


Fig. 10. Classification maps for different methods applied on the Houston 2013 data set. (a) False-color image. (b) Ground reference map. (c) SS-SVM. (d) PCA-EPFs. (e) OTVCA. (f) FPCA. (g) FDDA. (h) FR-EPFs.

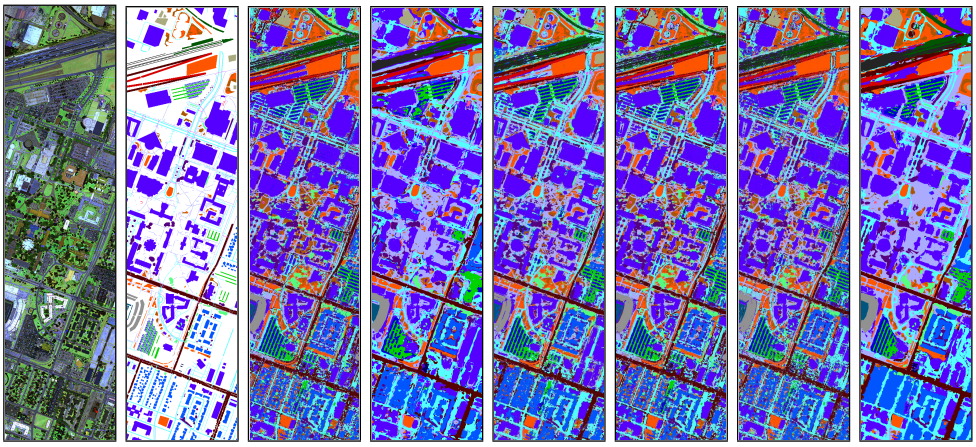


Fig. 11. Classification maps for different methods applied on the Houston 2018 data set. (a) False-color image. (b) Ground reference map. (c) SS-SVM. (d) PCA-EPFs. (e) OTVCA. (f) FPCA. (g) FDDA. (h) FR-EPFs.

1) In contrast to the classic feature extraction methods, the proposed method can fully use the continuous characteristics of the HSI data for overcoming the curse of dimensionality. This advantage will become more significant as the data dimension increases.
2) In this article, a rational approximation is introduced for the functional data learning framework in order to construct a more effective functional data fitting model. Compared with the data fitting models in FPCA and FDDA, this proposed model can produce a better fit with fewer basis functions.
3) In order to ensure that the spectral information from each class in an HSI is well fitted, a JRBFS is generated by AFD. Compared with fixed basis function systems, such as Fourier and B-spline basis function systems, the JRBFS has class adaptivity and needs fewer basis functions to achieve a similar fitting effect as those systems. Thus, the proposed functional data fitting model based on the JRBFS can obtain low-dimensional functional representations (FRs), which contains most of the energy of the spectral information based on the MPP.
4) For exploiting spatial information in HSIs, the functional spectral–spatial feature extraction method FR-EPFs based on the FRs by the edge-preserving filtering is proposed here. This method can well maintain the spatial smoothness of HSIs, thus effectively solving the spatial variability of the spectral information. However, the proposed method also has two main shortcomings and needs to be improved. One shortcoming is the construction of the rational basis functions. The TM system is obtained through consecutive optimal selections of the parameters. In our future work, we will study simultaneous optimization to directly obtain a set of basis functions. The second shortcoming is oversmoothing. Although the proposed method can fully mine the spatial structural information of HSIs in order to improve the classification accuracies, the structural information of samples of unknown classes cannot be maintained.

VI. Conclusion

In this article, an effort is made to produce advanced classification methods for HSIs based on functional data learning. The starting point of this article is based on the continuity of the spectral curves of hyperspectral data. Combining that with the theory of rational function approximation, a special functional feature extraction method based on adaptive rational

function approximation is proposed and tested in experiments. Although the functional fitting method based on the B-spline basis function system in our previous work has also achieved very good classification results, the dimension of its FRs has been demonstrated to be too high. Moreover, since the B-spline basis function system is not orthonormal, its FRs cannot be used as effective features for the spectral curves, which depends on the basis function system. However, the JRBFS based on class adaptivity, as proposed here, can effectively overcome these problems and obtain more efficient functional features. At the same time, considering the use of spatial information in functional feature extraction methods, this article introduces the DTRF on the basis of FRs in order to establish more discriminative spectral–spatial features for classification. The experimental results demonstrate that the proposed method provides a promising classification performance and outperforms in most cases other widely used methods in terms of OA, AA, and κ.

ACKNOWLEDGMENT

The authors would like to thank Prof. D. Landgrebe for providing the Indian Pines data set, Dr. J. A. Gualtieri for providing the Salinas data set, Prof. P. Gamba for providing the Pavia University data set, and the Hyperspectral Image Analysis Laboratory and the National Center for Airborne Laser Mapping (NCALM) at the University of Houston for providing the Houston 2013 and 2018 data sets.

REFERENCES

[1] J. Liang, J. Zhou, L. Tong, X. Bai, and B. Wang, "Material based salient object detection from hyperspectral images," *Pattern Recognit.*, vol. 76, pp. 476–490, Apr. 2018.

[2] B. Zhang *et al.*, "Remotely sensed big data: Evolution in model development for information extraction [point of view]," *Proc. IEEE*, vol. 107, no. 12, pp. 2294–2301, Dec. 2019.

[3] S. Li, W. Song, L. Fang, Y. Chen, P. Ghamisi, and J. A. Benediktsson, "Deep learning for hyperspectral image classification: An overview," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6690–6709, Sep. 2019.

[4] L. Zhou *et al.*, "Subspace structure regularized nonnegative matrix factorization for hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 4257–4270, 2020.

[5] X. Bai, F. Xu, L. Zhou, Y. Xing, L. Bai, and J. Zhou, "Nonlocal similarity based nonnegative tucker decomposition for hyperspectral image denoising," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 3, pp. 701–712, Mar. 2018.

[6] P. Ghamisi, J. Plaza, Y. Chen, J. Li, and A. J. Plaza, "Advanced spectral classifiers for hyperspectral images: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 1, pp. 8–32, Mar. 2017.

[7] 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.*, early access, Aug. 18, 2020, doi: 10.1109/TGRS.2020.3015157.

[8] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 480–491, Mar. 2005.

[9] N. Falco, J. A. Benediktsson, and L. Bruzzone, "Spectral and spatial classification of hyperspectral images based on ICA and reduced morphological attribute profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 11, pp. 6223–6240, Nov. 2015.

[10] F. Luo, L. Zhang, B. Du, and L. Zhang, "Dimensionality reduction with enhanced hybrid-graph discriminant learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 8, pp. 5336–5353, Aug. 2020.

[11] P. Ghamisi, J. A. Benediktsson, and J. R. Sveinsson, "Automatic spectral–spatial classification framework based on attribute profiles and supervised feature extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 9, pp. 5771–5782, Sep. 2014.

[12] B. Rasti, M. O. Ulfarsson, and J. R. Sveinsson, "Hyperspectral feature extraction using total variation component analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 6976–6985, Dec. 2016.

[13] B. Rasti *et al.*, "Feature extraction for hyperspectral imagery: The evolution from shallow to deep," 2020, *arXiv:2003.02822*. [Online]. Available: https://arxiv.org/abs/2003.02822

[14] P. Ghamisi *et al.*, "Advances in hyperspectral image and signal processing: A comprehensive overview of the state of the art," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 37–78, Dec. 2017.

[15] H.-G. Muller, "Functional modelling and classification of longitudinal data," *Scandin. J. Statist.*, vol. 32, no. 2, pp. 223–240, Jun. 2005.

[16] J. O. Ramsay and B. W. Silverman, *Functional Data Analysis*, 2nd ed. New York, NY, USA: Springer, 2005.

[17] H. Li, G. Xiao, T. Xia, Y. Y. Tang, and L. Li, "Hyperspectral image classification using functional data analysis," *IEEE Trans. Cybern.*, vol. 44, no. 9, pp. 1544–1555, Sep. 2014.

[18] Z. Ye, J. Chen, H. Li, Y. Wei, G. Xiao, and J. A. Benediktsson, "Supervised functional data discriminant analysis for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 2, pp. 841–851, Feb. 2020.

[19] M. Lv, J. E. Fowler, and L. Jing, "Spatial functional data analysis for the Spatial–Spectral classification of hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 6, pp. 942–946, Jun. 2019.

[20] A. Zullo, M. Fauvel, and F. Ferraty, "Experimental comparison of functional and multivariate spectral-based supervised classification methods in hyperspectral image," *J. Appl. Statist.*, vol. 45, no. 12, pp. 2219–2237, Sep. 2018.

[21] T. Qian, L. Zhang, and Z. Li, "Algorithm of adaptive Fourier decomposition," *IEEE Trans. Signal Process.*, vol. 59, no. 12, pp. 5899–5906, Dec. 2011.

[22] T. Qian and Y.-B. Wang, "Adaptive Fourier series—A variation of greedy algorithm," *Adv. Comput. Math.*, vol. 34, no. 3, pp. 279–293, Apr. 2011.

[23] Y. Li, L. Zhang, and T. Qian, "2D partial unwinding—A novel nonlinear phase decomposition of images," *IEEE Trans. Image Process.*, vol. 28, no. 10, pp. 4762–4773, Oct. 2019.

[24] L. Baratchart, W.-X. Mai, and T. Qian, "Greedy algorithms and rational approximation in one and several variables," in *Modern Trends in Hypercomplex Analysis*. Cham, Switzerland: Springer, 2016, pp. 19–33.

[25] W. Mai and T. Qian, "Rational approximation in hardy spaces on strips," *Complex Variables Elliptic Equ.*, vol. 63, no. 12, pp. 1721–1738, Dec. 2018.

[26] L. He, J. Li, C. Liu, and S. Li, "Recent advances on spectral–spatial hyperspectral image classification: An overview and new guidelines," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1579–1597, Mar. 2018.

[27] J. Liang, J. Zhou, Y. Qian, L. Wen, X. Bai, and Y. Gao, "On the sampling strategy for evaluation of spectral-spatial methods in hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 862–880, Feb. 2017.

[28] P. Ghamisi, M. Dalla Mura, and J. A. Benediktsson, "A survey on spectral–spatial classification techniques based on attribute profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2335–2353, May 2015.

[29] Y. Gu, T. Liu, X. Jia, J. A. Benediktsson, and J. Chanussot, "Nonlinear multiple kernel learning with multiple-structure-element extended morphological profiles for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3235–3247, Jun. 2016.

[30] D. Hong, X. Wu, P. Ghamisi, J. Chanussot, N. Yokoya, and X. X. Zhu, "Invariant attribute profiles: A spatial-frequency joint feature extractor for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 6, pp. 3791–3808, Jun. 2020.

[31] L. Fang, S. Li, W. Duan, J. Ren, and J. A. Benediktsson, "Classification of hyperspectral images by exploiting spectral–spatial information of superpixel via multiple kernels," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 12, pp. 6663–6674, Dec. 2015.

[32] J. Peng, L. Li, and Y. Y. Tang, "Maximum likelihood estimation-based joint sparse representation for the classification of hyperspectral remote sensing images," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 6, pp. 1790–1802, Jun. 2019.

[33] J. Peng, W. Sun, and Q. Du, "Self-paced joint sparse representation for the classification of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 1183–1194, Feb. 2019.

[34] X. Kang, S. Li, and J. Atli Benediktsson, "Spectral–spatial hyperspectral image classification with edge-preserving filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2666–2677, May 2014.
[35] S. Li, X. Kang, and J. Hu, "Image fusion with guided filtering," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2864–2875, Jul. 2013.
[36] X. Kang, X. Xiang, S. Li, and J. A. Benediktsson, "PCA-based edge-preserving features for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 12, pp. 7140–7151, Dec. 2017.
[37] 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.
[38] W. Mi, T. Qian, and F. Wan, "A fast adaptive model reduction method based on Takenaka–Malmquist systems," *Syst. Control Lett.*, vol. 61, no. 1, pp. 223–230, Jan. 2012.
[39] J. Wang and T. Qian, "Approximation of monogenic functions by higher order Szegö kernels on the unit ball and half space," *Sci. China Math.*, vol. 57, no. 9, pp. 1785–1797, 2014.
[40] M. Dalla Mura, J. Atli Benediktsson, B. Waske, and L. Bruzzone, "Extended profiles with morphological attribute filters for the analysis of hyperspectral data," *Int. J. Remote Sens.*, vol. 31, no. 22, pp. 5975–5991, Dec. 2010.
[41] G. Camps-Valls, L. Gomez-Chova, J. Muñoz-Marí, J. Vila-Francés, and J. Calpe-Maravilla, "Composite kernels for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 1, pp. 93–97, Jan. 2006.
[42] C. Debes *et al.*, "Hyperspectral and LiDAR data fusion: Outcome of the 2013 GRSS data fusion contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2405–2418, Jun. 2014.

Zhijing Ye received the B.Sc. and Ph.D. degrees in mathematics from the Huazhong University of Science and Technology, Wuhan, China, in 2011 and 2016, respectively.

He is with the School of Science, Wuhan University of Technology, Wuhan. His research interests include statistical learning, pattern recognition, and hyperspectral image processing.

Tao Qian received the Ph.D. degree in harmonic analysis from Peking University, Beijing, China, in 1984.

From 1984 to 1986, he was with the Institute of Systems Science, Chinese Academy of Sciences. He was also a Research Fellow with Macquarie University and held a post-doctoral position with the Flinders University of South Australia, Australia, until 1992, under the supervision of A. McIntosh and G. Gaudry. He was a Lecturer with New England University, Armidale, NSW, Australia, from 1992 to 2000. He has been an Associate Professor with the University of Macau, Macau, China, since 2000. He received the Full Professorship in 2003, and was the Head of the Department of Mathematics, from 2005 to 2011. From 2013 to 2018, he was appointed as the Distinguished Professor with the University of Macau. Since 2019, he has been a Full Professor with the Macau University of Science and Technology. His research interests include pure analysis, applied and computational harmonic analysis, complex and Clifford analysis, signal analysis, and related areas.

Liming Zhang (Member, IEEE) received the B.S. degree in computer software from Nankai University, Tianjin, China, the M.S. degree in signal processing from the Nanjing University of Science and Technology, Nanjing, China, and the Ph.D. degree in image processing from the University of New England, Armidale, NSW, Australia.

She is an Assistant Professor with the Faculty of Science and Technology, University of Macau, Macau, China. Her research interests include signal processing, image processing, computer vision, and multimedia computing.

Lei Dai received the B.Sc. degree in mathematics and the M.Sc. degree in statistics from the Huazhong University of Science and Technology, Wuhan, China, in 2015 and 2018, respectively. She is pursuing the Ph.D. degree with the Faculty of Science and Technology, University of Macau, Macau, China.

Her research interests include global optimization, time-frequency analysis, and image processing.

Hong Li (Member, IEEE) received the M.Sc. degree in mathematics and the Ph.D. degree in pattern recognition and intelligence control from the Huazhong University of Science and Technology, Wuhan, China, in 1986 and 1999, respectively.

She is a Professor with the School of Mathematics and Statistics, Huazhong University of Science and Technology. Her research interests include approximation theory, signal processing, and pattern recognition.

Jón Atli Benediktsson (Fellow, IEEE) 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 engineering from Purdue University, West Lafayette, IN, USA, in 1987 and 1990, respectively.

Since July 1, 2015, he has been the President and Rector of the University of Iceland. From 2009 to 2015, he was the Pro-Rector of Science and Academic Affairs and a Professor of electrical and computer engineering with the University of Iceland. His research interests are in remote sensing, biomedical analysis of signals, pattern recognition, image processing, and signal processing, and he has published extensively in those fields.

Dr. Benediktsson is a fellow of SPIE. He was a member of the 2014 IEEE Fellow Committee. He is a member of Academia Europea, the Association of Chartered Engineers in Iceland (VFI), Societas Scinetiarum Islandica, and Tau Beta Pi. 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 Council's Outstanding Young Researcher Award in 1997, in 2000 he was granted the IEEE Third Millennium Medal, in 2004, he was a co-recipient of the University of Iceland's Technology Innovation Award, in 2006, he received the yearly research award from the Engineering Research Institute of the University of Iceland, in 2007, he received the Outstanding Service Award from the IEEE GRSS, in 2020, the IEEE GRSS Education Award, in 2018, the IEEE GRSS David Landgrebe Award, and in 2016, the OECE Award from the School of ECE, Purdue University. He was co-recipient of the 2012 IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING Paper Award, and in 2013, he was co-recipient of the IEEE GRSS Highest Impact Paper Award. In 2013, he received the IEEE/VFI Electrical Engineer of the Year Award. In 2016 and 2018, he was a co-recipient of the International Journal of Image and Data Fusion Best Paper Award. In 2021, he was honored as a recipient of the Order of the Falcon from the President of Iceland. He is a Highly Cited Researcher (Clarivate Analysis, 2018–2020). He was the 2011–2012 President of the IEEE Geoscience and Remote Sensing Society (GRSS) and has been on the GRSS AdCom since 2000. He was an Editor in Chief of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (TGRS) from 2003 to 2008 and 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 was on the Editorial Board of the PROCEEDINGS OF THE IEEE from 2015 to 2020. He is on the International Editorial Board of the *International Journal of Image and Data Fusion*, the Editorial Board of *Remote Sensing*, and was the Chairman of the Steering Committee of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (J-STARS) from 2007 to 2010. He is a co-founder of the Biomedical Startup Company Oxymap (www.oxymap.com), Reykjavik.