

Bayesian gravitation bases classification for hyperspectral images.

ZHANG, A., SUN, G., PAN, Z., REN, J., JIA, X., ZHANG, C., FU, H. and YAO, Y.

2022

© 2022 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Bayesian Gravitation based Classification for Hyperspectral Images

Aizhu Zhang, *Member, IEEE*, Genyun Sun, *Member, IEEE*, Zhaojie Pan, Jinchang Ren, *Senior Member, IEEE*, Xiuping Jia, *Fellow, IEEE*, Chenglong Zhang, Hang Fu and Yanjuan Yao

Abstract—Integration of spectral and spatial information is extremely important for the classification of high-resolution hyperspectral images (HSIs). Gravitation describes interaction among celestial bodies which can be applied to measure similarity between data for image classification. However, gravitation is hard to combine with spatial information and rarely been applied in HSI classification. This paper proposes a Bayesian Gravitation based Classification (BGC) to integrate the spectral and spatial information of local neighbors and training samples. In the BGC method, each testing pixel is first assumed as a massive object with unit volume and a particular density, where the density is taken as the data mass in BGC. Specifically, the data mass is formulated as an exponential function of the spectral distribution of its neighbors and the spatial prior distribution of its surrounding training samples based on the Bayesian theorem. Then, a joint data gravitation model is developed as the classification measure, in which the data mass is taken to weigh the contribution of different neighbors in a local region. Four benchmark HSI datasets, i.e. the Indian Pines, Pavia University, Salinas, and Grss_dfc_2014, are tested to verify the BGC method. The experimental results are compared with that of several well-known HSI classification methods, including the support vector machines, sparse representation, and other eight state-of-the-art HSI classification methods. The BGC shows apparent superiority in the classification of high-resolution HSIs and also flexibility for HSIs with limited samples.

Index Terms—Gravitation, Hyperspectral image, Bayesian theorem, Image classification.

I. INTRODUCTION

THE recent advances of the hyperspectral remote sensing technology are capable of collecting hyperspectral sensing images (HSIs) with hundreds of spectral bands and reasonable spatial resolution simultaneously [1]. This detailed spectral and spatial information increases the possibility of more accurately discriminating materials of interest. Thus the hyperspectral remote sensing technology has played an increasingly important role in many remote sensing application,

such as plant assessment [2], mineral exploration [3], disaster monitoring [4], and astronomy [5]. In these applications, classification is one of the most fundamental while challenging tasks.

To obtain high quality classification results, complete utilization of Spectral and Spatial (SS) information is extremely important [6, 7]. In the past decades, researchers have developed numerous methods to synthetically utilize the SS information of HSIs [8-24]. Some studies first extract spatial features using texture filters on each spectral feature, such as gray-level co-occurrence matrix (GLCM) [8], Gabor [9], local binary pattern (LBP) [10], and Markov model [11]. The extracted SS features are then stacked together as the input of commonly used classifiers, such as support vector machines (SVM) [23], random forest (RF) [25], decision tree [26], artificial neural networks [27], and so on. Unfortunately, these kinds of methods always produce huge dimension of features and result in deterioration of classification accuracy. Meanwhile, the data projection based methods, such as sparse representation [14, 28] and manifold learning [15, 29], and graph-based methods [30] are also developed to mine the physically meaningful consensus in a low dimensional feature space in order to predict the classification labels [16]. Another way to deal with the high dimensional and big computational challenges in HSI filed is the cloud computing architectures [31]. In these algorithms, the spatial information is often incorporated by the joint of neighbors, like the joint sparse representation classification (JSRC) [17]. Nevertheless, the uses of spatial information are limited to that among the central pixel and its surrounding neighbors, and the spatial information of the training samples is not incorporated.

In addition to the aforementioned methods, some other studies combine the SS information by integrating different classifiers [12, 19-24]. One typical kind of method is the fusion of pixel-based classification and superpixel-based segmentation results by the majority voting [12, 32]. In these researches, the spectral information contributes mainly to the classification while the spatial adjacent relationships between pixels are considered to refine the results. The commonly used

Manuscript received 8 April, 2022; revised 2022. This work was supported by the Chinese Natural Science Foundation Projects (42271347, 41971292, 41871270), National Key Research and Development Program (2019YFE0126700), and the Opening Fund of the Key Laboratory of Poyang Lake Wetland and Watershed Research (Jiangxi Normal University), Ministry of Education (PK2020003). (*Corresponding author: Genyun Sun*)

Aizhu Zhang, Genyun Sun, Zhaojie Pan, and Hang Fu are with the College of Oceanography and Space Informatics, China University of Petroleum (East China), Qingdao, 266580, China. Genyun Sun is also the Laboratory for Marine Mineral Resources, Qingdao National Laboratory for Marine Science and Technology, Qingdao, 266237, China, and Aizhu Zhang is also with the Key

Laboratory of Poyang Lake Wetland and Watershed Research (Jiangxi Normal University), Ministry of Education, Nanchang, 330022, China. (email: genyunsun@163.com)

Jinchang Ren is with the National Subsea Centre, Robert Gordon University, Aberdeen, AB10 7AQ, United Kingdom.

Xiuping Jia is with the School of Engineering and Information Technology, University of New South Wales at Canberra, Canberra, ACT2600, Australia.

Chenglong Zhang is with the School of Control Science and Engineering, Shandong University, Jinan, 250012, China.

Yanjuan Yao with the Satellite Environment Center, Ministry of Environmental protection of China, Beijing, 100094, China.

classification methods are machine learning methods such as SVM and RF [25] while the superpixel-based segmentation methods include mean-shift [33], simple linear iterative clustering (SLIC) [34], and fractal net evolution approach (FNEA) [35]. The classification results are often varied due to the difficulty in determining the optimal segmentation scale. Another type of method is to design multiple classifiers to model different types of features and their combination produces the final classification [19]. One widely used example is the multi-kernel SVM classifier [20]. For example, in [21], two Gaussian radial basis function (RBF) based kernels, i.e. the spectral RBF kernel and texture RBF kernel, are combined by a weighting method. In [23], a series of kernel functions with different scale parameters are constructed to model spatial information at different scales. However, the selection of kernel functions, training of kernel parameters, and determination of the weighting parameters are still challenging [24, 36]. Although some intelligent optimization algorithms (IOAs) like genetic algorithm (GA) [37], particle swarm optimization (PSO) [38], and differential evolutionary (DE) [23] have been introduced to optimize the parameters, how to avoid an IOA fall into a local optimum is a big issue [24].

Most of the aforementioned classification methods are constructed based on the eager learning algorithms/principles (ELAs) [39]. The ELAs put significant effort in abstracting from the training samples by creating condensed representations during the learning phase [40]. The representations are considered approximate globally and applied directly to the new testing pixels. However, ELAs generally unable to provide good local approximations and show poor generalization ability when dealing with complex problem. Even the deep learning algorithms, which have recently become a hot topic and superior methods in HSI classification, also face series of issues [5, 13, 27, 41-45]. For example, although the 2D and 3D conventional neural networks (CNNs) [41, 46, 47] can simultaneously extract and abstract the SS information, they usually require significantly large training samples and more computation resources and time [48].

In contrast, lazy learning algorithms (LLAs), such as k-nearest neighbor (KNN) [14] and case-based reasoning [49], are non-abstracting local learning methods. LLAs classify a testing pixel by finding the most similar samples in its local neighborhood, and assign the majority class label of the neighborhood [40]. The most prominent advantage of LLAs is that they have the innate incremental learning ability and thus can be very effective under complex decision space [50]. In recent years, the LLAs have been combined with the sparse representation methods to promote classification accuracy of HSIs [51-53].

More recently, several LLAs inspired by Newton's law of gravity have been proposed [54-61]. These methods simulate the attractive gravitational force between data points, denoted as data gravitation, as the classification measure. Following the Newton's law of gravity, if the testing sample falls into a certain category, the distance between it and the corresponding training sample will be small and thus the data gravitation between them is large. A data point is assigned the class label whose training samples exerted the maximum resultant data gravitation to it. Obviously, different from the general distance-based clustering models such as KNN and K-means, the data gravitation based

methods can comprehensively utilize the information of all samples and these methods have proven their superiority in data classification [54, 57, 60]. For example, Peng et al. [54] firstly proposes a data gravitation classification (DGC) method for classification of UCI datasets. In [60], to process the imbalanced classification problem, an amplified gravitation coefficient (AGC) is introduced to develop an imbalanced DGC (IDGC) method. The effectiveness of IDGC has been further verified in the imbalanced traffic identification mission [61]. To deal with the classification problem with noisy data, Wen et al. [57] proposes a cognitive gravitation model (CGM) based on both the law of gravitation and the law of cognitive. Besides, the superiority of data gravitation in classification has motivated its application to edge detection of HSIs [62].

Motivated by these applications, we have tested the availability of data gravitation in HSI classification by developing a joint neighborhood learning based GDC method (JDGC) [63]. Even so, the integration of SS information is still a challenging problem for data gravitation. Especially, the spatial prior information of the training samples is omitted. Whereas, the spatial prior of training samples can be a beneficial constraint to the classifier [64]. Further, similar to the traditional joint KNN method, JDGC relies on Euclidean distance as a classification measure and holds that weight of each neighbor in a local region is identical. This is not reasonable because of each neighbor have different importance and distribution. How to find a way to obtain the different weight of each neighbor in a joint region is another major issue.

Aiming to address the above questions, we present a Bayesian Gravitation based Classification (BGC) method in this paper. In this method, each testing pixel is assigned to the class whose training samples exerted the largest data gravitation in a local joint region. The main contributions of BGC contain:

- (1) Proposed a Bayesian based data mass calculation method. Here the testing pixel with more serried neighboring pixels and denser training samples is assigned with larger mass. This process effectively promotes the information mining of the limited training samples and successfully integrates the SS information of HSIs and training samples;

- (2) Designed a novel joint data gravitation model as the classification measure, in which the joint Euclidean distance is weighted by the data mass. Thus each neighboring pixel in the joint region can contribute different weights to the central one synthetically based on their spectral and structural similarity.

The remainder of this paper is organized as follows. Section II reviews the principle of the data gravitation. Section III presents the details of the proposed BGC method. Section IV describes the experimental datasets, and the corresponding comparison results are discussed in Section V. Finally, Section VI concludes this study.

II. DATA GRAVITATION

In data gravitation model, each data point is assumed as a massive object. Following the Newton's law of gravity, any two massive objects in the universe can exert attractive gravitational force to each other. The strength of the gravitational force (F) between two objects is direct proportional to the product of their

masses (m_1 and m_2) and inverse proportional to the square of the distance (r) between them [65]:

$$F = G \frac{m_1 m_2}{r^2}, \quad (1)$$

where G is the gravitational constant, which is often defined as 1 in the data gravitation based methods.

Obviously, the larger mass and the shorter distance will produce the bigger gravitational force. This is used to define the data gravitation and is used as a scalar to express the similarity between data points [54]. For a data set, each data will be attracted by all the other data. Thus the larger accumulative data gravitation indicates the greater similarity. As shown in Fig. 1(a), for data point with unit mass, when the distance between data point P and the geometric center of *class 1* ($2r$) is two times of that of *class 2* (r), following Eq. (1), data gravitation exerted by *class 2* ($4F$) is four times larger than that of *class 1* (F). That is, P is more likely belonging to *class 2*. While for Fig. 1(b), although the distance between data point P and *class 1* and *class 3* is the same, the data gravitation exerted by *class 3* ($2F$) is two times larger than that of *class 1* because of the overall mass of *class 3* is twice of the *class 1*. In this situation, P is more likely belonging to *class 3*.

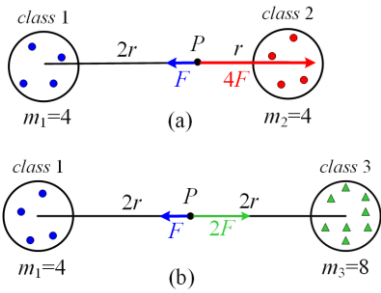


Fig. 1. Data gravitation exerted on data sample P .

Owing to its simple principle, the data gravitation theory has been utilized to solve the data clustering and classification problem [54, 57, 60-63]. These studies found that the definition

of the data mass and distance metric is crucial since they form the similarity metric. Basic DGC method sets the mass of one specific data to unit value and the classification measure relies on the Euclidian distance between data [54]. In [57], self-information is used to respect the data mass. In [59], reciprocal of the accumulative distance between a data and its k -nearest neighbors is set as the data mass and thus the densely distributed data will show larger data gravitation. This accords with the observation of Fig. 1(b) that the data density affects the classification result.

Nevertheless, until now, data gravitation rarely been applied in HSI classification because of gravitation is hard to combine with spatial information. This paper aims to design a novel data mass calculation method and thus develop a more effective data gravitation based classification measure by integrating the SS information of testing pixels and prior training samples.

III. BAYESIAN GRAVITATION BASED CLASSIFICATION (BGC)

Fig. 2 presents the overall framework of the proposed BGC method. BGC method contains two main steps: (1) *Bayesian based data mass calculation* in which the mass of each pixel is calculated based on the spectral distribution of its neighbors and the spatial distribution of its surrounding training samples; (2) *Joint data gravitation model* where the data mass is taken as the weight coefficient of joint Euclidean distance in the local joint region. Details of the two main steps are presented in the following subsections.

A. Bayesian based Data Mass calculation

Assuming an HSI with N pixels belongs to C classes, the first step of the BGC method is to calculate the data mass of each testing pixel y_i ($i \in [1, 2, \dots, N]$). In BGC, each testing pixel is assumed as a massive object with unit volume and particular density. As is well-known, the mass (M) of an object is the product of its density (ρ) and volume (V), i.e. $M = \rho V$ [66].

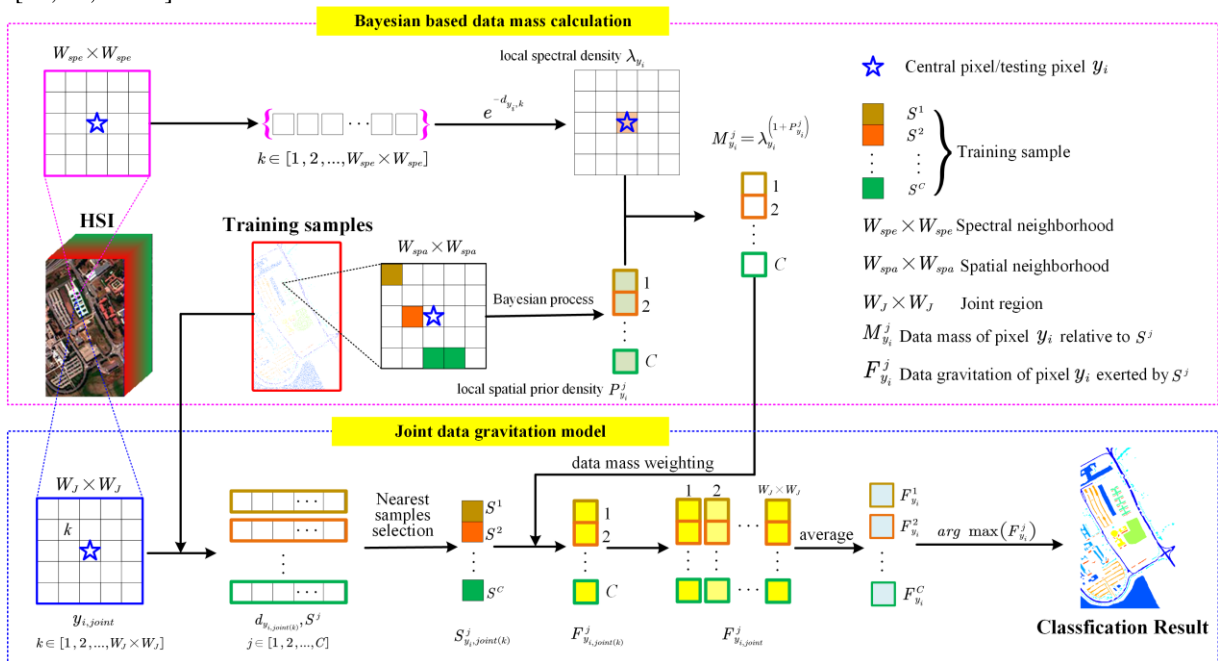


Fig. 2. Framework of the BGC method.

As the volume of a pixel is assumed as a unit value, the data mass of y_i is only determined by its density:

$$M_{y_i} = \rho_{y_i}. \quad (2)$$

Generally, the density of a data point is determined by its distribution in the attribute space. Given this, for each testing pixel y_i , we define a local spectral density parameter λ_{y_i} and a local spatial prior density parameter $P_{y_i}^j$ to jointly describe its distribution in the spectral-spatial space.

For the local spectral density, the λ_{y_i} is defined as follows:

$$\lambda_{y_i} = \sum_{k=1}^{W_{spe} \times W_{spe}} e^{-d_{y_i,k}} - 1, \quad (3)$$

where $d_{y_i,k}$ is the spectral Euclidean distance between y_i and its k -th neighbor ($k \in [1, 2, \dots, W_{spe} \times W_{spe}]$) as shown in Fig. 2. Obviously, if the similarity between the spectral features of the central pixel and its neighbors is strong, the corresponding density value is high.

For the local spatial prior density, $P_{y_i}^j$ is defined as the spatial prior distribution of the training samples in a $W_{spa} \times W_{spa}$ local neighborhood as shown in Fig. 2. In the local neighborhood, the denser the training samples of a certain class is, the higher the possibility that the pixel belongs to this class. Hence, the possibility that the pixel i is classified to the j -th class is set as $P_{y_i}^j$.

To calculate $P_{y_i}^j$, we firstly suppose two events:

Event $A_{y_i}^j$: the pixel y_i is a training sample of the j -th class;

Event $B_{y_i}^j$: the pixel y_i is classified to the j -th class, i.e.

$$P_{y_i}^j = P(B_{y_i}^j).$$

Obviously, events $A_{y_i}^j$ and $B_{y_i}^j$ are correlate to each other, and the prior knowledge of event $A_{y_i}^j$ can make the probability of event $B_{y_i}^j$ closer to the real probability. This happens to follow the Bayesian Theorem [67, 68]. The details for calculating the $P(B_{y_i}^j)$ are given in the follows.

Suppose the neighborhood of y_i contains NS_i training samples and $NS_i = \sum_{j=1}^C n_j$ (n_j is the training samples size of the j -th class in the local neighborhood), the probability of event $A_{y_i}^j$ can be written as:

$$P(A_{y_i}^j) = \frac{n_j}{W_{spa} \times W_{spa}}. \quad (4)$$

As the events $A_{y_i}^j$ and $B_{y_i}^j$ correlate to each other, following the Bayesian theorem, the probability of event $B_{y_i}^j$ can be obtained by:

$$P(B_{y_i}^j) = \frac{P(A_{y_i}^j \cap B_{y_i}^j)}{P(A_{y_i}^j | B_{y_i}^j)}. \quad (5)$$

In a classification model, a pixel will be classified to a specific class of it belongs to the class in the training data. We

thus can conclude that when event $A_{y_i}^j$ occurring, event $B_{y_i}^j$ will be true. Therefore, their joint probability can be expressed as:

$$P(A_{y_i}^j \cap B_{y_i}^j) = P(A_{y_i}^j) = \frac{n_j}{W_{spa} \times W_{spa}}. \quad (6)$$

For the conditional probability $P(A_{y_i}^j | B_{y_i}^j)$, when we classify the pixel y_i to the j -th class, the probability of event $A_{y_i}^j$ occurring is the proportion of j -th training samples in the total number of training samples in the local neighborhoods, i.e.:

$$P(A_{y_i}^j | B_{y_i}^j) = \frac{NS_i}{W_{spa} \times W_{spa}}. \quad (7)$$

Then we can substitute Eq. (6) and Eq. (7) into Eq. (5) and obtain:

$$P(B_{y_i}^j) = \frac{n_j}{W_{spa} \times W_{spa}} \bigg/ \frac{NS_i}{W_{spa} \times W_{spa}}. \quad (8)$$

Thus, the local spatial prior density of the training samples involved with the j -th class can be defined as:

$$P_{y_i}^j = P(B_{y_i}^j) = \frac{n_j}{NS_i}. \quad (9)$$

In this way, the category with more training samples in the local spatial neighborhood will be regarded to have a larger local spatial prior density.

Consequently, integrating the local spectral density and the local spatial prior density shown in Eq. (3) and Eq. (9) can make comprehensive use of the SS information of HSIs and the prior information of training samples. In this paper, the integrated density of y_i involved with the j -th category ($j \in [1, 2, \dots, C]$) is defined by:

$$\rho_{y_i}^j = \lambda_{y_i}^{(1+P_{y_i}^j)}. \quad (10)$$

Accordingly, the data mass of y_i can be rewritten as:

$$M_{y_i}^j = (\sum_{k=1}^{W_{spe} \times W_{spe}} e^{-d_{y_i,k}} - 1)^{(1+\frac{n_j}{NS_i})}. \quad (11)$$

As shown, the data mass is a vector associated with the spatial prior distribution of the training samples. In this way, the more similar the center pixel is to the surrounding pixels and the more training samples of j -th class in the neighborhood is, the heavier the center pixel will be.

B. Joint Data Gravitation Model

The HSIs with high-resolution usually have large homogeneous regions where the neighboring pixels within the regions are likely to be the same type target [14]. That is, pixels located in a small neighborhood often sharing the similar spectral character and spatial structure. This understanding has been utilized to present the joint Euclidean distance in joint KNN which holds that the importance of each testing sample in a local region is equal [53]. Apparently, this is not reasonable for it omit the different spectral and spatial distribution of testing pixels in a local region. Therefore, in this paper, we design a joint data gravitation model to calculate the data gravitation exerted on each testing pixel y_i ($i \in [1, 2, \dots, N]$) in

which the data mass is used to weigh the joint Euclidean distance of different neighboring pixel. The basic principle of the joint data gravitation model versus the basic data gravitation model is illustrated in Fig. 3.

As seen in Fig. 3(a), in the conventional data gravitation model, a testing sample will be affected by all the training samples and therefore it will be assigned to the class that exerts the maximum accumulative gravitational force. Note that, to use all the samples directly, not always better. In the proposed joint data gravitation model, on the contrary, as shown in Fig. 3(b), the category of the center pixel is determined only by its closely neighboring pixels in the local joint region and their nearest training samples. Especially, the data mass, $M_{y_i}^j$, is taken to weigh the contribution of different neighbors in the local joint region. Hence, the proposed BGC method can effectively improve the efficacy and robustness of HSI classification.

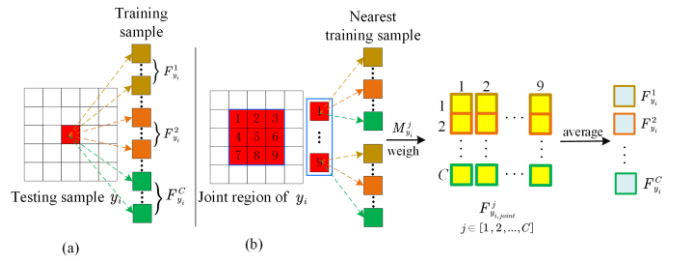


Fig. 3. The basic principle of (a): data gravitation model and (b) joint data gravitation model.

Details of the proposed joint data gravitation model are given below:

Step 1: For each central pixel y_i , set a $W_j \times W_j$ square neighborhood as its joint region, $y_{i,joint}$, as shown in the left bottom of Fig. 2.

Step 2: Calculate the data mass of pixel y_i using the Bayesian based data mass calculation strategy (Eq. (11)). Due to the mass is directly related to the spatial distribution of every class of training samples, the mass is a vector other than a scalar, which can be written as $M_{y_i} = [M_{y_i}^1, M_{y_i}^2, \dots, M_{y_i}^C]$.

Step 3: For each testing pixel y_i , calculate the Euclidean distance between its every neighboring pixel k ($k \in [1, 2, \dots, W_j \times W_j]$) in the joint region and all the training samples, denoted as $d_{y_{i,joint}(k), S^j}$ where $j \in [1, 2, \dots, C]$.

Step 4: For each neighboring pixel, select the nearest training sample of every class, denoted as $S_{y_{i,joint}(k)}^j$.

Step 5: Calculate the data gravitation exerted on the k -th neighbors, $F_{y_{i,joint}(k)}^j$, from the nearest training sample of the j -th class by:

$$F_{y_{i,joint}(k)}^j = \frac{M_{y_{i,joint}(k)}^j \cdot m_s}{(y_{i,joint}(k) - S_{y_{i,joint}(k)}^j)^2 + \varepsilon}, \quad (12)$$

where m_s is the mass of the training sample, ε is a small constant. In this paper, they are set to 1 and 10^{-6} , respectively.

Accordingly, for each testing pixel y_i , we can obtain the data gravitation between all its neighbors and the nearest training sample of the j -th class, denoted by $F_{y_{i,joint}}^j$ where $j \in [1, 2, \dots, C]$.

Here the data mass is taken as the weighting coefficient of the joint Euclidean distance of different neighboring testing pixel $y_{i,joint}(k)$.

Step 6: The average gravitation of the joint neighbors exerted by the j -th class samples, $F_{y_i}^j$, therefore can be calculated by:

$$F_{y_i}^j = \frac{\sum_{k=1}^{W_j \times W_j} F_{y_{i,joint}(k)}^j}{W_j \times W_j}. \quad (13)$$

In this way, the SS information of the joint neighbors is combined to produce the final data gravitation.

Step 7: The label of the testing pixel y_i can be determined by:

$$class(y_i) = \arg \max_{j=1,2,\dots,C} F_{y_i}^j. \quad (14)$$

The pseudo-code of the proposed BGC method is given in Algorithm 1.

Algorithm 1 shows the pseudo-code of the proposed BGC method.

Algorithm 1 BGC for HSIs classification

Initialization: N : number of pixels in an HSI; C : number of classes; W_{spe} , W_{spa} , W_j : sizes of the local neighborhoods for local spectral density, local spatial prior density of training samples, and joint region, respectively;

Bayesian based data mass calculation

- 1: **For** each testing pixel y_i where $i \in [1, 2, \dots, N]$
- 2: Calculate the local spectral density by Eq. (3).
- 3: **For** each class $j \in [1, 2, \dots, C]$
- 4: Calculate the local spatial prior density by Eq. (9).
- 5: **End for**
- 6: Calculate the data mass using Eq. (11).
- 7: **End for**

Joint data gravitation model

- 8: **For** each testing pixel y_i where $i \in [1, 2, \dots, N]$
- 9: Construct the joint region $y_{i,joint}$.
- 10: **For** each class $j \in [1, 2, \dots, C]$
- 11: Calculate the Euclidean distance between each neighbor and all the training samples.
- 12: **For** each neighboring pixel, select the nearest training sample.
- 13: Calculate the data gravitation exerted on neighbors by Eq. (12).
- 14: **End for**
- 15: Calculate the average data gravitation using Eq. (13).
- 16: **End for**
- 17: Predict the label of the i -th pixel using Eq. (14).
- 18: **End for**

IV. EXPERIMENTAL SETUP

In this section, we describe the details of four benchmark HSI datasets and experimental settings. The sensitivity of parameters in BGC method is also analyzed.

A. Datasets

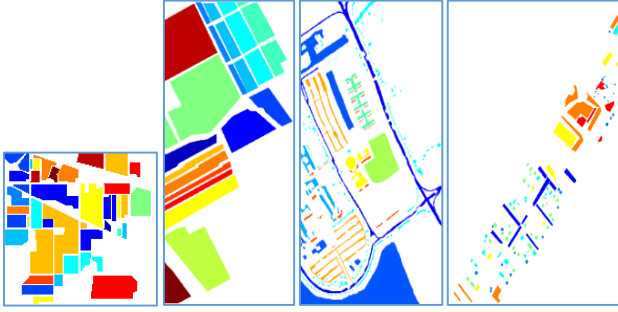
The four benchmark datasets are Indian Pines, Salinas, Pavia University, and Grss_dfc_2014. All the datasets are download from website^{1,2}. Details of these datasets are given below:

1) *Indian Pines (20-m resolution, 145×145)*: The dataset is acquired by the AVIRIS instrument over the agricultural area of Northwestern Indiana in 1992. The data has a wavelength range from 0.4 to 2.5 μm with 224 wavebands. Considering the

¹http://www.ehu.es/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes (Indian Pines, Salinas, and Pavia University)

²[http://www.grss-ieee.org/community/technical-committees/data-fusion/2014-ieee-grss-data-fusion-contest/\(Grss_dfc_2014\)](http://www.grss-ieee.org/community/technical-committees/data-fusion/2014-ieee-grss-data-fusion-contest/(Grss_dfc_2014))

atmospheric and water absorption, 24 bands are discarded as previous works [14, 69] and this paper preserves the other 200 bands for classification test. This agricultural area mainly contains 16 different types of crops as the ground truth (GT) image shown in Fig. 4(a).



(a) Indian Pines (b) Salinas (c) Pavia University (d) Grss_dfc_2014
Fig. 4. The ground truth images of the datasets.

2) *Salinas (3.7-m resolution, 512×217)*: The dataset is acquired by the AVIRIS sensor over Salinas Valley, California. Compared to Indian Pines data, spatial resolution of this data is significantly improved. After removing 20 water absorption bands from original 224 spectral bands, 204 bands are reserved. The area contains 16 land-cover classes as the GT image shown in Fig. 4(b).

3) *Pavia University (1.3-m resolution, 610 × 340)*: The dataset is acquired by the Reflective Optics Spectrographic Imaging System-03 (ROSIS-03) sensor over the university of Pavia, Italy. This sensor contains 115 spectral channels with a much higher spatial resolution than the first two datasets. After removing 12 noisy bands, 103 channels are kept for further processing. Nine classes of different materials are considered in this data as the GT image shown in Fig. 4(c).

4) *Grss_dfc_2014 (0.2-m resolution, 759×564)*: The data is originally used in the 2014 IEEE GRSS Data Fusion Contest [70], collected over an urban area near Thetford Mines in Quebec, Canada. This data contains 84 bands in the long wavelength range from 7.8 to 11.5 μm . Moreover, its spatial resolution up to 0.2 m, thus it is a very challenging dataset for the HSI classification. This area contains 7 different classes of land cover as shown in Fig. 4(d).

B. Parameters Sensitivity Analysis

In BGC method, the spatial contextual information is utilized in both the Bayesian based data mass calculation process and joint data gravitation model process by setting three square neighborhoods. Given that too small neighborhood cannot fully describe the structural information and too large neighborhood will increase the heterogeneity and computational time, we determine the values of W_{spe} , W_{spa} , and W_J on different datasets seriously based on individual spatial resolution and content of the scene.

For W_{spe} , it decided the value of local spectral density λ_{y_i} and the related formulation is an exponential function as shown in Eq. (3). It is important to maintain the homogeneity of local neighbors and thus keep the spectral density larger than 1. Only in this way the mass can be an increasing function of spatial prior density of training samples as illustrated in Eq. (9). Thus

the value of W_{spe} should be a very small value and set as 5 for all the tested datasets finally.

For W_{spa} and W_J , because they jointly affect the value of data gravitation, we conducted the combined test to select the parameters for each dataset by trial and errors. The results are listed in Table I. Note that the setting of training and testing samples follows that introduced in Section IV.C. For the Indian Pines, we tested the W_J and W_{spa} with the value of {1, 3, 5, 7, 9, 11, 13, 15, 17, 19} respectively. Fig. 5(a) presents the generated overall classification accuracy (OA in percentage) for each pair of parameters. As shown, when W_{spa} equals {7, 9, 11} and W_J is set to {3, 5}, the BGC method produce the best OA. To reduce the computational time, the W_{spa} and W_J are set to 7 and 3 for the Indian Pines as listed in Table I. For the other three datasets, because of their much higher spatial resolution, we tested the larger values of W_{spa} belonging to {11, 13, 15, 17, 19, 21, 23, 25, 27, 29}. The corresponding experimental results are displayed in Figs. 5(b)-(d). Following the principle that the larger OA with the smaller neighborhood is better, W_{spa} and W_J for Salinas, Pavia University, and Grss_dfc_2014 are set to {5, 21}, {5, 23}, and {9, 19}, respectively as illustrated in Table I. Table I summarized the parameter settings of BGC on different datasets.

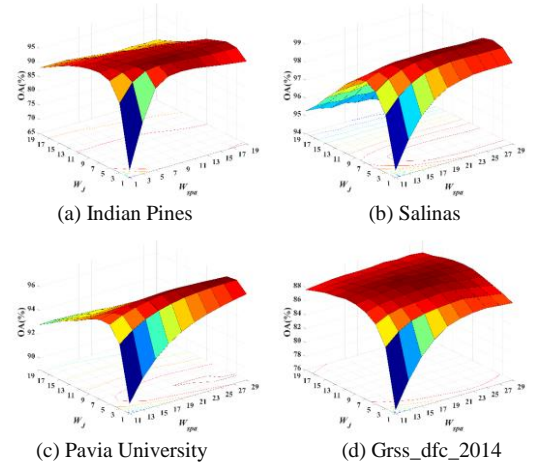


Fig. 5. Effects of parameter settings of BGC.

Methods	Indian Pines	Salinas	Pavia University	Grss_dfc_2014
SVM		RBF kernel, 5-fold cross validation sparsity=3		
JSRC		regularization parameter= 10^{-2}		
	$W_J=7$	$W_J=7$	$W_J=9$	$W_J=17$
IDGC	--	--	--	--
JDGC	$W_J=7$	$W_J=13$	$W_J=15$	$W_J=15$
			$W_{spe}=5$	
BGC	$W_{spa}=7$	$W_{spa}=21$	$W_{spa}=23$	$W_{spa}=19$
	$W_J=3$	$W_J=5$	$W_J=5$	$W_J=9$

C. Experimental Settings

In this paper, to fully evaluate the performance of the proposed BGC method, two classical HSI classification

algorithms, SVM and JSRC [14], and two improved variants of data gravitation based classification methods, IDGC [60] and JDGC [63], are adopted as comparisons. The SVM and IDGC only use the original spectral bands in this paper while JSRC and JDGC can synthetically utilize the SS information by combing with the joint neighborhood learning. To keep fair comparison, in the following experiments, the parameter settings of all the four compared methods were also decided by trial and errors, as summarized in Table I. For SVM, RBF was selected as the kernel function in which the penalty coefficient and gamma parameter were estimated by 5-fold cross validation. For JSRC, on all the datasets, the sparsity was set to 3 and the regularization parameter is empirically set as 10^{-2} . The size of the joint neighborhood was set to 7, 7, 9, and 17, respectively. IDGC does not involve any parameter. For JGDC, the values of W_j were set to 7, 13, 15, and 15, for testing. In addition, for all the five algorithms, the same settings of training samples were used as listed in Tables II-V. As shown, 10% of the samples were randomly selected as the training samples and the remaining 90% were chosen as the testing samples for the Indian Pines image. For the Salinas, Pavia University, and Grss_dfc_2014 datasets, the percent of the training samples and testing samples were set to 1% and 99%, respectively.

Table II
SAMPLE SETTINGS OF THE INDIAN PINES DATA

No.	Class	Training	Testing
1	Alfalfa	5	41
2	Corn-notill	143	1285
3	Corn-min	83	747
4	Corn	24	213
5	Grass/pasture	49	434
6	Grass/trees	73	657
7	Grass/ pasture-mowed	3	25
8	Hay-windrowed	48	430
9	Oats	2	18
10	Soybean-notill	98	874
11	Soybean-min	246	2209
12	Soybean-clean	60	533
13	Wheat	21	184
14	Woods	127	1138
15	Buildings-Grass-Trees-Drives	39	347
16	Stone-Steel Towers	10	83
Total		1031	9218

Table III
SAMPLE SETTINGS OF THE SALINAS DATA

No.	Class	Train	Test
1	Weed_1	21	1988
2	Weed_2	38	3688
3	Fallow	20	1956
4	Fallow_P	14	1380
5	Fallow_S	27	2651
6	Stubble	40	3919
7	Celery	36	3543
8	Grapes	113	11158
9	Soil	63	6140
10	Corn	33	3245
11	Lettuce_4wk	11	1057
12	Lettuce_5wk	20	1907
13	Lettuce_6wk	10	906
14	Lettuce_7wk	11	1059
15	Vineyard_U	73	7195
16	Vineyard_T	19	1788
Total		549	53580

Table IV
SAMPLE SETTINGS OF THE PAVIA UNIVERSITY DATA

No.	class	Training	Testing
1	Asphalt	67	6564
2	Meadows	187	18462
3	Gravel	21	2078
4	Trees	31	3033
5	Metal sheets	14	1331
6	Bare Soil	51	4978
7	Bitumen	14	1316
8	Bricks	37	3645
9	Shadows	10	937
Total		432	42344

Table V
SAMPLE SETTINGS OF THE GRSS_DFC_2014 DATA

No.	Class	Training	Testing
1	Road	45	4398
2	Trees	11	1082
3	Red roof	19	1835
4	Grey roof	22	2104
5	Concrete roof	39	3849
6	Vegetation	74	7283
7	Bare soil	18	1753
Total		228	22304

For the quantitative evaluation, Average classification Accuracy (AA) of each class, the Overall classification Accuracy (OA) and kappa coefficient (κ) were calculated to compare the performance of the five classification methods. Besides, the running time was utilized to compare the computation efficiency of the comparison methods. All the experiments were performed using MATLAB 2016b on a computer with 3.0-GHz CPU and 8-GB memory.

V. RESULTS AND DISCUSSIONS

A. Classification Results and Analysis

This section presents the qualitative and quantitative classification results as well as the running time of the four datasets yielded by the proposed BGC method and the other four compared methods. Figs. 6-9 show the classification maps of the four datasets. The corresponding classification accuracy is listed in Tables VI-IX. Table X summarizes the running time. For each test, the best results are marked in bold. Details and discussions of the experiments are given below.

1) *Indian Pines*: As shown in Fig. 6, the classification results of SVM and IGDC have many salt-and-pepper noise. In comparison, the results of JSRC, JGDC, and BGC methods are much better, due to the combination of SS information. Moreover, the joint neighborhood learning-based data gravitation methods, i.e. the JGDC and BGC, show better classification capability than that of JSRC as highlighted by the red circles in Fig. 6. This is because of that the neighbor average filtering mechanism of the joint model can alleviate the effects of noise, which has been widely used in image denoising applications. Further, the Bayesian based model helps to further promote the BGC method as shown in Fig. 6(e). In a word, the BGC shows the best qualitative results for Indian Pines. This is also confirmed by the quantitative results listed in Table VI. From Table VI we can conclude that in 15 out of the 16 kinds of crops, the BGC generated the best classification accuracy and obtained the best results in OA, AA, and κ . The utilization

of the data density distribution in spectral space and spatial distribution knowledge of the training samples may help to perform further utilization of the SS information. Besides, for class 7 and class 9 which only have 3 and 2 training samples respectively, the BGC method also got high classification accuracy. We therefore can infer that the BGC may also suit to classification problem with limited samples.

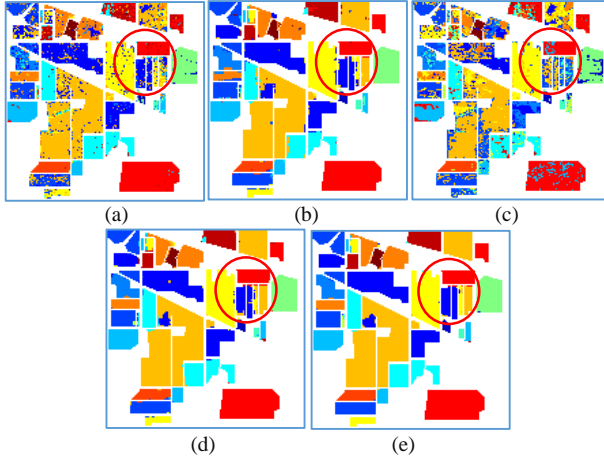


Fig. 6. Classification maps of the Indian Pines with 10% training samples per class: (a) SVM, (b) JSRC, (c) IGDC, (d) JGDC, and (e) BGC.

Table VI

CLASSIFICATION ACCURACY OF THE INDIAN PINES WITH 10% TRAINING SAMPLES PER CLASS

class	SVM	JSRC	IGDC	JGDC	BGC
1	0.4878	0.9268	0.8537	0.9268	0.9268
2	0.7385	0.9502	0.2467	0.9339	0.9782
3	0.6693	0.9465	0.7068	0.9759	0.9960
4	0.5540	0.8826	0.7042	0.9437	0.9624
5	0.9032	0.9585	0.6521	0.9862	0.9908
6	0.9665	0.9041	0.8174	0.9802	1.0000
7	0.5600	1.0000	0.9600	0.9200	1.0000
8	0.9581	0.9977	0.8163	0.9953	0.9953
9	0.3333	0.2778	1.0000	0.2222	0.7222
10	0.7620	0.9085	0.7197	0.9336	0.9771
11	0.7827	0.9683	0.5455	0.9909	0.9932
12	0.7598	0.8762	0.3133	0.9493	0.9812
13	0.9837	0.9185	0.9565	0.9511	0.9891
14	0.9156	0.9965	0.8383	0.9956	1.0000
15	0.5533	0.9049	0.1009	0.9625	0.9914
16	0.9518	0.8313	0.9157	0.8675	0.9277
OA	0.7963	0.9448	0.5950	0.9677	0.9882
AA	0.7425	0.8905	0.6967	0.9087	0.9645
κ	0.7679	0.9448	0.5468	0.9631	0.9865

2) *Salinas*: Fig. 7 shows the classification maps of the Salinas and the corresponding details of classification accuracy are summarized in Table VII. Comparing to the Indian Pines, the spatial resolution of Salinas is much high. Thus the spectral heterogeneity is more serious. Especially the Celery and Vineyard which are highlighted by the red circles are easily confused as shown in Fig. 7. As shown, SVM and IGDC are almost impossible to distinguish these two classes and the improvement of JSRC is limited. In comparison, the combination of SS information in JDGC and BGC can effectively promote their classification results and the BGC also obtained the best classification accuracy as illustrated in Table VII. This experiment demonstrated that the BGC can better process the HSI with high spatial resolution.

3) *Pavia University*: Fig. 8 displays the classification accuracy of the Pavia University. The corresponding details of

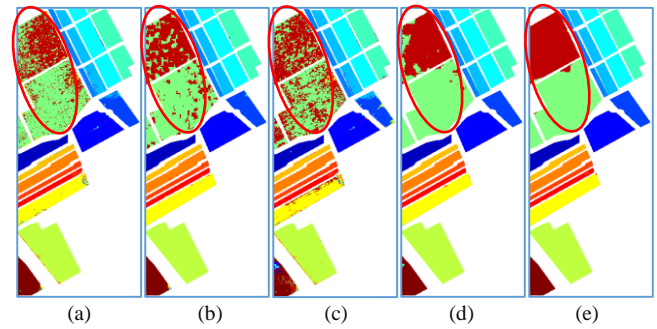


Fig. 7. Classification maps of the Salinas with 1% training samples per class: (a) SVM, (b) JSRC, (c) IGDC, (d) JGDC, and (e) BGC.

Table VII

CLASSIFICATION ACCURACY OF THE SALINAS WITH 1% TRAINING SAMPLES PER CLASS

class	SVM	JSRC	IGDC	JGDC	BGC
1	0.9925	1.0000	0.9904	1.0000	1.0000
2	0.9911	0.9986	0.9810	1.0000	1.0000
3	0.9054	0.9903	0.8088	0.9877	0.9949
4	0.9877	0.8949	0.9891	0.8877	0.9732
5	0.9725	0.9679	0.9736	0.9891	0.9936
6	0.9934	0.9967	0.9926	0.9867	0.9944
7	0.9944	0.9983	0.9879	1.0000	1.0000
8	0.8389	0.8654	0.5944	0.9876	0.9872
9	0.9811	1.0000	0.9619	1.0000	1.0000
10	0.9270	0.9735	0.7652	0.9837	0.9935
11	0.8865	0.9905	0.9461	0.9934	0.9943
12	1.0000	0.9722	0.9995	0.9974	0.9990
13	0.9735	0.9934	0.9735	0.9868	0.9823
14	0.9367	0.9594	0.9292	1.0000	0.9849
15	0.6853	0.7801	0.7104	0.8745	0.9933
16	0.9754	0.9877	0.7785	0.9989	0.9955
OA	0.9059	0.9322	0.8362	0.9742	0.9935
AA	0.9401	0.9257	0.8989	0.9796	0.9929
κ	0.8953	0.9606	0.8184	0.9713	0.9928

classification accuracy are listed in Table VIII. Like the Salinas, the spatial resolution of Pavia University is also very high. Thus, abundant details result in significant differences within the classification results. As shown in Fig. 8 (a)-(b), the SVM only using the spectral features shows a lot of misclassification which is also confirmed by the qualitative evaluation results listed in Table VIII. Besides, although the SS feature based JSRC and JDGC produced much better results, there are still existing numerous confusions between the Tress and the Meadows as highlighted in the red circles. Moreover, compared to the results of the Indian Pines and Salinas, the difference between JDGC and BGC is enlarged. This may because the BGC considers the local density of pixels to calculate their masses. It can be seen as a weighted process for JDGC. This overcomes the problem that the JDGC is greatly affected by its neighborhood heterogeneity [63]. Overall, the BGC method achieved the best classification results with an OA up to 0.97 as shown in Table VIII.

4) *Grss_dfc_2014*: Fig. 9 and Table IX show the qualitative and quantitative results of the Grss_dfc_2014 generated by the five comparisons, respectively. Due to the very high spatial resolution and long wavelength property, the classification difficulty of the Grss_dfc_2014 is much high than the other three datasets. As shown in Table IX, the overall classification accuracy of different algorithms is relatively lower. Nevertheless, the integration of SS information is still meaningful. As illustrated, the OA values of both SVM and

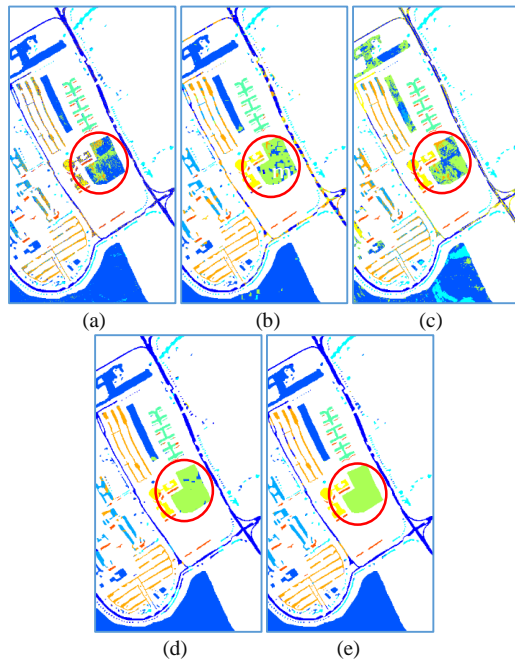


Fig. 8. Classification maps of the Pavia University with 1% training samples per class: (a) SVM, (b) JSRC, (c) IDGC, (d) JGDC, and (e) BGC.

Table VIII

CLASSIFICATION ACCURACY OF THE PAVIA UNIVERSITY WITH 1% TRAINING SAMPLES PER CLASS

class	SVM	JSRC	IDGC	JGDC	BGC
1	0.8419	0.5730	0.5672	0.9633	0.9759
2	0.9843	0.9802	0.6673	0.9957	0.9963
3	0.6487	0.8056	0.4418	0.8128	0.9644
4	0.7999	0.7751	0.9575	0.4857	0.7455
5	0.9722	0.9932	0.9902	0.9752	0.9992
6	0.3112	0.7987	0.6264	0.9620	1.0000
7	0.6429	0.9483	0.9187	0.9757	1.0000
8	0.7267	0.8883	0.7646	0.9871	0.9591
9	0.9082	0.5315	0.9989	0.7524	0.9669
OA	0.8285	0.8541	0.6904	0.9338	0.9704
AA	0.7595	0.8104	0.7703	0.8789	0.9564
κ	0.7509	0.8055	0.6129	0.9113	0.9605

IDGC are smaller than 0.7 and that of JSRC, IDGC, and BGC are all bigger than 0.8. The improve effect of the SS information can also be confirmed by the figures shown in Fig. 9 (b), Fig. 9(c) and Fig. 9(e). Especially for the circled area, although the very high spatial resolution leads to the image details cannot be fully displayed, the speckle problem of SVM and IDGC is obviously more serious. This further confirms that the proposed BGC is feasible for the classification of high-resolution HSIs.

5) *Running time*: Table X lists the running time of the five comparison methods on the four datasets. Although the IDGC is the most efficient, its classification performance is unable to meet the requirement as shown in Figs. 6-9 and Tables VI-IX. For the other four algorithms, as can be seen, the DGC based methods, i.e. JDGC and BDGC, have much higher efficiency than that of the SVM and JSRC. This is mainly because the SVM and JSRC methods are eager learning algorithms that require longer time to perform classification model training. In particular, there are the largest training samples in Indian Pines dataset (1031), thus SVM takes the longest running time in this case. By contrast, the DGC based methods are lazy learning algorithms that only need to calculate the data similarity between testing pixels and training samples. In addition,

although the BGC desires longer running time than JDGC, BGC obtained the best performance on all the four datasets and the time is much lower than SVM and JSRC.

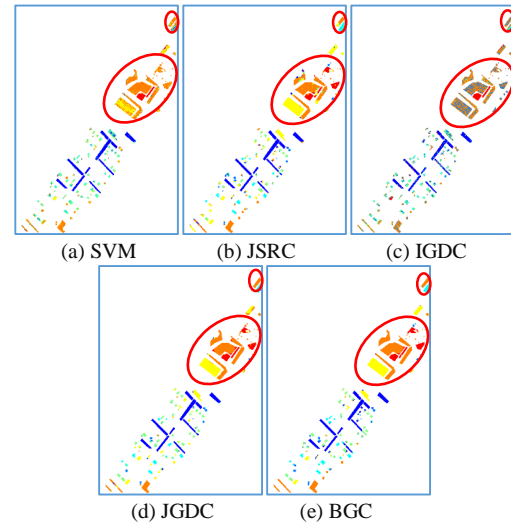


Fig. 9. Classification maps of the Grss_dfc_2014 with 1% training samples per class.

Table IX

CLASSIFICATION ACCURACY OF THE GRSS_DFC_2014 WITH 1% TRAINING SAMPLES PER CLASS.

class	SVM	JSRC	IDGC	JGDC	BGC
1	0.9659	0.9757	0.8481	0.9923	0.9527
2	0.0157	0.4455	0.2736	0.3919	0.5573
3	0.2594	0.6354	0.3041	0.6218	0.7297
4	0.5837	0.4815	0.5242	0.7376	0.6972
5	0.5214	0.9618	0.0057	0.9101	0.9623
6	0.8584	0.8439	0.6666	0.9470	0.9856
7	0.4404	0.7895	0.4398	0.8118	0.8945
OA	0.6725	0.8153	0.5082	0.8655	0.8989
AA	0.5207	0.7333	0.4375	0.7732	0.8256
κ	0.5784	0.7673	0.3895	0.8298	0.8724

TABLE X
COMPARISON OF RUNNING TIME (s)

datasets	SVM	JSRC	IDGC	JGDC	BGC
Indian Pines	101.077	63.203	0.547	1.175	1.952
Pavia University	20.384	161.989	0.942	4.654	5.669
Salinas	27.789	340.192	1.850	4.477	10.934
Grss dfc 2014	5.055	88.165	0.221	1.650	4.176

B. Comparison with State-of-the-Art Approaches

To further test the performance of the proposed algorithm, we compare the BGC algorithm with eight state-of-the-art approaches, including Generalized Tensor Regression (GTR) [71], Invariant Attribute Profiles (IAP) [72], Spatial-Spectral Manifold Reconstruction Preserving Embedding (SSMRPE) [29], Superpixelwise Adaptive SSA (SpaSSA) [73], mini-batch Graph Convolutional Networks (mini-GCNs) [74], 2DCNN [41], end-to-end Fusion Network (FuNet-C) [74], and CNN-Enhanced Graph Convolutional Network (CEGCN)[75]. These algorithms are all recently proposed for HSI classification with high precision. As to the parameter settings, for the first four algorithms, parameters are set by trial and errors. For the other four algorithms, we adopted the settings of parameters from the corresponding literatures [41, 74, 75]. The parameter settings for BGC are kept the same as the settings in Table I. The comparison results are listed in Table XI.

As shown in Table XI, the CEGCN with pixel- and

Table XI
COMPARISON WITH STATE-OF-THE-ART APPROACHES.

Datasets	accuracy	GTR	IAP	SSMRPE	SpaSSA	miniGCN	2DCNN	FuNet-C	CEGCN	BGC
Indian Pines (10%)	OA	0.9811	0.9630	0.9112	0.9777	0.7314	0.8679	0.8581	0.9904	0.9882
	κ	0.9785	0.9577	0.8986	0.9745	0.6996	0.8491	0.8377	0.9891	0.9865
	$T(s)$	19.803	6.415	28.129	27.256	12.884	216.200	121.277	9.54	1.952
Salinas (1%)	OA	0.9744	0.9671	0.9549	0.9729	0.8920	0.8944	0.8912	0.9933	0.9935
	κ	0.9715	0.9634	0.9498	0.9698	0.8798	0.8823	0.8788	0.9926	0.9928
	$T(s)$	205.410	14.831	44.826	150.815	56.434	1239.660	317.980	51.21	5.669
Pavia University (1%)	OA	0.9260	0.9324	0.8792	0.9423	0.8491	0.9034	0.9150	0.9944	0.9913
	κ	0.9013	0.9088	0.8384	0.9231	0.8021	0.8709	0.8871	0.9926	0.9883
	$T(s)$	117.000	15.340	19.648	212.572	35.520	437.115	163.848	93.36	10.934
Grss_dfc_2014 (1%)	OA	0.8680	0.8233	0.7397	0.8423	0.6299	0.6695	0.6778	0.8078	0.8989
	κ	0.8331	0.7733	0.6665	0.8021	0.5437	0.5777	0.5938	0.7135	0.8724
	$T(s)$	79.476	39.094	32.943	278.777	16.442	171.004	51.516	203.82	4.176

superpixel-level feature fusion obtains a close result to BGC in the first three datasets. The GTR, SpaSSA, IAP, and SSMRPE obtained relatively better classification performances than the other three deep leaning based algorithms in most cases. This is mainly because these algorithms deeply excavate the spectral features or comprehensively utilize the SS features, which further confirms the necessity of SS combination in HSI classification. In comparison, the BGC method does not involve with any feature extraction process but can effectively mine the SS information and achieved the better OA and kappa coefficient, especially on the complex scene of Grss_dfc_2014. Besides, the computational efficiency of BGC method has an absolute advantage as illustrated in Table XI, owing to the BGC method does not require training process.

C. Analysis of Training Sample Size

In view of the good performance of BGC in dealing with small-sample problem in previous experiments, we further tested its flexibility in cases with different sizes of small-sample. Specifically, we randomly selected training samples according to their number and percentage. For the number of sample size (num/class), the data is selected as 3, 5, 10, 12, and 15 per-class. For the percentage of sample size (per/class), the value is set to 1%, 3%, 5%, 8%, and 10% per-class. The Indian Pines dataset is selected as the testing data for it has 16 different kinds of targets and several of the classes only with few training samples. The parameter settings of BGC on each dataset are consistent with that listed in Table I.

We compared the experimental results with a recently presented model, denoted as 3D-Gabor method, which focuses on the HSI classification with small-sample [69]. The 3D-Gabor method also emphasizes the simultaneous utilization of the SS information. In this paper, the obtained results of BGC and that of the 3D-Gabor method are all summarized in Table XII. Note that the results of 3D-Gabor method are inherited from its original paper. As can be seen, for all the cases, the BGC algorithm is still superior to 3D-Gabor method. This comes from that the 3-D Gabor method in essential is a handcrafted feature-based method, which requires predefined image structures. In contrast, BDC does not need to design any feature extraction strategy using determined image structure while only explores the simple SS information from its local area. This may further verify that BGC has better data mining

capability than many other classification methods do.

Table XII
CLASSIFICATION ACCURACY (OA) OF THE INDIAN PINES WITH VARIABLE SIZE OF TRAINING SAMPLES

num/class	3	5	10	12	15
BGC	0.6916	0.7686	0.8458	0.8646	0.8940
3D-Gabor	0.6650	0.7400	0.8250	0.8450	0.8700
per/class	1%	3%	5%	8%	10%
BGC	0.9106	0.9545	0.9705	0.9815	0.9882
3D-Gabor	--	<0.9500	0.9600	0.9700	<0.9800

D. Discussions

The integration of SS information is essential for improving the HSI classification. Although the simplicity of data gravitation principle has motivated its application in HSIs, it is hard to combine with the spatial information. In contrast, our proposed BGC method can fully mine the SS information of the HSI and training samples. The BGC thus possesses three specific properties.

First of all, BGC method is a lazy learning algorithm with innate incremental learning ability. Hence it obtains outstanding classification performance on the Grss_dfc_2014 dataset as shown in Fig. 9 and Table XI. Moreover, BGC spend less running time as illustrated in Table X and Table XI since it does not require training process.

Secondly, the Bayesian based data mass calculation method combine the local spectral information of neighbors and the local spatial prior information of training samples, thus it effectively promotes the information mining of the limited training samples and improves the combination of SS information of testing pixels and training samples. In addition, the presentation of joint data gravitation model makes good use of the neighborhood pixels and alleviates the effects of heterogeneity in the high-resolution HSIs.

Lastly, the BGC method does not involve with any feature extraction process but can effectively mine the spectral and structural similarity of pixels. The classification accuracy can precede many of the state-of-the-art classification methods as reported in section V.B and Section V.C. This also show its high potential for the small sample classification problem.

VI. CONCLUSIONS

This paper presents a Bayesian gravitation-based classification (BGC) to further explore the SS information of HSIs. In BGC, the SS information of the neighboring pixels and training samples are firstly combined to calculate the mass of each pixel. Then, a joint gravitation model is designed based on the joint neighborhood learning strategy in which the joint Euclidean distance is weighted by the data mass. By this means, the SS similarity of each neighboring testing pixel can be measured. Finally, the testing pixel is labeled to the class that exerted the largest average data gravitation to the neighbors. To verify the proposed method, four benchmark hyperspectral datasets i.e. the Indian Pines, Salinas, Pavia University, and Grss_dfc_2014 data sets are tested. Comparison results with the SVM, JSRC, IDGC, JDC, and other eight state-of-the-art classification methods confirm the superiority of the proposed BGC. The BGC also shows feasibility on the HSI classification with limited samples. Nevertheless, its feasibility for extremely small size of training samples needs further research.

ACKNOWLEDGMENT

The authors would like to thank Telops Inc. (Québec, Canada) for acquiring and providing the data used in this study, the IEEE GRSS Image Analysis and Data Fusion Technical Committee and Dr. Michal Shimon (Signal and Image Centre, Royal Military Academy, Belgium) for organizing the 2014 Data Fusion Contest, the Centre de Recherche Public Gabriel Lippmann (CRPGL, Luxembourg) and Dr. M. Schlerf (CRPGL) for their contribution of the Hyper-Cam LWIR sensor, and Dr. M. D. Martino (University of Genoa, Italy) for her contribution to data preparation.

REFERENCES

- [1] J. Yue, W. Zhao, S. Mao et al., "Spectral-spatial classification of hyperspectral images using deep convolutional neural networks," *Remote Sens. Lett.*, vol. 6, no. 6, pp. 468-477, 2015.
- [2] A. AISuwaidi, B. Grieve, and H. Yin, "Feature-ensemble-based novelty detection for analyzing plant hyperspectral datasets," *IEEE J-STARS*, vol. 11, no. 4, pp. 1041-1055, 2018.
- [3] I. C. C. Acosta, M. Khodadadzadeh, L. Tusa et al., "A machine learning framework for drill-core mineral mapping using hyperspectral and high-resolution mineralogical data fusion," *IEEE J-STARS*, vol. 12, no. 12, pp. 4829-4842, 2019.
- [4] Y. Chang, L. Liang, C. Han et al., "multisource data fusion for landslide classification using generalized positive Boolean functions," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 6, pp. 1697-1708, 2007.
- [5] Y. Chen, X. Zhao, and X. Jia, "Spectral-spatial classification of hyperspectral data based on deep belief network," *IEEE J-STARS*, vol. 8, no. 6, pp. 2381-2392, 2015.
- [6] S. Xu, J. Li, M. Khodadadzadeh et al., "Abundance-indicated subspace for hyperspectral classification with limited training samples," *IEEE J-STARS*, vol. 12, no. 4, pp. 1265-1278, 2019.
- [7] C. Chen, W. Li, H. Su et al., "Spectral-spatial classification of hyperspectral image based on kernel extreme learning machine," *Remote Sens.*, vol. 6, no. 6, pp. 5795-5814, 2014.
- [8] M. Imani, and H. Ghassemian, "GLCM, Gabor, and morphology profiles fusion for hyperspectral image classification," in *Proc. ICEE*, Shiraz, Iran, 2016, pp. 460-465.
- [9] S. Jia, K. Wu, J. Zhu et al., "Spectral-spatial Gabor surface feature fusion approach for hyperspectral imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 1142-1154, 2019.
- [10] S. Jia, J. Hu, J. Zhu et al., "Three-dimensional local binary patterns for hyperspectral imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 4, pp. 2399-2413, 2017.
- [11] G. Rellier, X. Descombes, F. Falzon et al., "Texture feature analysis using a Gauss-Markov model in hyperspectral image classification," *IEEE Trans. Geosci. Remote.*, vol. 42, no. 7, pp. 1543-1551, 2004.
- [12] A. Zhang, S. Zhang, G. Sun et al., "Mapping of coastal cities using optimized spectral-spatial features based multi-scale superpixel classification," *Remote Sens.*, vol. 11, no. 9, pp. 998, 2019.
- [13] W. Zhao, and S. Du, "Spectral-spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4544-4554, 2016.
- [14] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification using dictionary-based sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3973-3985, 2011.
- [15] Y. Chen, Z. Lin, and X. Zhao, "Riemannian manifold learning based k-nearest-neighbor for hyperspectral image classification," in *Proc. IEEE-IGRSS*, Melbourne, VIC, Australia, 2013, pp. 1975-1978.
- [16] L. Zhang, Q. Zhang, B. Du et al., "Simultaneous spectral-spatial feature selection and extraction for hyperspectral images," *IEEE Trans. Cybernetics*, vol. 48, no. 1, pp. 16-28, 2018.
- [17] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification via kernel sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 217-231, 2013.
- [18] G. Sun, H. Huang, A. Zhang et al., "Fusion of multiscale convolutional neural networks for building extraction in very high-resolution images," *Remote Sens.*, vol. 11, no. 3, pp. 227, 2019.
- [19] M. Gönen, and E. Alpaydm, "Multiple kernel learning algorithms," *J. Mach. Learn. Res.*, vol. 12, pp. 2211-2268, 2011.
- [20] Y. Gu, J. Chanussot, X. Jia et al., "Multiple kernel learning for hyperspectral image classification: A review," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 11, pp. 6547-6565, 2017.
- [21] Y. Wang, Y. Zhang, and H. Song, "A Spectral-texture kernel-based classification method for hyperspectral images," *Remote Sens.*, vol. 8, no. 11, pp. 919, 2016.
- [22] Y. Gu, T. Liu, X. Jia et al., "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, 2016.
- [23] G. Sun, X. Rong, A. Zhang et al., "Multi-scale Mahalanobis kernel based support vector machine for classification of high-resolution remote sensing images," *Cognit. Comput.*, vol. 13, pp. 787-794, 2021.
- [24] F. Chen, B. Tang, T. Song et al., "Multi-fault diagnosis study on roller bearing based on multi-kernel support vector machine with chaotic particle swarm optimization," *Meas.*, vol. 47, pp. 576-590, 2014.
- [25] Y. Zhang, G. Cao, X. Li et al., "Cascaded random forest for hyperspectral image classification," *IEEE J-STARS*, vol. 11, no. 4, pp. 1082-1094, 2018.
- [26] P. Goel, S. Prasher, R. Patel et al., "Classification of hyperspectral data by decision trees and artificial neural networks to identify weed stress and nitrogen status of corn," *Comput. Electron. Agric.*, vol. 39, pp. 67-93, 2003.
- [27] J. Feng, J. Chen, L. Liu et al., "CNN-based multilayer spatial-spectral feature fusion and sample augmentation with local and nonlocal constraints for hyperspectral image classification," *IEEE J-STARS*, vol.

- 12, no. 4, pp. 1299-1313, 2019.
- [28] F. Luo, L. Zhang, X. Zhou et al., "Sparse-adaptive hypergraph discriminant analysis for hyperspectral image classification," *IEEE Geosci. Remote Sens.*, vol. 17, no. 6, pp. 1082-1086, 2020.
- [29] H. Huang, G. Shi, H. He et al., "Dimensionality reduction of hyperspectral imagery based on spatial-spectral manifold learning," *IEEE Trans. Cybernetics*, vol. 50, no. 6, pp. 2604-2616, 2020.
- [30] F. Luo, Z. Zou, J. Liu et al., "Dimensionality reduction and classification of hyperspectral image via multistructure unified discriminative embedding," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1-16, 2022.
- [31] Z. Wu, J. Sun, Y. Zhang et al., "Scheduling-guided automatic processing of massive hyperspectral image classification on cloud computing architectures," *IEEE Trans. Cybernetics*, vol. 51, no. 7, pp. 3588-3601, 2021.
- [32] P. Ghamisi, M. S. Couceiro, M. Fauvel et al., "Integration of segmentation techniques for classification of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 11, no. 1, pp. 342-346, 2014.
- [33] D. Comaniciu, and P. Meer, "Mean Shift: a robust approach toward feature space analysis," *IEEE Trans. Pattern Anal.*, vol. 24, no. 5, pp. 603-619, 2002.
- [34] R. Achanta, A. Shaji, K. Smith et al., "SLIC Superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal.*, vol. 34, no. 11, pp. 2274-2282, 2012.
- [35] G. Sun, Y. Hao, X. Chen et al., "Dynamic post-earthquake image segmentation with an adaptive spectral-spatial descriptor," *Remote Sens.*, vol. 9, no. 9, pp. 899, 2017.
- [36] T. Phientrakul, and B. Kijssirikul, "Evolutionary strategies for hyperparameters of support vector machines based on multi-scale radial basis function kernels," *Soft Comput.*, vol. 14, no. 7, pp. 681-699, 2010.
- [37] A. Mnassri, M. Bennis, and A. Cherif, "GA algorithm optimizing SVM multi-class kernel parameters applied in Arabic speech recognition," *Indian J. Sci. Technol.*, vol. 10, no. 27, pp. 1-9, 2017.
- [38] F. Kuang, S. Zhang, Z. Jin et al., "A novel SVM by combining kernel principal component analysis and improved chaotic particle swarm optimization for intrusion detection," *Soft Comput.*, vol. 19, no. 5, pp. 1187-1199, 2015.
- [39] D. P. Solomatine, M. Maskey, and D. L. Shrestha, "Eager and lazy learning methods in the context of hydrologic forecasting," in *Proc. Int. Jt. Conf. Neural. Netw.*, Vancouver, BC, Canada, 2006, pp. 4847-4853.
- [40] I. Hendrickx, and A. van den Bosch, "Hybrid algorithms with instance-based classification," in *Machine Learning: ECML 2005*, vol. 3720, Gama J., Camacho R., Brazdil P.B., Jorge A.M., Torgo L. eds., Berlin, Heidelberg: Springer, 2005, pp. 158-169.
- [41] Y. Chen, H. Jiang, C. Li et al., "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 6232-6251, 2016.
- [42] S. Piramanayagam, E. Saber, W. Schwartzkopf et al., "Supervised classification of multisensor remotely sensed images using a deep learning framework," *Remote sens.*, vol. 10, no. 9, pp. 1429, 2018.
- [43] S. Li, W. Song, L. Fang et al., "Deep learning for hyperspectral image classification: an overview," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6690-6709, 2019.
- [44] L. Dong, H. Du, F. Mao et al., "Very high resolution remote sensing imagery classification using a fusion of random forest and deep learning technique-subtropical area for example," *IEEE-JSTARS*, vol. 13, pp. 113-128, 2020, 2020.
- [45] X. Zhang, G. Sun, X. Jia et al., "Spectral-spatial self-attention networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1-15, 2022.
- [46] Y. Li, H. Zhang, and Q. Shen, "Spectral-spatial classification of hyperspectral imagery with 3D convolutional neural network," *Remote Sens.*, vol. 9, no. 1, pp. 67, 2017.
- [47] J. Yang, Y. Zhao, and J. C. Chan, "Learning and transferring deep joint spectral-spatial features for hyperspectral classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 8, pp. 4729-4742, 2017.
- [48] G. Sun, X. Zhang, X. Jia et al., "Deep fusion of localized spectral features and multi-scale spatial features for effective classification of hyperspectral images," *Int. J. Appl. Earth Obs.*, vol. 91, pp. 102157, 2020.
- [49] A. Aamodt, and E. Plaza, "Case-based reasoning: foundational issues, methodological variations, and system approaches," *AI Commu.*, vol. 7, no. 1, pp. 39-59.
- [50] Z. Hou, S. Liu, and T. Tian, "Lazy-learning-based data-driven model-free adaptive predictive control for a class of discrete-time nonlinear systems," *IEEE Trans. Neur. Net. Lear.*, vol. 28, no. 8, pp. 1914-1928, 2017.
- [51] B. Tu, X. Zhang, X. Kang et al., "Hyperspectral image classification via fusing correlation coefficient and joint sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 15, no. 3, pp. 340-344, 2018.
- [52] X. Ou, Y. Zhang, H. Wang et al., "Hyperspectral image target detection via weighted joint K-nearest neighbor and multitask learning sparse representation," *IEEE Access*, vol. 8, pp. 11503-11511, 2020.
- [53] B. Tu, S. Huang, L. Fang et al., "Hyperspectral image classification via weighted joint nearest neighbor and sparse representation," *IEEE JSTARS*, vol. 11, no. 11, pp. 4063-4075, 2018.
- [54] L. Peng, B. Yang, Y. Chen et al., "Data gravitation based classification," *Inform. Sciences.*, vol. 179, no. 6, pp. 809-819, 2009.
- [55] M. Indulska, and M. E. Orłowska, "Gravity based spatial clustering," in *Proc. GIS'02*, McLean, Virginia, USA, 2002, pp. 125-130.
- [56] A. Teng, L. Peng, Y. Xie et al., "Gradient descent evolved imbalanced data gravitation classification with an application on internet video traffic identification," *Inform. Sciences.*, vol. 539, pp. 447-460, 2020.
- [57] G. Wen, J. Wei, J. Wang et al., "Cognitive gravitation model for classification on small noisy data," *Neurocomputing*, vol. 118, pp. 245-252, 2013.
- [58] O. Reyes, C. Morell, and S. Ventura, "Effective lazy learning algorithm based on a data gravitation model for multi-label learning," *Inform. Sciences.*, vol. 340-341, pp. 159-174, 2016.
- [59] Z. Wang, Z. Yu, C. L. P. Chen et al., "Clustering by local gravitation," *IEEE Trans. Cybernetics*, vol. 48, no. 5, pp. 1383-1396, 2018.
- [60] L. Peng, H. Zhang, B. Yang et al., "A new approach for imbalanced data classification based on data gravitation," *Inform. Sciences.*, vol. 288, pp. 347-373, 2014.
- [61] L. Peng, H. Zhang, Y. Chen et al., "Imbalanced traffic identification using an imbalanced data gravitation-based classification model," *Comput. Commun.*, vol. 102, pp. 177-189, 2017.
- [62] G. Sun, A. Zhang, J. Ren et al., "Gravitation-based edge detection in hyperspectral images," *Remote Sens.*, vol. 9, no. 6, pp. 592, 2017.
- [63] A. Zhang, C. Zhang, J. Rong et al., "Gravitation based classification method for hyperspectral imagery," in *Pro. WHISPERS*. Amsterdam, Netherlands, Netherlands, 2019, pp. 1-4.
- [64] L. Sun, Z. Wu, J. Liu et al., "Supervised spectral-spatial hyperspectral image classification with weighted Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1490-1503, 2015.
- [65] P. Shafigh, S. Y. Hadi, and E. Sohrab, "Gravitation based classification," *Inform. Sciences.*, vol. 220, pp. 319-330, 2013.

- [66] I. Newton, *Mathematical principles of natural philosophy*, London, UK: A. Strahan, 1802.
- [67] W. Hsu, "Bayesian classification," in *Encyclopedia of Database Systems*, L. Liu and M. T. Özsu, eds., New York, USA: Springer, 2018, pp. 263-267.
- [68] D. Leman, "Expert system diagnose tuberculosis using Bayes Theorem method and Shafer Dempster method," in *Proc. CITSM*, Parapat, Indonesia, Indonesia, 2018, pp. 1-4.
- [69] S. Jia, L. Shen, J. Zhu et al., "A 3-D Gabor phase-based coding and matching framework for hyperspectral imagery classification," *IEEE Trans. Cybernetics*, vol. 48, no. 4, pp. 1176-1188, 2018.
- [70] W. Liao, X. Huang, F. V. Coillie et al., "Processing of multiresolution thermal hyperspectral and digital color data: outcome of the 2014 IEEE GRSS Data Fusion Contest," *IEEE JSTARS*, vol. 8, no. 6, pp. 2984-2996, 2015.
- [71] J. Liu, Z. Wu, L. Xiao et al., "Generalized tensor regression for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 2, pp. 1244-1258, 2020.
- [72] D. Hong, X. Wu, P. Ghamisi et al., "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, 2020.
- [73] G. Sun, H. Fu, J. Ren et al., "SpaSSA: Superpixel wise adaptive SSA for unsupervised spatial-spectral feature extraction in hyperspectral image," *IEEE Trans. Cybernetics*, vol. 52, no. 7, pp. 6158-6169, 2022.
- [74] D. Hong, L. Gao, J. Yao et al., "Graph convolutional networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5966-5978, 2021.
- [75] Q. Liu, L. Xiao, J. Yang et al., "CNN-enhanced graph convolutional network with pixel- and superpixel-level feature fusion for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 10, pp. 8657-8671, 2021.



Aizhu Zhang (Member, IEEE) received the B.Sc., M.Sc. and Ph.D. degrees from China University of Petroleum (East China), Qingdao, China, in 2011, 2014, and 2017, respectively.

She is currently a Lecturer with China University of Petroleum (East China), Qingdao, China. Her research interests are in multi-source remote sensing processing, wetland remote sensing and city environment remote sensing.



Genyun Sun (Member, IEEE) received the B.Sc. degree from Wuhan University, Wuhan, China, in 2003 and Ph.D. degree in Institute of Remote Sensing Applications, Chinese Academy of Sciences in 2008.

He is currently a Professor with China University of Petroleum (East China), Qingdao, China. His research interests cover remote sensing image processing, including hyperspectral and high resolution remote sensing and intelligent optimization.



Zhaojie Pan received the B.Sc. degree in geomatics engineering from China University of Petroleum (East China), Qingdao, China, in 2021. He is currently working toward the M.Sc. degree at China University of Petroleum (East China), Qingdao, China.

His research interests include hyperspectral classification and deep learning.



Jinchang Ren (Senior Member, IEEE) received the B.Eng., M.Eng. and D.Eng. degrees from the Northwestern Polytechnical University, Xi'an, China in 1992, 1997 and 2000, respectively, and the Ph.D. degree from the University of Bradford, Bradford, U.K., in 2019.

He is currently a Professor with the National Subsea Centre, Robert Gordon University, Aberdeen, U.K. His research interests include image processing, computer vision, machine learning, and big data analytics.

Dr. Ren acts as Associate Editor for several international journals, including *IEEE Transactions on Geoscience and Remote Sensing* (TGRS) and *Journal of the Franklin Institute*.



Xiuping Jia (Fellow, IEEE) received the B.Eng. degree from the Beijing University of Posts and Telecommunications, Beijing, China, in 1982 and the Ph.D. degree in Electrical Engineering from The University of New South Wales, Australia, in 1996.

Since 1988, she has been with the School of Engineering and Information Technology, The University of New South Wales at Canberra, Australia, where she is currently an Associate Professor. She has authored or coauthored more than 270 referred papers,

including over 160 journal papers, addressing various topics including data correction, feature reduction and image classification using machine learning techniques. She has coauthored of the remote sensing textbook titled *Remote Sensing Digital Image Analysis* [Springer-Verlag, 3rd (1999) and 4th eds. (2006)]. Her research interests include remote sensing, hyperspectral image processing and spatial data analysis.

Dr. Jia is a Subject Editor of the *Journal of Soils and Sediments* and an Associate Editor-in-Chief of the *IEEE Transactions on Geoscience and Remote Sensing* (TGRS).



Chenglong Zhang received the M.Sc. degree from China University of Petroleum (East China), Qingdao, China, in 2018. He is currently pursuing the Ph.D. degree in at Shandong University, Jinan, China.

His research interests include hyperspectral image processing and medical object identification.



Hang Fu received the B.Sc. degrees in geomatics engineering from China University of Petroleum (East China), Qingdao, China, in 2019. He is currently pursuing the PhD degree in computer technology and resource information engineering at China University of Petroleum (East China), Qingdao, China.

His research interests are related to hyperspectral image feature extraction and classification.



Yanjuan Yao received the Ph.D. degree in cartography and remote sensing from the Institute of Remote Sensing Applications, Chinese Academy of Sciences, Beijing, in 2007. She undertook the post-doctoral research at Peking University from 2007 to 2009. She is currently a Professor with the Satellite Environment Center (SEC), Ministry of Environmental Protection, Beijing.

Her research interests include radiation transfer modeling for optical remote sensing, terrestrial parameter inversion from multisource remote sensing data, and quantitative remote sensing application for e