Hyperspectral Band Selection Using Crossover based Gravitational Search Algorithm

Aizhu Zhang[1,2], Ping Ma[1,2], Sihan Liu[3*], Genyun Sun[1,2*], Hui Huang[1,2], Jaime Zabalza[4], Zhenjie Wang[1,2], Chengyan Lin[1,2]

[1] School of Geosciences, China University of Petroleum (East China), Qingdao, China;
[2] Laboratory for Marine Mineral Resources, Qingdao National Laboratory for Marine Science and Technology, Qingdao, China;
[3] Satellite Environment Center, Ministry of Environmental Protection of China, Beijing, China
[4] Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow, U.K.

[*] corresponding.liusihan1200@163.com (Sihan Liu); genyunsun@163.com (Genyun Sun)

Abstract: Band selection is an important data dimensionality reduction tool in hyperspectral images (HSIs). To identify the most informative subset band from the hundreds of highly corrected bands in HSIs, a novel hyperspectral band selection method using a crossover based gravitational search algorithm (CGSA) is presented in this paper. In this method, the discriminative capability of each band subset is evaluated by a combined optimization criterion, which is constructed based on the overall classification accuracy and the size of the band subset. As the evolution of the criterion, the subset is updated using the *V*-shaped transfer function based CGSA. Ultimately, the band subset with the best fitness value is selected. Experiments on two public hyperspectral datasets, i.e. the Indian Pines dataset and the Pavia University dataset, have been conducted to test the performance of the proposed method. Comparing experimental results against the basic GSA and the PSOGSA (hybrid PSO and GSA) revealed that all of the three GSA variants can considerably reduce the band dimensionality of HSIs without damaging their classification accuracy. Moreover, the CGSA shows superiority on both the effectiveness and efficiency compared to the other two GSA variants.

1. Introduction

Hyperspectral remote sensors can synchronously record hundreds of narrow spectral bands from the same scene. The obtained spectral data can characterise the properties of different materials and potentially be helpful for the analysis of different objects in the scene. However, due to many of the spectral bands are highly related, the hyperspectral images (HSIs) are always of high degree of information redundancy and requires a lot of storage space [1]. Although too few spectral bands are hard to produce acceptable accuracy, the serious information redundancy in HSIs also wrecks the data analysis accuracy, and causes the well-known *Curse of Dimensionality or Hughes Phenomenon* [2-3]. Consequently, extracting the most informative data from the original spectral bands and thereby reducing the information redundancy has become an essential problem for the analysis and application of HSIs [4].

Feature extraction is a typical kind of technique for mitigating the data dimensionality reduction problem [5-6]. Many of the feature extraction algorithm are constructed based on the geometric and affine transformations, such as Principal Component Analysis (PCA) [7], Maximum-Noise Fraction transformation (MNF) [8], Independent Component Analysis (ICA) [9], and wavelet-based transforms [10]. Although these aforementioned methods have been widely utilized in the data compression of HSIs, they may lead to physically non-interpretable results since they always realize the compression purpose by changing the original physical meaning of the original data [11].

In contrast, band selection methods select the most informative band subset from the original spectral bands based on statistical analysis and optimization criteria [12], which can keeps the original physical meaning of each sleeted band. That is to say, band selection can preserve useful information in a more complete way and reduce the data dimension of HSIs as well.

Exhaustive algorithm is the most basic method for selecting subset of bands on the base of the statistical analysis and optimization criteria. In this method, each kind of band combination needs to be verified and then the most suitable subset can be obtained. That is, if a HSIs has D spectral bands, the exhaustive algorithm will have to test 2^D times band combination to search for the most informative subset bands. If the D is a large number, exhaustive algorithm becomes impracticable. Thereby, many nature-inspired algorithms (NAs) have been introduced to reduce the computational time of band selection in recent years. For example, classical NAs including Genetic Algorithm (GA) [13], Particle Swarm Optimization (PSO) [1, 14], and Ant Colonization Optimization (ACO) [15] etc. have been adapted to the area of band selection for HSIs.

Gravitational Search Algorithm (GSA) is a recently proposed NA inspired by the law of Newton's gravity and mass interactions [17]. Owing to its simple concept and superior performance, GSA has attracted much attention from researchers in different application areas [17-19]. Various experimental results have demonstrated the high computational efficiency and the competitive convergence performance over many other NAs [17, 20-21]. Thanks to these advantages, GSA has attracted increasing interest in the field of engineering optimization, such as parameter

identification [22], data clustering [23], image classification [24], and thresholding [25].

Also, these aforementioned advantages and successful application of GSA make it a promising choice for the feature/band selection problems. For example, in [26], the optimization behaviours of GSA are combined together with the speed of Optimum-Path Forest (OPF) classifier to provide a fast and accurate framework for feature selection. In [27], an improved version of the binary GSA is proposed and used as a tool to select the best subset of features with the goal of improving classification accuracy. In [28], GSA is utilized to perform feature subset selection for intrusion detection system. In [29], a GSA based automatic unsupervised feature selection method which requires no prior knowledge of the data to be classified is developed. The chaotic maps based GSA also has been applied to the band selection of airborne hyperspectral image [30].

Nevertheless, due to the fact that GSA cannot maintain and utilize the global best position achieved until now (***Gb***) in the search process, the basic GSA is inclined to confront weak exploitation when handling complex problems [31-32]. In this paper, to alleviate the aforementioned problem, a crossover based GSA (CGSA) is proposed and extended to recognize the most informative band subset for HSIs. In the proposed method, a ***Gb*** based crossover is randomly inserted to GSA based on a crossover probability. Therefore, the CGSA can randomly inherit some promising search directions from ***Gb*** and largely enhance its exploitation ability. When extending CGSA to band selection, we first code the position of each particle in CGSA within a binary space. Each particle represents a candidate band subset. Subsequently, each candidate subset is evaluated based on a combined optimization criterion constructed by the overall classification accuracy and the size of the candidate subset. Finally, the band subset with the smallest fitness values, i.e. the subset with less bands and more discriminative spectral information is obtained.

The remainder of this paper is organized as follows. The general processing of band selection and the basic GSA is briefly described in Section 2. Section 3 introduces the details of the proposed CGSA-based band selection method. In Section 4, the experimental data, comparison results, and analysis are presented. At last, Section 5 provides a conclusion for this work.

2. Band selection and basic GSA

2.1. Band selection based on NAs

In the band selection methods based on NAs, the problem of band selection is modelled as an optimization problem in a *D*-dimensional space, where *D* stands for the number of spectral channels. In such a case, each binary coded candidate solution is associated with a subset of bands in the *D*-dimensional space. The candidate solutions are then updated and optimized following the optimization of NAs. The main frameworks of NAs based band selection methods include four main steps: initial subset generation, subset evaluation, search strategy for subset update, and stopping criteria. The initialization and stopping are two common processes in NAs while the other two steps perform important role for the effectiveness of the band selection method. Although search strategies of different NAs are various, two key and general issues included in subset evaluation and search strategy are the optimization criterion for subset evaluation and the transfer function for mapping a continuous search space to a discrete search space.

2.1.1 Optimization criterion: The optimization criterion is used as the fitness function to evaluate the quality of the selected bands. For the supervised band selection, the most widely applied optimization criterion is the maximum of classification overall accuracy (*OA*). For a candidate solution, the corresponding *OA* is calculated by:

$$OA = \frac{\sum_{i=1}^{Nc} C_{ii}}{\sum_{j=1}^{Nc}\sum_{i=1}^{Nc} C_{ij}} \times 100 \quad (1)$$

where N_c is the number of classes, C_{ii} is the number of pixels correctly assigned to class *i*, C_{ij} is the number of pixels assigned to class *j*, which belongs to class *i*.

Indeed, for each candidate solution, we need to train and test a classifier to compute the *OA*. A candidate solution with a higher *OA* are always considers as a more informative subset with higher separability.

2.1.2 Transfer functions: Most of the NAs are proposed originally for solving the continuous search space other than the discrete search space. Thus for solving the band selection problem, a transfer function to construct the binary version of a NA and preserve the concepts of the search process is very important. The capability of the transfer function is to map velocity values of each candidate solution to probability values and force particles to move in a binary space [33]. Two of the main families of transfer functions are *S*-shaped and *V*-shaped transfer functions [34], as shown in Eq. (2) and Eq. (3), respectively. The equations and figures of four *S*-shaped and four *V*-shaped transfer functions are given in Fig. 1.

$$\begin{cases} S1: T(y) = \frac{1}{1+e^{-2y}}, \\ S2: T(y) = \frac{1}{1+e^{-y}}, \\ S3: T(y) = \frac{1}{1+e^{-\frac{y}{2}}}, \\ S4: T(y) = \frac{1}{1+e^{-\frac{y}{3}}}, \end{cases} \quad (2)$$

$$\begin{cases} V1: T(y) = \left|\mathrm{erf}(\frac{\sqrt{\pi}}{2}y)\right|, \\ V2: T(y) = |\tanh(y)|, \\ V3: T(y) = \left|\frac{y}{\sqrt{1+y^2}}\right|, \\ V4: T(y) = \left|\frac{2}{\pi}\arctan(\frac{\pi}{2}y)\right|. \end{cases} \quad (3)$$

where *y* is the value of a velocity vector's element in a dimension, *T*(*y*) is the corresponding probability calculated based on the transfer functions as shown in Eq. (2)-Eq. (3).

As shown in Fig. 1, when the value of velocity vector's elements bigger than 0, although the shapes of the curves are different, both the *S*-shaped and *V*-shaped transfer functions assign an increased probability of position vector's elements change (from 0 to 1 or vice versa) as the value of velocity increased. When the value of velocity vector's elements are smaller than 0, the *S*-shaped transfer functions assign a decreased probability of position vector' elements change as the value of velocity increased as shown in Fig.

1(a). In contrast, the *V*-shaped transfer functions assign an increased probability of position vector' elements change as the value of velocity increased as illustrated in Fig. 1(b). In [34], the properties and effectiveness of the two families of transfer functions have been investigated. It is demonstrated that the *V*-shaped transfer functions, especially the V4 functions performed much better than the *S*-shaped transfer functions in binary PSO algorithms.

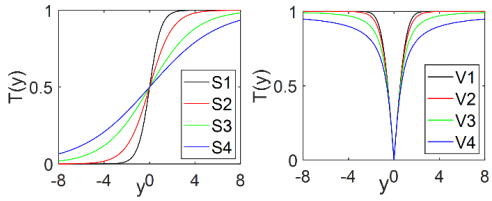


Fig. 1 *The S-shaped and V-shaped families of transfer functions.* ***(a)*** *S-shaped transfer function* ***(b)*** *V-shaped transfer functions.*

2.2. Basic GSA

In the processing of GSA, each particle $\boldsymbol{X}_i = [x_{i1}, x_{i2}, \dots, x_{iD}]$ (i={1, 2,…, *NP*}) is defined as a mass object moving through the *D*-dimensional search space with a velocity $\boldsymbol{V}_i = [v_{i1}, v_{i2}, \dots, v_{iD}$. *NP* denotes the size of the population. The velocity of each particle is initialized to zeros and the update relies on the gravitational forces exerted by its neighbours following the law of gravity [17]. According to the law of gravity, the gravitational force between two particles is directly proportional to their masses and inversely proportional to their distance. Therefore, we can follow that with the gravitational force, the lighter mass will be attracted and moves towards the heavier ones. For a population with *NP* particles in GSA, all the particles will move towards those particles that have heavier masses, and ultimately realize the convergence of all the particles [17].

Due to the mass of particle performing a very important role in the processing of GSA, the masses of particles are calculated from their fitness values as follows:

$$nmfit_i^t = \frac{fit_i^t - worst^t}{best^t - worst^t} \quad (4)$$

$$Mass_i^t = \frac{nmfit_i^t}{\sum_{j=1}^{N} nmfit_j^t} \quad (5)$$

where t is the current iteration, fit_i^t is the fitness value of the particle i at current time, $Mass_i^t$ represents the mass of particle i, $worst^t$ and $best^t$ denotes the worst and best fitness values of a population in the current time. For a maximization problem, $worst^t$ and $best^t$ are defined by:

$$worst^t = \min_{j\in\{1,\dots,N\}} fit_j^t \quad (6)$$

$$best^t = \max_{j\in\{1,\dots,N\}} fit_j^t \quad (7)$$

For a minimum problem, the definition of $worst^t$ and $best^t$ is the other way round.

For the gravitational force, the force acting on the particle i from the particle j in each dimension d at the t-th iteration is calculated follows

$$F_{id,jd}^t = G^t \frac{Mass_i^t \times Mass_j^t}{R_{ij}^t + \varepsilon}\left(x_{jd}^t - x_{id}^t\right) \quad (8)$$

where $Mass_i^t$ and $Mass_j^t$ are the masses of the particles i and j in the current iteration, R_{ij}^t is the Euclidian distance between the particles i and j in iteration t; ε is a small positive constant, which is defined as 10^-6 in this paper, x_{id}^t and x_{jd}^t represents the position of the i-th and j-th particles in the d-th dimension in iteration t, G^t is a decreasing gravitational constant for controlling the search accuracy, which is defined as

$$G^t = G_0 \times \exp\left(-\alpha \times \frac{t}{T_{max}}\right) \quad (9)$$

where G_0 is the initial value of gravitational constant, α is a decrease coefficient, t is the current iteration, and $T_{\max}$ is the maximum number of iterations. In the basic GSA, the G_0 and α is set to 20 and 100, respectively.

Generally, in the iteration t, the total gravitational force acts on the particle i in the d-th dimension, F_{id}^t, should be the sum of all the gravitational forces exerted from other N-1 particles. In the basic GSA, to promote the balance between exploration and exploitation as well as give a stochastic characteristic to GSA, the F_{id}^t is defined as the randomly weighted sum of the forces exerted from K_{best} particles as given below:

$$F_{id}^t = \sum_{j\in K_{\text{best}}, j\neq i}^{NP} rand_j \cdot F_{id,jd}^t \quad (10)$$

where $rand_j$ represents a random number between interval [0,1], K_{best} is an archive that stores the particles ranked in the first K position after fitness sorting in each iteration, the value of K is initialized as *NP* in the beginning and linearly decreased with time down to one. Obviously, with the K_{best} model, each particle attracted by less and less particles in the iterations. That is, the exploration fades out while the exploitation fades in as time goes by. Finally, all the particles tend to refine the local area around the global best particle. This operation plays a crucial role in the balance of exploration and exploitation in basic GSA.

Following the obtained gravitational force and the law of motion, the acceleration of the particle i in the d-th dimension at iteration t, a_{id}^t, can be obtained by

$$a_{id}^t = \frac{F_{id}^t}{Mass_i^t} \quad (11)$$

Therefore, based on the obtained acceleration, the velocity and the position of the particle i in iteration t can be updated as follows:

$$v_{id}^{t+1} = rand_i \times v_{id}^t + a_{id}^t \quad (12)$$

$$x_{id}^{t+1} = x_{id}^t + v_{id}^{t+1} \quad (13)$$

where $rand_i$ is a uniform random variable in the interval [0, 1].

3. CGSA-based band selection

3.1. The proposed CGSA

In CGSA, a $\boldsymbol{Gb}$ guided crossover operator is introduced to promote the exploitation ability of the basic GSA by:

$$x_{id}^{t+1} = Gb_d^t + rand \cdot (x_{id}^{t+1} - pb_{jd}^t) \quad (14)$$

where Gb_d^t denotes the d-th dimension of the global best position of the population achieved until now ($\boldsymbol{Gb} = [gb_1, gb_2, \dots, gb_D]$), pb_{jd}^t is the d-th dimension of the personal best position of the particle j (randomly selected

from the *NP* particles) achieved until now ($\boldsymbol{Pb}_j = [pb_{j1}, pb_{j2}, \ldots, pb_{jD}]$), and rand is a uniform random variable in the interval [0, 1]. Obviously, the promising information from the both the ***Gb*** and $\boldsymbol{Pb}_j$ are all combined into the new position of the particle to perform a more refined exploitation around the promising areas.

In the evolution process, after calculating the velocity of each particle, CGSA executes the proposed crossover operation to constitute a new trial solution. The new position update equations in CGSA are formulated as follows:

$$\begin{cases} x_{id}^{t+1} = x_{id}^{t} + v_{id}^{t+1}, & if\ rand < p_c, \\ x_{id}^{t+1} = Gb_d^t + rand \cdot \left(x_{id}^t - pb_{jd}^t\right), & Otherwise. \end{cases} \quad (15)$$

where p_c is the crossover rate which controls the probability of inheriting from the ***Gb***. For a healthy search process, the optimization algorithm should emphasize on the exploration in the earlier search stages while paying more attention to the exploitation in the latter search stages. Therefore, the value of p_c is adaptively adjusted along with the iteration following:

$$p_c = 1 - \left({}^{t}/_{T_{max}}\right) \quad (16)$$

With the adaptively adjusted p_c, particles gains increased probability to learn from the ***Gb*** with evolution of the population. The flowchart of the proposed CGSA is given in Fig. 2.

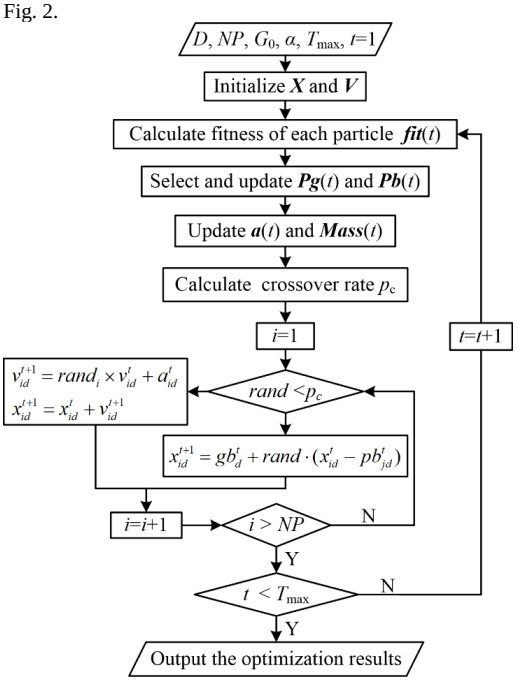


Fig. 2 *Flowchart of CGSA.*

3.2. CGSA based band selection

To adapt CGSA to the problem of hyperspectral band selection, some modifications involve population initialization and subset generation should be done. Accordingly, the CGSA based hyperspectral band selection includes a four step routine: (1) population initialization and band mapping, (2) subset evaluation based on supervised classification, (3) subset update based on CGSA, and (4) stopping criteria. Detailed description of each step is presented in the followings subsections 3.2.1-3.2.4.

3.2.1 Population initialization and band mapping: For a HSIs with *D* bands, we need to initialize a population with *NP* candidate band subsets first. The value in each dimension is randomly set to 0 or 1. That is, each particle $\boldsymbol{X}_i = [x_{i1}, x_{i2}, \ldots, x_{iD}]$ stands for a candidate band subset with *D* dimension. For each particle, if the value of x_{ij} (j={1,2,…,D}) is 0, the *j*-th band of the original HSIs is abandoned. Otherwise, if the value of x_{ij} is 1, the *j*-th band of the original HSIs is selected. Obviously, the population initialization process is also a band mapping step for the HSIs. An illustration of the band mapping is given in Fig. 3.

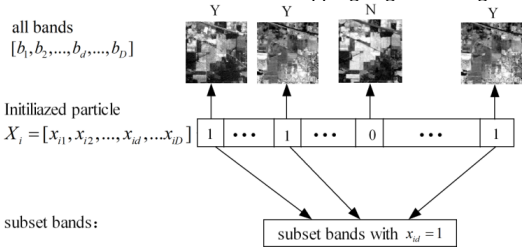


Fig. 3 *Illustration of the band mapping.*

3.2.2 Subset evaluation based on supervised classification: The evaluation of band subset, i.e. the fitness evaluation of each particle, relies on the objective function or optimization criterion. Because the goal of band selection is to identify the most informative bands from the original bands of HSIs, a better band subset should contribute as much as possible to the classification accuracy while containing as few bands as possible. Accordingly, an objective function that combines the overall classification accuracy of the Support Vector Machine (SVM) classifier and number of bands is utilized in this paper follows

$$fit(\boldsymbol{X}_i) = \mathrm{OA}(\boldsymbol{X}_i) - \omega \times \frac{D_b}{D} \quad (17)$$

where $OA(\boldsymbol{X}_i)$ is the overall classification accuracy, ω is a weight factor for balancing the classification accuracy and the size of the *i*-th band subset. Note that the value of D_b is the sum of each dimension of the particle $\boldsymbol{X}_i$, i.e. the number of selected bands.

From the objective function we can conclude that a larger ω will make the band selection method emphasize more on the dimensionality reduction while a smaller ω makes the band selection method concentrate more on the classification accuracy. In this paper, the parameter ω is experimentally set to 0.6.

3.2.3 Subset update based on CGSA: After obtaining the fitness of each candidate solution, the velocity of them can be updated following Eqs. (4)-(16). Then we need to update the position of each particle based on the transfer functions. Following the introduction in Section 2.2.2, the *V*-shaped transfer functions V2 is adapted in this paper. That is, the velocity of a particle can be associated to the probability of changing its state as

$$x_{id}^{t+1} = \begin{cases} x_{id}^{t+1}, & if\ \Delta x_{id}^{t+1} > rand, \\ complement(x_{id}^{t}), & Otherwise. \end{cases} \quad (18)$$

$$where \quad \Delta x_{id}^{t+1} = |tanh(v_{id}^{t+1})|$$

where $complement(x_{id}^{t})$ denote the complement of the original binary value of x_{id}^{t}, i.e., if the original value of x_{id}^{t} is 0 the $complement(x_{id}^{t})$ is set to 1, vice versa.

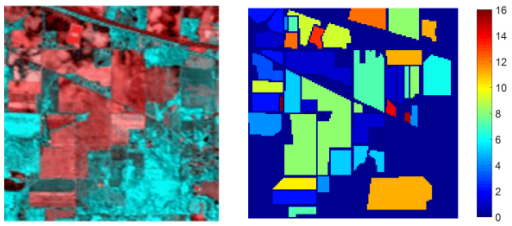
3.2.4 Stopping criteria: Following the process of CGSA, the population keeps iterative evolution and the band subset gradually optimizes until a predesigned stopping criterion is reached. Typical stopping criteria include maximum number of iterations (T_{max}), maximum fitness evaluations times, and so on. In this study, the T_{max} is chosen as the stopping criterion. Finally, when the algorithm reaches the maximum number of iterations, the particle that possesses the minimum fitness values is outputted as the optimal band subset.

4. Experiment results and discussions

To validate the proposed CGSA for hyperspectral band selection, the binary GSA and Binary PSOGSA (hybrid PSO and GSA) are utilized to perform compared band selection on two famous hyperspectral remote sensing image, i.e. the "Indian Pines" and "Pavia University". Both of the HSIs can be obtained from [35].

4.1. Data Description

4.1.1 Indian Pines. The Indian Pines is built by the Airborne Visible Infrared Imaging Spectrometer (AVIRIS) sensor in North-western Indiana. The AVIRIS has 224 bands with wavelength range from 400nm to 500nm. Due to the fact that the values of 4 spectral bands of the AVIRIS are 0 and 20 spectral bands of the sensor are easily affected by the water absorption band, these 24 spectral bands have been removed. Accordingly, the tested Indian Pines image in this paper contains only 200 bands. The pseudo-colour image composed by bands 27 (R), 50 (G) and 127 (B) with 145×145 pixels is shown in Fig. 4 (a). The corresponding ground truth reference image that contains 16 different classes is shown in Fig. 4(b). As Fig. 4(b) illustrates the fact that Indian Pines dataset is very complex and not all of the pixels belong to the 16 classes, many pixels not related to any class were divided into the background with dark blue colour. The number of samples utilized in this paper is given in Table 1.



(a) (b)

Fig. 4 *Indian Pines scene. (**a**) Original image (**b**) sample image of Indian Pines*

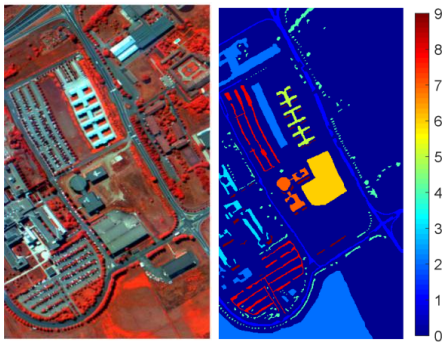
4.1.2 Pavia University. The Pavia University dataset is acquired by the Reflective Optics System Imaging Spectrometer (ROSIS) sensor during a flight campaign over Pavia, northern Italy. Pavia University scene is 610*340 pixels with a number of spectral bands 103. The geometric resolution is 1.3 meters. The ground truths differentiate 9 classes. The pseudo-colour image composed by bands 97 (R), 28 (G) and 5 (B) is shown in Fig. 5(a). The corresponding ground truth reference image that contains 9 different classes is shown in Fig. 5(b). As Fig. 5(b) illustrated, due to the fact that Pavia University dataset is very complex and not all of the pixels belong to the 9 classes, many pixels not related to any class were divided into the background with dark blue colour. The number of samples utilized in this paper is given in Table 2.

Table 1 Samples of Indian Pines.

Number	Class	GT	Trainning	Validation	Test
1	Alfalfa	54	8	7	39
2	Corn-notill	1434	25	25	1384
3	Corn-mintill	834	25	25	784
4	Corn	234	25	25	184
5	Grass-pasture	497	25	25	447
6	Grass-trees	747	25	25	697
7	Grass-pasture-mowed	26	8	7	11
8	Hay-windrowed	489	25	25	439
9	Oats	20	8	7	5
10	Soybean-notill	968	25	25	918
11	Soybean-mintill	2468	25	25	2418
12	Soybean-clean	614	25	25	564
13	Wheat	212	25	25	162
14	Woods	1294	25	25	1244
15	Buildings-Grass-Trees-Drives	380	25	25	330
16	Stone-Steel-Towers	95	25	25	45

Table 2 Samples of Pavia University.

Number	Class	GT	Training	Validation	Test
1	Asphalt	6631	331	331	5969
2	Meadows	18649	932	932	16785
3	Gravel	2099	104	104	1891
4	Trees	3064	153	153	2758
5	Painted metal sheets	1345	67	67	1211
6	Bare Soil	5029	251	251	4527
7	Bitumen	1330	66	66	1198
8	Self-Blocking Bricks	3682	184	184	3314
9	Shadows	947	47	47	853



(a) (b)

Fig. 5 *Pavia University scene.* ***(a)*** *Original image* ***(b)*** *sample image*

4.2. Comparison results

4.2.1 Parameter Settings: To perform fair experiments, all the basic GSA, PSOGSA, and CGSA based band selection methods utilize the same objective function shown in Eq. (15). In addition, the initial gravitational constant G_0, the decrease coefficient α, the population size (*NP*), and the maximum number of iterations (T_{max}) of the basic GSA, PSOGSA, and CGSA were set to 20, 100, 10, and 10, respectively. Moreover, to decrease the influence of randomicity, all the three compared algorithms perform 30 independent runs on each of the datasets.

4.2.2 Experimental results and analysis: The performance of GSA, PSOGSA, and CGSA are compared based on five measures including CPU processing time for selecting optimal subset (ST_{CPU}), the number of bands in the optimal subset (N_{sel}), CPU processing time for image classification based on the optimal subset (CT_{CPU}), the overall classification accuracy (*OA*) and the Kappa Coefficient (*Kappa*). For the two tested public datasets, the average values of the five measures and the corresponding error bar figures produced by the three compared algorithms are reported in Table 3 and Figs. 5-6. Moreover, the CPU processing time for image classification based on the optimal subset (CT_{CPU}), the overall classification accuracy (*OA*) and the Kappa Coefficient (Kappa) of SVM classifier using all the hyperspectral bands are also reported in Table 3. The best results in each row are bolded.

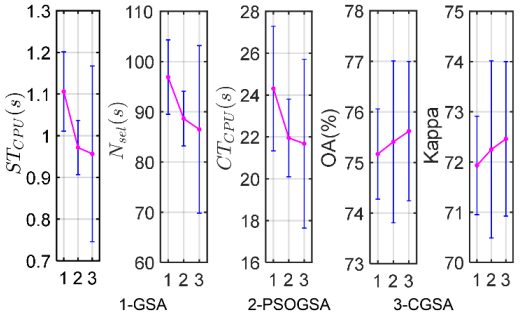


Fig. 6 *Statistical analysis of the 5 measures using error bar in Indian Pines dataset.*

From Table 3, we can conclude that all of the three GSA variants based band selection methods can effectively reduce the dimension and improve the classification accuracy of the HSIs on both the Indian Pines and Pavia University datasets. For example, for the Indian Pines image,

Table 3 The results of hyperspectral band selection.

Dataset	method	ST_{CPU}(s)	N_{sel}	CT_{CPU}(s)	OA(%)	Kappa
Indian Pines	all bands	--	--	45.961	73.392	70.022
	GSA	1.106	97	24.308	75.167	71.934
	PSOGSA	0.972	89	21.948	75.408	72.251
	CGSA	**0.957**	**87**	**21.678**	**75.620**	**72.461**
Pavia University	all bands	--	--	333.88	92.703	90.145
	GSA	26.116	57	195.264	92.705	90.154
	PSOGSA	18.717	56	194.916	92.729	90.190
	CGSA	**19.663**	**54**	**192.644**	**92.744**	**90.205**

the overall classification accuracy (*OA*) has been increased from 73.392% to 75.167%, 75.408%, and 75.620% whilst the size of the optimal band subset has been reduced from 200 to 97, 89, and 87 after the band selection operation based on the basic GSA, PSOGSA, and CGSA, respectively. Moreover, because of the fact that the size of the bands were largely reduced, these GSA based methods have considerably reduced the CPU times for image classification (CT_{CPU}). In addition, compared to the basic GSA and PSOGSA based methods, the CGSA based method produced the highest overall classification accuracy and obtained a band subset with the least bands. The mean values of each measure shown in Fig. 6 and Fig. 7 also confirmed the superiority of the proposed CGSA. This may come from the utilization the ***Gb*** guided crossover operation which can promote the exploitation ability of the basic GSA.

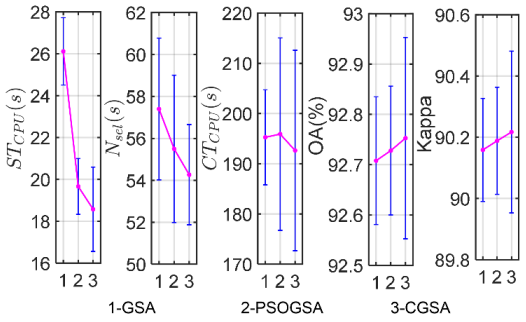


Fig. 7 *Statistical analysis of the 5 measures using error bar in Pavia University dataset.*

5. Conclusion

In this paper, a crossover based GSA (CGSA) is developed to construct a novel band selection method for HSIs. In the proposed CGSA, the global best experience of the whole population is maintained and utilized to guide the evolution of the CGSA and thereby promote the exploitation ability of the basic GSA. When extending CGSA for band selection, the optimization of band subset is performed based on an objective function constructed based on the overall classification accuracy of the SVM classifier and the size of the band subset. While the generation and optimization of the band subset mainly rely on the utilization of a V-shaped transfer function based CGSA. At last, the particle with the best fitness value is regarded as the optimal band subset. We conducted experiments with the Indian Pines and Pavia University datasets and the obtained band selection results were compared with that of the basic GSA and PSOGSA. The experimental results confirmed that all of the three GSA variants based band selection methods can efficiently identify the most informative spectral band subset with high classification accuracy and considerably reduce the band dimensionality of HSIs as well. Moreover, the CGSA based method displays obvious superiority compared to the basic GSA and PSOGSA based methods.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (41471353), the Natural Science Foundation of Shandong Province (ZR201709180096, ZR201702100118), the Fundamental Research Funds for the Central Universities (18CX05030A, 18CX02179A), and the Postdoctoral Application and Research Projects of Qingdao (BY20170204).

References

[1] Su, H., Du, Q., Chen, G., *et al.*: 'Optimized Hyperspectral Band Selection Using Particle Swarm Optimizatio', IEEE J. Sel. Top. Appl., 2014, 7(6), pp. 2659-2670.

[2] Hughes, G.: 'On the mean accuracy of statistical pattern recognizers', IEEE T. Inform. Theory, 1968, 14(1), pp. 55-63.

[3] Plaza, A., Martinez, P., Plaza, J., *et al.*: 'Dimensionality reduction and classification of hyperspectral image data using sequences of extended morphological transformations', IEEE T. Geosci. Remote, 2005, 43(3), pp. 466-479.

[4] Zhang, A.Z., Sun G.Y., Wang Z.J.: 'Optimized hyperspectral band selection using hybrid genetic algorithm and gravitational search algorithm'. Proc. Ninth Int. Conf. Multispectral Image Processing and Pattern Recognition, Enshi, China, November, 2015, pp. 981403-1-981403-6.

[5] Zabalza, J., Ren, J., Zheng, J., *et al.*: 'Novel segmented stacked autoencoder for effective dimensionality reduction and feature extraction in hyperspectral imaging', Neurocomputing, 2016, 214(C), pp.1062.

[6] Ren, J., Zabalza, J., Marshall, S., *et al.*: 'Effective feature extraction and data reduction with hyperspectral imaging in remote sensing', IEEE Signal Proc. Mag., 2014, 31(4), pp.149-154.

[7] Keshava, N., Mustard, J.F.: 'Spectral unmixing', IEEE signal Proc. Mag., 2002, 19(1), pp. 44-57.

[8] Green, A.A, Berman, M., Switzer, P., *et al.*: 'A transformation for ordering multispectral data in terms of image quality with implications for noise removal', IEEE T. Geosci. Remote, 1988, 26(1), pp. 65-74.

[9] Wang, J., Chang, C.I.: 'Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis', IEEE T. Geosci. Remote, 2006, 44(6), pp. 1586-1600.

[10] Bruce, L.M., Koger, C.H., Li, J.: 'Dimensionality reduction of hyperspectral data using discrete wavelet transform feature extraction', IEEE T. Geosci. Remote, 2002, 40(10), pp. 2331-2338.

[11] Nakamura, R.Y.M., Fonseca, L.M.G., Santos, J.A.D., *et al.*: 'Nature-Inspired Framework for Hyperspectral Band Selection', IEEE T. Geosci. Remote, 2014, 52(4), pp. 2126-2137.

[12] Keshava, N.: 'Distance metrics and band selection in hyperspectral processing with applications to material identification and spectral libraries', IEEE T. Geosci. Remote, 2004, 42(7), pp. 1552-1565.

[13] Vafaie, H., De Jong, K.: 'Genetic algorithms as a tool for feature selection in machine learning'. Proc. Fourth Int. Conf. Tools with Artificial Intelligence, Arlington, VA, USA, November 1992, pp. 200-203.

[14] Firpi, H.A., Goodman, E.: 'Swarmed feature selection'. Proc. Int. Conf. Information Theory, Washington DC, USA, October, 2004, pp. 112-118.

[15] Al-Ani, A.: 'Feature subset selection using ant colony optimization'. Int. J. Comput. Int., 2005, 2(1), pp. 53-58.
[16] Rashedi, E., Nezamabadi-Pour H., Saryazdi S.: 'GSA: a gravitational search algorithm', Inform. Sciences, 2009,179(13), pp. 2232-2248.
[17] Jiang, S., Wang, Y., Ji, Z.: 'Convergence analysis and performance of an improved gravitational search algorithm', Appl. Soft Comput., 2014, (24), pp. 363-384.
[18] Sun, G., Ma, P., Ren, J., *et al*.: 'A stability constrained adaptive alpha for gravitational search algorithm', Knowledge-Based Systems, 2018, 139, pp. 200-213.
[19] Zhang, A., Sun, G., Ren, J., *et al*.: 'A Dynamic Neighborhood Learning-Based Gravitational Search Algorithm', IEEE T. Cybernetics, 2018, 48(1), pp. 436 -447.
[20] Kumar, J.V., Kumar, D.M.V., Edukondalu, K.: 'Strategic bidding using fuzzy adaptive gravitational search algorithm in a pool based electricity market', Appl. Soft Comput., 2013, 13(5), pp. 2445-2455.
[21] Mirjalili, S.A., Hashim, S.Z.M., Sardroudi, H.M.: 'Training feedforward neural networks using hybrid particle swarm optimization and gravitational search algorithm', Appl. Math. Comput., 2012, 218(22), pp. 11125-11137.
[22] Zhang, N., Li, C., Li, R., *et al*.: 'A mixed-strategy based gravitational search algorithm for parameter identification of hydraulic turbine governing system', Knowl.-Based Syst., 2016, 109, pp. 218-237.
[23] Kumar, V., Chhabra, J.K., Kumar, D.: 'Automatic cluster evolution using gravitational search algorithm and its application on image segmentation', Eng. Appl. Artif. Intel., 2014, 29(3), pp. 93-103.
[24] Razavi, S.F., Sajedi, H.: 'Cognitive discrete gravitational search algorithm for solving 0-1 knapsack problem', J. Intell. Fuzzy Syst., 2015, 29(5), pp. 2247-2258.
[25] Sun, G., Zhang, A., Yao, Y., et al.: 'A novel hybrid algorithm of gravitational search algorithm with genetic algorithm for multi-level thresholding', Appl. Soft Comput., 2016, 46, pp. 703-730.
[26] Rashedi, E., Nezamabadi-Pour, H.: 'Feature subset selection using improved binary gravitational search algorithm', J. Intell Fuzzy Syst., 2014, 26(3), pp.1211-1221.
[27] Papa, J.P., Pagnin, A., Schellini, S.A., *et al*.: 'Feature selection through gravitational search algorithm'. Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing, Prague, Czech Republic, May 2011, pp. 2052-2055.
[28] Behjat, A.R., Mustapha, A., Nezamabadi-Pour, H., *et al*.: 'Feature subset selection using binary gravitational search algorithm for intrusion detection system'. Proc. Asian Conf. Intelligent Information and Database Systems, Kuala Lumpur, Malaysia, March 2013, pp. 377-386.
[29] Kumar, V., Chhabra, J.K., Kumar, D.: 'Automatic Unsupervised Feature Selection using Gravitational Search Algorithm', IETE J. Res., 2015, 61(1), pp. 22-31.
[30] Wang, M., Wan, Y., Ye, Z., *et al*.: 'A band selection method for airborne hyperspectral image based on chaotic binary coded gravitational search algorithm', Neurocomputing, 2018, (273), pp. 57-67.
[31] Mirjalili, S., Lewis, A.: 'Adaptive gbest-guided gravitational search algorithm', Neural Comput. Appl., 2014, 25(7-8), pp. 1569-1584.
[32] Yin, B., Guo, Z., Liang, Z., *et al*.: 'Improved gravitational search algorithm with crossover', Comput. Electr. Eng., 2017, pp. 1-12.
[33] Rashedi, E., Nezamabadi-Pour, H., Saryazdi, S.: 'BGSA: binary gravitational search agorithm', Nat. Comput., 2010, 9(3), pp. 727-745.
[34] Mirjalili, S., Lewis, A.: 'S-shaped versus V-shaped transfer functions for binary Particle Swarm Optimization', Swarm Evol. Comput., 2013, 9, pp. 1-14.
[35] 'HSIs image datasets', http://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes, accessed March 2018.