Research Article

A New Feature Selection Method for Hyperspectral Image Classification Based on Simulated Annealing Genetic Algorithm and Choquet Fuzzy Integral

Hongmin Gao, Lizhong Xu, Chenming Li, Aiye Shi, Fengchen Huang, and Zhenli Ma

College of Computer and Information Engineering, Hohai University, Nanjing 211100, China

Correspondence should be addressed to Lizhong Xu; lzhxu@hhu.edu.cn

Received 1 June 2013; Revised 14 September 2013; Accepted 15 September 2013

Academic Editor: Gianluca Ranzi

Copyright © 2013 Hongmin Gao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Hyperspectral remote sensing technology is a rapidly developing new integrated technology that is widely used in numerous areas. Rich spectral information from hyperspectral images can aid in the classification and recognition of the ground objects. However, the high dimensions of hyperspectral images cause redundancy in information. Hence, the high dimensions of hyperspectral data must be reduced. This paper proposes a hybrid feature selection strategy based on the simulated annealing genetic algorithm (SAGA) and the Choquet fuzzy integral (CFI). The band selection method is proposed from subspace decomposition, which combines the simulated annealing algorithm with the genetic algorithm in choosing different cross-over and mutation probabilities, as well as mutation individuals. Then, the selecting bands are further refined by CFI. Experimental results show that the proposed method can achieve higher classification accuracy than traditional methods.

1. Introduction

Hyperspectral remote sensors peculiarly provide measurements of the Earth's surface with very high spectral resolution, usually resulting in tens of channels. Unlike multispectral sensors, the high spectral resolution renders hyperspectral remote sensors very powerful in applications requiring the identification of subtle differences in ground covers (e.g., material quantification and target detection). On the other hand, the large-dimensional data spaces generated by these sensors introduce challenging methodological problems. In the context of supervised classification, the most important methodological issue raised by these sensors is the so-called curse of dimensionality (also known as the Hughes effect) that occurs when the numbers of features and of available training samples are unbalanced [1].

Meanwhile, hyperspectral remote sensing images have nonlinear properties. These nonlinear properties originate from the multiscattering between photons and ground targets, within pixel spectral mixing, and from scene heterogeneity. In addition, given that the pixel size in most remote sensing systems is sufficiently large to include different types

of land cover, classification error arises and produces unreliable classification results. In this case, traditional classifiers may fail completely.

In remote sensing literature, numerous methods have been developed to solve the hyperspectral data classification problem. A successful approach to hyperspectral data classification is based on the support vector machine (SVM). SVM determines two classes by identifying the optimal separating hyperplane that maximizes the margin between the closest training sample and the separating hyperplane. Data samples located at the hyperplane border are referred to as support vectors and are used to create a decision surface. The properties of SVM for both full-dimensional and reduced-dimensional data have been investigated, while multi-class SVM strategies have been considered in [2]. Hyperspectral image classification using different kernel-based approaches has been analyzed and compared, and SVM has been found to be more useful than other kernel-based methods in [3]. SVM classification performance is compared with other well-known neural approaches in [4], which exhibited that SVM provides simplicity, robustness, and increased classification accuracy compared with neural

networks. In addition, some improved SVM methods have also been successfully used in hyperspectral image classification. The proposed method, called contextual SVM using Hilbert space embedding showed significant improvement over other methods on several hyperspectral images in [5]. A semisupervised method for addressing a domain adaptation problem based on multiple-kernel SVMs in the classification of hyperspectral data was presented in [6]. Thus, SVM is very suitable for hyperspectral image classification. However, dimension reduction is not sufficiently considered in SVM.

Commonly used dimension reduction methods fall into two categories, namely, feature selection and feature extraction. Since every band of hyperspectral data has its own corresponding image, the feature extraction approach maps a high-dimensional feature space to low-dimensional space via linear or nonlinear transformation. However, the original physical interpretation of the image cannot be retained. Thus, feature extraction approaches are unsuitable for the dimension reduction of hyperspectral images. Given that the spectral distance between adjacent bands in the hyperspectral data is only 10 nm and because the correlation between them is extremely high [7], a considerable redundancy is observed, which should be largely reduced by the feature selection or band selection methods to improve classification efficiency and accuracy. A semisupervised feature-selection technique for hyperspectral image classification was developed in [8]. A method for unsupervised band selection by transforming the hyperspectral data into complex networks was presented in [9]. Therefore, a new dimension reduction method is proposed that combines the simulated annealing genetic algorithm (SAGA) with the Choquet fuzzy integral (CFI).

A population and temperature ladder-based new genetic algorithm (GA) or the so-called SAGA was recently proposed to examine a sample from a distribution defined on a space of finite binary sequence. The feature selection strategy of hyperspectral images based on GA and SVM was proposed in [10, 11]. A GA-based feature selection and local-Fisher's discriminant analysis-based feature projection are performed for effective dimensionality reduction in [12]. But SAGA method works by simulating a parallel population of samples with different temperatures. The population is updated via selection, mutation, cross-over, and exchange operations that are highly similar with GA. SAGA has the learning capability of GA, as well as the fast-mixing capability of parallel tempering (simulated tempering). In most cases, classification accuracy is only used as the fitness function, but internal relations between bands and classes have not been taken into account. Considering the above problem, a correction method based on CFI is proposed. The CFI does not assume the independence of one element from another and, based on any fuzzy measure, it is employed to perform the overall evaluation of an input pattern [13]. Moreover, the fuzzy measure defined on an attribute is used as the relative degree of importance of this attribute such that the connection weights can be interpreted as the fuzzy measure values or the degrees of importance of the respective input variables. The band selection method of this paper that is based on SAGA and CFI (SAGA-CFI) cannot only improve the accuracy of classification, but also effectively reduce the

uncertainty of the information in order to further improve the accuracy.

Since hundreds of bands in the hyperspectral imagery exist, the direct search space for SAGA and CFI on the original band space becomes extremely huge. An adaptive subspace decomposition (ASD) method for hyperspectral data dimensionality reduction was proposed in [14]. To avoid the impact of enormous data sets on traditional statistical classification techniques, the ASD scheme is used. Thus, the differences between global and local statistical characteristics have been fully considered, and the problem presented by a limited number of training samples is then alleviated.

In this paper, we use SAGA and CFI in every subspace to choose suitable bands based on ASD which differs from the previous work [5, 6, 8–12] in three aspects. First of all, ASD is employed to divide the bands into disjoint subspace rather than mutual information. Although mutual information may make better performance than ASD, it also cannot be chosen in this paper because mutual information is interconnected with entropy of information, and it can be directly formulated by entropy. It is better to keep independence between ASD and CFI. Furthermore, based on GA, SAGA is used in band selection which includes a schedule of temperatures and approaches the global minimum when the temperatures change gradually. Last but not least, CFI is first employed to further optimize the band selection method. Thus, we reduce the search space and computational complexity, while avoiding the selection of an excessive number of adjacent bands.

The remainder of this paper is organized as follows. Section 2 introduces subspace decomposition. Section 3 presents the proposed SAGA. In Section 4, a brief description of three related elements and fuzzy measure followed by CFI is given. Section 5 provides the SVM classification adopted in this paper. Section 6 describes the proposed method. Experiments and analysis are demonstrated in Section 7. Finally, Section 8 concludes the paper.

2. Subspace Decomposition

The main characteristics of hyperspectral remote sensing data are a large quantity of imaging channels (approximately 220 bands) and a narrow band spectrum. The spectrum of hyperspectral data is highly concentrated, rendering overall and local characteristics quite different. We may lose some important local characteristics if we select the bands from the total space. In terms of the overall situation, the bands are notably characterized by groups. We can divide all bands into several groups as long as a lower correlation exists between adjacent bands. Subspace decomposition not only reduces the dimension of the images, but also significantly improves the efficiency of data processing. Division of data sources based on ASD and fusion classification based on consensus theory is proposed in [15]. So the commonly used method continues to be ASD. According to the correlation matrix of hyperspectral images between bands, the full data space with dimensionality is adaptively decomposed into numerous subspaces with different dimensionalities. In each subspace, the bands have very strong correlation, while the energy is

more concentrated. Hence, full data dimensionality can be logically reduced.

Since different bands have different correlations, all subspaces do not have the same dimensionality. Therefore, the goal is to match the features of each subspace with one or few classes. For this purpose, the new method primarily depends on the correlation matrix R between different bands. The element of the correlation matrix R is defined as

$$R_{ij} = \frac{E[(x_i - \mu_i)(x_j - \mu_j)]}{\sqrt{E(x_i - \mu_i)^2} \sqrt{E(x_j - \mu_j)^2}}. \quad (1)$$

The value R_{ij} of the matrix R ranges between 0 and 1. The closer R_{ij} is to 1, the more correlation exists between the two bands. μ_i and μ_j are the mean values of x_i and x_j , respectively. $E[\bullet]$ is the value of the mathematical expectation.

3. Simulated Annealing Genetic Algorithm

Traditional selection, cross-over, and mutation operator, as well as the selection of fitness proportion in GA, allow the superior chromosome to maintain its predominance or strengthen it in the subsequent generations. The convergent chromosome may not be the overall optimal chromosome. SAGA combines the simulated annealing algorithm with GA. Thus, SAGA can perform the temperature-control function in the simulated annealing algorithm by controlling selection probability [16]. If we want to sample from a distribution defined on a space of finite binary sequence, we employ the following:

$$\pi(X) \propto \exp\left[-\frac{H(X)}{\tau}\right], \quad (2)$$

where X is the d -dimensional binary vector $X = \{\beta_1, \beta_2, \dots, \beta_d\}$ with $\beta_i \in \{0, 1\}$, τ is the scale parameter (a so-called temperature that can be any value of interest), and $H(X)$ is the fitness function in terms of GA.

First, a sequence of distributions $\pi_1(x_1), \pi_2(x_2), \dots, \pi_N(x_N)$ is constructed as follows:

$$\pi_i(x_i) = \frac{1}{Z(t_i)} \exp\left[-\frac{H(x_i)}{t_i}\right], \quad (3)$$

where, for $i = 1, 2, \dots, N$, $Z(t_i) = \sum_{x_i} \exp[-H(x_i)/t_i]$. The temperatures $T = (t_1, t_2, \dots, t_N)$ form a ladder with the order $t_1 > t_2 > \dots > t_{N-1} > t_N$. For convenience, we denote the ladder by $T = (t_1, t_2, \dots, t_N)$. Note that we always set $t_N = \tau$ as $\pi_N(x) = \pi(x)$ to correspond to the target distribution from which we obtain the sample. $X = \{x_1, x_2, x_3, \dots, x_N\}$ denotes a population of samples where $x_i = \{\beta_1^i, \beta_2^i, \dots, \beta_d^i\}$ is a sample from $\pi_i(x)$ and is called a chromosome or an individual in terms of GA, while N represents the population size. In SAGA, the Boltzmann distribution of the population is expressed as

$$\pi(X) = \prod_{i=1}^N \pi_i(x_i) = \frac{1}{Z(T)} \cdot \exp\left[-\sum_{i=1}^N \frac{H(x_i)}{t_i}\right], \quad (4)$$

where $Z(T) = \prod_{i=1}^N Z_i(t_i)$. The population is updated by selection, cross-over, mutation, and exchange operators.

3.1. Selection. The probability of having the chromosome chosen first is

$$p(x_i) = \frac{\exp[-H(x_i)/t]}{Z(X)}, \quad (5)$$

and probability of (x_i, x_j) is

$$p[(x_i, x_j) | X] = \frac{\{\exp[-H(x_i)/t] + \exp[-H(x_j)/t]\}}{[(N-1)Z(X)]}, \quad (6)$$

where $Z(X) = \sum_{i=1}^N \exp[-H(x_i)/t]$.

3.2. Cross-Over. One chromosome pair, such as x_i and x_j ($i \neq j$), is selected from the current population X through the roulette wheel. Two offspring, y_i and y_j , are generated according to a specific cross-over operator. A new population Y is proposed as $Y = \{x_1, x_2, x_3, \dots, y_i, \dots, y_j, \dots, x_N\}$ and is accepted with probability $\min(1, p_c)$ according to the Metropolis-Hastings rule that is expressed as follows:

$$p_c = \exp\left\{-\frac{[H(y_i) - H(x_i)]}{t_i} - \frac{[H(y_j) - H(x_j)]}{t_j}\right\} \times \frac{P[(y_i, y_j) | Y]}{P[(x_i, x_j) | X]}, \quad (7)$$

where $P[(y_i, y_j) | Y]$ denotes the selection probability of (y_i, y_j) from the population Y and $P[(x_i, x_j) | X]$ denotes the selection probability of (x_i, x_j) from the population X .

3.3. Mutation. We define the mutation operator as an additional move of the Metropolis-Hastings rule. One chromosome, such as x_k , is uniformly chosen from the current population X . A new chromosome is generated by the addition of a random vector e_k , such that

$$y_k = x_k + e_k, \quad (8)$$

where e_k is usually chosen to achieve moderate acceptance probability for the mutation operation. The new population $Y = \{x_1, x_2, x_3, \dots, y_k, \dots, x_N\}$ is accepted with the probability $\min(1, p_m)$ according to the following Metropolis-Hastings rule:

$$p_m = \exp\left\{-\frac{[H(y_k) - H(x_k)]}{t_k}\right\} \frac{T(X | Y)}{T(Y | X)}. \quad (9)$$

3.4. Exchange. A straightforward implementation of relative parallel tempering can outperform simulated annealing in several crucial respects, and parallel tempering can offer a powerful alternative to simulated annealing for combinatorial optimization problems [17]. Given the current population

X and the attached temperature ladder T in $\{X, T\} = \{x_1, t_1, x_2, t_2, \dots, x_m, t_m, x_{m+1}, t_{m+1}, \dots, x_N, t_N\}$, we propose to obtain a new population Y by making an exchange between x_m and x_{m+1} without changing the T . That is, $\{Y, T\} = \{x_1, t_1, x_2, t_2, \dots, x_{m+1}, t_m, x_m, t_{m+1}, \dots, x_N, t_N\}$. The new population is then accepted with probability $\min(1, p_e)$ according to the Metropolis-Hastings rule below:

$$p_e = \exp \left\{ [H(x_m) - H(x_{m+1})] \left(\frac{1}{t_m} - \frac{1}{t_{m+1}} \right) \right\}. \quad (10)$$

3.5. Fitness Function. In addition, another key of SAGA is the design of the fitness function. We use only the classification accuracy obtained from the training feature subset as the fitness function. The purpose of the iterative repetition is to determine the optimal feature subset and to maximize classification accuracy. The adopted classifier is SVM, which is described in Section 5.

4. Choquet Fuzzy Integral

Based on subspace decomposition, CFI method is used to further refine the selecting bands. The definition of fuzzy measure and Choquet integral are shown in [18, 19].

Definition 1 (fuzzy measure (see [18])). Denote the Borel set as B which is obtained from the domain $X = \{x_1, x_2, \dots, x_n\}$, and then define a fuzzy measure $g : B \rightarrow [0, 1]$ on B , it must satisfy the following conditions:

- (1) $g(X) = 1$; $g(\phi) = 0$, ϕ is null set;
- (2) given two subsets $A, B \subset B$, if $A \subset B$, so $g(A) \leq g(B)$;
- (3) If $A_1 \subset A_2 \subset \dots \subset A_n \subset \dots$, then $\lim_{i \rightarrow \infty} (g(A_i)) = g(\lim_{i \rightarrow \infty} A_i)$.

According to the definition of fuzzy measure, Sugeno introduces the g -fuzzy measure.

Definition 2 (g -fuzzy measure (see [19])). For all the sets $A, B \subset X$, $A \cap B = \phi$, there exists $\lambda > -1$ to satisfy

$$g(A \cup B) = g(A) + g(B) + \lambda g(A)g(B). \quad (11)$$

Obviously, when $\lambda = 0$, g -fuzzy measure is the probability measure.

Given a finite set $X = \{x_1, x_2, \dots, x_n\}$ and orders $g^i = g(\{x_i\})$, the mapping $g : x_i \rightarrow g^i$ as fuzzy density function, also as single-point importance. If $A = \{x_{i1}, x_{i2}, \dots, x_{im}\} \subseteq X$, according to (20), the following formula can be deduced as

$$g(A) = \frac{\prod_{x_i \in A} (1 + \lambda g^i) - 1}{\lambda}, \quad (12)$$

where $\lambda > -1, \lambda \neq 0$. Because $g(X) = 1$, the value of λ can be achieved by solving $\lambda + 1 = \prod_{i=1}^n (1 + \lambda g^i)$. It can be proved that given the fixed set $\{g^i\}$, $0 < g^i < 1$, there exists one and only one $\lambda \in (-1, \infty)$ and $\lambda \neq 0$. So, if the fuzzy density g^i ($i = 1, 2, \dots, n$) is given, it can get the unique g -fuzzy measure.

With regards to the theory on information fusion, fuzzy density g^i server as the importance or the contribution of the source x_i . The group of source A can determine a unique g -fuzzy measure in the process of data fusion. Based on the g -fuzzy measure, Choquet proposed a fuzzy integral method.

Definition 3 (Choquet integral (see [19])). Given a function $h : X \rightarrow [0, 1]$, and its Choquet integral on fuzzy measure g is defined as

$$\begin{aligned} C_\mu(h(x_1), h(x_2), \dots, h(x_n)) \\ = \sum_{i=1}^n (h(x_i) - h(x_{i-1})) g(A_i). \end{aligned} \quad (13)$$

In the equation, the value of the function $h(x_i)$ can be interpreted as a credibility estimation of the source x_i for specific target. Note that the function $h(x_i)$ is increasing, $0 = h(x_0) \leq h(x_1) \leq h(x_2) \leq \dots \leq h(x_n) \leq 1$; fuzzy measure g is the importance or contribution of information source with respect to the ultimate decision-making or estimation, $A_i = \{x_i, x_{i+1}, \dots, x_n\}$.

According to (13), CFI can be seen as the weighted sum of $h(x_1), h(x_2), \dots, h(x_n)$, and the weights depend on of the rank of $\{x_i\}$, and the value of $h\{x_i\}$ decides the rank of $\{x_i\}$; so the CFI is a nonlinear function of function h . It is clear that when $\lambda = 0$, the g -fuzzy measure is the probability measure, and the CFI is a linear function of h . The CFI is used in data fusion if we regard the $h(x_i) \in [0, 1]$ as a result of goal judgment and g —as the degree of importance or contribution. Obviously the CFI is the nonlinear combination of the result of information source with the importance of information source.

Before computing the fuzzy integral we must compute the value of λ . From (12), we know that the solution to λ of the fuzzy integral is the root of high-order polynomial. If there are many sources, there is computation burden to get the value of parameter λ , blocking the online and real-time of algorithm.

5. SVM Classification Methods

Training data are required to train the SVM model. However, these data cannot be separated without errors. The data points that are closest to the hyperplane are used to measure the margin, while SVM attempts to identify the hyperplane that maximizes the margin and minimizes a quantitative proportion to the number of misclassification errors [20, 21]. SVM derives the optimal hyperplane as the solution of the following convex quadratic programming problem [22]:

$$\begin{aligned} \min_{w, b, \xi} \quad & \frac{1}{2} w^T w + C \sum_{i=1}^n \xi_i, \\ \text{s.t.} \quad & y_i (w^T \phi(x_i) + b) \geq 1 - \xi_i, \\ & \xi_i \geq 0, \quad i = 1, 2, \dots, n, \end{aligned} \quad (14)$$

where $\{(x_1, y_1), \dots, (x_i, y_i)\}$ are the labeled training datasets, $x_i \in R^d$ and $y_i \in \{-1, 1\}$; w^* , and b^* defines a linear classifier

in the feature space; C is the regularization parameter defined by the user; and ξ_i is a positive slack variable that handles permitted errors.

The aforementioned optimization problem can be reformulated through a Lagrange function, where Lagrange multipliers can be found via dual optimization to generate a convex quadratic programming solution as follows [23–25]:

$$\begin{aligned} \max \quad Q(\alpha) &= \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j K(x_i, x_j), \\ \text{s.t.} \quad \sum_{i=1}^n \alpha_i y_i &= 0, \quad 0 \leq \alpha_i \leq C, \quad i = 1, 2, \dots, n, \end{aligned} \quad (15)$$

where $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_n]$ is the vector of Lagrange multipliers, while $K(\bullet, \bullet)$ is a kernel function which is introduced as follows [26]:

$$K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j). \quad (16)$$

Thus, the final result is a discrimination function $F(x)$ conveniently expressed as a function of the data in the original (lower) dimensional feature space [27]:

$$\begin{aligned} F(x) &= \text{sgn} \left[(w^*)^T \phi(x) + b^* \right] \\ &= \text{sgn} \left(\sum_{i=1}^n \alpha_i^* y_i K(x_i, x) + b^* \right). \end{aligned} \quad (17)$$

6. Proposed Method

6.1. Adaptive Subspace Decomposition. In the beginning, adaptive subspace decomposition is used to divide into seven subspace according to (1). All R_{ij} values are identified, and then the proper threshold T_b is set. The continuous bands of $|R_{ij}| \geq T_b$ in the same subspace are subsequently placed. We can dynamically control the number of subspaces and the number of bands in each subspace by changing the threshold T_b .

6.2. The Band Order Method in Subspace. SAGA is used in order to find out the optimal bands in each subspace. Here we choose common binary coding method as the genetic coding mode, and the iteration times of SAGA is 50. Generally, a subspace has many bands, and all the suitable bands should be chosen. Meanwhile, if the subspace has only one band, it must be chosen.

6.3. The Band Reorder Method in Subspace. After the bands are chosen according to SAGA, they also can be further optimized based on the CFI method. CFI takes into account the factors of entropy of information, correlation coefficient, and standard distance between the means.

6.3.1. Entropy of Information and Variance. According to Shannon's information theory, entropy measures information content in terms of uncertainty. The entropy of the hyperspectral components represents the information content of

each component. Thus, the higher the entropy, the richer the information content, resulting in a more meaningful representation. The entropy or total information [28] is defined as

$$H(X) = \sum_x p_X(x) \ln p_X(x), \quad (18)$$

where $p_X(x)$ is the probability of pixel value x .

Variance represents deviation from mean value to the gray-scale value of pixel. The formulae of computing mean value μ_i and variance σ_i^2 are as follows [29]:

$$\mu_i = \frac{1}{M \times N} \sum_{x=1}^M \sum_{y=1}^N f_i(x, y), \quad (19)$$

$$\sigma_i^2 = \frac{1}{M \times N} \sum_{x=1}^M \sum_{y=1}^N [f_i(x, y) - \mu_i]^2, \quad (20)$$

where i and $i+1$ are the numbers of two adjacent bands. M and N represent the width and the height of image, and $f_i(x, y)$ is the gray-scale value of pixel (x, y) .

6.3.2. Correlation Coefficients. In statistics, the correlation coefficient denotes the accuracy of a least square fitting to the original data. It is a normalized measure of the strength of the linear relationship between two variables. Correlation is employed in many types of applications, such as hyperspectral image processing where it is used to measure and to quantitatively compare the similarity between bands [30]. The two-dimensional normalized correlation function for image processing is shown below:

$$\begin{aligned} \text{CC} &= \left(\sum_{x=1}^M \sum_{y=1}^N [f_i(x, y) - \mu_i] [f_{i+1}(x, y) - \mu_{i+1}] \right) \\ &\times \left(\left(\sum_{x=1}^M \sum_{y=1}^N [f_i(x, y) - \mu_i]^2 \right) \right. \\ &\times \left. \left(\sum_{x=1}^M \sum_{y=1}^N [f_{i+1}(x, y) - \mu_{i+1}]^2 \right) \right)^{-1/2}, \end{aligned} \quad (21)$$

where CC is a real number between -1 and 1 .

6.3.3. Standard Distance between the Means. Object classes need to be analyzed in depth in which the band is easy to be distinguished [31] that is, the statistical distance between object classes in the band. Standard distance between the means d is defined as

$$d = \frac{|\mu'_1 - \mu'_2|}{\sigma'_1 + \sigma'_2}, \quad (22)$$

where μ'_1 and μ'_2 are spectrum means of corresponding regions of the two samples. σ'_1 and σ'_2 are variances of corresponding regions of the two samples. d reflects separability of the two samples in each band.

Then, the procedure of the band reorder method using CFI is as follows.

FIGURE 1: Overall structure of the proposed method.

- (1) According to (18), entropies of information in each subspace are computed and recorded as u_1 .
- (2) According to (21), the correlation coefficients in each subspace are computed and recorded as u_2 .
- (3) According to (22), standard distances between the means in each subspace are computed and recorded as u_3 .
- (4) Belief function is constructed and domain is $U = \{u_1, u_2, u_3\}$. The relations between index value of each factor and band reorder are described below. (1) The bigger the entropy is, the rich the information is. (2) The smaller the correlation coefficient is, the

more independent the band is. (3) The bigger the standard distance between the means is, the easier to distinguish the two samples is. So the belief functions of CFI are listed as follows:

$$\begin{aligned}
 h(u_1) &= \frac{u_1 - u_{1\min}}{u_{1\max} - u_{1\min}}, \\
 h(u_2) &= \frac{u_{2\max} - u_2}{u_{2\max} - u_{2\min}}, \\
 h(u_3) &= \frac{u_3 - u_{3\min}}{u_{3\max} - u_{3\min}},
 \end{aligned} \tag{23}$$

TABLE 1: Correlation matrix R between 5 and 15.

Bands	5	6	7	8	9	10	11	12	13	14	15
5	1.0000	0.9793	0.9823	0.9813	0.9809	0.9800	0.9785	0.9769	0.9750	0.9710	0.9667
6	0.9793	1.0000	0.9889	0.9892	0.9886	0.9881	0.9869	0.9857	0.9835	0.9798	0.9755
7	0.9823	0.9889	1.0000	0.9920	0.9930	0.9926	0.9918	0.9911	0.9892	0.9856	0.9819
8	0.9813	0.9892	0.9920	1.0000	0.9938	0.9941	0.9935	0.9929	0.9911	0.9875	0.9836
9	0.9809	0.9886	0.9930	0.9938	1.0000	0.9950	0.9954	0.9949	0.9934	0.9897	0.9861
10	0.9800	0.9881	0.9926	0.9941	0.9950	1.0000	0.9959	0.9959	0.9944	0.9907	0.9872
11	0.9785	0.9869	0.9918	0.9935	0.9954	0.9959	1.0000	0.9966	0.9959	0.9926	0.9894
12	0.9769	0.9857	0.9911	0.9929	0.9949	0.9959	0.9966	1.0000	0.9968	0.9943	0.9915
13	0.9750	0.9835	0.9892	0.9911	0.9934	0.9944	0.9959	0.9968	1.0000	0.9965	0.9949
14	0.9710	0.9798	0.9856	0.9875	0.9897	0.9907	0.9926	0.9943	0.9965	1.0000	0.9977
15	0.9667	0.9755	0.9819	0.9836	0.9861	0.9872	0.9894	0.9915	0.9949	0.9977	1.0000

TABLE 2: Dimensions and bands of each subspace.

Subspace	1	2	3	4	5	6	7
Dimensions	1–15	16–35	36	37–38	39–76	77–97	98–220
Bands	15	20	1	2	38	21	123

TABLE 3: New dimensions and bands of each subspace.

Subspace	1	2	3	4	5	6	7	8	9	10	11
Dimensions	5–15	16–35	36	37–38	39–76	77	79	87–97	98–102	111–148	166–216
Bands	11	20	1	2	38	1	1	11	15	38	51

where $0 \leq h(u_1) \leq h(u_2) \leq \dots \leq h(u_m) \leq 1$. Equation (23) is reordered and a new equation (24) is generated:

$$\begin{aligned} h'(u_1) &= \min \{h(u_1), h(u_2), h(u_3)\}, \\ h'(u_2) &= \text{mid} \{h(u_1), h(u_2), h(u_3)\}, \\ h'(u_3) &= \max \{h(u_1), h(u_2), h(u_3)\}. \end{aligned} \quad (24)$$

$h'(u_1)$, $h'(u_2)$, and $h'(u_3)$ are minimum, median, and maximal values of the three, respectively.

- (5) Another important problem that needs to be fixed is fuzzy measure g . Belief function is arranged in ascending order, and the biggest one is of primary importance. In each subspace, $S = h'(u_1) + h'(u_2) + h'(u_3)$,

$$\begin{aligned} g(u_1) &= \frac{h'(u_1)}{S}, \\ g(u_2) &= \frac{h'(u_2)}{S}, \\ g(u_3) &= \frac{h'(u_3)}{S}. \end{aligned} \quad (25)$$

- (6) The formula of computing fuzzy integral value is as follows:

$$C = \sum_{i=1}^3 g(h_{\partial i}) (h'(u_i) - h'(u_{i-1})), \quad (26)$$

where $h_{\partial i} = \{u_i, u_{i+1}, \dots, u_n\}$, $h'(u_0) = 0$.

TABLE 4: Band number and index value in each subspace.

Subspace	Band No.	Index value	Band No.	Index value
1	11	0.9419		
2	17	0.9732	18	0.9706
	16	0.9604	19	0.9502
3	36	1		
4	37	0.9868		
	71	0.9667	69	0.9658
5	70	0.9652	73	0.9642
	67	0.9638	72	0.9636
	74	0.9636		
6	77	1		
7	79	1		
8	88	0.9813	89	0.9659
9	101	0.9608		
	119	0.9689	120	0.9689
	121	0.9687	122	0.9685
	118	0.9683	123	0.9681
10	124	0.9675	125	0.9671
	126	0.9669	127	0.9663
	128	0.9663	129	0.9660
	130	0.9659		
11	184	0.9659		

6.4. Flowchart of the Proposed Method. The flowchart of this paper is illustrated in Figure 1.

TABLE 5: Numbers of training samples and of test samples.

Classes	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7
Number of training samples	48	485	632	86	777	749	163
Number of test samples	150	1435	2192	235	2417	2469	516

TABLE 6: Error matrix of SAGA-SVM.

Class	Reference data							Row total	UA
	1	2	3	4	5	6	7		
Classified data									
1	137	0	35	3	51	44	0	270	0.5074
2	0	1328	42	4	41	0	9	1424	0.9326
3	1	0	2024	0	45	107	0	2177	0.9297
4	4	22	38	222	0	0	13	299	0.7425
5	5	28	0	0	2235	47	10	2325	0.9613
6	0	36	53	3	0	2271	8	2371	0.9578
7	3	21	0	3	45	0	476	548	0.8686
Column total	150	1435	2192	235	2417	2469	516	9414	
PA	0.9133	0.9254	0.9234	0.9446	0.9247	0.9198	0.9225		0.9249

FIGURE 2: AVIRIS false color image.

7. Experiments and Analysis

7.1. Hyperspectral Images. Experiments were conducted on a hyperspectral data set from the Northwest Indiana Indian Pine test site 3 (2×2 mile portion of Northwest Tippecanoe County, Indiana) on June 12, 1992. These data include 145 by 145 pixels and 220 bands. The false color image is shown in Figure 2, which is composed of band 89, band 5, and band 120.

7.2. Subspace Decomposition Experiment. The ASD scheme is used to obtain the correlation value between the bands. Table 1 gives the values of the parts of the correlation matrix R according to (1).

As presented in Table 1, the autocorrelation coefficient of each band is equal to 1, and the correlation value is very

high. In this paper, the ASD method is performed using the correlation criterion of a given threshold, which is 0.8. The full data space is decomposed into seven subspaces. The dimensions of each subspace are shown in Table 2.

From the 220 spectral channels acquired by the AVIRIS sensor, 41 bands were discarded because they were affected by atmospheric problems. The discarded bands were as follows: 1–4, 78, 80–86, 103–110, 149–165, and 217–220. As a result, the new dimensions of each subspace are shown in Table 3.

7.3. SAGA in Each Subspace. The hyperspectral image is categorized into seven classes according to the real data on the ground. The ratio of training and test samples is 1:3 because SVM is suitable for small samples. SAGA is used in each subspace, while fitness is computed and illustrated in Figure 3. We select the most optimum band in each subspace. SAGA in subspace numbers. 3, 6, and 7 is unnecessary because each of these subspaces contains only one band. The kernel function used is a radial basis function, while the two SVM parameters (i.e., C and γ) are selected based on fivefold cross-validation during the training phase. The search range for C is in $[2^{-3}, 2^{10}]$ and $[2^{-8}, 2^2]$ for γ .

7.4. Index Value of CFI. Entropy, correlation coefficient, and standard distance between the means of each band are computed. The index values of CFI are then obtained and sorted in descending order in each subspace. The bigger the index value is, the more important the band is. T_c is a given threshold of index value. Table 4 shows the index values of the bands when threshold T_c is 0.940.

SAGA is used to determine which band/bands shall be selected in each subspace, but it cannot indicate which bands have higher priority than others. The index values of CFI are then further refined the selecting bands, and the more effective optimizations come into being.

TABLE 7: Error matrix of CFI-SVM ($T_c = 0.94$).

Class	Reference data							Row total	UA
	1	2	3	4	5	6	7		
Classified data									
1	134	0	69	5	66	70	0	344	0.3895
2	0	1276	55	9	75	0	15	1430	0.8923
3	5	0	1937	0	75	135	0	2152	0.9001
4	1	41	62	210	0	0	14	328	0.6402
5	6	39	0	0	2114	63	20	2242	0.9429
6	0	35	69	6	0	2201	15	2326	0.9463
7	4	44	0	5	87	0	452	592	0.7635
Column total	150	1435	2192	235	2417	2469	516	9414	
PA	0.8933	0.8892	0.8837	0.8936	0.8746	0.8915	0.8760		0.8841

TABLE 8: Error matrix of SAGA-CFI-SVM ($T_c = 0.94$).

Class	Reference data							Row total	UA
	1	2	3	4	5	6	7		
Classified data									
1	136	0	32	1	53	49	0	271	0.5018
2	0	1323	43	2	35	0	4	1407	0.9403
3	5	0	2058	0	50	95	0	2208	0.9321
4	3	23	27	227	0	0	13	293	0.7747
5	2	24	0	0	2239	29	13	2307	0.9705
6	0	35	32	1	0	2296	9	2373	0.9676
7	4	30	0	4	40	0	477	555	0.8595
Column total	150	1435	2192	235	2417	2469	516	9414	
PA	0.9067	0.9220	0.9389	0.9660	0.9264	0.9299	0.9244		0.9301

TABLE 9: Total accuracy and Kappa value ($T_c = 0.94$).

Index	Total accuracy	Kappa value
SVM	86.72%	0.8315
GA-SVM	90.28%	0.8752
SAGA-SVM	92.49%	0.9030
CFI-SVM	88.41%	0.8539
SAGA-CFI-SVM	93.01%	0.9114

7.5. Computational Time Complexity. There is one issue that needs to be considered. The proposed procedure constructs and analyses probably consume considerable time. Thus, we compare the time complexity of the four methods GA, SAGA, CFI, and SAGA-CFI in this part. The time complexity of SAGA-CFI is $O(n^2)$, just the same as the other three methods. This means that the processing cost of SAGA-CFI is no more than the others.

7.6. Classification Experiment. The hyperspectral image is also categorized into seven classes, while the ratio of training and test samples remains 1:3. The numbers of training samples and of test samples are shown in Table 5.

In this work, we implement another two similar classification methods for hyperspectral images to compare with

the algorithm proposed in the paper. One method is based on SAGA and SVM classification (SAGA-SVM). The other method is based on CFI and SVM classification (CFI-SVM). Similarly, the method of this paper is based on SAGA, CFI, and SVM classification (SAGA-CFI-SVM). The error matrices of the three methods are presented in Tables 6, 7, and 8, while the total accuracy and Kappa value are exhibited in Table 9. The threshold T_c is 0.94 in all of the above methods. Table 10 shows the different total accuracy and Kappa value by changing the threshold T_c .

In the error matrix, the product's accuracy (PA) is defined as

$$PA_i = \frac{x_{i,i}}{x_{+i}}, \quad (27)$$

and the user's accuracy (UA) is defined as

$$UA_i = \frac{x_{i,i}}{x_{i+}}, \quad (28)$$

where $x_{i,i}$ is the value on the major diagonal of the i th row in the error matrix, x_{i+} is the total number of the i th row, and x_{+i} is the total number of the i th column.

FIGURE 3: (a) The fitness curve of subspace 1; (b) the fitness curve of subspace 2; (c) the fitness curve of subspace 4; (d) the fitness curve of subspace 5; (e) the fitness curve of subspace 8; (f) the fitness curve of subspace 9; (g) the fitness curve of subspace 10; and (h) the fitness curve of subspace 11.

TABLE 10: Total accuracy and Kappa value with changing T_c .

Index	Total accuracy				Kappa value	
T_c	0.940	0.965	0.970	0.940	0.965	0.970
CFI-SVM	88.41%	89.12%	68.75%	0.8539	0.8654	0.6173
SAGA-CFI-SVM	93.01%	93.91%	72.52%	0.9114	0.9223	0.6355

To measure the agreement between the classification and the reference data, we compute the kappa coefficient based on the following equation:

$$\text{Kappa} = \frac{\left[N \sum_{i=1}^k x_{i,i} - \sum_{i=1}^k (x_{i+} \times x_{+i}) \right]}{\left[N^2 - \sum_{i=1}^k (x_{i+} \times x_{+i}) \right]}, \quad (29)$$

where N is the number of total pixels.

The original reference image, the SAGA-SVM classification image, the CFI-SVM classification image, and the SAGA-CFI-SVM are illustrated in Figure 4.

8. Conclusions

An innovative band selection algorithm called SAGA-CFI has been developed and combined with the classification method

FIGURE 4: (a) The original reference image; (b) the SVM classification image; (c) the GA-SVM classification image; (d) the SAGA-SVM classification image; (e) the CFI-SVM ($T_c = 0.94$) classification image; (f) the SAGA-CFI-SVM ($T_c = 0.94$) classification image; (g) the CFI-SVM ($T_c = 0.965$) classification image; and (h) the SAGA-CFI-SVM ($T_c = 0.965$) classification image.

SVM to classify hyperspectral remote sensing images. On the basis of subspace decomposition, SAGA was used in each subspace to lower the computational complexity and select the suitable bands, and CFI method was adopted to further modificate the selecting bands in order to increase classification accuracy. SAGA-CFI-SVM has been implemented

to achieve improved classification methods compared with conventional algorithms. Comparison results show that the proposed method is superior in terms of classification accuracy.

The classification of hyperspectral remote sensing images based on SAGA-CFI-SVM in this paper is far from complete

and thus requires further research. One problem cited is the further reduction of the computational complexity of SAGA and the acceleration of the searching procedure faster. Another problem is the thorough improvement of the kernel function to obtain significantly higher classification accuracy. Last but not least, we need to study the classification method based on selective ensemble support vector machine, for it may further improve the accuracy.

Acknowledgments

This study is supported by the National Natural Science Foundation of China (no. 61271386) and funded by the CRSRI Open Research Program (no. CKWV2013215/KY) and the Industrialization Project of Universities in Jiangsu Province (no. JH10-9).

References

- [1] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 8, pp. 1778–1790, 2004.
- [2] I. T. Jolliffe, *Principal Component Analysis*, Springer Series in Statistics, Springer, New York, NY, USA, 1986.
- [3] A. A. Green, M. Berman, P. Switzer, and M. D. Craig, "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 26, no. 1, pp. 65–74, 1988.
- [4] G. Camps-Valls, L. Gómez-Chova, J. Calpe-Maravilla et al., "Robust support vector method for hyperspectral data classification and knowledge discovery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 7, pp. 1530–1542, 2004.
- [5] P. Gurram and H. Kwon, "Contextual SVM using hilbert space embedding for hyperspectral classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 5, pp. 1031–1035, 2013.
- [6] Z. Sun, C. Wang, H. Wang, and J. Li, "Learn multiple-kernel SVMs for domain adaptation in hyperspectral data," *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 5, pp. 1224–1228, 2013.
- [7] B. Tso and P. Mather, *Classification Methods For Remotely Sensed Data*, CRC Press, New York, NY, USA, 2001.
- [8] C. Yang, S. Liu, L. Bruzzone, R. Guan, and P. Du, "A feature-metric-based affinity propagation technique for feature selection in hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 5, pp. 1152–1156, 2013.
- [9] W. Xia, B. Wang, and L. Zhang, "Band selection for hyperspectral imagery: a new approach based on complex networks," *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 5, pp. 1229–1233, 2013.
- [10] S. Li, H. Wu, D. Wan, and J. Zhu, "An effective feature selection method for hyperspectral image classification based on genetic algorithm and support vector machine," *Knowledge-Based Systems*, vol. 24, no. 1, pp. 40–48, 2011.
- [11] M. Pal, "Hybrid genetic algorithm for feature selection with hyperspectral data," *Remote Sensing Letters*, vol. 4, no. 7, pp. 619–628, 2013.
- [12] M. Cui, S. Prasad, W. Li, and L. M. Bruce, "Locality preserving genetic algorithms for spatial-spectral hyperspectral image classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 6, no. 3, pp. 1688–1697, 2013.
- [13] M. Qinglin, "A gas outburst prediction algorithm based on Choquet fuzzy integral," in *Proceedings of the WRI Global Congress on Intelligent Systems (GCIS '09)*, pp. 3–7, May 2009.
- [14] Y. Zhang, M. D. Desai, J. Zhang, and M. Jin, "Adaptive subspace decomposition for hyperspectral data dimensionality reduction," in *Proceedings of the International Conference on Image Processing (ICIP '99)*, pp. 326–329, October 1999.
- [15] J. Zhang, Y. Zhang, B. Zou, and T. Zhou, "Fusion classification of hyperspectral image based on adaptive subspace decomposition," in *Proceedings of the International Conference on Image Processing (ICIP '00)*, pp. 472–475, September 2000.
- [16] F. Liang and W. H. Wong, "Evolutionary Monte Carlo: applications to Cp model sampling and change point problem," *Statistica Sinica*, vol. 10, no. 2, pp. 317–342, 2000.
- [17] C. Wang, J. D. Hyman, A. Percus, and R. Caflisch, "Parallel tempering for the traveling salesman problem," *International Journal of Modern Physics C*, vol. 20, no. 4, pp. 539–556, 2009.
- [18] S. Auephanwiriyakul, J. M. Keller, and P. D. Gader, "Generalized Choquet fuzzy integral fusion," *Information Fusion*, vol. 3, no. 1, pp. 69–85, 2002.
- [19] N. Hao and G. J. Wang, "Double-null asymptotic additivity of generalized fuzzy valued Choquet integrals," *Sichuan Shifan Daxue Xuebao*, vol. 30, no. 1, pp. 62–65, 2007.
- [20] Y. Tarabalka, M. Fauvel, J. Chanussot, and J. A. Benediktsson, "SVM- and MRF-based method for accurate classification of hyperspectral images," *IEEE Geoscience and Remote Sensing Letters*, vol. 7, no. 4, pp. 736–740, 2010.
- [21] H. B. Wang, Z. Chen, X. Wang, and Y. Ma, "Random finite sets based UPF-CPHD multi-object tracking," *Journal on Communications*, vol. 33, no. 12, pp. 147–153, 2012.
- [22] P. Du, K. Tan, and X. Xing, "Wavelet SVM in reproducing kernel hilbert space for hyperspectral remote sensing image classification," *Optics Communications*, vol. 283, no. 24, pp. 4978–4984, 2010.
- [23] C. Huang, K. Song, S. Kim et al., "Use of a dark object concept and support vector machines to automate forest cover change analysis," *Remote Sensing of Environment*, vol. 112, no. 3, pp. 970–985, 2008.
- [24] G. P. Tan, X. Y. Ni, X. Q. Liu, C. Y. Qu, and L. Y. Tang, "Real-time multicast with network coding in mobile ad-hoc networks," *Intelligent Automation and Soft Computing*, vol. 18, no. 7, pp. 783–794, 2012.
- [25] D. Tuia, F. Ratle, F. Pacifici, M. F. Kanevski, and W. J. Emery, "Active learning methods for remote sensing image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 7, pp. 2218–2232, 2009.
- [26] J. Ham, Y. Chen, M. M. Crawford, and J. Ghosh, "Investigation of the random forest framework for classification of hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 492–501, 2005.
- [27] B. Zhang, S. Li, X. Jia, L. Gao, and M. Peng, "Adaptive Markov random field approach for classification of hyperspectral imagery," *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 5, pp. 973–977, 2011.
- [28] V. Tsagaris, V. Anastassopoulos, and G. A. Lampropoulos, "Fusion of hyperspectral data using segmented PCT for color representation and classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 10, pp. 2365–2374, 2005.

- [29] D. Ming, J. Luo, L. Li, and Z. Song, "Modified local variance based method for selecting the optimal spatial resolution of remote sensing image," in *Proceedings of the 18th International Conference on Geoinformatics (Geoinformatics '10)*, June 2010.
- [30] Y. Zhu, P. K. Varshney, and H. Chen, "Evaluation of ICA based fusion of hyperspectral images for color display," in *Proceedings of the 10th International Conference on Information Fusion (FUSION '07)*, July 2007.
- [31] Y. S. Zhao, "Methods on optimal bands selection in hyperspectral remote sensing data interpretation," *Journal of Graduate School, Academia Sinica*, vol. 16, no. 2, pp. 153–161, 1999 (Chinese).