Sensors & Transducers, Vol. 157, Issue 10, October 2013, pp. 6-13

www.sensorsportal.com

Sensors & Transducers

http://www.sensorsportal.com

Object-Oriented Classification of Hyperspectral Remote Sensing Images Based on Genetic Algorithm and Support Vector Machine

Hongmin GAO, Lizhong XU, Chenming LI, Xin WANG, Minggang XU
College of Computer and Information Engineering, Hohai University,
Nanjing, 211100, China
E-mail: banjamin2006@163.com, lcm@hhu.edu.cn

Received: 11 July 2013 /Accepted: 25 October 2013 /Published: 31 October 2013

Abstract: This paper proposes a method of reducing dimensions based on genetic algorithm and object-oriented classification based on support vector machine (SVM). The basic idea is subspace decomposition of hyperspectral images at first, then selecting suitable bands in each subspace by using genetic algorithm and putting all selected bands of each subspace together. Furthermore, the hyperspectral image is segmented into a series of objects and then the spectral features and spatial features of objects in the selected bands are extracted. Finally, SVM classification is used according to features of the objects. The algorithm proposed is more effective and superior in dimension reduction and classification of hyperspectral image. *Copyright © 2013 IFSA.*

Keywords: Hyperspectral image classification, Genetic algorithm, Support vector machine, Band selection

1. Introduction

In recent years, a variety of remote sensing satellite has been successfully launched, which promotes remote sensing data acquisition technology towards the directions [1] of three high (high spatial resolution, high spectral resolution and high time resolution) and three multiple (multiple platforms, multiple sensors and multiple perspectives). Compared with the traditional multi-spectral remote sensing images, the hyperspectral image is imaged within the certain scope of narrow bands, which provides people with rich spectral information. It expands the application range of the remote sensing technology. However, with the increase of spectral information, the data dimension of hyperspectral image has also increased, which brings loads of data of the image. It not only makes the operation complicated and processing speed greatly decreasing, but also leads to decrease the accuracy of classification in the case of limited samples [2]. This means that it is necessary to reduce data dimension before processing and analyzing hyperspectral image. Dimension reduction methods mainly include two forms: feature extraction and feature selection. Feature extraction means that the original data is transformed to another space according to certain rules. The most information of original data after the transformation in the space is concentrated in the low dimension, therefore low dimensional data can be used for subsequent processing instead of the original data. Feature extraction mainly includes principal component analysis (PCA) [3], minimum noise fraction (MNF) [4] and so on. Feature selection means that we select some subset of the original feature space which is a simplified feature space and

Article number P_1377

contains the main spectral features. Feature selection methods mainly include band selection method based on the amount of information and band selection method based on inter-class separability. Feature extraction is convenient and quick, but it is implemented through a certain transformation. So it will damage physical properties of original band. For hyperspectral image with a number of bands, feature selection is a good method for dimension reduction.

In order to extract useful information from remote sensing image, complex phenomenon is simplified to general category by classifying [5]. In the process of classification, hyperspectral data has some problems such as a lot of bands, more data, uncertain data and easy to be influenced by Hughes phenomenon. Nowadays there are a variety of intelligent classification methods for hyperspectral image classification, such as fuzzy classification, decision tree classification, neural network and Support Vector Machine (SVM) classification [6]. SVM is the most effective method which has also most widely used in the statistical learning. It has strict theoretical basis. In terms of classification ability SVM is superior to some classifiers such as maximum likelihood and neural networks in the small sample learning, noise resistance, the learning efficiency and extension. It can effectively overcome the Hughes phenomenon in hyperspectral classification owing to the lack of sample.

The characteristics of object-oriented classification are the most basic classification objects, which is the different from pixel. The image objects defined as a single zone with same shapes and same spectral features [8]. The core of object-oriented classification is hyperspectral image segmentation. Spectral information and spatial information are combined together at this stage. In many cases, the image objects extracted can provide more meaningful information for image classification. Pixel only has spectral features, but image object contains abundant spatial information such as shape. The establishment of image object can improve the effect of classification.

This paper researches the dimension reduction and classification of hyperspectral remote sensing image .Firstly, we separate adaptively feature space into subspace. Then we select suitable bands according to genetic algorithm. We carry out image segmentation and extract spectral and spatial features from the image parcel object. At last we use the SVM classification. Its effectiveness and superiority can be verified by simulation.

2. Band Selection

2.1. 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 to be 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. The commonly used method continues to be ASD [9]. 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:

$$R_{ij} = E\left[(x_i - \mu_i)(x_j - \mu_j)\right] \Big/ \sqrt{E(x_i - \mu_i)^2}\sqrt{E(x_j - \mu_j)^2} \quad (1)$$

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.

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.

2.2. Band Selection Based on Genetic Algorithm

Genetic algorithm is a kind of search optimization algorithm based on evolutionary biology and molecular genetics. The search process of genetic algorithm is as follows: at first the initial individuals randomly generate a number of codes constitute the original group. Moreover, according to the principle of survival of the fittest, we decode the individual and restore the coding parameters to the actual parameters. The object function is used to calculate the fitness and choose the individuals with high

fitness to be a new group. Finally, these excellent characteristics of new groups can be inherit and transmit to the next generation by cross and mutation. So, we repeat this process "select-cross-mutation-reselect" again and again in order to gradually accumulate good genes coming from each group and increase fitness of the optimal individual, until the iteration converges.

2.2.1. Initial Population

The first step to solve optimization problem with genetic algorithm is to generate an initial population. The standard method of generating initial population is to produce random values in the feasible domain and assign them to each gene of each chromosome. Random choice is to ensure that the initial population can express equably for the whole search space. If the initial population does not cover the search space, it may ignore the zones which are not covered.

The size of the initial population will affect the computational complexity and the space search ability. The large number of individuals can increase the diversity, and further increase space search ability of the population. However, the more individuals are, the slower the search speed is. On the other hand, if the number of individuals in population is too small, it can only search a small part of the space. It need more iteration to converge, even it can't find the optimal solution.

2.2.2. Fitness Function

In Darwin's evolutionary model, the individual with best traits can have greater chance of survival and reproduction. In order to determine an individual's ability to survive, we use a mathematical function to characterize good or bad degree of the solution which chromosomes represent, called fitness function. For optimization problem with constraints, it should add a penalty function besides original objective function. This paper define the classification accuracy which is obtained through training feature subset as fitness function, because by iterating again and again our goal is to find optimal feature subset and maximize the classification accuracy. Classification accuracy can be calculated by SVM.

2.2.3. Cross

Cross means the parent produces offspring by sexual reproduction. It is expected that the offspring can inherit the good genes coming from their parent. The offspring is produced by two parents, or it can be produced by one or more. Single-point cross means we randomly select a cross in parent chromosomes, then exchange two pieces of parent chromosomes to produce offspring. There are also two points cross and homogeneous cross.

One of the earliest floating point cross operator is linear operator put forward by Wright. Three offspring (xi(t)+X2(t)), 1.5xi(t)-0.5x2(t)), (-0.5xi(t)+1.5x2(t)) can be produced by the parent Xi(t)+X2(t). The best one can be selected as offspring.

Compared with parent operator, main goal of multiple parents operators is to strengthen exploration ability. Through integrating information of multiple parents individual, the similarity between the offspring and the parent is lower compared with parents operator on average. It is damaged more.

2.2.4. Mutation

Mutation can avoid falling into local minimum near global minimum. Mutation is randomly change offspring. For binary code string, we can randomly choose bit and reserve corresponding bit. For binary representation of floating-point decision variables, Hinterding come up with a Gauss mutation. We can transform a decision variable into a floating-point value and use Gauss noise to mutation. For each chromosome, it generates a random number in Poisson, which decides number of variant genes in this chromosome. Then, we transform the bit strings represent these genes. Each float-point value is added step length N(0,aj), aj is 0.1 times than domain of corresponding float point. Later we transform variant float-point value back to bit string. Hinterding pointed out that Gauss variation representing decision variables as float point can obtain better result. Jones put forward a macro variation operator, which is also called the headless chicken operator. This operator as above uses cross operator restructure a parent individual and an individual generating randomly create an offspring. Although we restructure each individual and the individual generating randomly by crossing, this process can not be called cross operator. Because the concept of heredity does not exist. This operator is still called variation owing to introducing new genetic material which generates randomly.

2.2.5. Selection

After obtaining a new population, it need to select according to survival of the fittest. Only the individuals with high fitness can be selected. Holland comes up with a proportional selection to select the individual with high fitness. We can apply a probability distribution of proportional to fitness to individual selection.

$$\varphi(x_i) = \frac{f_{\max} - f(x_i)}{\sum_{k=1}^{N} (f_{\max} - f(x_k))} \tag{2}$$

In the formula, N is the size of population, $\varphi(x_i)$ is the probability of x_i selection, $f_{\max}$ is the maximum fitness function in the population, $f(x_i)$ is the fitness function of x_i.

Boltzmann Selection is based on thermodynamics laws of simulated annealing. Selection probability is:

$$\varphi(x_i)=\frac{1}{1+e^{f(x_i)/T(t)}} \tag{3}$$

T is temperature parameter and its initial value is very big, which can ensure that each individual has the same probability to be selected .With decrease of T(t), the choice will be focused on better individual.

When new offspring replaces the parent, it will lead to lose the parent with high fitness. Elite selection can avoid this problem and it copies directly the best individual in current population to next generation. The way in this paper is elite selection.

2.2.6. Termination Conditions

When the fitness function of the optimal individual does not improve through continuous certain generations, it will be stopped. When population doesn't change, during a period of time, the average genetic variation is too little, so the genetic method of IT can be terminated. It is stopped at the time of getting an acceptable solution. When the optimal solution obtained is closed to the optimal solution in theory, it is in accord with the requirement and then it can be terminated.

3. SVM Classification

SVM is a kind of machine learning algorithm based on statistical learning theory, by means of the criterion of structural risk minimization (SRM), which minimizes sampling error at the same time reduces the model of the upper bound of the generalization error, so as to improve the generalization ability of the models [10].

For the training sample (x_i, y_i), i =1，2…n, $x_i \in R^d$, $y_i \in \{-1,1\}$, we can get the optimal classification plane standardly by solving the following quadratic programming problem, so as to the two kinds samples are separated as possible:

$$\begin{cases} \min\limits_{w,b,\xi} \dfrac{1}{2}w^T w + C\sum\limits_{i=1}^{n}\xi_i \\ s.t. \quad y_i(w^T\phi(x_i)+b) \geq 1-\xi_i \\ \xi_i \geq 0, i=1,2,\cdots,n \end{cases} \tag{4}$$

In the formula, C is a normal number, for the punishment factor, also known as the regularization parameters of SVM.

Introduction of Lagrange multiplier we can transform the optimal classification plane into the dual problem of a convex quadratic programming problem:

$$\begin{cases} \max Q(\alpha)=\sum\limits_{i=1}^{n}\alpha_i - \dfrac{1}{2}\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n}\alpha_i\alpha_j y_i y_j K(x_i,x_j) \\ s.t. \quad \sum\limits_{i=1}^{n}\alpha_i y_i = 0 \\ 0 \leq \alpha_i \leq C, i=1,2\cdots n \end{cases} \tag{5}$$

After solving the above problem, we can get the optimal classification function is:

$$F(x)=\text{sgn}[(w^*)^T\phi(x)+b^*]=\text{sgn}(\sum_{i=1}^{n}\alpha_i^* y_i K(x_i,x)+b^*) \tag{6}$$

The sum of the formula (6) is in fact only for support vector. b^* is classification threshold, we can use any one of the support vector to obtain, or obtain it by any pair of values of two types. The kernel function can take many forms:(1)linear kernel function: $K(x,x_i)=(x\bullet x_i)$; (2) the form of polynomial kernel function: $K(x,x_i)=[(x^T x_i)+1]^q$,the corresponding SVM is a q rank order polynomial classifier; (3) the form of kernel function of Gaussian RBF (radial basis function, RBF): $K(x,x_i)=\exp\left\{-\dfrac{\|x-x_i\|^2}{\gamma^2}\right\}, \gamma>0$, Corresponding SVM is a kind of radial basis function classifier; (4) S kernel function (sigmoid), such as $k(x,x_i)=\tanh[v(x^T x_i)+c]$, in the formula, parameters $v>0, c<0$. This paper chooses Gaussian radial basis kernel function for classification.

4. Simulation Experiments and Analysis

4.1. The Experimental Images

In this section, we can conduct the analysis and evaluation on the performance of the algorithm mainly by the simulation. This study choose AVIRIS hyperspectral remote sensing images taken by remote sensing experimental area in northwest Indiana which is released by purdue university on June 12, 1992, it has a total of 220 bands. Experiment selected 89, 5, 120 band and the false color image formed is shown in Fig. 1.

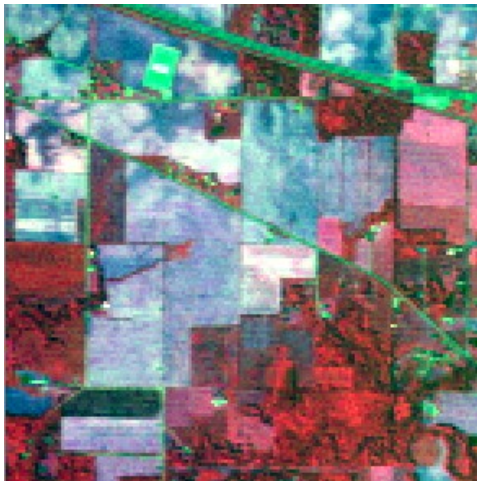

Fig. 1. AVIRIS false color synthesized image.

4.2. Subspace Decomposition

We use adaptive subspace decomposition based on correlation filtering to obtain related coefficients between each band, due to excessive value of number, only lists some the related coefficients of the band, shown in Table 1.

Seen from Table 1, the correlation coefficient of hyperspectral data of each band is 1, the correlation coefficient between the band and the band is large, it verified the previously mentioned conclusion that correlation between bands is very strong, there is a large amount of data redundancy, dimension reduction is very necessary. By setting the corresponding threshold, the band number of each subspace and the number of subspace can be determined. In this paper, the threshold is set as 0.8, 7 sub space result from the decomposition and the subspace dimension is shown in Table 2.

Because some bands are polluted by water vapor absorption and noise seriously in the original band, must to be eliminated, these bands are: 1~4, 78, 80~86103~110149~, 165217~165217. The band dimension subspace decomposition and band contained after removing severe noise pollution as shown in Table 3.

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

4.3. Band Selection in Subspace

On the basis of surficial truth, the classification experiment is carried out according to the seven class object, training samples and testing samples is about 1:1 proportion to select, using genetic algorithm in each subspace, to calculate the fitness value of each subspace, namely the classification accuracy rate, as shown in Fig. 3-10. Due to 3, 6, 7 three subspace are single band, so we don't have to use genetic algorithm in band selection. Support vector machine select and use Gaussian radial basis kernel function, the search scope of C and γ respectively: $[2^{-3}, 2^{10}]$ and $[2^{-8}, 2^{2}]$.

4.4. Image Segmentation and Feature Extraction

By using the segmentation algorithm based on watershed algorithm [13], the hyperspectral image is segmented into several objects, as shown in Fig. 2.

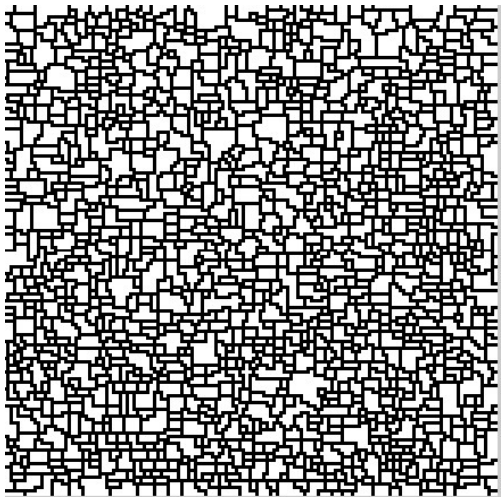

Fig. 2. Image of segmentation result.

The normalized difference vegetation index (NDVI), the normalized difference water index (NDWI) are used in the extraction of the information such as vegetation, water body, therefore, with the normalized difference vegetation index, the normalized difference water index are used as band object characteristics acquisition. Describing operator used to describe area texture content is based on pixel statistics of the mean, standard deviation, third order moment, entropy, partial degrees, etc. A total of seven kinds of characteristics, in 11 bands selected, each band has seven characteristics, so that each object is represented as a single point in the 77-dimensional feature space.

4.5. Classification Experiments

After selected the corresponding band, choose 7 types of surface features to classify, support vector machine (SVM) method is still used here, the training samples and testing samples are also selected according to approximately 1:3 proportion, the specific selection as shown in Table 4.

After confirmed the training samples and testing samples, we conduct the compare respectively from the following three aspects of support vector machine classification result (SVM), the genetic algorithm and support vector machine classification result (GA-SVM), and the genetic algorithm and support vector machine object-oriented classification result (GA-SVM-O). Total classification accuracy and Kappa value are shown in Table 5.

Table 4. 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 5. Total accuracy and Kappa value.

Index	Total accuracy	Kappa value
SVM	88.41%	0.8539
GA-SVM	90.67%	0.8826
GA-SVM-O	92.13%	0.9056

1) Error Matrix

Error matrix is also called the confusion matrix, matrix needs to point to point comparison between evaluate image and calibration images, accurately determine each point on the calibration image corresponds to the position a classified image. The main diagonal of the matrix means the number of pixels assigned to the correct category. Outside of diagonal elements are remote sensing classification relative to the error classification tree of the ground reference point, called the error. Among each element in the line respects the intersection pixel number of certain types of collections in classification results and each type of collections of the ground reference number; each element in the row respects the intersection pixel number of certain types of collections of the ground reference number and each type of collections in classification results. The column total respects all pixels number of certain types of the collection in ground reference information.

Product's accuracy (PA) is:

$$PA_i = \frac{x_{i,i}}{x_{+i}} \tag{7}$$

User's accuracy (UA) is:

$$UA_i = \frac{x_{i,i}}{x_{i+}} \tag{8}$$

2) Kappa Analysis

Kappa analysis is a kind of method that quantitative evaluates the consistency between remote sensing classification figure and reference data, which can reflect the classification error of overall image, it uses discrete multivariate method, overcomes the problem that the integral precision is too dependent on category numbers and samples. The formula is expressed as:

$$Kappa = \frac{\text{overall accuracy- exp ected accuracy}}{1 - \text{exp ected accuracy}} \tag{9}$$

This measure is through the conformance between error matrix of the diagonal and the total number of all lines and rows to express, which can be calculated by following formula:

$$Kappa = \frac{N\sum_{i=1}^{k} x_{i,i} - \sum_{i=1}^{k}(x_{i+} \times x_{+i})}{N^2 - \sum_{i=1}^{k}(x_{i+} \times x_{+i})} \tag{10}$$

The original reference image, the SVM classification image, the GA-SVM classification image and the GA-SVM-O classification image are illustrated in Fig. 3-6.

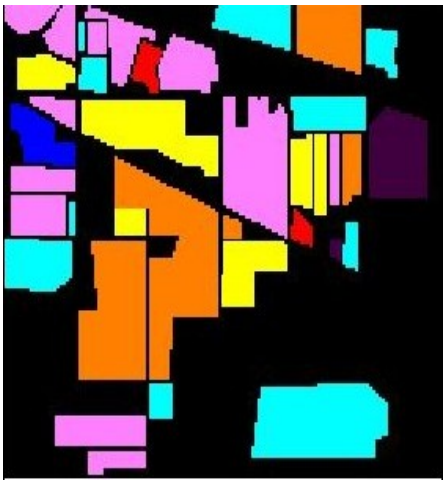

Fig. 3. Original reference image.

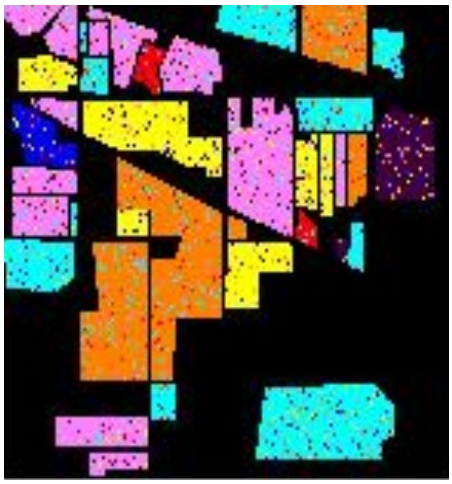

Fig. 4. SVM classification image.

Fig. 5. GA-SVM classification image.

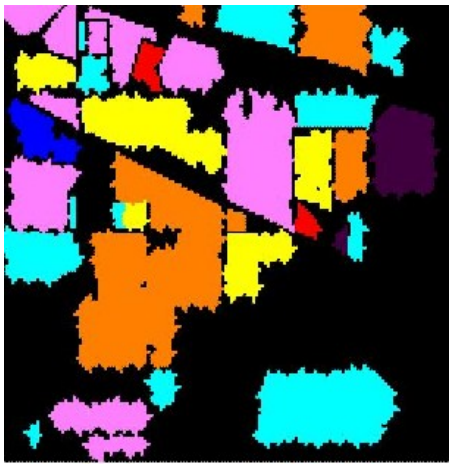

Fig. 6. GA-SVM-O classification image.

5. Conclusion

Hyperspectral images have some peculiarities such as large amount of data and high inter-band redundancy, through the process of dimension reduction for hyperspectral images can not only reduce the computational complexity, but also it can greatly improve the accuracy of classification. We will give priority to the band selection method between two main types in dimension reduction method – feature extraction and band selection. Because although feature extraction is convenient,

quick, but it is implemented by some transformation, which can erode the physical properties of the original band; And band selection is to deal with the original hyperspectral data process directly, retain the characteristic and sequence of the original data, which is a truly effective dimensionality reduction method. Support vector machine has the advantages of high-dimensional features, small samples and uncertain problems, which is suitable for hyperspectral image. And it can make full use of space characteristics of spot figure by the object-oriented classification at the same time .This paper puts forward its own solution based on the analysis of the basis of dimensionality of hyperspectral classification. Experimental results show that this method has certain advantages on the band selection and classification, which is a kind of effective classification method of hyperspectral data.

Based on genetic algorithm and support vector machine, object-oriented classification of hyperspectral remote sensing image remains further researches in the following aspects: (1) Artificial intelligence methods such as the genetic algorithm in dealing with so many data quantity is very time consuming, We can whether explore a better method in such aspects as parameters selection and set up when artificial intelligence methods such as the genetic algorithm in dealing with so much data that is very time consuming; (2) The kernel function of support vector machines uses the Gaussian radial basis form so far, can we make a great even fundamental improve on the kernel function in order to improve the classification ability more effectively.

Acknowledgements

This study is supported by the National Natural Science Foundation of China (No. 61271386), the Industrialization Project of Universities in Jiangsu Province (No. JH10-9).

References

[1] Deren Li, Discussion about the development of remote sensing and GIS in the 21st century, *Wuhan University Journal of Natural Sciences (information science section)*, Vol. 28, Issue 2, 2003, pp. 127-131.

[2] Yamin Zheng, Research on improvement of feature selection method based on genetic algorithm, *Chongqin University*, 2008.

[3] I. Jolliffe, Principal component analysis, New York: *Springer-Verlag*, 1986.

[4] A. Green, M. Berman, et al., A transformation for ordering multispectral data in terms of imagequality with implications for noise removal, *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 26, Issue 1, 1988, pp. 65-74.

[5] B. Tso, P. Mather, Classification methods for remotely sensed data, New York: *CRC Press*, 2001.

[6] J. Guo, J. Zhang, Y. Zhang, et al., Study of the comparison of land cover classification for multidate MODIS images, *Academic Journal of Surveying and Mapping*, Vol. 38, Issue 1, 2009, pp. 88-92.

[7] G. Hughes, On the mean accuracy of statistical pattern recognizers, *IEEE Transactions on Information Theory,* Vol. 14, Issue 1, 1968, pp. 55-63.

[8] F. Du, Q. Tian, X. Xia, et al., Analysis and evaluation of the object-oriented surface feature classification method, *Remote Sensing Technology and Application*, Vol. 19, Issue 1, 2004, pp. 20-23.

[9] J. Zhang, Y. Zhang, et al., Fusion classification of hyperspectral image based on adaptive subspace decomposition, in the *Proceeding of ICIP2000, IEEE Signal Processing Society*, Canada, Vol. 3, 2000, pp. 472-475.

[10] X. Zhang, Introduction to statistical learning theory and support vectormachines, Act Automation Sinica, Vol. 1, Issue 26, 2000, pp. 32-42.

[11] V. Vapnik, Statistical learning theory, New York: *Wiley,* 1998.

[12] C. Hsu, C. Lin, A comparison of methods for multi-class support vector machines, *IEEE Transactions on Neural Networks*, Vol. 13, Issue 2, 2002, pp. 415-42.

[13] M. Sonka, V. Hlavac, R. Boyle, Image processing, analysis and machine vision, Beijin, *The People's Posts and Telecommunications Press,* 2003, 83-127.