An automatic hyperspectral image classification method based on Subspace Partition and SVM

Hongwei She, Weidong Yan
School of Science, Northwestern Polytechnical University
Xi'an, China
Email: shehw@nwpu.edu.cn

Abstract—**Since the traditional Support Vector Machine (SVM) algorithm has high classification accuracy at the expense of huge training samples, an automatic hyperspectral image classification method based on subspace partition and SVM (ASP-SVM) algorithm is proposed. In this proposed method, the endmembers were extracted as the representative samples for each class, and a coarse classification result is obtained based on minimum distance clustering. General Sphere Criterion is introduced and applied to the coarse result, and the testing samples is divided into identified samples and unidentified samples. Then, the subspace partition is accomplished according to the probable mixing. Samples which have the highest category confidence in the subspace are selected as the training samples to subdivide the subspace of the unidentified samples to get the final classification. Classification experiment of hyperspectral data illustrates that the proposed approach is satisfied.**

Keywords-hyperspectral image classification; subspace partition; support vector machines; general sphere criterion

I. INTRODUCTION

Multispectral/hyperspectral image classification methods can be generally divided into supervised classification[1,2] and unsupervised classification[3,4]. Supervised classification require labeled samples to train the classifier, and the classification results rely on the quality of the labeled samples used for learning. However, the available training samples are often not enough for an adequate learning of the classifier. The classification, which synthesizes information of known and unknown samples to improve accuracy under little prior knowledge, is named semi-supervised classification[5]. In fact, the semi-supervised classification is supervised. Unsupervised classification is merely based on statistical characteristics of data without any prior knowledge. K-means and iterative self-organizing data (ISODATA) are two widely used unsupervised classification algorithms.

SVM is widely used in RS image classification and performs pretty well, but it needs enough samples to provide guarantee. Most of remote sensing (RS) images use unsupervised or semi-supervised classification due to the lack of real ground data. In this paper, an automatic SVM classification based on subspace partition is proposed. The remainder of the paper is organized as follows. Section II provides a basic principle of SVM theory. Section III proposes an automatic SVM classification based on subspace partition. In Section IV, the experiment for hyperspectral image is described, and the experimental result and analysis are provided. Finally, the conclusion is provided in Section V.

II. SVM THEORY

The Support Vector Machine was first proposed by Prof. Vapnik and his group[6,7]. It tends to find optimal classification surface if linear separable. This problem is the same as following duality problem: finding maximum value of Eq.1 under restrained condition:

$$\sum_{i=1}^{N} y_i\alpha_i = 0,\ \alpha_i \ge 0 \quad (i=1,2,\cdots,N).$$

$$Q(\alpha) = \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{N}\alpha_i\alpha_j y_i y_j (x_i \cdot x_j) \tag{1}$$

If α_j^* has optimal solution, then we get Eq.2.

$$w^* = \sum_{i=1}^{N}\alpha_i^* y_i x_i \tag{2}$$

Since most of practical problem are nonlinear separable, nonlinear SVM introduce kernel function $K(X_i, X_j)$ to substitute dot product in Eq.1. Thus α_i can be solved by Eq.3:

$$\text{Maximize} \quad Q(\alpha) = \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{N}\alpha_i\alpha_j y_i y_j K(x_i, x_j) \tag{3}$$

$$\text{s.t.} \quad \sum_{i=1}^{N}\alpha_i y_i = 0,\ 0 \le \alpha_i \le C, \quad i = 1,2,...,N \tag{4}$$

Here, C is penalty factor. Eq.5 is used to make decision after training.

$$f(x) = \text{sgn}(\sum_{i=1}^{N} \alpha_i y_i K(x_i, x) + b) \quad (5)$$

The SVM can be widely used in high-dimensional data classification Since the Kernel function not only needs no concrete form of nonlinear map but also avoid curse of dimensionality problem [8]. The general kernel functions are as follows [9]:

(1) Polynomial kernel: $K(x, y) = [(x \bullet y) + 1]^d$, if $d = 1$, it will be linear kernel;

(2) Radial basis function: $K(x, y) = e^{-\|x-y\|^2/2\sigma^2}$;

(3) Sigmoid function: $K(x, y) = \tanh(k(x \bullet y) + \delta)$.

III. AUTOMATIC CLASSIFICATION METHOD BASED ON SUBSPACE PARTITION AND SVM

A. Subspace Partition Theory

For target detection or classification of n pattern samples $X \subset R^m, Y \subset R^n$, we should find a mapping [10]:

$$f : R^m \to R^n, Y = f(x) \quad (6)$$

Training time will be very long when applying SVM to complex high dimension data. The subspace partition can effectively reduce the complexity of classification.

Therefore, for a complex classification, the mapping function f can be decomposed into combination of sub-maps $f = f_1 + f_2 + \cdots + f_k$ to reduce complexity. Here we have Eq.10:

$$\left.\begin{array}{l} f_1 : R^m \to R^{l_1} \\ f_2 : R^m \to R^{l_2} \\ \vdots \\ f_k : R^m \to R^{l_k} \end{array}\right\} \quad (7)$$
$$1 \le l_i \le n, \quad i = 1, 2, \cdots, k$$

Based on above theory, to realize classification, initial data can be firstly classified coarsely. Then, according to given rules, samples satisfied classification requirements will be labeled while the unsatisfied ones will be divided into subspace to do subdivision. To introduce the theory of subspace partition, the following definitions are given:

Definition 1: Class Center of Samples. Assume there are m unclassified samples $x_j, j = 1, 2, ..., m$ in n classes, the class center of class i is defined as Eq.8:

$$C_i = \frac{1}{m_i} \sum_{j=1}^{m_i} x_j^{(i)} \quad (8)$$

Here, m_i is the total samples of class i and $x_j^{(i)}$ is sample j of class i 。

Definition 2: Distance between Sample and Class Center. The distance between sample x and center of class i is defined as Eq.9:

$$D_i(x) = \|x - C_i\|, i = 1, 2, ..., n \quad (9)$$

We can also use other distance to measure $D_i(x)$ according to the actual conditions. Additionally, spectral similarity is another measurement for hyperspectral image processing. But it should be noted that for some measurements like spectral angle and coefficient, the larger the value the shorter the distance. Thus when using these measurements, we should take some process like reciprocal.

Definition 3: Class Radius. It is the maximum distance between class center and samples of the same class. Let R_i is the class radium of class i , as shown in Eq.10:

$$R_i = \max D_i(x_j^{(i)}), j = 1, 2, ..., m_i \quad (10)$$

According to above definition, the classification rule is Eq.11.

$$if \ D_i(x) = \min_{j=1,\cdots,n} D_j(x), then \ x \in \omega_i \quad (11)$$

That is, put x into the class which the center is closest to x .

Definition 4: General Sphere Criterion. Assuming that R_i is class radius of class i , if $D_i(x) \le R_i$, we say that x is in the general sphere of class i . Otherwise, x is out of the general sphere of class i **.**

Definition 5: Mix Classification Discriminate Rule. Let R_i, R_j are class radius of class i and class j , if each sample $x^{(i)}$ of class i satisfies $D_j(x^{(i)}) > R_j$ while each sample $x^{(j)}$ of class j satisfies $D_i(x^{(j)}) > R_i$, we believe that class i and class j are not mix classified. Otherwise, two classes have mix classification.

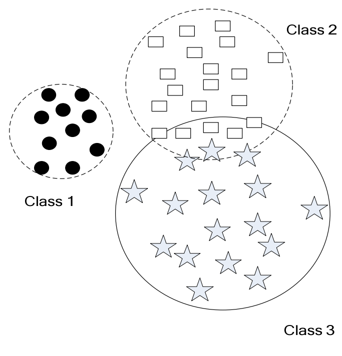

Figure 1. General sphere of three classes' samples

Figure1 indicate that class 1 can be easily classified since it has a clear boundary and it's general sphere not coincides with other two classes'. However, general sphere of class 2 and class 3 overlap each other and the overlap part should be focused on. Because a simple classification way cannot get right result, this part will be subdivided based on following classification algorithm.

B. Endmember extraction method

The equations are an exception to the prescribed specifications of this template. You will need to determine whether or not your equation should be typed using either the Times New Roman or the Symbol font (please no other font). To create multileveled equations, it may be necessary to treat the equation as a graphic and insert it into the text after your paper is styled.

The endmember extraction technology is the research focus of the hyperspectral image processing. Most of the endmember extraction methods are based on the conception of convex geometry[11,12]. An endmember extraction method based on maximum distance is given as follows [13]:

Step1: Calculate the average spectral vector of all pixels in the image, and denoted by $\mathbf{u}$.

Step2: Find the first endmember $\mathbf{e}_1$ which has the maximum distance with $\mathbf{u}$.

Step3: The pixel has the maximum distance with $\mathbf{e}_1$ is the second endmember $\mathbf{e}_2$.

Step4: The pixel has the maximum distance with the line formed by $\mathbf{e}_1$ and $\mathbf{e}_2$ is the third endmember $\mathbf{e}_3$. Then, find the pixel who has the maximum distance with the surface formed by $\mathbf{e}_1$, $\mathbf{e}_2$ and $\mathbf{e}_3$, and set it to fourth end-member $\mathbf{e}_4$. And so on, we can find all endmembers in the image.

The key of above method is how to calculate the distance from a pixel to a hyperplane. An effective way is by Schmitt orthogonalization to find the unit normal vector with the hyperplane[14].

C. Automatic SVM classification algorithm

The procedure of the ASP-SVM algorithm is described as follows:

Step1: Use the endmember extraction method given in section 3.2 to extract endmembers from the hyperspectral image.

Step2: Choose the endmember as the representative sample for each class and take it as initial class center. $C_i(i=1,2,...,n)$;

Step3: For each unclassified sample x , calculate $D_i(x)$, and then put x into the class which has minimum distance. Revise class center as follows:

$$C_i = \frac{m_i}{m_i+1}C_i + \frac{1}{m_i+1}x$$

and set $m_i = m_i + 1$.

Step4: Calculate class radius, label the samples and divide subspace by rule as following:

(1) Initial subspace is built for each class at first;

(2) For each sample x , if it is in general sphere of only one class, give it final class label of this class. Otherwise, if it is in general sphere of several classes, all these classes form a subspace and x become an unclassified sample of this subspace.

(3) To reduce classifier, subspace can be merged if a subspace is subset of another one. For instance, a subspace includes three classes {A、B、C}, then the subspaces formed by {A, B}, {A, C} or {B, C} can be merged into the subspace {A, B, C}.

Step5: For each subspace which included several classes, select the labeled samples which closest to each class center to form training samples to train SVM classifier. Then, SVM classifier is used to subdivide initial subspace and label unclassified samples.

Step6: Final classification is achieved after above five steps.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

The performance of the proposed approach is presented using a HYDICE aerial hyperspectral data of Washington area and an AVIRIS hyperspectral image data of Moffett Field of Indian Pines. The experiment is obtained on Core Duo, 2.2GHz and 2G memory PC with MATLAB codes.

The HYDICE aerial hyperspectral data of Washington area is gotten from Website of Purdue University and consists of 210 bands. First, it initially reduced to 191 bands by preprocessing. The image covers parts of Washington commercial district, including 1280×307 pixels and seven typical classes: Roofs, Street, Path, Grass, Trees, Water and Shadow. Sample information are shown in TABLE I.

TABLE I. SAMPLE INFORMATION OF WASHINGTON DATA

Class	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7
Objects	Roofs	Street	Path	Grass	Trees	Water	Shadow
Pixels	3834	416	175	1928	405	1124	97

Experiments have been carried out and compared with the following two well known unsupervised classification methods: ISODATA and KMEANS. The results are shown in Figure2.

In the experiment, we take spectral coefficient as distance measurement. Figure 2(a) shows the pseudo color image of original one (band 60, 27 and 17). Figure2(b), Figure2(c) and Figure2(d) are the classification results of ISODATA, KMENS and our method ASP-SVM. Figure2(e) shows the testing samples site. TABLE II shows that the given method ASP-SVM is significantly better than the classical unsupervised classification method ISODATA and KMEANS in classification accuracy. However, the given method is more time cost than the other two. It is noted that TABLE II only shows statistical result of testing (labeled data) and cannot give exact assessment for the whole RS image, but it can basically reflect whole classification performance on some extent. Here, run-time is the time to classify all pixels. This experiment takes radial basis function as kernel function, and training time spent mainly in choosing best parameters. Choosing penalty factor C and parameter σ beforehand with experience can greatly reduce run-time.

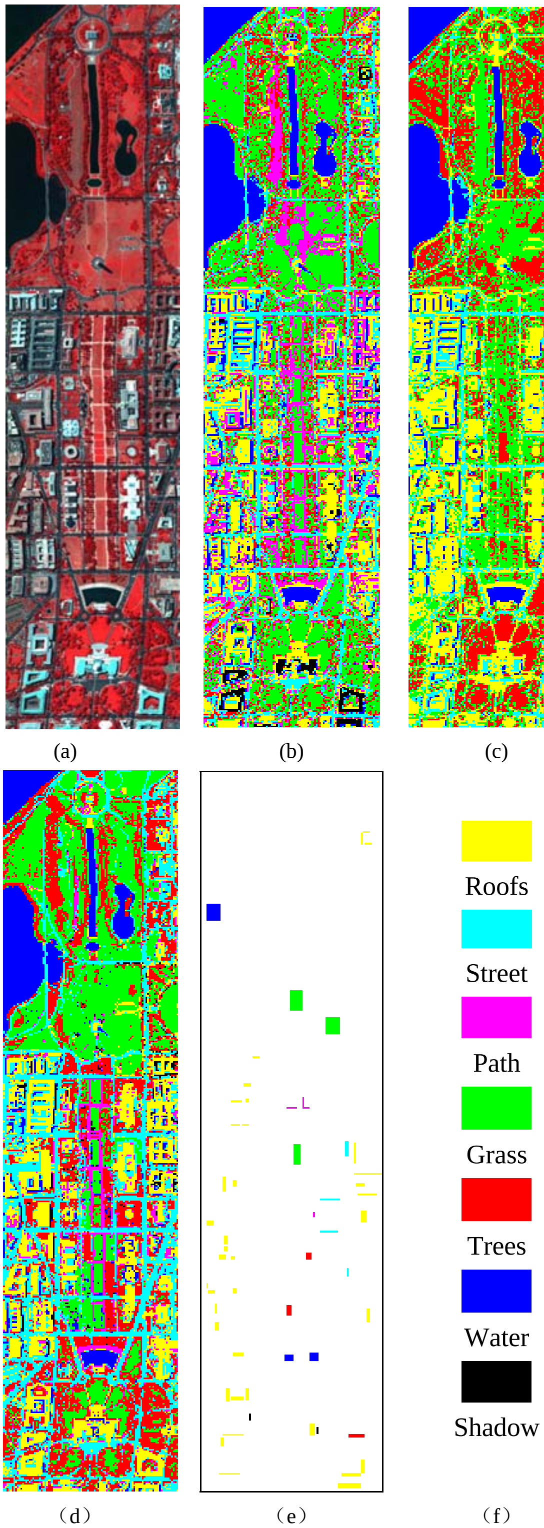

Figure 2. Classification Result of Washington Data

TABLE II. CLASSIFICATION PERFORMANCE STATISTICS

Classification method	ISODATA	KMEANS	ASP-SVM
Accuracy (%)	55.45	61.12	82.61
Kappa	0.5483	0.6217	0.8526
Run-time（s）	139.218312	93.123284	657.152796

Figure3 shows another experiment on hyperspectral image data of Moffett Field of Indian Pines. The original scene with size of 614 × 512 pixels and 224 bands is available online at htt:://aviris.jpl.nasa.gov/html/aviris.freedata.html. The main ground objects contain: vegetation、grass and field、buildings、slough and water、salt flats、empty terra. Due to lack of ground truth, the Quantitative evaluation can not be given. However, we can see that the given method is effective for unsupervised hyperspectral image classification.

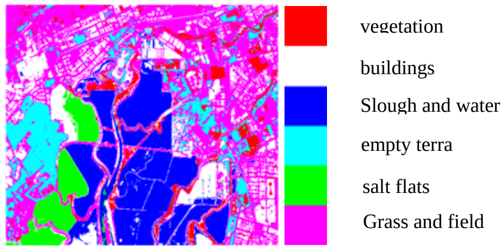

Figure 3. Classification Result of Moffett Field data by ASP-SVM

V. CONCLUSIONS

This paper presents a novel ASP-SVM classification method without any manual assistance. Choose the endmember extracted from the given hyperspectral image as the representative sample for each class. Then identified samples and unidentified samples are obtained from preliminary classification according to general sphere criterion. Subspace partition is accomplished because of probable mixing. Samples which have the highest category confidence in the subspace are selected as training samples to subdivide subspace to get final classification. Classification experiment of hyperspectral datasets illustrates that the proposed method is satisfied.

ASP-SVM is of great significance since it avoids huge training samples in traditional SVM classification. However, the result is affected by endmember extraction result.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (No. 61201323), Northwestern Polytechnical University Foundation of Fundamental Research (No.JC20110277).

REFERENCES

[1] I. Dopido, M. Zortea, A. Villa, et al., “Unmixing Prior to Supervised Classification of Remotely Sensed Hyperspectral Images”, IEEE Geoscience and Remote Sensing Letters, 2011, 8(4): 760~764.

[2] G. Moser, S. B. Serpico, “A Markovian generalization of support vector machines for contextual supervised classification of hyperspectral

images", 2010 2nd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), 2010, pp: 1~4.

[3] G. Bilgin, S. Erturk, T. Yildirim, "Unsupervised Classification of Hyperspectral Image Data Using Fuzzy Approaches That Spatially Exploit Membership Relations", IEEE Geoscience and Remote Sensing Letters, 2008, 5(4): 673~677.

[4] H. Z. Jiao, Y. F. Zhong, L. P. Zhang, P. X. Li, "Unsupervised Remote Sensing ImageClassification Using an Artificial DNA Computing Artificial DNA Computing", 2011 Seventh International Conference on Natural Computation (ICNC), July, 2011, 1341~1345.

[5] W. Z. Liao, A. Pižurica, P. Scheunders, W. Philips, Y. G. Pi, "Semisupervised Local Discriminant Analysis for Feature Extraction in Hyperspectral Images", IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, JAN., 2013, Vol. 51(1): 184~198.

[6] X. P. Wen, X. F. Yang, "An Unsupervised Classification Method for Hyperspectral Image Using Spectra Clustering", IEEE International Symposium on Knowledge Acquisition and Modeling Workshop, 2008, 1117 ~1120.

[7] C. Cortes, V. Vapnik, "Support Vector Networks", Machine Learning, 1995,20: 273~297.

[8] B. Scholkopf, K. Sung, C. Burges, et al., "Comparing Support Vector Machines with Gaussian Kernels to Radial Basis Function Classifiers", IEEE Transactions on Signal Processing, 1997, 45(11): 2758 ~2765.

[9] C. J. C. Burges, "Atutorial on Support Vector Machines for Pattern Recognition", Data Mining and Knowledge Discovery, 1998, 2(2): 955~974.

[10] K. Tan, P. J. Du, "Hyperspectal remote sensing image classification based on support vector machine", J. InfraredMillim. Waves, 2008, 27(2): 123~128.

[11] Y. N. Zhang, L. C. Jiao, "A local adaptive wavelet and gauss neural network synthesis classification system", Journal of Electronics, 1999,21(3):236~331.

[12] S. S. Li, Q. J.Tian, "An endmember progressive extraction algorithm for hyperspectral remote sensing image", Journal o f Remote Sensing , 2009, 13(2) : 269~275.

[13] M. Shoshany, F. Kizel, N. S. Netanyahu, et al., "An Iterative Search in End-Member Fraction Space for Spectral Unmixing", IEEE Geoscience and Remote Sensing Letters, 2011, 8(4): 706~709.

[14] X. R. Geng, "Target Detection and Classification for Hyperspectral Imagery", Institute of remote sensing applications, Chinese academy of sciences, 2005.