HYPERSPECTRAL IMAGE FEATURE EXTRACTION BASED ON GENERALIZED DISCRIMINANT ANALYSIS

Guopeng Yang [a,] *, Xuchu Yu [b], Xin Zhou [c,]

[a] Zhengzhou Institute of Surveying and Mapping,450052, Henan, China - yangguopeng@hotmail.com
[b] Zhengzhou Institute of Surveying and Mapping,450052, Henan, China - xc_yu@yahoo.com
[c] Institute of Information Engineering,450052, Zhengzhou, Henan, China - zx007_0_0@126.com

Commission VII, WG VII/3

KEY WORDS: Hyperspectral Image, Feature Extraction, Generalized Discriminant Analysis, Kernel Function

ABSTRACT:

The hyperspectral image enriches spectrum information, so compared with panchromatic image and multispectral image; it can classify the ground target better. The feature extraction of hyperspectral image is the necessary step of the ground target classification, and the kernel method is a new way to extract the nonlinear feature. In this paper, First the mathematical model of the generalized discriminant analysis was described, and then the processing method of this model was given, finally, we did two experiments. Through the tests, we can see that, in the feature space extracted by generalized discriminant analysis, the samples of the same class are near with each other; the samples of the different classes are far away. It can be concluded that the method described in this paper is suitable to hyperspectral image classification, and it can do better job than the method of linear discriminant analysis.

1. INTRODUNCTION

Hyperspectral remote sensing technology, which firstly comes out in the early 1980s, organically hangs the radiation information which relates to the targets' attribute, and the space information which relates to the targets' position and shape together. The spectrum information, which the hyperspectral image enriches, compared with panchromatic remote sensing image and multi-spectral remote sensing image, can be used to classify the ground target classification better. Hyperspectral remote sensing has very wide electromagnetic wave range, from visible light to shortwave red, even to medium infrared and thermal infrared. It has high spectral resolution, and has lots of bands, so can get the ground target's spectral feature curve, and recognize the targets by selecting and extracting the bands. We can get the target's spectral radiant parameters, and the quantitative analysis of the earth's surface target and extraction become possible. Because of the advantages of hyperspectral remote sensing, at present, lots of countries in the world have respect for this type of remote sensing. Hyperspectral remote sensing craft is form aerial to space aerospace. It' will become an important path of map cartography, vegetation investigation, ocean remote sensing, agriculture remote sensing, atmosphere research, environment monitoring, military information acquiring (Tong et al., 2006).

The hyperspectral images have so high dimension and the ground targets are so complicated, that it's difficult to obtain enough training samples (Hoffbeck et al., 1996). However, the traditional image classification method, such as the statistical pattern recognition and neural networks methods, which are based on large number samples hypothesis, need to get enough training samples to evaluate the prior classes' information which often cause the "Hughes" phenomenon. So, the feature extraction is one of the most important steps when we analyze the hyperspectral images (Zhang, 2003).

In the mid 1990s, with the kernel method applied to support vector machine successfully, people try to extend the ordinary linear methods of feature extraction and classification to nonlinear situation by using kernel function. Kernel methods for pattern analysis are developing so fast that there are so many achievements in the applied fields. It is named as the third revolution of pattern analysis algorithms following the linear analysis algorithms, neural networks and decision trees learning algorithms. Kernel methods have become focus of machine learning, application statistic, pattern recognition, and data mining, successfully applied in face recognition, speech recognition, character recognition, machine malfunction classification and so on (John et al., 2005).

We don't need to know the concrete form and parameters of the nonlinear mapping, the changes of form and parameters of kernel function can change the mapping from the input space to feature space, and change the performance of kernel methods. We can avoid dimension disasters phenomenon which exits in traditional mode analysis methods by using the kernel function, and it also can simplify computation, therefore, Kernel methods can precede the input with high dimensions. The kernel methods can combine with the different analysis algorithms, design the different kernel algorithms, and the two parts can be designed separately, so we can select different kernel function and analysis algorithm in different application fields.

In order to improve classification accuracy of hyperspectral remote sensing image, we can use the special classifier, such as SVM and KFDA. If we extract suitable feature of the hyperspectral image, the common classifier also can be used. One of the research trends in hyperspectral image is the

* Corresponding author. Tel.:+86-13733179927; E-mail address:.yangguopeng@hotmail.com.

nonlinear methods, and the kernel methods provide a new approach to the feature extraction. Some research scholars have studied the feature extraction methods of hyperspectral based on kernel function, such as kernel principal components analysis (KPCA) and kernel Bhattacharyya feature extraction (KBFE) (Lu, 2005).

In 2000, the Generalized Discriminant Analysis (GDA) was brought forward by Baudat (Baudat et al., 2000), which is the nonlinear extraction of Linear Discriminant Analysis, has been successfully used in face recognition (Gao et al., 2004) and mechanical failure classification (Li, 2003). In this paper, we first introduced the mathematical model and the solution of the GDA, applied this method to extract features from the hyperspectral image. Then we made experiments with two groups of the hyperspectral images which were obtained by different kinds of hyperspectral imaging system. At last the result was analyzed. The main contents were described in detail as follow.

2. GENERALIZED DISCRIMINANT ANALYSIS

Through mapping samples from the input space to the feature space with high dimensions, we carry on the liner methods of feature extraction in this feature space. Because of the dimension in the feature space is very large, and it may be infinitude, in order to avoid deal with the samples perceptibly ,we use the kernel functions to compute the inner product in the feature space.

2.1 Theory of Feature Extraction Based on GDA

Suppose there are C classes of samples, which are belong to $\omega_1, \omega_2, \ldots, \omega_m$, and the original sample $\boldsymbol{x}$ has n dimensions, so $\boldsymbol{x} \in \mathrm{R}^n$. If we map the sample $\boldsymbol{x}$ to feature space H with higher dimensions by the mapping ϕ, in the feature space, $\boldsymbol{x}$ will be $\phi(\boldsymbol{x}) \in H$ .If all the samples are mapped to the future space H, the intraclasses scatter matrix $\mathbf{S}_w^\phi$, the interclasses scatter matrix $\mathbf{S}_b^\phi$ and the total scatter matrix $\mathbf{S}_t^\phi$ of the training samples, will be described as follows:

$$\boldsymbol{S}_w^\phi = \frac{1}{N}\sum_{i=1}^{c}\sum_{j=1}^{N_i}\left(\phi(\boldsymbol{x}_j^i) - \boldsymbol{m}_i^\phi\right)\left(\phi(\boldsymbol{x}_j^i) - \boldsymbol{m}_i^\phi\right)^T \tag{1}$$

$$\boldsymbol{S}_b^\phi = \frac{1}{N}\sum_{i=1}^{c}\frac{N_i}{N}\left(\boldsymbol{m}_i^\phi - \boldsymbol{m}_0^\phi\right)\left(\boldsymbol{m}_i^\phi - \boldsymbol{m}_0^\phi\right)^T \tag{2}$$

$$\boldsymbol{S}_t^\phi = \frac{1}{N}\sum_{j=1}^{N}\left(\phi(\boldsymbol{x}_j) - \boldsymbol{m}_0^\phi\right)\left(\phi(\boldsymbol{x}_j) - \boldsymbol{m}_0^\phi\right)^T \tag{3}$$

where N_i is the amount of training samples belonging to the class ω_i, N is the amount of all the training samples. In the feature space H, $\phi(\boldsymbol{x}_j^i)$ is the sample j ($j = 1, \ldots N_i$) of class i ($i = 1, \ldots, C$), $\phi(\boldsymbol{x}_j)$ is the sample j ($j = 1, \ldots, N$) of all the samples, $\boldsymbol{m}_i^\phi = E\{\phi(\boldsymbol{x}) \mid \omega_i\}$ is the mean of samples in the class i, $\boldsymbol{m}_0^\phi = \sum_{i=1}^{C} P(\omega_i)\boldsymbol{m}_i^\phi$ is the mean of all the samples. $\boldsymbol{S}_b^\phi$, $\boldsymbol{S}_w^\phi$ and $\boldsymbol{S}_t^\phi$ are all nonnegative matrixes.

In the feature space H, the Fisher discriminant function can be defined as

$$J_1(\boldsymbol{w}) = \frac{\boldsymbol{w}^{\mathrm{T}}\boldsymbol{S}_b^\phi\boldsymbol{w}}{\boldsymbol{w}^{\mathrm{T}}\boldsymbol{S}_w^\phi\boldsymbol{w}} \tag{4}$$

where $\boldsymbol{w}$ is a nonzero vector.

In the feature space H, Generalized Discriminant Analysis (GDA) is to find a group of discriminant vectors ($\boldsymbol{w}_1, \ldots \boldsymbol{w}_d$), which can maximize the Fisher discriminant function (4), and all the vectors are orthogonal.

$$\boldsymbol{w}_i^T\boldsymbol{w}_j = 0, \forall i \neq j; i, j = 1, \ldots, d$$

The first discriminant vector $\boldsymbol{w}_1$ of GDA is also the fisher discriminant vector, which is the eigenvector corresponding to maximal eigenvalue of eigenfunction $\boldsymbol{S}_b^\phi\boldsymbol{w} = \lambda\boldsymbol{S}_w^\phi\boldsymbol{w}$ .If we know the first r discriminant vectors $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_r$, the $r+1$ discriminant vector $\boldsymbol{w}_{r+1}$ can be gotten through resolving the follow optimization problem.

$$\textbf{\textit{Model I}}: \begin{cases} \max(J_1(\boldsymbol{w})) \\ \boldsymbol{w}_j^T\boldsymbol{w} = 0, j = 1, \ldots, r \\ \boldsymbol{w} \in H \end{cases} \tag{5}$$

According to the theory of the reproducing kernel Hilbert space, the eigenvectors are linear combinations of H elements, so $\boldsymbol{w}$ can be expressed as

$$\boldsymbol{w} = \sum_{i=1}^{N}\alpha^i\phi(\boldsymbol{x}_i) = \boldsymbol{\phi}\boldsymbol{\alpha} \tag{6}$$

where $\boldsymbol{\phi} = (\phi(\boldsymbol{x}_1), \ldots, \phi(\boldsymbol{x}_N))$, $\boldsymbol{\alpha} = (\alpha^1, \ldots, \alpha^N)^T$, $\boldsymbol{\alpha}$ is optimal kernel discriminant vector, which can map the sample $\phi(\boldsymbol{x})$ in the feature space to the direction $\boldsymbol{w}$

$$\boldsymbol{w}^T\phi(\boldsymbol{x}) = \boldsymbol{w}^T\boldsymbol{\phi}^T\phi(\boldsymbol{x}) = \boldsymbol{\alpha}^T\xi_{\boldsymbol{x}} \tag{7}$$

where $\xi_{\boldsymbol{x}} = (k(\boldsymbol{x}_1, \boldsymbol{x}), k(\boldsymbol{x}_2, \boldsymbol{x}), \ldots, k(\boldsymbol{x}_N, \boldsymbol{x}))^T$ .For the sample $\boldsymbol{x} \in R^n$, $\xi_{\boldsymbol{x}}$ is the kernel sample vector which relates to $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N$, so the kernel matrix is

$$\boldsymbol{K} = (\xi_{\boldsymbol{x}_1}, \xi_{\boldsymbol{x}_2}, \ldots, \xi_{\boldsymbol{x}_N})$$

In the feature space H, the mean of each classes and the mean of all the samples can also be mapped to the direction $\boldsymbol{w}$

$$\boldsymbol{w}^T\boldsymbol{m}_i^\phi = \boldsymbol{\alpha}^T\boldsymbol{\phi}^T\frac{1}{N_i}\sum_{i=1}^{N_i}\phi(\boldsymbol{x}_k^i) = \boldsymbol{\alpha}^T\boldsymbol{\mu}_i \tag{8}$$

$$\boldsymbol{w}^T\boldsymbol{m}_0^\phi = \boldsymbol{\alpha}^T\boldsymbol{\phi}^T\frac{1}{N}\sum_{i=1}^{N_i}\phi(\boldsymbol{x}_k^i) = \boldsymbol{\alpha}^T\boldsymbol{\mu}_0 \tag{9}$$

where

$$\boldsymbol{\mu}_i = \left(\frac{1}{N_i}\sum_{k=1}^{N_i}\left(\phi(\boldsymbol{x}_1)\cdot\phi(\boldsymbol{x}_k^i)\right), \ldots, \frac{1}{N_i}\sum_{k=1}^{N_i}\left(\phi(\boldsymbol{x}_N)\cdot\phi(\boldsymbol{x}_k^i)\right)\right) \tag{10}$$

$$\boldsymbol{\mu}_0 = \left(\frac{1}{N}\sum_{k=1}^{N}\left(\phi(\boldsymbol{x}_1)\cdot\phi(\boldsymbol{x}_k^i)\right), \ldots, \frac{1}{N}\sum_{k=1}^{N}\left(\phi(\boldsymbol{x}_N)\cdot\phi(\boldsymbol{x}_k^i)\right)\right) \tag{11}$$

According to the Equation (8),(10)and (11),there are

$$\boldsymbol{w}^T\boldsymbol{S}_b^f\boldsymbol{w} = \boldsymbol{\alpha}^T\boldsymbol{K}_b\boldsymbol{\alpha} \tag{12}$$

$$\boldsymbol{w}^T\boldsymbol{S}_w^f\boldsymbol{w} = \boldsymbol{\alpha}^T\boldsymbol{K}_w\boldsymbol{\alpha} \tag{13}$$

$$\boldsymbol{w}^T \boldsymbol{S}_t^f \boldsymbol{w} = \boldsymbol{\alpha}^T \boldsymbol{K}_t \boldsymbol{\alpha} \tag{14}$$

where

$$\boldsymbol{K}_b = \sum_{i=1}^{c} \frac{N_i}{N} (\boldsymbol{\mu}_i - \boldsymbol{\mu}_0)(\boldsymbol{\mu}_i - \boldsymbol{\mu}_0)^T \tag{15}$$

$$\boldsymbol{K}_w = \frac{1}{N} \sum_{i=1}^{c} \sum_{j=1}^{N_i} \left(\xi_{\boldsymbol{x}_j^i} - \boldsymbol{\mu}_i\right)\left(\xi_{\boldsymbol{x}_j^i} - \boldsymbol{\mu}_i\right)^T \tag{16}$$

$$\boldsymbol{K}_t = \frac{1}{N} \sum_{j=1}^{N} \left(\xi_{\boldsymbol{x}_j} - \boldsymbol{\mu}_i\right)\left(\xi_{\boldsymbol{x}_j} - \boldsymbol{\mu}_i\right)^T \tag{17}$$

where $\boldsymbol{K}_b$ is the kernel interclasses scatter matrix, $\boldsymbol{K}_w$ is the kernel intraclasses scatter matrix, and $\boldsymbol{K}_t$ is the total scatter matrix. All of three matrixes are nonnegative matrixes, and their sizes are $N \times N$.

From Equation (12) and (13), Fisher discriminant function (4) can be expressed as

$$J_1'(\boldsymbol{\alpha}) = \frac{\boldsymbol{\alpha}^T \boldsymbol{K}_b \boldsymbol{\alpha}}{\boldsymbol{\alpha}^T \boldsymbol{K}_w \boldsymbol{\alpha}} \tag{18}$$

where $\boldsymbol{\alpha}$ is a nonzero vector. The orthogonal constraint condition can be expressed as

$$\boldsymbol{w}_i^T \boldsymbol{w}_j = \boldsymbol{\alpha}_i^T \boldsymbol{\phi}^T \boldsymbol{\phi}_j \boldsymbol{\alpha}_j = \boldsymbol{\alpha}_i^T \boldsymbol{K} \boldsymbol{\alpha}_j = 0, \forall i \neq j; i, j = 1, \ldots, d$$

So, the *Model I* can be expressed by kernel matrixes as

$$\textbf{\textit{Model II}}: \begin{cases} \max(J_1'(\boldsymbol{\alpha})) \\ \boldsymbol{\alpha}_j^T \boldsymbol{K} \boldsymbol{\alpha} = 0, j = 1, \ldots, r \\ \boldsymbol{\alpha} \in R^N \end{cases} \tag{19}$$

That is to say that, if we know the first r discriminant vectors $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_r$, the $r+1$ discriminant vector $\boldsymbol{\alpha}_{r+1}$ can be got through resolving the above optimization problem. $\boldsymbol{\alpha}_1$ is the eigenvector corresponding to the maximal eigenvalue of eigenfunction $\boldsymbol{K}_b \boldsymbol{\alpha} = \lambda \boldsymbol{K}_w \boldsymbol{\alpha}$. If $\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_d\}$ is from the *Model II* and $\{\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_d\}$ is from *Model I*, the relationship between them is

$$\boldsymbol{w}_i = \sum_{k=1}^{N} \alpha_i^k \boldsymbol{\phi}(\boldsymbol{x}_k) = \boldsymbol{\phi} \boldsymbol{\alpha}_i, \ i = 1, \ldots, d \tag{20}$$

where $\boldsymbol{\phi} = (\phi(\boldsymbol{x}_1), \ldots, \phi(\boldsymbol{x}_N))$.

In Baudat's literature (Baudat et al., 2000), instead of $J_1(\boldsymbol{w})$, they used $J_2(\boldsymbol{w})$

$$J_2(\boldsymbol{w}) = \frac{\boldsymbol{w}^{\mathrm{T}} \boldsymbol{S}_b^{\phi} \boldsymbol{w}}{\boldsymbol{w}^{\mathrm{T}} \boldsymbol{S}_t^{\phi} \boldsymbol{w}}$$

Correspondingly, the *Model I* of GDA can be rewritten as

$$\textbf{\textit{Model I}}: \begin{cases} \max(J_2(\boldsymbol{w})) \\ \boldsymbol{w}_j^T \boldsymbol{w} = 0, j = 1, \ldots, r \\ \boldsymbol{w} \in H \end{cases} \tag{21}$$

and the *Model II* of GDA can be rewritten as

$$\textbf{\textit{Model II}}: \begin{cases} \max(J_2'(\boldsymbol{\alpha})) \\ \boldsymbol{\alpha}_j^T \boldsymbol{K} \boldsymbol{\alpha} = 0, j = 1, \ldots, r \\ \boldsymbol{\alpha} \in R^N \end{cases} \tag{22}$$

For *Model I* with $J_2'(\boldsymbol{\alpha})$, if we have known the first r $(r \geq 1)$ discriminant vectors, the $\boldsymbol{\alpha}_{r+1}$ can be gotten by resolving the following eigenfunction.

$$\boldsymbol{\Gamma} \boldsymbol{K}_b \boldsymbol{\alpha}_{r+1} = \lambda \boldsymbol{K}_t \boldsymbol{\alpha}_{r+1} \tag{23}$$

where $\boldsymbol{\Gamma} = \boldsymbol{I} - \boldsymbol{K}\boldsymbol{\Lambda}^T (\boldsymbol{\Lambda}\boldsymbol{K}\boldsymbol{K}_t^{-1}\boldsymbol{K}\boldsymbol{\Lambda}^T)^{-1} \boldsymbol{\Lambda}\boldsymbol{K}\boldsymbol{K}_t^{-1}$, $\boldsymbol{I}$ is an identity matrix. $\boldsymbol{\Lambda} = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_r)^T$. Because $\boldsymbol{w}$ is an identity vector in *Model I*, $\boldsymbol{w}^T \boldsymbol{w} = \boldsymbol{\alpha}^T \boldsymbol{K} \boldsymbol{\alpha} = 1$. If $\boldsymbol{\alpha}$ has been known, $\boldsymbol{\alpha}$ should be standardized by dividing $\sqrt{\boldsymbol{\alpha}_i^T \boldsymbol{K} \boldsymbol{\alpha}_i}$.

In the feature space H, if a group of discriminant vectors $\{\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_d\}$ have been known, for the sample $\phi(\boldsymbol{x})$, its discriminant feature is

$$\boldsymbol{w}_i \phi(\boldsymbol{x}) = \sum_{k=1}^{N} \alpha_i^k \phi(\boldsymbol{x}_k) \phi(\boldsymbol{x}) = \sum_{k=1}^{N} \alpha_i^k k(\boldsymbol{x}_k, \boldsymbol{x}) = \boldsymbol{\alpha}_i^T \xi_{\boldsymbol{x}} \tag{24}$$

where $\xi_{\boldsymbol{x}}$ is kernel vector of the input sample $\boldsymbol{x}$.

The transformation function of GDA is

$$\begin{aligned} \boldsymbol{y} = \boldsymbol{W}^T \phi(\boldsymbol{x}) &= [\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_d]^T \phi(\boldsymbol{x}) \\ &= [\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_d]^T \xi_{\boldsymbol{x}} \end{aligned} \tag{25}$$

where $\boldsymbol{y}$ is the feature extracted by GDA which has d dimensions.

2.2 Kernel Function

Basing on the theory of kernel function, once a kernel function $k(\boldsymbol{x}, \boldsymbol{y})$ accords with Mercer theorem, then it corresponds to a inner product kernel function, mapping function and feature space in a certain space. In fact, to change kernel parameter is to implicitly change mapping function in order to change the complexity of distribution in sample sub-space. There are three kinds of kernel that are usually used.

(1) Dimensional polynomial kernel of degree d

$$k(\boldsymbol{x}, \boldsymbol{y}) = [(\boldsymbol{x} \cdot \boldsymbol{y}) + p]^d$$

where p and d are custom parameters. If $p = 0$ and $d = 1$, it will be called linear kernel function.

(2) Radial basis function (RBF) kernel

$$k(\boldsymbol{x}, \boldsymbol{y}) = \exp\left(-\frac{\|\boldsymbol{x} - \boldsymbol{y}\|^2}{\sigma^2}\right)$$

where $\sigma^2 > 0$.

(3) Neural Network kernel function

$$k(\boldsymbol{x}, \boldsymbol{y}) = \tanh(\mu(\boldsymbol{x} \cdot \boldsymbol{y}) + v)$$

where μ and v are parameters. Different from polynomial kernel and RBF kernel, the neural network kernel accords with the Mercer theorem only when (μ, v) are certain values.

2.3 Flow of Feature Extraction based on GDA

According to Baudat's literature (Baudat et al., 2000), we select $J_2(\boldsymbol{w})$ as the Fisher discriminant function, through the analysis above, the steps of feature extraction based on generalized discriminant are described as follows.

(1) Select the kernel function $k(\cdot, \cdot)$ and its parameters, and the amount d of the feature will be extracted.

(2) Calculate the kernel matrix $\boldsymbol{K}$, and calculate $\boldsymbol{K}_b$ and $\boldsymbol{K}_t$ according to the Equation (15) and (17).
(3) Resolve the Equation $\boldsymbol{K}_b\boldsymbol{\alpha} = \lambda\boldsymbol{K}_w\boldsymbol{\alpha}$ in order to get the eigenvector $\boldsymbol{\alpha}_1$ corresponding to the maximum eigenvalue.
(4) Get other discriminant vectors $\boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3, \mathrm{L}\ , \boldsymbol{\alpha}_d$ by Equation (22), and standardize them by dividing $\sqrt{\boldsymbol{\alpha}_i^T \boldsymbol{K}\boldsymbol{\alpha}_i}$.
(5) Extract the feature using Equation (25) for any input sample $\boldsymbol{x}$.

3. EXPERIMENT

In order to know whether the feature extraction based on GDA could improve the classification precision of hyperspectral image, we did two experiments. The experiments data are obtained by different remote sensors (AVIRIS and PHI). We also compared the GDA with other feature extraction methods, including Principal Component Analysis (PCA), Kernel PCA (KPA), and Linear Discriminant Analysis (LDA).

3.1 Experiment Flow

The steps of the experiments we have done are given below:

(1) Collect the samples of different ground types according the spectral library or the known ground cover information. And then, divide the samples into training samples and test samples.
(2) Using the training samples, calculate the transform matrixes of different feature extraction methods separately, including PCA, KPCA, LDA and GDA.
(3) From the transform matrixes which we got in Step 2 we extracted the feature of the hyperspectral images.
(4) Train the Minimum Distance Classifier (MDC) through training samples with feature extracted by Step 3. And then, evaluate the classification result of the testing samples.

3.2 Experiment 1

Experiment Data: The NASA AVIRIS (Airborne Visible/ Infrared Imaging Spectrometer) instrument acquired data over the Cuprites mine field, Nevada, USA. AVIRIS acquired data in 224 bands of 10 nm width with centre wavelengths from 400 - 2500 nm. The image of this data is shown in Figure 1. There are eight kinds of ores in this area; the samples of them are described in Table 1.

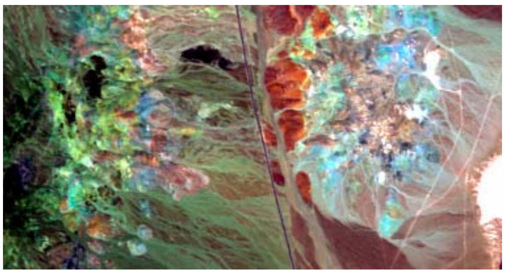


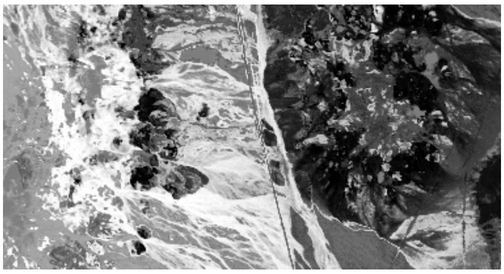
Figure 1. Hyperspectral image from AVIRIS (R:178,G:111,B:33)

Class Name	Samples Number
Alunite	604
Buddingtonite	89
Dickite	395
Kaolinite	290
Lite	762
Quartz	285
Salt	381
Tuff	1033

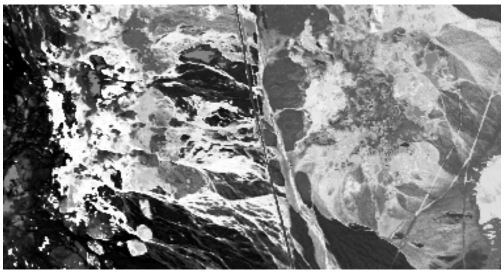
Table 1. Samples of this hyperspectral image

Atmospheric radiation correction based on ATREM has been applied to the AVIRIS image. After eliminating the bands which have too much noise and which are absorbed by the vapour, we used 190 bands in the experiment.

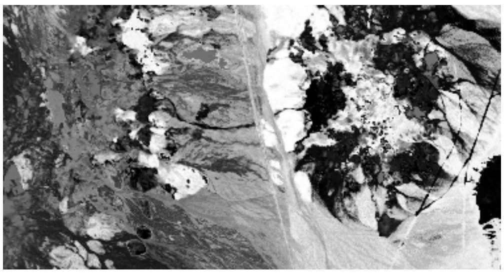
We selected 50 samples each class randomly as the training samples, and talked the others as testing samples. In the test, we selected the Poly kernel and RBF kernel for KPCA and GDA. The feature images extracted based on RBF-GDA is shown in Figure 2.



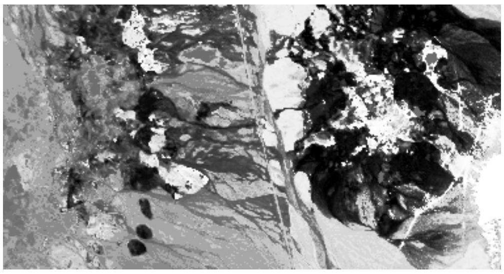
(1) Image of the first feature



(2) Image of the second feature



(3) Image of the third feature



(4) Image of the forth feature

Figure 2. The feature images extracted by RBF-GDA ($\sigma^2 = 10^7$)

We evaluated the classification precision with the testing samples, using the minimum distance classifier, and the result was shown in Table 2. The classification result with the feature extracted by RBF-GDA was shown in Figure 3.

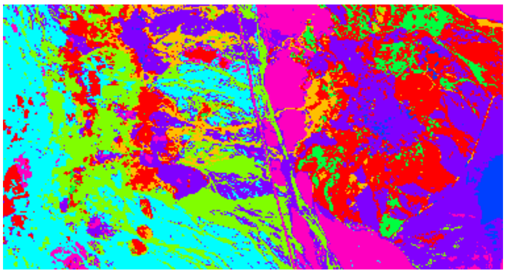


Figure 3. The classification result with feature extracted by RBF-GDA ($\sigma^2 = 10^7$)

Feature extracted Methods	Miss classification (%)
All bands	23.87
PCA	25.7
LDA	23.84
Ploy-KPCA $d = 1, p = 0$	40.1
RBF-KPCA $\sigma^2 = 10^7$	19.22
Ploy-GDA $d = 2, p = 0$	7.53
RBF-GDA $\sigma^2 = 10^7$	3.75
RBF-GDA $\sigma^2 = 10^8$	4.83

Table 2. The precision of classification with features extracted with different methods.

3.3 Experiment 2

Experiment Data: The PHI instrument, created in Shanghai Institute of Technology and Physics, acquired data over Changzhou, Jiangsu, China, (E119°22′11″, N31°41′44″). PHI acquires data in 80 bands width with centre wavelengths from 0.42–0.85μm, and the size of the image is 346 X 512.

Six kinds of objects exist in the image: (Colour-Class of the target-Amount of sample): 1-house-221, 2-water-222, 3-soil-205, 4-tree-228, 5-vegetation-266, 6-road-238, the results are visualized in figure 4.

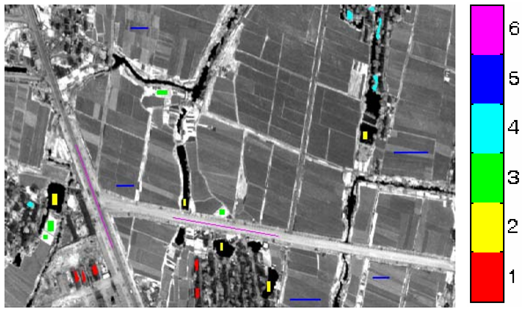
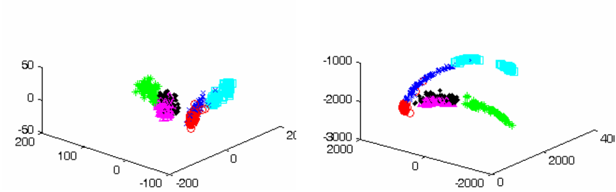
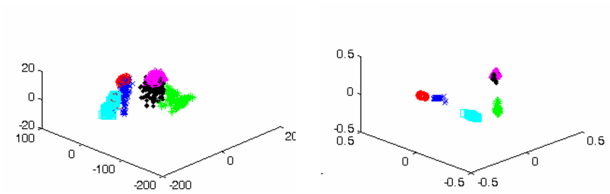


Figure 4. Samples distribution in this PHI image

We assigned the samples each class randomly as the training samples and testing samples equally. The feature was extracted by different feature extraction methods. In the feature space, the distribution of samples was shown in Figure 5.



(1) PCA (2) Ploy-KPCA ($d = 1, p = 0$)



(3) LDA (4) RBF-GDA($\sigma^2 = 10^3$)

Figure 5. Samples distribution in different feature space

We assigned the samples each class randomly as the training samples and testing samples equally. The feature was extracted by different feature extraction methods. In the feature space, the distribution of samples was shown in Figure 5.

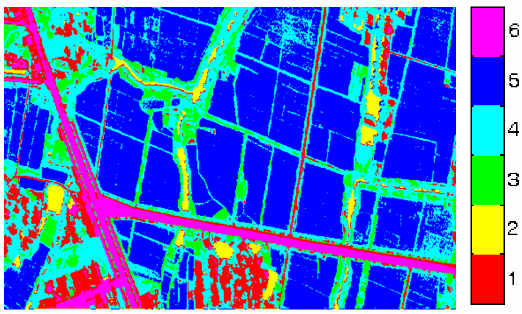


Figure 6. The classification result with feature extracted by RBF-GDA ($\sigma^2 = 10^3$)

Feature extracted Methods	Miss classification (%)
All bands	8.77
PCA	9.04
LDA	8.05
Ploy-KPCA $d = 1, p = 0$	11.84
RBF-KPCA $\sigma^2 = 10^7$	16.81
Ploy-GDA $d = 2, p = 0$	7.46
RBF-GDA $\sigma^2 = 10^7$	1.46
RBF-GDA $\sigma^2 = 10^8$	3.51

Table 3. The precision of classification with features extracted with different methods.

We evaluated the classification precision with the testing samples, using the minimum distance classifier. The classification result with the feature extracted by RBF-GDA was shown in Figure 6. The classification result of different feature extraction methods is shown in Table 3.

4. CONCLUSION

Through the experiments of feature extraction with AVIRIS and PHI images we made some conclusions.

The PCA is to find project directions, which can make the samples variance maximized. The KPCA, using the kernel function, can realize the information compression to a great extent, but it is not good for classification.

When the kernel function and its parameters are correctly selected, in the feature space extracted by GDA, the samples of the same class are near with each other; the samples of the different classes are far away. The GDA is a feature extracting method which is more suitable to classification than the LDA.

When the kernel function and its parameters are correctly selected, the classification precision is much better with the features extracted by GDA, than the features extracted by other methods. How to select the kernel function and find suitable parameter is our further research.

REFERENCES

Tong, Q.X., Zhang, B., Zheng, L.F., 2006. *Hyperspectral Remote Sensing*. Higher Education Press, Beijing, pp. 363-470.

Hoffbeck J.P., Landgrebe D.A., 1996, *Classification of Remote Sensing Images having High Spectral Resolution*. Remote Sensing of Environment, 57(3), pp.119-126.

Zhang, L.P., 2003. *Study of Feature Extraction and Classification of Hyperspectral Remote Sensing Image Based on Projection Pursuit and Nonlinear Principle Curves*. Shandong University of Science and Technology, pp.1-8.

John S.T., Nello C., 2004. *Kernel Methods for Pattern Analysis*. Cambridge University Press. Cambridge, pp.1-24.

Lu, W., 2005. *A Research of Feature Extraction and Classification Techniques for Target Detection in Hyperspectral image*. Information engineering University. Zhengzhou, pp.121-137.

Baudat, G., Anouar, F. 2000.*Generalized discriminant analysis using a kernel approach*. Neural Computation, 12(10), pp. 2385-2404.

Gao X.M., Yang, J.Y., Jin, Z., 2004, *Kernel-Based Foley-Sammon Discriminant Analysis and Face Recognition*. Journal of Computer-Aided Design & Computer Graphics, pp. 962-967.

Li W.H., 2003, Mechanical Fault Feature Extraction and Classification Based on Kernel Methods. Huazhong University of Science & Technology, pp.30-40.