International Journal of Database and Theory and Application
Vol.9, No.2 (2016), pp.183-192
http://dx.doi.org/10.14257/ijdta.2016.9.2.20

Hyperspectral Image Classification by Fusion of Multiple Classifiers

Yanbin Peng[1], Zhigang Pan[1], Zhijun Zheng[1] ,Xiaoyong Li[1]

[1] *School of Information and Electronic Engineering, Zhejiang University of Science and Technology, Hangzhou 310023, China*
E-mail: pyb2010@126.com

Abstract

Hyperspectral image mostly have very large amounts of data which makes the computational cost and subsequent classification task a difficult issue. Firstly, to solve the problem of computational complexity, spectral clustering algorithm is imported to select efficient bands for subsequent classification task. Secondly, due to lack of labeled training sample points, this paper proposes a new algorithm that combines support vector machines and Bayesian classifier to create a discriminative/generative hyperspectral image classification method using the selected features. Experimental results on real hyperspectral image show that the proposed method has better performance than the other state-of-the-art methods.

Keywords*: hyperspectral image, band selection, classification, support vector machine, Bayesian.*

1. Introduction

Hyperspectral sensors record reflection and emittance information in hundreds of narrowly spaced spectral bands with wavelengths ranging from the visible spectrum to the infrared region. In a hyperspectral image, each pixel is represented by a feature vector whose entries correspond to reflection of various bands. The obtained three-dimensional image cube contains large amounts of discriminative information for hyperspectral image classification.

Nowadays, a considerable amount of research has been done on hyperspectral image classification using machine learning algorithms during the past decade. Melgani [1] propose a theoretical discussion and experimental analysis aiming at understanding and assessing the potentialities of SVM classifier in hyper-dimensional feature spaces, and studies the potentially critical issue of applying binary SVMs to multi-class problems in hyperspectral data. Chen [2] sparsely represents a test sample in terms of all of the training samples in a feature space induced by a kernel function. The recovered sparse representation vector is then used directly to determine the class label of the test pixel. Li [3] constructs a new family of generalized composite kernels which exhibit great flexibility when combining the spectral and the spatial information contained in hyperspectral data. And then propose a multinomial logistic regression classifier for hyperspectral image classification. Liu [4] present a post processing algorithm for a kernel sparse representation based hyperspectral image classifier, which is based on the integration of spatial and spectral information. Qian [5] propose a hyperspectral feature extraction and pixel classification method based on structured sparse logistic regression and three-dimensional discrete wavelet transform texture features. Chang [6] propose a nearest feature line embedding transformation for the dimension reduction of a hyperspectral image. Three factors, including class separability, neighborhood structure preservation, and feature line embedding measurement, are considered simultaneously to determine an effective and discriminating transformation in the Eigen spaces for land

ISSN: 1975-0080 IJDTA

cover classification. Ye [7] propose a fusion-classification system to alleviate ill-conditioned distributions in hyperspectral image classification. This method combines a windowed 3-D discrete wavelet transform with a feature grouping algorithm to extract and select spectral-spatial features from the hyperspectral image dataset, and then, employs a multi classifier decision-fusion approach for the final classification.

However, there are two major problems in the process of hyperspectral image classification and recognition. On the one hand, the computational cost is too high in classifying pixels when all bands are used. Due to high correlation between spectral bands, band selection algorithm should be imported to remove the redundant bands which do not contribute to the classification task. According to non-linear relation between band data of hyperspectral image, spectral clustering can be introduced to cluster and select effective bands for classification. In spectral clustering based band selection algorithm, neighbor graph and similarity matrix are generated from histogram of band image, then spectral bands are divided in to k clusters by spectral clustering algorithm. At last, k selected representative bands are generated for subsequent classification and recognition task. On the other hand, the labeling cost in hyperspectral image is too high, resulting in the lack of labeled training sample points. Therefore, new classification algorithm should be proposed to improve the classification accuracy under the environment of small sample set.

Therefore, based our former research works [8-11], this paper propose a new framework for hyperspectral image classification, firstly, spectral clustering is used to select effective bands for the subsequent classification task. Secondly, a new algorithm is proposed which combines support vector machines and Bayesian classifier to create a discriminative/generative hyperspectral image classification method using the selected features. Experimental results on real hyperspectral image show that the proposed method has better performance than the other state-of-the-art methods.

2. Band Selection Algorithm

As displayed in Figure 1, hyperspectral image is a three dimensional image array with I, J, and K dimension in sequence. Wherein, I and J dimension correspond to width and length spatial dimensions. Dimension K corresponds to spectral dimension. H stands for the whole image array in which every band H_k is an image matrix. Therefore, each band in image cube can be regarded as a data point with $I \times J$ dimensions. In spectral clustering based band selection algorithm, bands with high correlations are grouped into a cluster, after that, a representative band is selected from each cluster. At last, the selected bands perform the subsequent classification task on behalf of all bands.

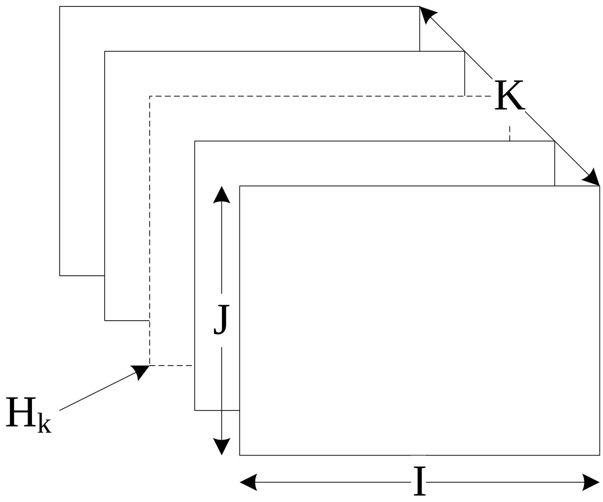

Figure 1. Hyperspectral Image Cube H

Spectral clustering starts with the construction of a similarity matrix $W=\{w_{ij}\}$, in which the component $w_{ij} \in R^{L\times L}$ measures how well band i is similar to band j. There are several measurements which can be used to build the similarities between band images, such as Mahalanobis distance, Euclidean distance, mutual information based distance, etc. However, those similarity measures are not suitable for directly used in spectral clustering algorithm. Therefore, KL divergence based measure [12, 13] is used. We assume that $q_i(c)$ and $q_j(c)$ are the probability distributions of image of spectral band i and j respectively. The KL divergence of between $q_i(c)$ and $q_j(c)$ is defined as: $KL(i,j)= \Sigma\ q_i(c)log(q_i(c)/\ q_j(c))+ \Sigma\ q_j(c)log(q_j(c)/\ q_i(c))$. However, divergence describes the difference between two bands. Therefore, similarity between $q_i(c)$ and $q_j(c)$ is defined as: w_{ij} =(1- $KL(i,j)/2$). Obviously, when $i=j$, w_{ij}=1, it means that the same band has the largest similarity. Based on the former definition, spectral clustering based band selection algorithm is as follows [14, 15]:

Algorithm 1: spectral clustering based band selection algorithm

Input: bands vector $H_1,H_2,\ldots,H_k$.

1) Compute histogram (probability distribution): $q_1(c)$, $q_2(c),\ldots,q_k(c)$ for all bands.

2) Compute similarity between bands using KL divergence formula, obtaining similarity matrix W.

3) Compute the unnormalized Laplacian matrix L of W.

4) Compute the first m eigenvectors $z_1,z_2,\ldots z_m$ of L.

5) Let $Z\in R^{n\times m}$ be the matrix containing the vectors $z_1,z_2,\ldots z_m$ as colums.

6) For i=1,2,…,n, let $x_i\in R^m$ be the vector corresponding to the i-th row of Z.

7) Cluster the points $x_1,x_2,\ldots,x_n$ with the k-means algorithm into m clusters: $B_1,B_2,\ldots B_m$.

8) select a band from every cluster randomly, obtaining m representative bands.

Output: m representative bands

3. Fusion of SVM and Bayesian

3.1. Hyperspectral Image Classifier based on SVM

SVM was found in 1997 by Marshall Reavis which combines linear modeling and instance based learning [16]. SVM model selects a small number of support vectors (instances in the boundary) from every class and builds a linear discriminative function which separates the training set as far as possible. In the case of nonlinear separable, kernel function (such as RBF kernel) is used to project the training instances into a high dimensional space where the training instances become linear separable. In the environment of small training sample sets, SVM has better prediction precision against other classifiers in solving classification and regression problem. In this paper, the labeled training sample set is $Samples=\{(O_1,p_1), (O_2,p_2),\ldots, (O_l,p_l)\}$, $O_i \in R^m$, $p_i \in \{+1, -1\}$ is the class label, *+1* stands that sample point belongs to the class, *-1* stands that sample point doesn't belongs to the class. SVM algorithm finds classification hyper plane which maximize classification margin based on training sample set. Then, SVM use the classification hyper plane to classify new sample points. Specific algorithm is as follows:

1) Assumes that we have training sample set: $Samples=\{(O_1,p_1), (O_2,p_2),\ldots, (O_l,p_l)\}$, wherein $O_i \in R^n$, $p_i \in \{+1, -1\}$, $i=1,2,\ldots l;$

2) Selects kernel function $K(O, O)$ and penalty parameter C, constructs and solves Optimization problem: $\min_\alpha \quad \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} p_i p_j \alpha_i \alpha_j K(O_i,O_j) - \sum_{j=1}^{l}\alpha_j$,

$$s.t. \quad \sum_{i=1}^{l} p_i\alpha_i = 0,$$

$$0 \le \alpha_i \le C, \ i = 1,\ldots,l$$

Gets the optimal solution: $\alpha^* = (\alpha_1^*,\ldots,\alpha_l^*)^T$;

3) Selects a positive component α_j^* from α^* which is less than C, according to α_j^* we calculate $b^* = p_j - \sum_{i=1}^{l} p_i\alpha_i^* K(O_i,O_j)$;

4) decision function is: $f(O) = \operatorname{sgn}\{\sum_{i=1}^{l} p_i\alpha_i^* K(O_i,O) + b^*\}$

In optimal solution α^*, the corresponding samples of nonzero components are support vectors. Only support vectors can influence decision function. Therefore, decision function can be rewritten as follows:

$$f(O) = \operatorname{sgn}\{\sum_{O_i \in SV} \alpha_i^* p_i K(O_i,O) + b^*\} = \operatorname{sgn}\{\sum_{j}\sum_{O_i \in SV_j} \alpha_i^* p_i K(O_i,O) + b^*\}$$

wherein, SV is the set of support vectors; SV_j is the set of the support vectors in class j, $j \in \{+1, -1\}$; O_i is support vector; p_i is the corresponding class label of sample O_i; α_i^* is the correspongding weight of sample O_i; b is bias term; $K()$ is kernel function. In the case of nonlinear separable, feature space is mapped into a high dimensional space through kernel function, while the training samples in high dimensional space are linear separable.

The support vector machine is essentially a two-class classifier. In our framework, we face a multiclass problem. Various algorithms have been proposed for combining multiple two-class SVMs to build a multiclass classifier. One commonly used method is to construct M separate SVMs, in which the M_{th} classification function $f_m(O)$ is trained using the data from class C_m as the positive training samples and the data from the other M-1 classes as the negative training samples. However, this method lead to the predicament that one pixel will be assigned to multiple classes simultaneously. This is solved by making prediction for a new sample O using the following decision functions: $f(O) = \max_m f_m(O)$.

3.2. Hyperspectral Image Classification based on Bayesian

In machine learning, Bayesian classifiers are a family of generative classifiers based on applying Bayes theorem with strong independence assumptions between the features. Bayesian classifier uses Bayesian formula to minimize the error rate of classification rules through calculating the Maximum a posteriori (MAP for short). Bayesian classifiers are highly scalable, requiring a number of linear parameters in a learning problem. Maximum-likelihood training can be done by calculating a closed-form expression, which takes linear time, rather than by expensive iterative approximation as used for many other classifiers. The Bayesian classification function of hyperspectral image is as follows:

$$f(O) = \arg\max_{j\in\{+1,-1\}}\{P(O\,|\,j)P(j)\}$$

$$P(O\,|\,j) = \sum_{i=1}^{num_j} P(O\,|\,i,j)P(i\,|\,j) = \sum_{i=1}^{num_j} N(O\,|\,\mu_{ji},\Sigma_{ji})P(i\,|\,j)$$

Wherein, $P(j)$ is prior probability, we use the proportion of all kinds of samples points to approximate prior probability, $P(j) = |sample_j| / |sample|$, $sample_j$ is sample set of class j, $sample$ is the whole sample set. $P(O\,|\,j)$ is likelihood function. $f(O)$ is maximum a posteriori, which represents the most possible class in the current sample value *O*. the jth class is expressed as a Gaussian mixture distribution with num_j clusters. $P(i\,|\,j)$ is the weight of ith cluster in jth class. $N(O\,|\,\mu_{ji},\Sigma_{ji})$ is the distribution of ith cluster, wherein μ_{ji} is mean value and $\sum_{ji}$ covariance matrix. The parameters of Gaussian mixture distribution ($P(i\,|\,j),\mu_{ji},\Sigma_{ji}$) are calculated by Expectation Maximization (EM for short). In EM algorithm, a new constraint condition $\sum_{ji} = \sigma I$ is added, wherein, σ is variance, I is unit matrix. When sample points is not uniform, we can assure $\sum_{ji} = \sigma I$ through down sampling and up sampling. The reason why we add this condition is to promote classifier fusion in the next section.

3.3. Classifier Fusion

As mentioned above, SVM is a discriminative classifier, and Bayesian is a generative classifier. Discriminative classifier only needs a small amount of training sample points, while suffer from a lack of robustness with respect to noise and over fitting. Generative classifier is not sensitive to noise, but needs a large amount of training sample points to reach a good classification precision. This paper proposes the fusion of two kinds of classifier, combining their advantages, implemented a classifier which is not sensitive to noise while need less training sample points.

The decision function of SVM classifier with RBF kernel function is as follows:

$$f(O) = \arg\max_{j\in\{+1,-1\}}\{\sum_{O_i\in SV_j}\alpha_i^* p_i K(O_i,O)+b^*\}$$
$$=\arg\max_{j\in\{+1,-1\}}\{\sum_{O_i\in SV_j}\alpha_i^* p_i \exp(-\gamma\|O-O_i\|^2)\} \quad (1)$$

Wherein, SV_j is support vector set of jth class, α_i is corresponding weight of support vector, γ is parameter of kernel function, b^* is support vector O_0 with weight α_0^*.

The decision function of Bayesian classifier is as follows:

$$f(O) = \arg\max_{j\in\{+1,-1\}}\{\sum_{i=1}^{num_j} P(j)P(i\mid j)\frac{1}{(2\pi\sigma^2)^{n/2}}\exp(-\frac{\|O-\mu_{ji}\|^2}{2\sigma^2})\} \quad (2)$$

Compare formula (1) and (2), can be found that as long as we make:

$$\alpha_i^* p_i = P(j)P(i\mid j)\frac{1}{(2\pi\sigma^2)^{n/2}} \quad (3)$$

$$\gamma = \frac{1}{2\sigma^2} \quad (4)$$

$$O_i = \mu_{ji}, \text{ for all } O_i \in SV_j \quad (5)$$

Then, the two formulas have uniform expression. Thus, the two kinds of classifier can be mutual conversion. The mean value of cluster in Bayesian corresponding to support vector in SVM; the variance of cluster in Bayesian can be used to calculate parameter γ in RBF kernel related to support vector; $P(j)P(i\mid j)\frac{1}{(2\pi\sigma^2)^{n/2}}$ can be used to calculate weight related to support vector. Seen from formula (1), SVM classifier is defined with the set of support vector ($O_{S1},O_{S2},\ldots,O_{Sn}$), the weight set of support vector ($\alpha_{S1}, \alpha_{S2},\ldots, \alpha_{Sn}$), and parameter set ($\gamma_{S1}, \gamma_{S2},\ldots, \gamma_{Sn}$), therefore can be represented as:

SVM = $\{(O_{S1},O_{S2},\ldots,O_{Sn}),(\alpha_{S1}, \alpha_{S2},\ldots, \alpha_{Sn}),(\gamma_{S1}, \gamma_{S2},\ldots, \gamma_{Sn})\}$

According to the correspondence between Bayesian and SVM classifier, Bayesian classifier can be expressed as the style of SVM classifier as follows:

Bayes = $\{(O_{B1},O_{B2},\ldots,O_{Bm}),(\alpha_{B1}, \alpha_{B2},\ldots, \alpha_{Bm}),(\gamma_{B1}, \gamma_{B2},\ldots, \gamma_{Bm})\}$

Given the scale factor $\eta \in [0,1]$, the fused classifier is as follows:

SVM/Bayes=$\{(O_{B1},O_{B2},\ldots,O_{Bm},O_{S1},O_{S2},\ldots,O_{Sn}),(\eta\alpha_{B1}, \eta\alpha_{B2},\ldots, \eta\alpha_{Bm},(1-\eta)\alpha_{S1}, (1-\eta)\alpha_{S2},\ldots, (1-\eta)\alpha_{Sn}),(\gamma_{B1}, \gamma_{B2},\ldots, \gamma_{Bm}, \gamma_{S1}, \gamma_{S2},\ldots, \gamma_{Sn})\}$

$\eta=0$ and $\eta=1$ Correspond to simple SVM and Bayesian classifier respectively. The fused classifier converts middle point of cluster in the Bayesian into support vector in SVM. In the new set of support vectors ($O_{B1},O_{B2},\ldots,O_{Bm},O_{S1},O_{S2},\ldots,O_{Sn}$), predicting the class label for new pixels using SVM prediction algorithm. As described in figure 2, solid circular points and hollow circular points represent two classes of samples. HP is separating hyper plane. The points in HP1 and HP2 are support vectors generated by SVM training algorithm. In the process of predicting, SVM only consider support vector, which declines the predicting precision when support vector is noised. Bayesian consider every class as mixture Gaussian distribution, and calculating mean value and covariance of every Gaussian distribution using EM algorithm. The resulting Gaussian distribution is shown as ellipse in figure 2. The fusion of classifiers is to expand support vector set with mean value of Gaussian distribution. The expanded support vector set is then used in predicting the class of new pixel, which will improve classification performance.

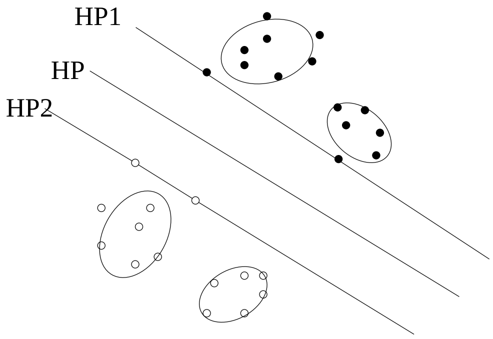

Figure 2. The Schematic Diagram of Classifier Fusion

4. Experiments and Results

Having presented the method of fusing SVM and Bayesian classifier (SB for short) in the previous sections, we now demonstrate the effect of our new method through several comparative experiments. These experiments are done in a real-world hyperspectral image data set which is a section of the scene taken over Washington DC mall by the sensor of HYDICE (hyperspectral digital imagery collection experiment). This data set has 500×307pixels, 210 spectral bands and seven classes composed of water, vegetation, manmade structures and shadow etc. Figure 3 shows the 60th band image of Washington DC mall data set.

In order to assess the performance of the new method proposed in this paper, we choose two models for comparison: 1) ID algorithm proposed in literature [17] with SVM classifier; 2) MVPCA algorithm proposed in literature [18] with Bayesian classifier.

Figure 3. Image of 60th Band in Washington DC Mall Data Set

4.1. The Comparison of Average Classification Precision in Noiseless Environment

The experiment result is as follows, wherein, the weight of SVM and Bayesian is 0.5.

Table 1. Comparison Table of Average Classification Precision in Noiseless Environment

Number of training samples	Average classification precision		
	ID+SVM	MVPCA+Bayesian	SB
20	0.721	0.591	0.736
40	0.757	0.662	0.843
60	0.779	0.731	0.889
80	0.792	0.780	0.912
100	0.796	0.825	0.924

It can be seen from the above experimental data, as the number of training samples increase, the average classification precision of all algorithms are gradually increasing. This is because that the increase of the sample points improves the classification accuracy of classifiers. Wherein, ID+SVM algorithm achieves good classification precision (0.721) under the condition of small sample set (20 sample points), it is because of the SVM is a discriminative classifier which has good performance in small sample set. With the increase number of sample points, the improving effect of ID+SVM algorithm is not obvious. This is because the new sample points are not taken as support vectors, where only support vector can influence the final classification. MVPCA+Bayesian algorithm have bad classification precision (0.591) under the condition of small sample set (20 sample points), it is because of the Bayesian is a generative classifier. Generative classifier classifies samples according to the statistical law of sample points. Small sample set cannot reflect statistical law of sample points. Thus this algorithm has low classification precision. With the number of sample points increase, the classification performance of Bayesian improves, which in turn increase the classification precision of pixels. Due to the fusion of SVM and Bayesian classifier, SB algorithm has merits of these two classifiers. Therefore, in the small sample set (20 sample points), SB achieves good classification precision (0.736). In general, the effect of SB algorithm is higher than the other two algorithms.

4.2. The Comparison of Average Classification Precision in Noise Environment

In order to validate the robustness of new method in the noise environment, we randomly choose 5 percent of sample points, and change their label. The repeated experimental result is as follows:

Table 2. Comparison Table of Average Classification Precision in Noise Environment

Number of training samples	Average classification precision		
	ID+SVM	MVPCA+Bayesian	SB
20	0.506	0.561	0.696
40	0.621	0.629	0.789
60	0.698	0.708	0.856
80	0.748	0.764	0.897
100	0.763	0.812	0.919

Seen from above experimental data, in small sample set (20 sample points), classification precision of ID+SVM algorithm fell sharply. It is because discriminative classifier is sensitive to noise. The sample set is too small, where noise points have great influence to the result. As the number of sample point increase, the influence of noise points to the result decrease, but the classification precision is still lower than the noiseless environment. MVPCA+Bayesian algorithm has relatively small influence by noise point as to that of the ID+SVM algorithm. The classification precision is a little lower than that of the noiseless environment. This is because Bayesian is a generative classifier, which classifies samples according statistical parameters, and is not sensitive to noise points. SB algorithm has merits of these two classifiers, showing good performance in small sample set and noise environment.

5. Conclusions

In this paper, we present an approach of fusing SVM and Bayesian classifier for hyperspectral data classification. Firstly, according to non-linear relation between band data of hyperspectral image, spectral clustering is introduced to cluster and select effective bands for subsequent classification task. Secondly, Aiming at solving the problem of small training sample set, a new algorithm is proposed which combines support vector machines and Bayesian classifier to create a discriminative/generative hyperspectral image classification method using the selected features. Experimental results on real hyperspectral image show that the proposed method has better performance than the other state-of-the-art methods.

Acknowledgements

Project supported by the Zhejiang Provincial Natural Science Foundation of China (No. LQ13F020015).

References

[1] F. Melgani and L. Bruzzone, “Classification of hyper-spectral remote sensing images with support vector machines”, IEEE Transactions on Geoscience and Remote Sensing, vol. 42, no. 8, **(2004)**, pp. 1778-1790.
[2] Y. Chen, N. M. Nasrabadi and T. D. Tran, “Hyper-spectral image classification via kernel sparse representation”, IEEE Transactions on Geoscience and Remote Sensing, vol. 51, no. 1, **(2013)**, pp. 217-231.
[3] J. Li, P. R. Marpu and A. Plaza, “Generalized composite kernel framework for hyper-spectral image classification”, IEEE Transactions on Geoscience and Remote Sensing, vol. 51, no. 9, **(2013)**, pp. 4816-4829.
[4] J. Liu, Z. Wu and L. Sun, “Hyper-spectral Image Classification Using Kernel Sparse Representation and Semi-local Spatial Graph Regularization”, Geoscience and Remote Sensing Letters, vol. 11, no. 8, **(2014)**, pp. 1320 – 1324.
[5] Y. Qian, M. Ye and J. Zhou, “Hyper-spectral image classification based on structured sparse logistic regression and three-dimensional wavelet texture features”, IEEE Transactions on Geoscience and Remote Sensing, vol. 51, no. 4, **(2013)**, pp. 2276-2291.
[6] Y. L. Chang, J. N. Liu and C. C. Han, “Hyper-spectral image classification using nearest feature line embedding approach”, IEEE Transactions on Geoscience and Remote Sensing, vol. 52, no. 1, **(2014)**, pp. 278-287.
[7] Z. Ye, S. Prasad and W. Li, “Classification based on 3-D DWT and decision fusion for hyper-spectral image analysis”, Geoscience and Remote Sensing Letters, vol. 11, no. 1, **(2014)**, pp. 173-177.
[8] Z. Zheng and Y. Peng, “Semi-Supervised Based Hyper-spectral Imagery Classification”, Sensors and Transducers, vol. 156, no. 9, **(2013)**, pp. 298-303.
[9] P. Y. Bin and A. I. J. Qing, “Hyper-spectral Imagery Classification Based on Spectral Clustering Band Selection”, Opto-Electronic Engineering, vol. 39, no. 2, **(2012)**, pp.63-67.
[10] J. Li, S. Jia and Y. Peng, “Hyper-spectral Data Classification with Spectral and Texture Features by Co-training Algorithm”, Opto-Electronic Engineering, vol. 39, no. 11, **(2012)**, pp.88-94.
[11] Y. Peng, Z. Zheng and C. Yu, “Automated negotiation decision model based on classifier fusion”, Journal of Shanghai Jiaotong University, vol. 47, no. 4, **(2013)**, pp. 644-649.
[12] J. R. Hershey and P. A. Olsen, “Approximating the Kullback Leibler Divergence Between Gaussian Mixture Models”, ICASSP, no. 4, **(2007)**, pp. 317-320.
[13] R. Neher and A. Srivastava, “A Bayesian MRF framework for labeling terrain using hyper-spectral imaging”, IEEE Transactions on Geoscience and Remote Sensing, vol. 43, no. 6, **(2005)**, pp.1363-1374.

[14] A. Y. Ng, M. I. Jordan and Y. Weiss, “On spectral clustering: Analysis and an algorithm”, Advances in neural information processing systems, **(2002)**, pp. 849-856.
[15] U. V. Luxburg, “A tutorial on spectral clustering”, Statistics and computing, vol. 17, no. 4, **(2007)**, pp.395-416.
[16] C. M. Bishop, “Pattern recognition and machine learning”, New York: springer, **(2006)**.
[17] C. I. Chang and S. Wang, “Constrained band selection for hyper-spectral imagery”, IEEE Transactions on Geoscience and Remote Sensing, vol. 44, no. 6, **(2006)**, pp. 1575-1585.
[18] C. I. Chang, Q. Du and T. L. Sun, “A joint band prioritization and band-decor-relation approach to band selection for hyper-spectral image classification”, IEEE Transactions on Geoscience and Remote Sensing, vol. 37, no. 6, **(1999)**, pp. 2631-2641.