Hyperspectral Image Classification based on Co-training

Zhijun Zheng[1] and Yanbin Peng[1]

[1]School of Information and Electronic Engineering, Zhejiang University of
Science and Technology, Hangzhou 310023, China
zjzheng9999@163.com

Abstract

The abundant information available in hyperspectral image has provided important opportunities for land-cover classification and recognition. However, "Curse of dimensionality" and small training sample set are two difficulties which hinder the improvement of computational efficiency and classification precision. In this paper, we present a co-training based method on hyperspectral image classification. Firstly, two views of samples are generated through two kinds of dimensionality reduction methods. After that, the co-training process is viewed as combinative label propagation over two independent views. 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 classification, dimensionality reduction, Feature selection, Feature abstraction, co-training

1. Introduction

Hyperspectral sensors simultaneously collect hundreds of continuous spectral bands with the resolution of the nanoscale. Such a large number of spectral bands contain rich information, providing significant challenges to hyperspectral image classification. In general, the dimensionality of samples strongly affects the performance of many classification methods. For hyperspectral image, on the one hand, there are redundant bands, and some bands maybe has weak discriminative ability. On the other hand, as the number of dimension improved, the number of training sample points needed increase exponentially, which is so called "Curse of dimensionality" problem. In the meanwhile, due to the cost of labeling sample points, the number of training sample points is few. It is therefore advantageous to decrease the dimension of hyperspectral data and import more efficient classification method.

Nowadays, a considerable amount of research has been done on hyperspectral image classification using machine learning algorithms during the past decade. Chen [1] propose a novel dictionary learning method for hyperspectral image classification. Its proposed method, linear regression fisher discrimination dictionary learning, obtains a more discriminative dictionary and a classifier by incorporating linear regression term and the fisher discrimination into the objective function during training, which improves the performances of classification. Li [2] 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. Ji [3] propose a new method to address both the pixel spectral and spatial constraints, in which the relationship among pixels is formulated in a hypergraph structure. A spatial-based hyperedge is generated to model the layout among pixels. Both the learning on the combinational hypergraph

is conducted by jointly investigating the image feature and the spatial layout of pixels to seek their joint optimal partitions. 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. Huang [5] propose a new sparse manifold learning method, called sparse manifold preserving for hyperspectral image classification. This method constructs the affinity weight using the sparse coefficients which reserves the global sparsity and manifold structure of hyperspectral image data, which it doesn't need to choose any model parameters for the similarity graph. Qian [6] propose a hyperspectral feature extraction and pixel classification method based on structured sparse logistic regression and three-dimensional discrete wavelet transform texture features. However, there are still some spaces to improve precision of hyperspectral image classification.

In this paper, based on our former research works [7-10], we present a co-training based method on hyperspectral image classification. Firstly, two views of samples are generated through two kinds of dimensionality reduction methods. The research on the dimensionality reduction can be divided into two main directions. One is feature selection method, which seeks a subset of original feature space which contains most of the discriminative abilities. This paper adopts spectral clustering [11-16] based band selection method. The other is feature abstraction method, which transforms the original sample points into the destination feature space which has less number of features while keep the discriminative abilities. This paper adopts linear discriminant analysis (LDA for short) and manifold learning based feature abstraction method. After that, the co-training process is viewed as combinative label propagation over two independent views. Experimental results on real hyperspectral image show that the proposed method has better performance than the other state-of-the-art methods.

2. Dimensionality Reduction Methods

The whole hyperspectral image cube H is shown in Figure 1. I, J and K corresponds to three dimensions of the data cube. I and J stand for width and length dimension and K stands for the spectral dimension. One band image is expressed as H_k, which is a data matrix with $I \times J$ dimensions. One sample point is a pixel vector x_i with K dimensions.
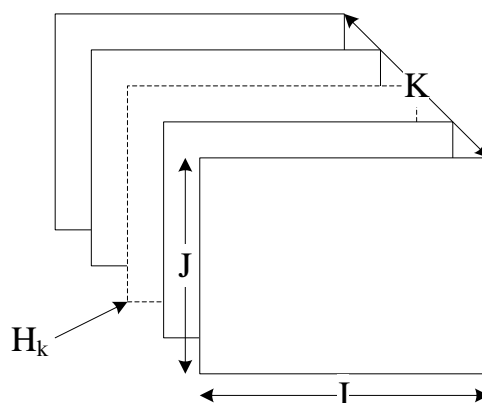
Figure 1. Hyperspectral Image Cube

The development of image acquisition technology has made it possible to capture hyperspectral data in hundreds of bands. The high dimensionality of hyperspectral

image increases the precision to classify and recognize the land-cover materials. However, the number of labeled sample points is few due to high labeling cost. At the same time, high dimensionality of hyperspectral image improves computational complexity. What's worse is there are redundancy bands which may decrease classification precision. Therefore, it is an important issue to reduce the dimensionality without degrading the classification precision. Dimensionality reduction can reduce computational complexity and improve classification precision. There are two major kinds of dimensionality reduction methods: feature selection and feature abstraction. The former method seeks a subset of original feature space which contains most of the discriminative abilities. This paper adopts spectral clustering based band selection method. The latter transforms the original sample points into the destination feature space which has less number of features while keep the discriminative abilities. This paper adopts LDA and manifold learning based feature abstraction method. These two methods are detailed in the following two subsections.

2.1. First View: Spectral Clustering based Band Selection Method

Spectral clustering based band selection method is an unsupervised feature selection technique. In practice, unsupervised band selection technique can be considered as a clustering problem, which is a process of grouping a set of objects in such a way that objects in the same group are more similar to each other than to those in other groups. In clustering based band selection method, each band is regarded as an object. The similarity between two bands is measured using a kind of measurement, based on which the bands are grouped into several clusters. The dimensionality reduction can be done by selecting a representative band from every band clusters. This paper adopts mutual information to measure similarity between bands, and spectral clustering algorithm to cluster band data.

Suppose we have similarity matrix $W = \{w_{i,j}\}$ of band data, wherein, $w_{i,j} \in R^{L \times L}$ stands for similarity degree between band i and band j. This paper adopts mutual information [17, 18] based method to calculate similarity degree. We assume that $q(c_i)$ and $q(c_j)$ are the probability distributions of spectral band i and j respectively. $q(c_i, c_j)$ represents the joint probability distribution. The mutual information between band i and band j is defined as: $MI(band_i, band_j) = \sum_{c_i=1}^{c} \sum_{c_j=1}^{c} (q(c_i, c_j) \log \frac{q(c_i, c_j)}{q(c_i)q(c_j)})$.

Based on the former definition, spectral clustering based band selection algorithm is as follows [19, 20]:

Algorithm 1: spectral clustering based band selection algorithm

Inputs: bands vector $band_1$, $band_2$,…, $band_n$, number k of clusters to construct

Outputs: k representative bands

1) Compute probability distribution of all bands: q (c_1), q (c_2),…,q (c_n).

2) Compute similarity degree between bands using mutual information formula, obtaining similarity matrix W.

3) Compute the unnormalized Laplacian matrix L of W.

4) Compute the first m eigenvectors v_1, v_2,…, v_m of L.

5) Let $V \in R^{n \times m}$ be the matrix containing the vectors v_1, v_2,…, v_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 V.

7) Cluster the points x_1, x_2,…, x_n with the k-means algorithm into k clusters: C_1, C_2,…, C_k.

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

2.2. Second View: LDA and Manifold Learning based Feature Abstraction Method

In hyperspectral image classification, unlabeled training samples are readily available but labeled ones are fairly expensive to obtain. Therefore, it is an interesting idea to extract features with the help of unlabeled data, which belongs to semi-supervised learning category. This section combines LDA and manifold learning to extract features through an optimization procedure. Linear discriminant analysis is a well-known dimensionality reduction method whose projection matrix is obtained by minimizing the within class covariance and at the same time maximizing the between class covariance [21, 22]. LDA is a supervised method, whose performance depends on the number of labeled training sample points. Therefore, manifold learning is imported to improve the performance by make full use of the unlabeled data [23, 24].

Given n_1 labeled sample points $\{(x_i, y_i)\}_{i=1}^{n_1}$ and n_2 unlabeled sample points. $\{(x_i)\}_{i=1}^{n_2}$. Wherein, $x_i \subset R^K$ is original feature vector of pixel, y_i is class label of pixel. Suppose w is the projection matrix which transforms original feature vector into new feature space with lower number of features. We obtain w by minimizing the objective function: $\arg\min_w \{\alpha M(w) + (1-\alpha)L(w)\}$. Wherein, $\alpha \in [0,1]$ is weight coefficient. $M(w)$ stands for manifold learning regularization, which incurs a heavy penalty if two neighboring points in the original space lie far away in the output space. The corresponding weight matrix is as follows:

$$MM_{ij} = \begin{cases} 1 & if \ x_i \in knn(x_j) \ or \ x_j \in knn(x_i) \\ 0 & otherwise \end{cases}$$

Wherein, knn() stands for k nearest neighbor algorithm. To apply knn algorithm, we should construct a similarity degree calculating method to measure how well sample x_i is similar to sample x_j. This section use mutual information based method to calculate similarity degree. Firstly, we assume that $p(c_i)$ and $p(c_j)$ represent the probability distributions of sample points x_i and x_j respectively. $p(c_i, c_j)$ represents the joint probability distribution. The mutual information is defined as:

$$MI(x_i, x_j) = \sum_{c_i=1}^{c} \sum_{c_j=1}^{c} (p(c_i, c_j) \log \frac{p(c_i, c_j)}{p(c_i)p(c_j)}).$$

Therefore, the corresponding manifold regularization item is as follows:

$$M(w) = \sum_{i=1}^{n_1+n_2} \sum_{j=1}^{n_1+n_2} \| w^T x_i - w^T x_j \|^2 MM_{ij} = Y^T (D_{MM} - MM)Y = Y^T L_{MM} Y$$

To integrate LDA and manifold learning into a unified framework, we adjust the objective function of LDA according to the objective function of manifold learning. Within-class weight matrix is defined as follows:

$$WI_{ij} = \begin{cases} 1 & if \ both \ point \ x_i \ and \ x_j \ have \ the \ same \ label \\ 0 & otherwise \end{cases}$$

Between-class weight matrix is defined as follows:

$$BE_{ij} = \begin{cases} 1 & if \ point \ x_i \ and \ x_j \ are \ different \ labeled \\ 0 & otherwise \end{cases}$$

LDA trys to minimize the within class covariance to keep the compactness in class, therefore, the following objective function should be minimized:

$$T_{wi} = \sum_{i=1}^{n_1+n_2} \sum_{j=1}^{n_1+n_2} || w^T x_i - w^T x_j ||^2 \, WI_{ij} = Y^T (D_{wi} - WI)Y = Y^T L_{wi} Y$$

Meanwhile, LDA also trys to maximize the between class covariance to keep the separability between classes, therefore, the following objective function should be maximized:

$$T_{be} = \sum_{i=1}^{n_1+n_2} \sum_{j=1}^{n_1+n_2} || w^T x_i - w^T x_j ||^2 \, BE_{ij} = Y^T (D_{be} - BE)Y = Y^T L_{be} Y$$

Therefore, the LDA objective function is defined as follows:

$$L(w) = \beta T_{wi} - (1-\beta)T_{be} = Y^T(\beta L_{wi} - (1-\beta)L_{be})Y \quad , \text{ Wherein, } \beta \in [0,1] \text{ is weight}$$
coefficient.

3. Co-training based Hyperspectral Image Classification Method

In order to solve the problem of small labeled training sample set, based on the two views obtained in Section 2, this paper enlarges the scale of labeled training sample points through importing co-training technique, thereby improves the classification precision of hyperspectral image recognition.

Co-training algorithm employs two classifiers, each of which labels the unlabeled sample point for the other during the learning process. Suppose we have labeled training sample set L={$(V_1,T_1,p_1),(V_2,T_2,p_2),\ldots,(V_{|L|},T_{|L|},p_{|L|})$}, wherein V_i is feature 1 of ith instance, T_i is the feature 2 of ith instance. p_i is class label of ith instance, $|L|$ is the number of labeled samples. Unlabeled training sample set is U={$(V_1,T_1),(V_2,T_2),\ldots,(V_{|U|},T_{|U|})$}, wherein $|U|$ is the number of unlabeled samples. This paper use two support vector machine (SVM for short) to train and predict instances expressed as feature 1 and feature 2 respectively. Two SVMs provide reliable new labeled sample points to each other iteratively. Eventually, they joint predict the class label of pixels on the enlarged labeled training sample set. In order to choose appropriate unlabeled sample points to label, the distance of sample point to the classification hyperplane is used to compute the labeling confidence. The co-training based hyperspectral image classification algorithm is as follows:

Algorithm 2. Co-training based hyperpsectral image classification
Inputs: labeled sample set L, unlabeled sample set U,
    Maximum number of iterations T
Outputs: classifier $g_1(x)$ and $g_2(x)$
$L_1=L$; $L_2=L$; $U_1=U$; $U_2=U$;
$g_1 \leftarrow SVM.Train(L_1, feature_1)$; //SVM trained in feature 1, returns classifier $g_1(x)$

$g_2 \leftarrow SVM.Train(L_2, feature_2)$; //SVM trained in feature 2, returns classifier $g_2(x)$
Repeat for T rounds   // Loop exit criteria (1)
    For $k \in \{1,2\}$ do

      $x^* = argmax(g_k(x))$ //find the most confident unlabeled sample point
         $x \in U_k$

     If $(|g_k(x^*)| > 1)$ //the confidence must be larger than 1

     $y^* = sgn(g_k(x^*))$

     $U_{(k-2)\%2+1} = U_{(k-2)\%2+1} - \{(x^*)\}$ //delete from the other's unlabeled sample set

     $L_{(k-2)\%2+1} = L_{(k-2)\%2+1} \cup \{(x^*, y^*)\}$ //add to the other's labeled sample set
   End if
   End for

If neither L_1 or L_2 changes // Loop exit criteria (2)
     Exit
  Else

$$g_1 \leftarrow SVM.Train(L_1, feature_1);$$

$$g_2 \leftarrow SVM.Train(L_2, feature_2);$$

  End if
End repeat

4. Experiments and Results

Having presented the method of co-training based hyperspectral image classification (CHSIC for short) in the previous section; 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 was acquired by AVIRIS sensor over the Kennedy Space Center (KSC for short), Florida, on March 23, 1996 [25]. KSC data set have original 224 bands, among which 48 bands are consider as water absorption and low SNR bands. These noise bands are removed from original data in this experiment. For classification purpose, 13 classes representing the various land cover materials were defined in this environment. The class number, class name, the number of training and testing samples of KSC data is listed in Table 1. Figure 2 shows the 31th band image of KSC data set.
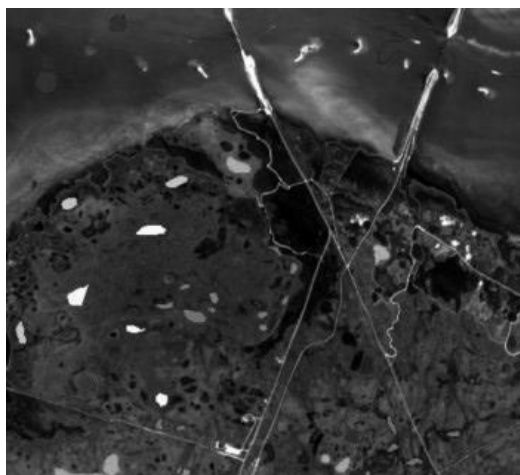
**Figure 2. Image of 31th Band in KSC Data Set**

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 [26] with SVM classifier; 2) MVPCA algorithm proposed in literature [27] with SVM classifier.

Table 1. Class Number, Class Name, the Number of Training and Testing Samples of KSC Data

Class number	Class name	Train	Test
1	Scrub	40	400
2	Willow swamp	20	200
3	Cabbage palm hammock	20	200

4	Cabbage palm/oak hammock	20	200
5	Slash pine	15	120
6	Oak/broadleaf hammock	20	200
7	Hardwood swamp	12	60
8	Graminoid marsh	30	300
9	Spartina marsh	35	300
10	Cattail marsh	30	300
11	Salt marsh	30	300
12	Mud flats	35	300
13	Water	40	400

Figure 3 displays the classification precision of algorithm ID, MVPCA and CHSIC. As is shown in Figure 3, the classification precision improves along with the increase number of features. Among them, CHSIC algorithm achieves the best performance.
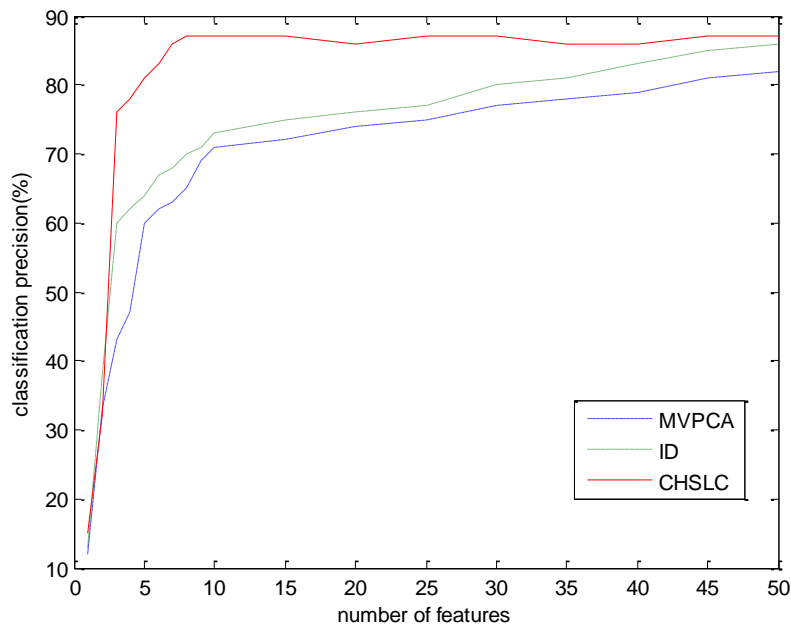
Figure 3. Comparison of Classification Precision

5. Conclusions

In this paper, we present a co-training based hyperspectral image classification method. In order to apply co-training algorithm, two independent views of original data were obtained through dimensionality reduction process. After that, two SVM classifiers iteratively label the unlabeled sample point for the other during the learning process. Based on the enlarged training sample set, classification precision is improved. A comparison with two popular hyperspectral image classification methods ID and MVPCA is conducted. SVM classifiers are utilized to evaluate the performance of relevance algorithms. The statistical experimental results on KSC data show that the proposed method has better performance than the other state-of-the-art methods.

Acknowledgments

References

[1] L. Chen, M. Yang and C. Deng, "Linear Regression Fisher Discrimination Dictionary Learning for Hyperspectral Image Classification", Springer International Publishing, **(2014)**, pp. 247-254.

[2] J. Li, P. R. Marpu and A. Plaza, "Generalized composite kernel framework for hyperspectral image classification", IEEE Transactions on Geoscience and Remote Sensing, vol. 51, no. 9, **(2013)**, pp. 4816-4829.

[3] R. Ji, Y. Gao and R. Hong, "Spectral-spatial constraint hyperspectral image classification", IEEE Transactions on Geoscience and Remote Sensing, vol. 52, no. 3, **(2014)**, pp. 1811-1824.

[4] J. Liu, Z. Wu and L. Sun, "Hyperspectral Image Classification Using Kernel Sparse Representation and Semilocal Spatial Graph Regularization", Geoscience and Remote Sensing Letters, vol. 11, no. 8, **(2014)**, pp. 1320 – 1324.

[5] H. Huang, F. Luo and J. Liu, "Sparse Manifold Preserving for Hyperspectral Image Classification", Pattern Recognition. Springer Berlin Heidelberg, **(2014)**, pp. 210-218.

[6] Y. Qian, M. Ye and J. Zhou, "Hyperspectral 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.

[7] P. Yan-Bin and A. I. Jie-Qing, "Hyperspectral Imagery Classification Based on Spectral Clustering Band Selection", Opto-Electronic Engineering, vol. 39, no. 2, **(2012)**, pp. 63-67.

[8] Z. Zheng and Y. Peng, "Semi-Supervised Based Hyperspectral Imagery Classification", Sensors and Transducers, vol. 156, no. 9, **(2013)**, pp. 298-303.

[9] Y. B. Peng, B. S. Liao and Z. J. Zheng, "Negotiation model based on semi-supervised opponent's negotiation preference learning", Computer Integrated Manufacturing Systems, vol. 18, no. 9, **(2012)**, pp. 2082-2090.

[10] J. Li, S. Jia and Y. Peng, "Hyperspectral Data Classification with Spectral and Texture Features by Co-training Algorithm", Opto-Electronic Engineering, vol. 39, no. 11, **(2012)**, pp. 88-94.

[11] P. Yang, Q. Zhu and B. Huang, "Spectral clustering with density sensitive similarity function", Knowledge-Based Systems, vol. 24, no. 5, **(2011)**, pp. 621-628.

[12] F. Shang, L. C. Jiao and J. Shi, "Fast density-weighted low-rank approximation spectral clustering", Data Mining and Knowledge Discovery, vol. 23, no. 2, **(2011)**, pp. 345-378.

[13] J. K. Thurlow, C. L. P. Murillo and K. D. Hunter, "Spectral clustering of microarray data elucidates the roles of microenvironment remodeling and immune responses in survival of head and neck squamous cell carcinoma", Journal of Clinical Oncology, vol. 28, no. 17, **(2010)**, pp. 2881-2888.

[14] L. Wang and M. Dong, "Multi-level Low-rank Approximation-based Spectral Clustering for image segmentation", Pattern Recognition Letters, vol. 33, no. 16, **(2012)**, pp. 2206-2215.

[15] H. Wang, J. Chen and K. Guo, "A genetic spectral clustering algorithm", Journal of Computational Information Systems, vol. 7, no. 9, **(2011)**, pp. 3245-3252.

[16] J. Cao, P. Chen and Q. Dai, "Local information-based fast approximate spectral clustering", Pattern Recognition Letters, vol. 38, no. 1, **(2014)**, pp. 63-69.

[17] J. B. Kinney and G. S. Atwal, "Equitability, mutual information, and the maximal information coefficient", Proceedings of the National Academy of Sciences, vol. 11, no. 9, **(2014)**, pp. 3354-3359.

[18] J. V. Koski, V. F. Maisi and T. Sagawa, "Experimental Observation of the Role of Mutual Information in the Nonequilibrium Dynamics of a Maxwell Demon", Physical review letters, vol. 13, no. 3, **(2014)**.

[19] V. U. Luxburg, "A tutorial on spectral clustering", Statistics and computing, vol. 17, no. 4, **(2007)**, pp. 395-416.

[20] A. Y. Ng, M. I. Jordan and Y. Weiss, "On spectral clustering: Analysis and an algorithm", Advances in neural information processing systems, vol. 2, no. 1, **(2002)**, pp. 849-856.

[21] M. Li and B. Yuan, "2D-LDA: A statistical linear discriminant analysis for image matrix", Pattern Recognition Letters, vol. 26, no. 5, **(2005)**, pp. 527-532.

[22] C. H. Li, H. H. Ho and B. C. Kuo, "A Semi-Supervised Feature Extraction based on Supervised and Fuzzy-based Linear Discriminant Analysis for Hyperspectral Image Classification", Appl. Math, vol. 9, no. 1, **(2015)**, pp. 81-87.

[23] D. Lunga, S. Prasad and M. Crawford, "Manifold-Learning-Based Feature Extraction for Classification of Hyperspectral Data: A Review of Advances in Manifold Learning", Signal Processing Magazine, IEEE, vol. 31, no. 1, **(2014)**, pp. 55-66.

[24] A. Talwalkar, S. Kumar and M. Mohri, "Large-scale SVD and manifold learning". The Journal of Machine Learning Research, vol. 14, no. 1, **(2013)**, pp. 3129-3152.

[25] J. Ham, Y. Chen and M. M. Crawford, "Investigation of the random forest framework for classification of hyperspectral data", Geoscience and Remote Sensing, IEEE Transactions on, vol. 43, no. 3, **(2005)**, pp. 492-501.

[26] C. I. Chang and S. Wang, "Constrained band selection for hyperspectral imagery", IEEE Transactions on Geoscience and Remote Sensing, vol. 44, no. 6, **(2006)**, pp. 1575-1585.

[27] C. I. Chang, Q. Du and T. L. Sun, "A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification", IEEE Transactions on Geoscience and Remote Sensing, vol. 37, no. 6, **(1999)**, pp. 2631-2641.

Authors

Zhijun Zheng, received his Ph.D. degree from Xi'an Jiaotong University. He has been associate professor at Zhejiang University of Science and Technology. His research interests include machine learning, computer architecture, software engineering, multimedia analysis retrieval, computer animation image retrieval and statistical learning.

Yanbin Peng, received his Ph.D. degree from the College of Computer Science and Technology, Zhejiang University. He has been associate professor at Zhejiang University of Science and Technology. His research interests include artificial intelligence, machine learning, data mining, pattern recognition and image retrieval.