

Research on Pixel Unmixing based on Support Vector Machine

Yunfeng Liu¹ and Laijun Lu²

^{1,2}College of Earth Sciences, Jilin University, Jinlin 130021, china
¹Road2000@126.com, ²luj1956@163.com

Abstract

Through the research of the application of support vector machine theory in the pixel unmixing, the advantages of the weighted posterior probability support vector machine theory in pixel unmixing is presented. Considering the difference of each support vector machine classifier, posterior probability pixel is used as weight coefficient of sub-pixel classification for pixels unmixing. This paper presents a weighted posterior probability support vector machine mixed pixel unmixing method. This method not only has the nonlinear model decomposition characteristics of high precision, but also reduces the standard support vector machine calculating the amount of multi classifier, and it has strong adaptability.

Keywords: *pixel unmixing, end-member extraction, data field, weighted posterior probability support vector machine*

1. Introduction

Pixel unmixing is a fundamental but important task in remote sensing data processing. Due to the limitation of spatial resolution of imaging system and complexity of land cover, one pixel in the image often contains various types of ground objects. The remote sensing image mixed pixels are therefore formed. Mixed pixel issue has become the main obstacle for remote sensing technology to develop in a quantitative manner. Regarding the issue, lots of researches have been made by domestic and foreign scholars [1-2]. Different neural network models were developed, as well as randomly mixed model, nonlinear stepwise regression model and optimization-based search mode [3-4]. Out of them all, support vector machine (SVM) is a new machine learning method raised by Vapnik *et al.* [5-6]. Since then on, people probed into SVM about its pattern recognition, machine learning, decision analysis etc. [7-9].

Due to the ability to process high-dimensional feature spatial data, SVM has favorable learning performance, promotion ability and strong generalization capability, especially outstanding superiority in remote sensing image processing [10-11]. SVM has been widely applied for remote sensing image compression, retrieval, categorization, image feature extraction and image fusion etc., achieving satisfactory effects. Posterior probability can fetch effectively sub-pixel information [12]. Eric [13] thought sub pixel classification (*i.e.* pixel decomposition) is special and it bases on posterior probability classification. So he discussed further why Bayes posterior probability value was replaced by mixed pixel components. To categorize sub pixels (*i.e.* decompose pixels) of remote sensing images, we propose SVM multi-class classification algorithm based on weighted posterior probability. For in-depth analysis we investigate the application of SVM in the nonlinear decomposition model of mixed pixels.

2. The Principle of Support Vector machine

Unlike other methods such as traditional artificial neural networks, SVM makes use of structural risk minimization principle to substitute the classic empirical risk minimization

principle. SVM considers difficulties of collecting large samples in real practice. In the design beginning, it's aimed at machine learning methods for processing small samples. Structural risk minimization principle ensures reliability and precision of data classification of small samples. In actual computation, the core idea is to map sample points onto high-dimensional feature space to minimize the classifying risk by constructing the optimal hyper plane of classification.

The primitive SVM method can only cope with two-class classification problem. Remote sensing data classifying belongs to the multi-class classification problem. In order to categorize remote sensing data, the two-class classification problem should be extended to support multi-class classification problem. Many algorithms were developed to extend SVM for multi-class classifying. Those algorithms have three types: one-vs-more combination mode, one-vs-one combination mode, and SVM decision tree probability model. Of them, the third type gets multi-class probability model by combining several two-class classification probability results. There have many classifier algorithms for categorizing remote sensing data. SVM has the following merits if being applied for classifying remote sensing data:

Small Sample Training: regarding the massiveness of remote sensing data, samples for training are far fewer than remote sensing image pixels. In terms of small samples, it's significant to acquire good classifying results. SVM takes into account the difficulty of obtaining large samples at the beginning of design. It's dedicated to studies on small sample classifiers. Therefore, it is rather fit for the classification of remote sensing data.

Ability to Deal with "Dimensional Curse": for the classification of high/multiple spectrum remote sensing data, traditional classifier technologies based on normal distribution hypotheses generally can't reach desirable results. So it's better to choose SVM-based classifier to cope with "dimensional curse".

3. Application of Support Vector Machine in the Processing of Remote Sensing Image

The characteristic of the remote sensing image pixels is a mixture of several land cover classes, rarely from a single uniform object composition. Thus the image pixel spectra characteristics is a mixture of several spectral features and spectral characteristics of reflection is not a single object. Pixel unmixing model usually has the following kinds: linear model, fuzzy model, geometrical optics model, random geometric model, probability model. The linear model is expressed by the reflectance of each single pixel spectral bands in the expressed as a linear combination of its characteristics of end member of reflectivity and their abundance. By using the probability model of linear discriminant analysis and EM produced a discriminant value, according to the scope of discriminant values of the pixel is divided into different categories. But the probability model in only two features mixing conditions to use. The fuzzy model is the various features of categories as fuzzy sets, fuzzy sets of pixel elements, each pixel is corresponding with a set of membership values, memberships represents the fuzzy mean vector pixels contained in this object classes and fuzzy covariance matrix. In [14], the author puts forward the approach for pixel unmixing support vector machine based on probability output. Using multi class posterior probability to estimate image mixed pixel components, using support vector machine training good estimate of two paired images of the a posteriori probability, were converted to the corresponding class posterior probability. If the pixel is just one of the mixed pixel, then the posterior probability describes the components of the component information, so it can be directly to the multi class posterior probability value as the mixed pixel component indexes. And then uses simulation to form test sample, to evaluate the accuracy of the method of decomposition of mixed pixel SVM multi class posterior probability. In ^[14], puts forward the approach for pixel unmixing nonlinear based on

support vector regression. The steps are: automatic extraction of all the typical spectral image with the method of unsupervised, using simulation methods combined training and test sample. Determination of parameters of support vector regression. Each pixel in the image is obtained by a regression component using support vector training well, and with the test samples for support vector regression method of pixel unmixing results accuracy assessment. Although, it has had on the support vector machine algorithm for remote sensing image mixed pixel decomposition, but the sub-pixel classification after each pixel within the percentage is simply based on the posterior probability. Did not consider the difference between two-class support vector machine classifier. This paper fully take into account the various differences between the two-class support vector machine classifiers, sub-pixel classifier classification using a posteriori probability as the individual weights of two-class support vector machine classifiers, this method can reduce the sub-pixel classification error rate.

4. Support Vector Machine based on Weighted Posteriori Probability

4.1. Support Vector Machine based on Weighted Posteriori Probability

In various SVM-based algorithms, different two-class classifiers choose different samples and accordingly the distribution of samples differs from one another. In this case, it's necessary to find out the differences among those two-class SVM classifiers. SVM can solve pattern recognition problem, *e.g.* recognizing physical objects on the ground and after analysis defining objects pending for recognition as one land cover class. Besides, it can get the posterior probability of such pending objects belonging to one class. In the real application of SVM for remote sensing image processing, as the acquired image information is polluted by some information like noises, each sample affects differently the classification. In some cases, samples are not positively regarded as one class but longed to some class according to certain probability or membership degree. Hence it's not appropriate to represent type/class with A. Obviously, the posterior probability becomes an important measuring indicator. For all two-class SVM classifiers, to estimate the posterior probability of one sample belonging to one class is helpful to extract information of sub pixels in remote sensing images. The SVM model based on weighted posterior probability used to decompose mixed pixels of remote sensing images can decide the type of samples and simultaneously obtains posterior probability of each land cover type, ultimately fetching the information about mixed pixel components.

4.2. Method for Determining the Weighted Posterior Probability of Sample Points

The SVM mixed pixel decomposition model based on weighted posterior probability combines output probability of a few SVM classifiers. At the same time, the model refers to differences of each two-class SVM classifier and calculates the posterior probability of each land cover class in the image with conditional probability of sample points. With posterior probability as weight coefficient of each two-class SVM classifier, the model determines the percentage of land cover type. Moreover, SVM based on weighted posterior probability gets every land cover type. After comparing experiment results with standard SVM mixed pixel decomposition results, we gain more accurate mixed pixel decomposing results.

By analyzing output probability combinations of some two-class SVM classifiers, SVM based on weighted posterior probability finds two-class SVMs in different combinations are not identical. It uses posterior probability of samples in one class of the two as the weight coefficient of SVM classifier. The type of end members (EMs) of remote sensing image mixed pixels and the percentage can be attained by reckoning posterior probability. In the following we introduce the application of weighted posterior

probability method discussed in [14] for decomposition of remote sensing image mixed pixels.

The k class $a_i, 1 \leq i \leq k, k \geq 3$, need to $k(k-1)/2$ Classifier, $i < j$ uses for a_i and a_j zone classification. For the test sample X given belongs to the i class posterior probability can be expressed as:

$$P(i|x) = \sum_{j=1, j \neq i}^N P_{ij}(i|j;x)P(j|x) \quad (1)$$

Where, $P_{ij}(i|j;x)$ is expressed by the i class and j class to constitute the two class support vector machine classifier. Calculated x belongs to the posterior probability of i class. $P(j|i)$ and $P(i|j)$ are expressed as in the two class classifier the i class and j class, sample X belongs to the i class and the j class conditional posterior probabilities.

Where, satisfy the following relations between $P(j|i)$ and $P(i|j)$:

$$P(i|j) = 1 - P(j|i) \quad (2)$$

To constitute all pairwise probability posteriori probability set: $p_i = \text{Pr ob}(x \in a_j | x)$, To obtain many kinds of probability value. Each remote sensing image mixed pixel in the end element solution, using weighted a posteriori probability explains the composition information points and its proportion in the mixed pixel, multi class weighted posterior probability value can be used as the mixed pixel of each end-member percentage. Mixed pixel weighted support vector machine a posteriori probability decomposition method is more to reduce influence of end-member point type fuzzy land cover types of sample points than on classifier based on support vector machine. So, the improved nonlinear pixel unmixing model precision. The work is completed on the Libsvm toolkit.

5. Experimental Analysis and Results

5.1. Experimental Data

In this paper, experimental data is the resolution of the 30m TM images in 2014. It is located in the north-central segment area of Great Khingan in our Northeastern, track number is 119/26, receiving period is April 29, 2011, the image spatial resolution is 30*30m2(TM1~TM7),no cloud is cover. The coverage is 48°51'36"-48°51'43"N, 125°14'30"-125°20'45"E, the size is 400x400.

The study regions include Nen River, Black River, the Five Connective Lakes and Mount Ke. Nen River situates in the Northeast of Inner Mongolia autonomous region, bordering Russia by Ergun River in Northwestern corner. The administrative division includes Hulunbeir League of Inner Mongolia autonomous region, Great Khingan and Black River area of Heilongjiang province, and all or part of totally 16 counties (banner or city) of Qiqihar municipality. The land area reaches totally 190,200 km², *i.e.* 0.2% of the national land area.

The region sits in the north-central segment area of Great Khingan in our Northeastern part. Great Khingan passes through the central part of the region from the south to the north. The main watershed altitude reaches 1000-1400m. The mountain top is perfectly round. Slope gradients are gentle. Valleys are wide and flat. Forests are thick. The forest coverage rate approximates 62%. It is one of the biggest forest zones in our country. In the east, it's Nen River valley and the eastern part of Songnen Plain. The plain elevation is about 200m. Soil layers are thick. The soil is fertile. Plough lands are concentrated to large areas. Due to the good tractor-plowing condition, it is rice production base in the

region. On the west side, it's the waving Hulun Buir Plateau, vast with abundant aquatic plants. It's a famed natural pasture in our country.

The region belongs to semi-humid temperate climate zone. Winter is long with less snow while summer is short and moist. Affected by the terrain, the climate differs greatly in different areas. In Great Khingan mountainous region, the annual average temperature is minus 2-4°C. The frost-free season lasts 80-100 days. The annual precipitation is 350-500mm. In Songnen Plateau area, the annual average temperature is below 1-2.5°C; the annual precipitation reaches 450-530mm in mainly summer and autumn. In the two seasons, it's hot and rainy, conducive to crop's growth. Influenced by the Mongolia-Siberian air, the western part of Great Khingan is drier and colder than the eastern part. The annual average temperature is below 2-5°C and annual precipitation reaches 300-400mm.

This paper selects the image synthesis image color 3, 5, 7 channel, it is shown in Figure 1, including the types of farmland, water, residential land, wasteland by images.

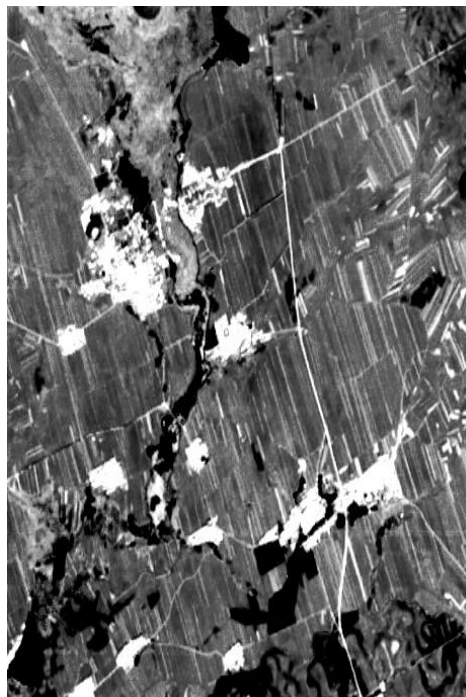
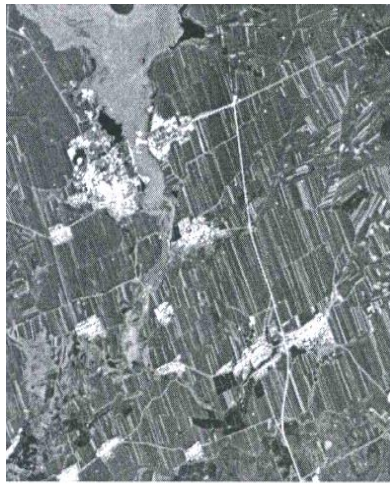


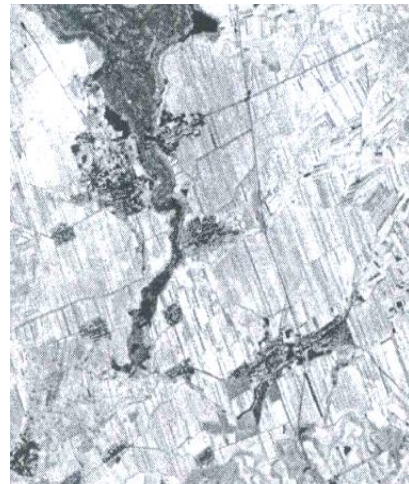
Figure 1. Study Regional Remote Sensing Image

5.1.1. Determine the Sample Data and Parameter Selection: Four main types of land including farmland, water, residents, badlands cover image, selected 349,338,357,364 training samples, accounting for 2.11% of the total amount of experimentation area data. Selection of Radial Basis Function (RBF) in this experiment for experimental nuclear function. The penalty coefficient parameter C and Gauss function parameters of G Libsvm software through the training sample data values determined for: 2.5, 0.8.

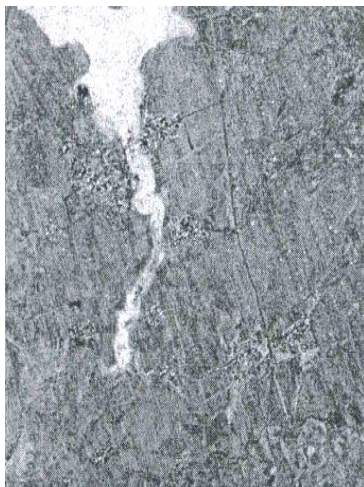
5.1.2. The Decomposition Results: According to the training sample data weighted posterior probability support vector machine to complete the calculation can be non-sample pixel posteriori probability. Figure 2 is a weighted posterior probability support vector machine method to obtain the experimental data of each land cover type abundance images. Figure 3 is a standard support vector machine pixel unmixing abundance image results.



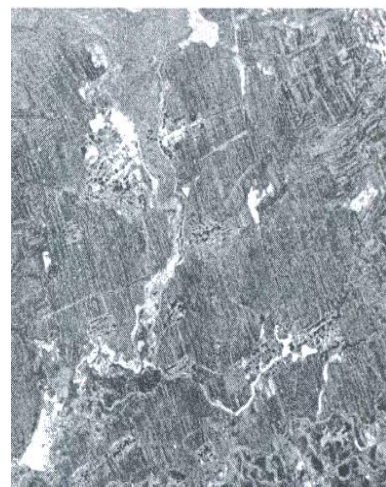
The residents



Farmland



Water



Wasteland

Figure 2. Support Vector Machine Pixel Unmixing based on Weighted Posterior Probability

Component diagram gray brighter part explains the abundance of decomposition of the image corresponding to the endmember land cover types content higher, on the contrary, containing the land less cover types, evaluation from qualitative angle, posterior probability support vector machine method abundance image in Figure 2, visual effect is better than the standard support vector machine model decomposition abundance image Figure 3 weighted. The reason is the proposed weighted posterior probability support vector machine based on the pixel unmixing, weight coefficient by posterior probability as each of the two class support vector machine classifier using the classification of sub pixel. The method is better than using only the standard support vector machine model decomposition of remote sensing image improves accuracy of sub-pixel classification. The following quantitative accuracy evaluation results of the decomposition.

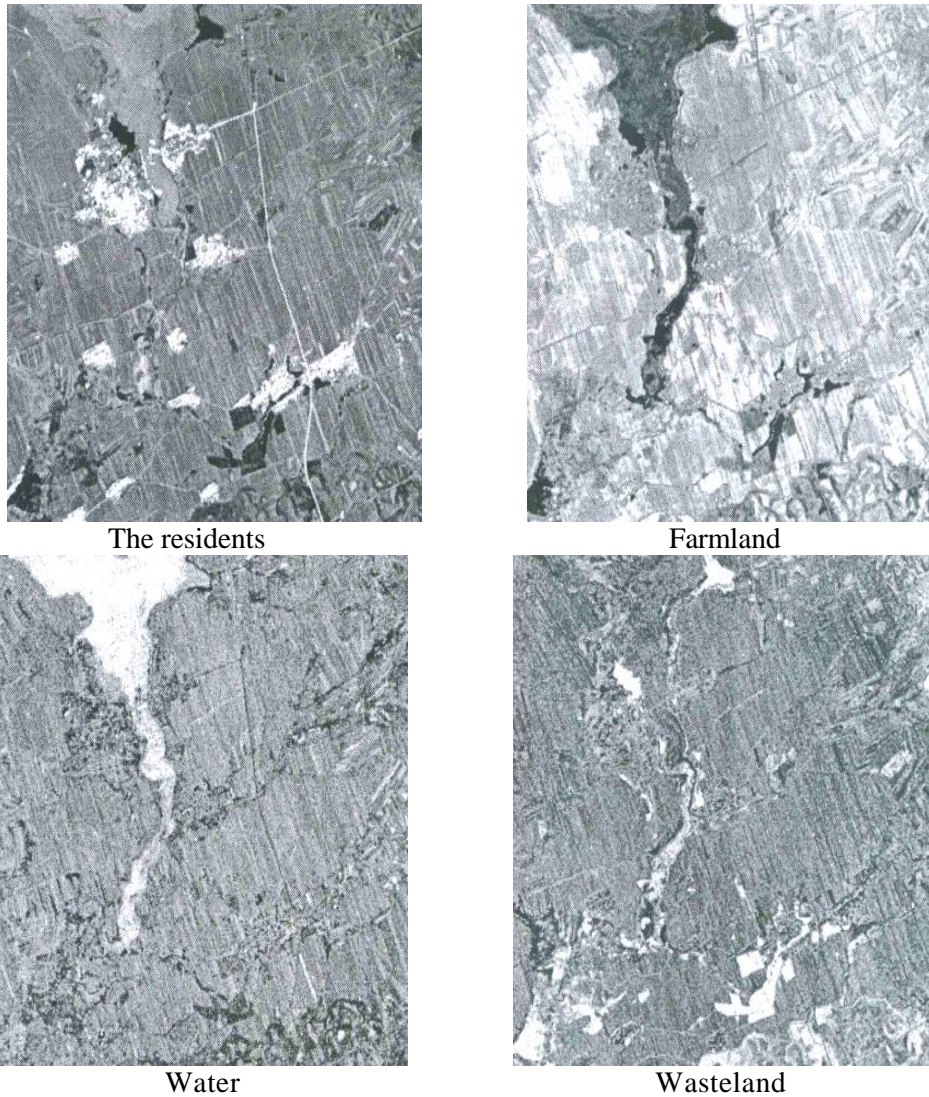


Figure 3. Pixel Unmixing Map based on Support Vector Machine

5.2. Accuracy Evaluation

To further verify the evaluation experiment decomposition accuracy this study using simulated image data decomposition error. Taking the average spectral coverage type sample of each type of land as the End-member vectors, by formula (3) generates test pattern

$$g(x, y) = [1 + \frac{N(0,1)}{SNR}] [\sum_{j=1}^5 r_j a_j(x, y)] \quad (3)$$

Where, $g(x, y)$ shows the mixed pixel, $a_j(x, y) \geq 0$ shows the percentage of each land cover type. r_j shows the end element of each land cover type. $SNR = 20:1$ shows the signal to noise ratio.

In the research, it's very hard to get the actual percentage of land-surveyed mixed pixels. So we used the simulation data of study regions to appraise the mixed pixel decomposition errors based on the weighted SVM. The evaluation method uses average spectrum of each class of ground object samples as EM spectral vector to form test simulation data by combination. With the simulation data, we compared decomposition results got by standard SVM mixed pixel decomposing method and the proposed one.

Then we determine goodness and badness of such results in virtue of the widely adopted RMSE indicator. Set S_i the percentage of some class of ground objects in simulated images and S the result after decomposition. By the following equation, we calculate errors of percentage results got respectively by the two methods and from the simulation images.

$$RMSE_k = \sqrt{\frac{1}{N} \sum_{i=1}^N (s_i - s_j)^2} \quad (4)$$

Using the root mean square error decomposition of four categories of land cover types of pixel error probability, it can be seen from table1, calculation of farmland, wasteland, water, residential pixel error rate respectively, 16.3%, 17.6%, 16.8%, is less than 18%, based on support vector machines posteriori probability weighted pixel unmixing extraction method simulation image pixel error rate is less than 0.18.

Table 1. Comparison of Two Kinds of Mixed Pixel Decomposition Method for Table RMSE

Endelement Decomposition model	The residents	Farmland	Water	Wasteland
Support vector machine model based weighted posterior probability	0.163	0.176	0.168	0.180
Standard support vector machine based on the decomposition of type	0.192	0.187	0.186	0.195

Using modified criteria of pixel unmixing methods for support vector machine most pixel error rate is higher than 0.18, the method presented below based on the criteria of mixed pixel decomposition model of support vector machine, using nonlinear decomposing model for mixed pixel decomposition of remote sensing images can be broken down more precise results.

6. Conclusion

Remote sensing image mixed pixels existing nonlinear phenomena and to the majority, the remote sensing image was unmixed calculated results than linear model by using nonlinear model well.

This paper presents weighted posterior probability support vector machine remote sensing image pixel unmixing method based on nonlinear model not only has the characteristics of high precision but also reduce the decomposition using the standard SVM multi classifier when the amount of calculation, and has strong adaptability.

Acknowledgement

This work was supported by National Basic Research Program of China (973 Program) No. 2015CB453005

References

- [1] W G. Liu, E. Y. Wu and S. Gopal, "ARTMMAP: A Neural Network Approach to Subpixel Classification", IEEE Transactions on Geoscience and Remote Sensing, vol. 42, no. 9, (2004), pp. 1976-1983.
- [2] C.-I. Chang, S.-S. Chiang and J. A. Smith, "Linear Spectral Random Mixture Analysis for Hyperspectral Imagery", IEEE Transactions on Geosciences and Remote Sensing, vol. 40, no. 2, (2002), pp. 375-392.
- [3] C. Huang and J. Townshend, "A Stepwise Regression Tree for Nonlinear Approximation: Applications to Estimating Subpixel Land Cover", International Journal of Remote Sensing, vol. 24, no. 1, (2003), pp. 75-90.
- [4] S. Tang, Q. Zhu and G J. Yan, "Analysis of the effect of genetic algorithm and its application in nonlinear inversion, remote sensing linear model", Journal of Beijing Normal University (NATURAL SCIENCE EDITION), vol. 38, no. 2, (2002), pp. 266-272.
- [5] V. Vapnik, "The Nature of Statistical Learning Theory", New York: Springer Verlag, (1995), pp. 10-20.
- [6] G M. Foody and A. Mathur, "A Relative Evaluation of Multiclass Image Classification by Support Vector Machines", IEEE Transactions on Geoscience and Remote Sensing, vol. 42, no. 6, (2004), pp. 1335-1343.
- [7] Luo sword by, C. Zhou and L. Yi, "Application of support vector machine and the remote sensing image spatial feature extraction and classification", Remote sensing newspaper, vol. 6, no. 1, (2002), pp. 50-55.
- [8] J. Zhang, C. He and Y. Pan, "Study on the classification of SVM composition of multi-source information of high spatial resolution remote sensing data based on", Journal of remote sensing, vol. 10 no. 1, (2006), pp. 49-56.
- [9] J. Weston and C. Watkins, "Support Vector Machines for Multi-class Pattern Recognition", Proc. of Seventh European Symposium on Artificial Neural Networks. Bruges: D-Facto Press, (1999), pp. 219-224.
- [10] Z.-y. Li and S.-w. Tang, "Face Recognition Using Improved Pairwise Coupling Support Vector Machines", Proc. of the Ninth International Conference on Neural Information Processing, Cambridge: MIT Press, (2002), pp. 876-880.
- [11] F. Melgani and L. Bruzzone, "Classification of Hyperspectral Remote Sensing Images with Support Vector Machines", IEEE Transactions on Geoscience and Remote Sensing, vol. 42, no. 8, (2004), pp. 1778-1790.
- [12] G M. Foody, "Relating the Land Cover Composition of Mixed Pixels of Artificial Neural Network Classification Output", Photogramtry Engineering and Remote Sensing, vol. 62, no. 5, (1996), pp. 491-499.
- [13] D. Eric, "On the Use of Prior and Posterior Information in the Subpixel Proportion Problem", IEEE Transactions on Geoscience and Remote Sensing, vol. 41, no. 11, (2003), pp. 2687-2691.
- [14] W. Bo, L. Zhang and P. Li, "The hyperspectral unmixing based on nonlinear support vector regression", Journal of remote sensing, vol. 10, no. 3, (2006), pp. 312-318.

Authors



Yunfeng Liu, He received his B.S degree from Chengdu University of Technology. He received his M.S degree from China University of Geosciences. He is a senior engineer in Jilin University. His research interests include digital earth science.



Laijun Lu, He received his B.S degree from Changchun geological College. He received his M.S degree from Changchun geological College. He got his PH.D from Changchun geological College. He is a professor in college of earth sciences of Jilin University. His research interests include resources and environment information system, digital geological science.

