The Self-adaptive Adjustment Method of Clustering Center in Multi-spectral Remote Sensing Image Classification of Land Use

Shujing Wan, Chengming Zhang, Jiping Liu, Yong Wang, Hui Tian, Yong Liang, Jing Chen

HAL Id: hal-01361030
https://inria.hal.science/hal-01361030

Submitted on 6 Sep 2016

The Self-adaptive Adjustment Method of Clustering Center in Multi-spectral Remote Sensing Image Classification of Land Use

Shujing Wan[1], Chengming Zhang(*)[1,2], Jiping Liu[2], Yong Wang[2], Hui Tian[1],
Yong Liang[1], Jing Chen[3]

[1] School of Information Science and Engineering,

Shandong Agricultural University, Taian, China

[2] Chinese Academy of Surveying and Mapping, Beijing, China;

[3]Academy of Shandong BaoLai-LeeLai bioengineering Co. Ltd

Abstract. As one kind of remote sensing images of land use composed by various categories of surface objects difficult to obtain multi-distribution model of class spectral feature, analyzing the spectral characteristics of LU of multispectral RS imagery, this paper presents a self-adaptive adjustment of clustering center method. Depending on the intercepted situation of the cluster centers between different features to conduct split, the sub-centers obtained are as the sub-category features and the cluster centers assemble to characterize category model which is better to deal with the problems of LU category composed by various surface objects and category model not satisfying multivariate normal distribution. As there are much differences between the many centers features in the unit of category area, so the selection of training area and the determinants of rules are easy. The results of experiment indicate that the LU classification accuracy is increased between 4% and 6% with this method.

Key words: multispectral remote sensing imagery, land use classification.

1. Introduction

The LU is classified by the influence degree and the purpose of human activities [1] which refers to the activities of the human imposed on the surface, and the land use categories belong to information categories. The object classification of multi-spectral remote sensing images is based on the spectral characteristics of ground objects, and the objects categories belong to spectral classes. Using multispectral RS imagery for LU classification, a convenient multivariate statistical model is in general not available for the multispectral characteristics of LU class [2-4].

Looking from the existing literature，   there are mainly three solutions to solve this problem: firstly, try best to make the samples satisfy the multivariate statistical distribution, such as with the unsupervised clustering method, to choose the dominant and continuous pixels as training samples, then classify with statistical pattern recognition method[4]. Though this solution reduces the differences of samples, the training samples don't contain all ground objects' information. So the difference between classificatory pattern and the ground object is rather great to result in classification error [5, 6].

Secondly, make use of elevations, textures, boundaries and all other secondary information to reduce the dependence on the spectral information, whereas the secondary information add the dimension of

(*):corresponding author

the classification feature vector, and the vector multivariate statistical model is more difficult available So we commonly adopt multi-classifier confusion method to classify different types of information, to reduce the classification vector dimensions of single classifier and improving the clustering performance of the classification vector, then fuse the results of several classifiers[7-9]. However, the problem of spectral information statistical distribution is not taken account in these methods, the classification accuracy increases is limited, and even it may lower to some classes [10].

Thirdly, take the un-statistical pattern recognition method to solve multivariate statistical distribution, such as decision tree method. This method makes use of the class subset to form the boundary of branch decision tree [11, 12], but it is difficult to divide subset and to form rules for the method.

Analyzing the multi-spectral feature of LU in RS image, this paper presents a new method, Region Multi-Center (RMC), for LU classification of multi-spectral RS image. In this method, the classification cell is region, classificatory pattern is formed by clustering centers of training samples, and adopt rule-based method to classify.

2 Multi-spectral remote sensing image analyses

2.1 distribution of Within-class distance

The theory basis of land use remote sensing image classification is: in the same condition, the same objects should be able to show some internal similarities of theirs, and performing in remote sensing image, the pixels formed by same objects should have the same or similar spectrum information features and space information features. The pixels characteristic vectors which represent the same objects of will distribute in the same feature space area. In fact, although with existing of same objects in different spectral the same kind objects may have greater diversity in the spectrum, that a small class always presents similar spectral features.

Therefore, it is necessary to study the characteristic vector and category of the distribution center distance vector in order to construct a new classifier. In order to give facilities to analysis, confirmed a distance interval analysis unit, first as a threshold to distance of the sample clustering and calculating the clustering center to category center distance, obviously they are consistent, and more meet statistical conditions.

Suppose c class pattern set composed of land-use type $\{x_1, x_2, \ldots\ldots, x_N\}$, the pattern set of class ω_j is $\{x_i^{(j)}; j=1,2\ldots, c; i=1,2,\ldots, n_j\}$, among this superscripts j indicates category, subscripts i indicates the sample model number within the class, n_j is the pattern number of ω_j set, $\sum_{j=1}^{c} n_j = N$, ω_j class category center vector with mean vector ith the model indicated, that: $m_j = 1/n_j \sum_{i=1}^{j} x_i(j)$, $j=1,2, \ldots$, c, to ω_j mode of class sample, clustering it by a certain distance threshold, form k Within-class centers (fig.1), they are $\{\omega_{j1}, \omega_{j2}, \ldots, \omega_{jl}\}$,
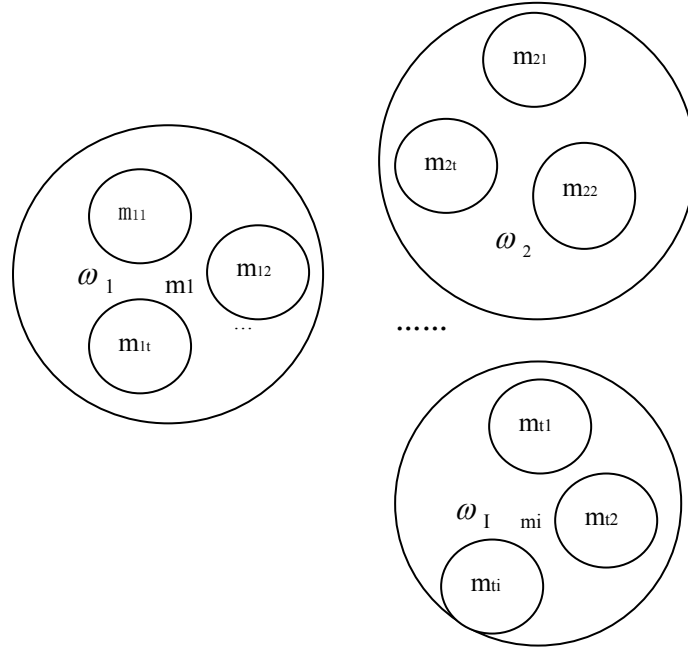
Fig.1. Category and within-class center schematic

l=1,2,…，k，form the pattern set of within-class center ω_{jl} is $\{x_i^{(jl)}; j=1,2,…,c; l=1,2,…,n_l\}, \omega_{jl}$ indicates the first within-class center of ω_j, n_l is the number of pattern set，$\sum_{l=1}^{k} nl = n_j$, the of mean vectors of within-class center are $\{m_{j1}, m_{j2}, …, m_{jl}\}, l=1,2, …, k, m_{jl} = 1/n_l \sum_{i=1}^{ni} xi(jl)$, j=1,2，…，c；

l=1,2，…，k，the distance of within-class is $d_{jl}=\|m_{jl}-m_j\|$，mapping with the ordinate of n_l and n_j's percentage and the abscissa of d_{jl}, called the within-class distance distributions.

2.2 The Analysis of the Land use classification of multi-spectral

The experiments use the area of Shandong province Laiwu city Xueye reservoir with the ETM + images in spatial resolution of 30m, a total of eight bands, the size of 1500 x 1200 pixel and the time is on May 31, 2007.

The experiments choose samples to analyze the infra-class distance distribution of multi-spectrum in four kinds of land use category as water, residential areas, forest land and bare soil. The rule of selecting samples is that samples should cover as many kinds of pixels in ground objects as possible. The samples should be more and the distribution range of samples should be wide to comply with the above principle. According to the rule, each class respectively chooses the sample 16384 pixels. Fig. 2 shows the rough region of the choosing samples.
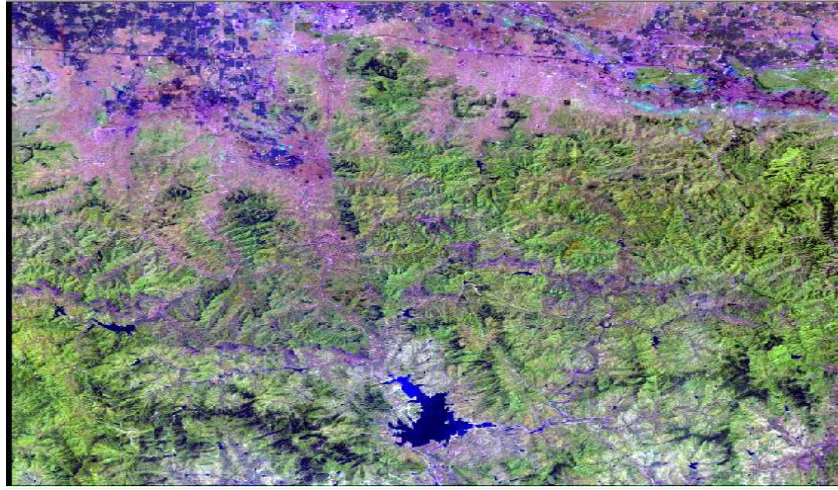
Fig.2. The remote image of Xueye Reservoir in experiment

Fig. 2 shows the graph of infra-class distance distributing of LU class, and the distance threshold of max-minimum clustering is 30. From Fig. 2, we can know: (1) The spectral value of pixels in every class has many infra-class centers. (2) Each class exists several relatively concentrated distance intervals, such as 30-40 and 60-70 of water. It represents that one class of LU is composed of several mainly ground material. (3) Spatial distance distribution of water is the most concentrated, and the distribution of residential area is more concentrated than that of woodland and plantation land, but the continuity of urban land is less than woodland and plantation land.
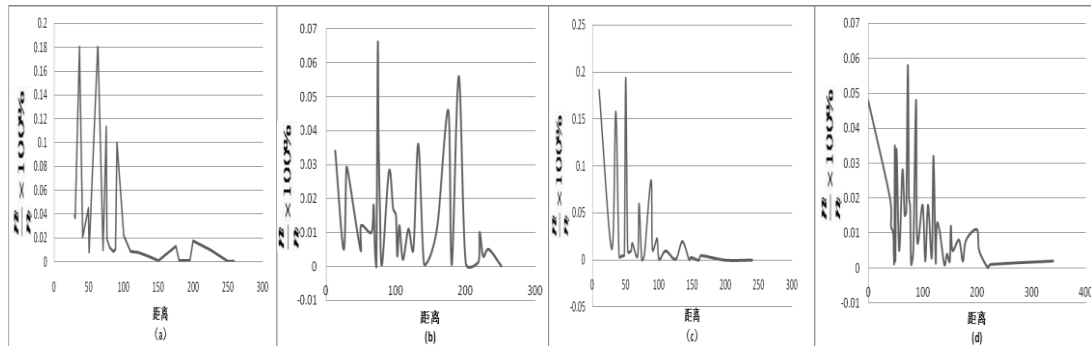

Fig.3. Graph of special distance distributing of

(a) Water (b) Woodland (c) bare soil (d) residential areas

3. Adaptive method of cluster centers

3.1 The accuracy analysis of the effects of the phenomenon of the " same objects in different spectral"

In essence, the " same objects in different spectral " phenomenon is due to in a variety of reasons the value deviation of similar features in some bands or all bands is too large resulting that feature vectors of the class in the feature space distribute excessively decentralization and it is difficult to attribute to the same cluster as we expect. Although it can reach the goals mandatorily by changing the cluster center

vectors and the effective radius vectors that it may result in that classify the pixels which not belong to the class to assign to which may reduce the classification accuracy. To better illustrate this problem, we take a classification of the two-band image and for example, assume that the image will be divided into A, B two classes that the distribution of each pixel is as Figure 4.
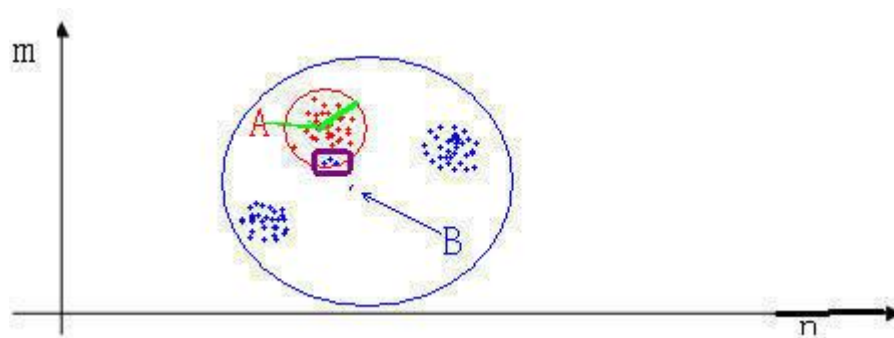

Fig.4. The schematic of the wrong classifications result

In Figure 4, as the pixels of the class B are distributed in the two regions, the resulting cluster centers and the distance between the two regions is too large, resulting in identification range (ie, blue circle) is too large. As a more extreme case, but another class A area was surrounded in the identification of the B range, surrounded by an orange rectangle should be divided into Class A of the pixel, is divided into Class B wrong. To solve this problem, cluster centers can be used to split ways. After the end of the study sample, the intersection of the cluster center to determine if the two cluster centers according to the radius of the resulting range of the identification part of the intersection, then the cluster centers need to be split.

3.2 Adaptive step

Adaptive cluster center is split by continued implementation of cluster centers，concrete steps to split is ：

(1)During the study of sample, the same category in the normal way to generate cluster centers, cluster centers in the sample vector and the maximum distance as the radius of the cluster centers identified.

(2)Calculate the result cluster centers two by two to determine whether the two cluster centers intersects , in another words, two cluster centers are identified areas of overlap. If they intersected, treat as (3).

(3)Split a larger range between the two cluster centers. Class from the sample to be split to choose one sample, randomly select a distance less than the original value of the largest identified as the threshold, the sample is divided into different subsets, separately for each study, resulting in the respective cluster centers and not split the class with the intersection of the judge, if there is still a subset of the cluster centers with which they have intersected, then re-elect a smaller threshold, re-split. After the split, the original sample a large set of complete small sample to be broken down into multiple sets, as a sub-category treatment. In this way, each "Spectrum with different material"of the surface features can be divided into several sub-categories, each sub-class corresponds to a pattern. The same sub-classes of spectral characteristics are similar.

4. Experiment

Experimental data is shown in figure 2.First, select the body of water, wood land and bare soil this three kinds of quite different categories of land use classification experiment. Various methods using the same training sample selection and accuracy evaluation. Training sample selection principle: the sample should cover as much as possible in a variety of land-use categories pixel surface features as much as possible with the visual way to select a known area of the pixel; accuracy assessment methods: random image classification 100 points, based on visual interpretation of the results as statistical classification error.

4.1 Common method of classification

(a) Parallelepiped method

(b) Minimum distance method

(c) Maximum likelihood

(d) K Neighbor method（k=19）

(e) BP Neural network

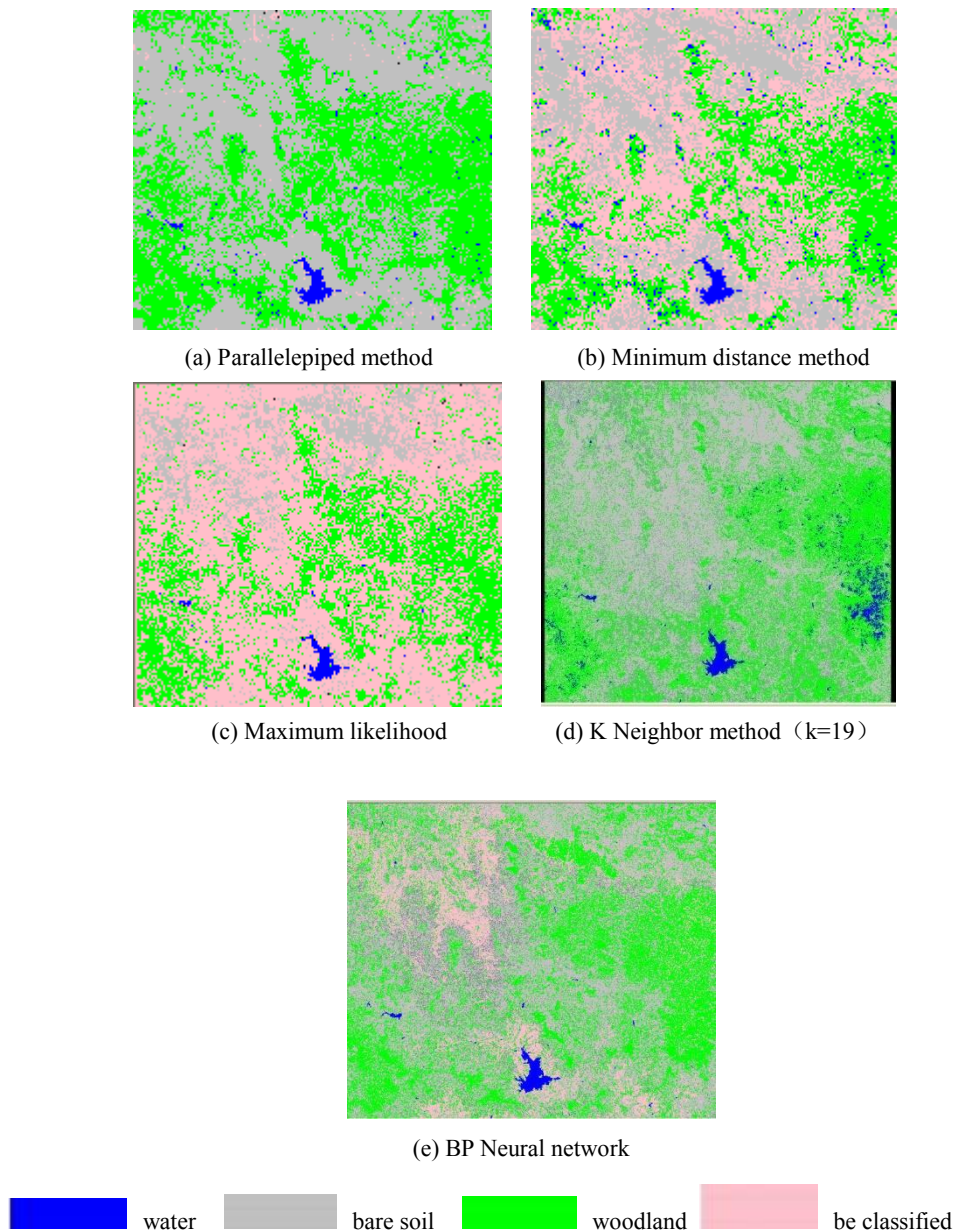water   bare soil   woodland   be classified

Fig.5. The results of classification

Looking at the results from the classification point of view, parallelepiped classification method is often a part of the misclassification phenomenon, mainly due to the similarity measure threshold T

impact. Through investigation and interpretation, the method divides a number of unclassified woodland into forest land and bare soil incorrectly. Using minimum distance classification method, the number of pixels to be classified significantly increased, to be classified and sub-phenomenon of mixed woodland is more serious, there is water affected by the shadow of the phenomenon of misclassification. K neighbor method can be a good way to distinguish watershed land and naked body while could not recognize forest, some forest land is divided into body of water mistakenly.

Maximum likelihood method, while classifies water bodies better would make serious mistake in the classification of the forest and bare soil. K-nearest neighbor method can be a good way to distinguish watershed land and naked body while could not recognize forest, some forest land is divided into body of water mistakenly. BP neural network method has better classification of various types of effects, but there is still some vegetation types not distinguished.

Table 1. Accuracy of common classification method (%)

Classification method	water	forest	town	Arable land	bare field	Average
Maximum likelihood	85.6	84.1	76.3	82.9	86.5	83.03
Minimum distance method	87.7	86.8	72.4	85.6	85.3	83.03
K Neighbor method	90.2	85.7	74.6	84.2	85.1	83.9
Neural network	87.0	85.2	81.7	84.1	85.9	84.78

4.2 Adaptive process of cluster centers and the results of classification

Using the cluster center adaptive classification method, training samples through the formation of water bodies, residential areas, woodland, bare soil, respectively, the optimal number of model centers 50, 20, 50 and 10, the experimental results shown in Figure 6.

The classification accuracy of different models and types of training samples that shown in Table 2.

Table 2. Classification accuracy of different models and types of training samples

water	forest	town	Arable land	unclassified	Average
88.5	87.8	86.8	89.3	86.2	87.72
88.2	87.6	87.0	88.9	85.3	87.40
87.9	88.3	86.4	89.2	86.0	87.56
86.4	88.5	88.7	89.0	84.1	87.34

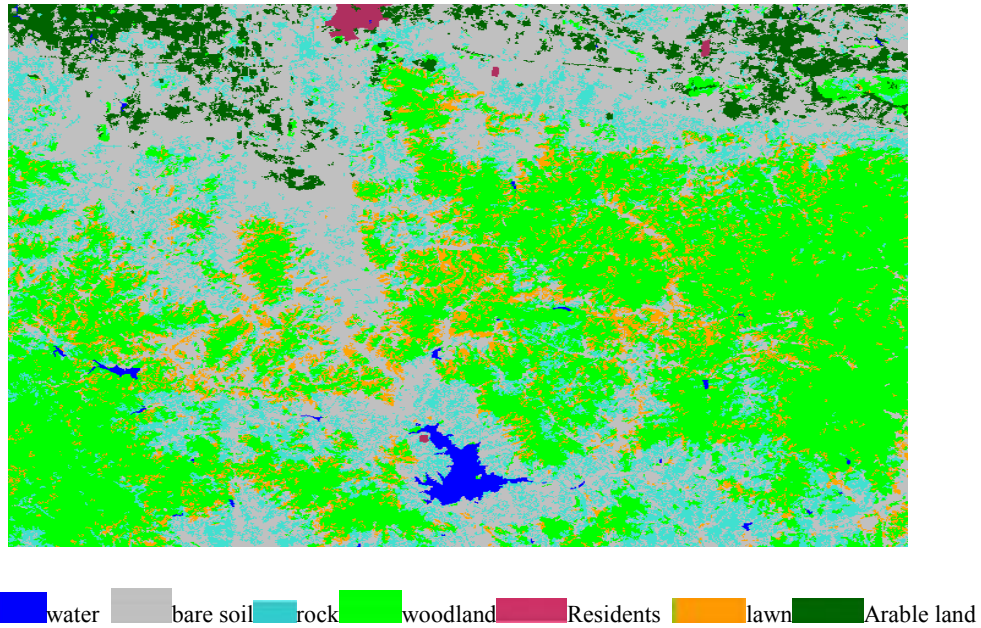water   bare soil   rock   woodland   Residents   lawn   Arable land

Fig. 6. The extraction results

Comparing different methods of classification results, we can see, the use of this method of classification, the results of the spots decreased; Table 7 Comparison Table 6 Classification accuracy increased by 4 percent.

5 Conclusions

This paper presents a RMC method for LU classification of multispectral RS imagery. The classification cell is region, and the classificatory pattern is represented by infra-class center set of the training samples. The method fully considers the characteristics that a kind of LU class is covered with several ground objects, and the classificatory pattern does not meet multivariate statistical distribution. To RMC method, the multispectral feature is more dispersal, the classification accuracy is improved greatly. There are two classification rules of RMC method, one is the type amount of classificatory infra-class centers which the pixel of the region cell belonged to, the other is the percentage of the pixels belonged to the class from the whole region cell pixels. The rules express sufficiently the basic idea of LU classification system and classification based on the area ratio of ground objects to whole region. Meanwhile, it fuses the region information and the spectral information. Owing to the large difference of RMC feature, it is simple to establish the classification rules and to select training samples.

RMC method which multi-centers confirmed by clustering solve preferably the problems of the division of subset and the confirmation of boundary of decision tree method. Furthermore, the method can be applied to single class recognition and can adopt different classification rules for different classes. RMC method can increase the LU classification accuracy of multispectral RS imagery between 4% and 6%.

Acknowledgements

References

1. Cingolania, A. M., Renisona, D.: Mapping vegetation in a heterogeneous mountain rangeland using landsat data: an alternative method to define and classify land-cover units. J. Remote Sensing of Environment. 84--97 (2004)

2. Jian Lin: Region multi-center method for land use classification of multispectral RS imagery. J. Journal of Remote Sensing. 165--172 (2010)

3. Xuehong Chen: Study on Combined Classifier Based on Error Analysis. J. Journal of Remote Sensing. 683--691 (2008)

4. Min Dong, Hong Zhu: The Calculation Method of Inverse Camera Response Function of a Single Image based on Kernel Estimation. J. Computer Engineering and Applications. 1--6 (2010)

5. Jie Chen: Rough set theory based object-oriented classification of high resolution remotely sensed imagery. J. Journal of Remote Sensing. 1139--1146 (2010)

6. Xiaoping LIU, Xia LI: Classification of Remote Sensing Images based on Ant Colony Optimization. J. Journal of Remote Sensing. 253--262(2008)

7. Joon Hur, JongWoo Kim: A Hybrid Classification Method Using Error Pattern Modeling. J. Expert Systems with Applications. 231--241(2008)

8. Lucas, R., Rowlands, A.: Rule-based classification of multi-temporal satellite imagery for habitat and agricultural land cover mapping. J. ISPRS Journal of Photogrammetry & Remote Sensing. 165--185 (2007)

9. Dianhui Mao: Remote sensing image classification based on formal concept analysis. J. Journal of Remote Sensing. 90--103 (2010)

10. Bian, F. L. and Wan ya: A novel spatial co-location pattern mining algorithm based on k-Nearest feature relationship. J. Geomatics and Information Science of Wuhan University. 331--334(2009)

11. Qiao Cheng: Remote sensing image classification method supported by spatial adjacency. J. Journal of Remote Sensing.88--99 (2011)

12. Hanhui Dan: An improved K –means cluster algorithm. J. Journal of Chongqing Technology and Business University (Natural Science Edition). 144—147 (2009)