The Classification Method of Multi-spectral Remote Sensing Images Based on Self-adaptive Minimum Distance Adjustment

Junhua Liu[1, 2], Chengming Zhang[2, 3,*] , Shujing Wan[3]

[1] Chinese Academy of Surveying and Mapping, Beijing, China

[2] Shandong Agricultural Science Colleague , Jinan, China

[3] School of Information Science &Engineering, Shandong Agricultural University， Taian, China

Abstract. The phenomenon of "Same Object with Different Spectra" in the issue of multi-spectral remote sensing images land use classification makes major effects on improving accuracy. The paper based on the analysis of modeling on classification problems, proposed a method based on minimum distance self-adaptive adjustment to realize the split of cluster centers and solved the problem of identified scope intersection leading to improving the accuracy in the classifying methods difficultly. By experiments compared with the traditional methods, it can improve classification accuracy about 4% and the results prove the validity of this method.

1 Introduction

Multi-spectral remote sensing images are the main data source in the land use applications. Studying the appropriate classification algorithm according to the fundamental characteristics of the multi-spectral remote sensing images is the primary means to obtain high-precision land-use information.

In the 1980s, it was mainly used statistical pattern recognition methods to classify remote sensing images by computer [1]. Since the 1990s, it emerged a large number of remote sensing image classification methods, such as the artificial intelligence classification, combination of remote sensing and GIS, object-oriented classification, composite classification and so on which achieved better results [2]. Dixon et al. took the support vector machine method to do land use classification of TM images that obtained the best classification accuracy [3].

Wardlow et al. used the decision tree method to classify crop in the central Plains of United States, which gained the overall classification accuracy of better than 80% [4]. Taochao et al. [5] came up with a high-resolution remote sensing image classification method based on probabilistic platen semantic model for the phenomenon of "same spectrum with different objects" in high-resolution remote sensing image. Li Gang proposed an uncertainty classification technology based on high-dimensional cloud model and RBF neural network modified [6].

It wasn't satisfied with the traditional remote sensing image classification methods in the aspects of automation, intelligent, and classification accuracy [7]; although the technology based on neural network has the advantages of strong fault tolerance and adaptability, it existed structure selected difficult, too fast of local convergence and study process difficulty controlling in the practical application, which resulted that was not wholly better than the traditional technologies; the method based on decision tree was not ideal in the treatment of the boundary [8,9].

For the fundamental characteristics of multi-spectral remote sensing images, with comprehensive considering the speed and accuracy of land use classification, the paper proposed a classification method of introducing self-adaptive minimum distance adjustment in order to better meet the needs of the actual production.

2 Self-adaptive Minimum Distance Adjustment Classification Method Principle

The existing remote sensing classification methods research more on how to improve the formation of the cluster center, but in judgment the attribution problem of the pixel to be classified it usually adopts the way that calculation the distance of pixel to be classified to the cluster center first, then taking it into the category with minimum distance, which results in the difficult to improve the classification accuracy.

2.1 Remote Sensing Image Classification Problem Modeling

Under the ideal conditions, the eigenvector of pixels representing similar surface features in remote sensing images will distribute in the same feature space regions; however, as different surface features with different eigenvectors, they should distribute in different surface features, which means that any kind of feature in images has only one gray value scope in any band for a remote sensing image to be classifies with b bands and for some feature c, it can obtain the value set of the pixels in every band, as Figure 1 shown:

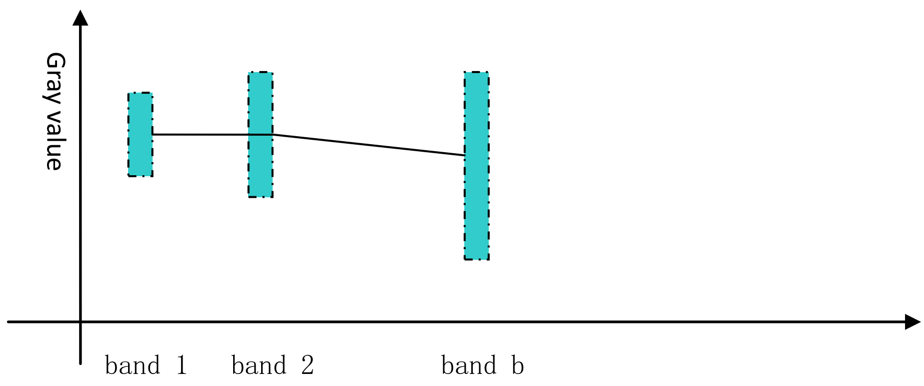


Fig. 1. The schematic diagram of spectral range under the ideal conditions

To the band i of class c, if use li as the lower bound of the gray value and hi as the upper, in the band i, the value of its center mi is: $m_i = \frac{(h_i + l_i)}{2}$ and based on it, it can define m=(m1,m2,…,mb)T as the center feature vector of the class which can describe the adaptive changes of the cluster center better. Use ri as the permissible error radius surrounded, by the analysis above, the value of ri is |hi-li|/2 and it can name all the vectors made by maximum permissible error of all the bands as effective radius vector. For pixel x to be classified in class c, compare the distance of gray value of each band to the surrounding component of the center feature vector, and if all the distances don't exceed the surrounding permissible radius, take it into class c.

2.2 The Processing Method of "Same Object with Different Spectra" Problem

The "Same Object with Different Spectra" phenomenon for a variety of reasons makes the error value in some bands or all bands too large, which lead to the feature vector of the class distributing too scattering in the feature space and difficult to attribute to same cluster as excepted. To better illustrate this problem, now explain it with an example of a two-band image classification, assuming that the image will be divided into two classes A, B and the pixels distribution of each class are shown in Figure 2.

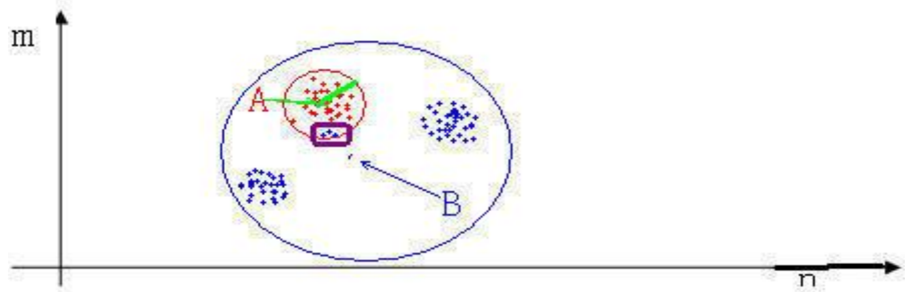


Fig. 2. The sketch map of "Same Object with Different Spectra" resulting in classification error

In Figure 2, for the pixel of class B distributing in two regions, the distance of cluster centers generated with two regions is too large leading to the range identified (blue circles) too large. As a more extreme case, the area of another class A is surrounded in the scope identified by class B and the part surrounded by an orange rectangle should be divided into pixels of class A but to class B.

From the analysis above, we can see that the root cause of classification accuracy reducing is too large of cluster centers and identified radius leading to the identified scope of two cluster centers intersect. For this, it can split cluster centers by adaptive minimum distance adjustment to eliminate intersection of identification range which can improve the classification accuracy effectively.

3 The Method Realization of Self-adaptive Minimum Distance Adjustment Classification

3.1 The Process of Self-adaptive Minimum Distance Adjustment

Based on the second analysis, it gives the method of self-adaptive minimum distance adjustment: doing the intersecting judgment to the cluster centers after the end of the sample study. If there are the intersecting parts generated on the basis of the two cluster centers' identification radius, it needs to split the cluster centers. The split steps are:

(1)Calculate the cluster centers generated one to another to determine whether they intersect. If intersect, process according to (2).

(2)Split the bigger one of identified scope in the two cluster centers. Select any sample from the classes to be split, choose a value of less than the biggest identified distance as threshold randomly, divide the samples into different subsets, study separately for each resulting in each cluster center and estimate with the classes not split. If there are still intersecting with the cluster center of some subset, select a less threshold and split again.

3.2 Algorithm Design

The algorithm uses K-Means algorithm to split the cluster center into two subsets each time and do sample set division in the way of binary tree. Each class generates a binary tree named as the subset tree of the class. A node of binary tree is corresponding to a sphere. Record each radius of sphere centre and the surrounding sample subset. The center of the sphere is defined as the center of the sample subset in this node and the radius is the maximum of Euclidean distance from the sample point on that node to the center of the sphere. The sphere obtained by the subdivision of a node's subset is said in two child nodes of it.

The classification steps of adaptive minimum distance classifier towards points to be classified i are shown in Figure 3, the specific steps:

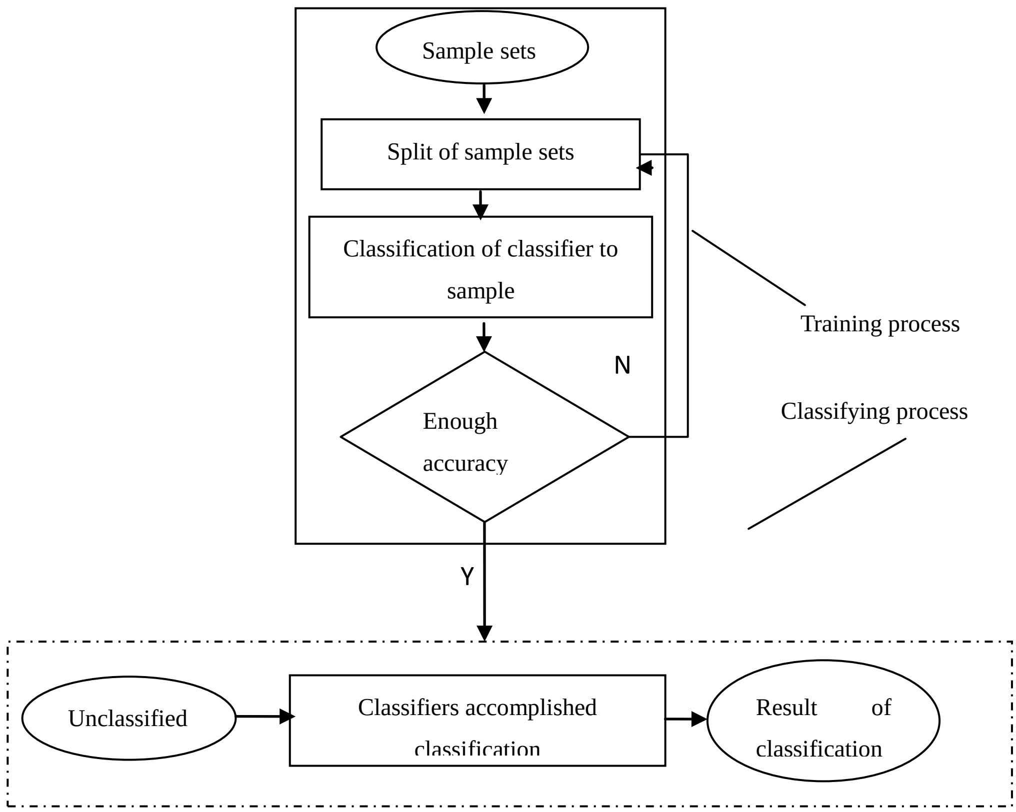


Fig. 3. The flowchart of self-adaptive minimum distance classification algorithm

(1) Find out the distance D of the point to subset tree corresponded by each class. The distance D(T,P) of unclassified point P to the subset tree is defined as:

A. If the centre of sphere Euclidean distance d corresponds to the root node of P to T is greater than twice sphere radius of the node, ignore all globules subdivided by the node and let D = d;

B. If the root of T is already the leaf node, let D=d;

C. If it is not satisfied with A and B, definite recursively as the less distance D1 and D2 of P to the left and right sub-tree T1 and T2 of T.

(2) Give the points to be classified to the class of a subset tree with minimum distance to D.

With using binary tree search in the distance calculation, there are a part of nodes joined in finding out the distance D to reduce the number of joining in the real calculation which guarantees that calculated amount will not increase much under the condition of sample number or sphere subdivided increasing.

Distance calculation using the only part of the nodes participate in the strike of the distance D, can effectively reduce the actual participation of computing d, the number does not guarantee the number of samples increases, or broken down to get the sphere increased computation does not increase.

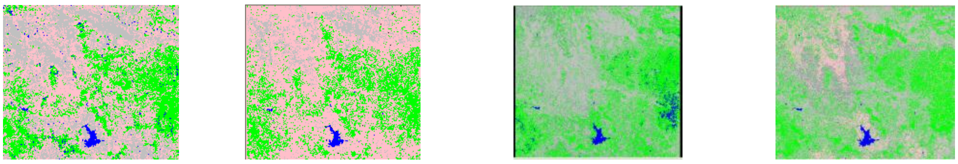
4 Experiment and Discussion

The paper selected Xueye basin of Laiwu City, Shandong Province as the study area. It selected three different kinds land use categories of water, forest and bare soil to classify and compare.

4.1 The Experimental Results and Analysis of Different Classification Methods in XueYe Basin

The data sources used in the experiment are ETM+ images OF 30m with a total of eight bands and the size of 1500 pixels × 1200 pixels acquired in May 31, 2007.

4.1.1 The classification of common ways



(a)Minimum distance (b)Maximum likelihood (c) KNN（K=19） (d) BP natural network

Fig. 4. The classification results of XueYe Reservoir

From the classification results of figure 4, it can be seen that with using the minimum distance method, the number of pixels to be classified increased significantly, the phenomenon of mixed classification with forest is more serious and the misclassification of the water affected by shadow was existence. K-nearest neighbor method can distinguish water and bare soil well, but recognize forest not well with classifying a part as water. Although the maximum likelihood classification method can classify water well, it exits serious mistake phenomenon of classifying the forest and bare soil into the kind to be classified. K-nearest neighbor method can distinguish water and bare soil well, but it cannot recognize forest commendably with classifying a part to water. BP neural network has better effects to classify each kind, but a part of vegetation still not distinguishes. The classification accuracy is shown in Table 1.

Table 1. Classification accuracy of common ways

Classification	water	forest	Bare soil	Average
Maximum likelihood	85.6	84.1	86.5	85.4
Minimum distance	87.7	86.8	85.3	86.5

KNN（K=19）	90.2	85.7	85.1	87.0
BP natural network	87.0	85.2	85.9	86.0

4.1.2 The Process and Result of Self-adaptive Minimum Distance Classification

While using the self-adaptive minimum distance classification, the experimental results are shown in the Figure 5.

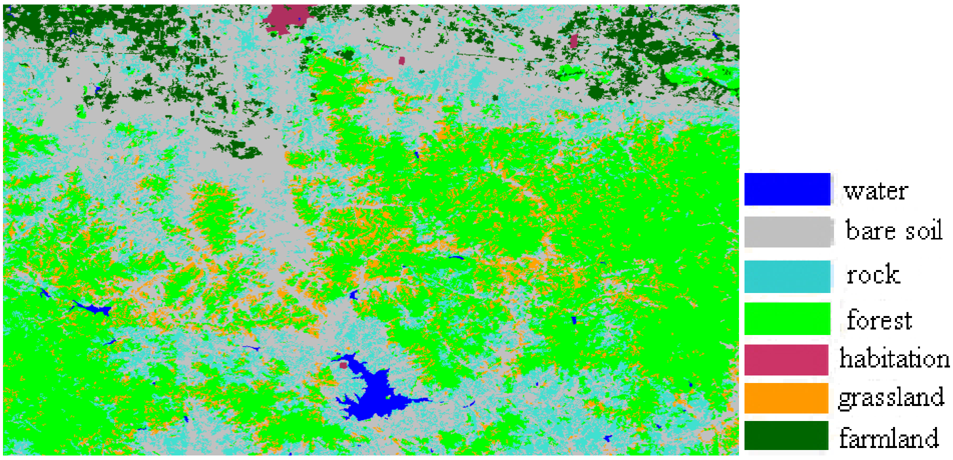


Fig. 5. The extraction results of XueYe reservoir

The classification accuracy of different training samples and category schema is shown in Table 2.

Table 2. The classification accuracy of different training samples and category schema

Water	Forest	Bare soil	Average
91.3	92.8	89.2	91.1

Contrasting the classification results of different methods, it showed that using the method of paper to classify, the pixel in the results obviously reduced and the accuracy increase 1.06% in average.

Acknowledgments. This study has been Funded by (1) Basic Research Fund of Chinese Academy of Surveying and Mapping under Grant (Contract Number: 7771109), (2) The Research Award Fund for Outstanding Young and Mid-career Scientists in Shandong (Contract Number: BS2011DX031), (3) Technology and Evolving Countries Program in Shandong (2012GGB01093), (4) Basic Research Fund of Chinese Academy of Surveying and Mapping under Grant (Contract Number: 7771211). Among them, (1) and (4) are the basic business founding of CASM.

References

1. Laiping Luo. The Research of Fuzzy Pattern Recognition and Decision Tree Methods Application in Remote Sensing Image Classification [D]. Capital Normal University，2006.
2. Shi Zepeng, Ma Youhua, Wang Yujia. Review on the Classification Methods of Land Use/Cover Based on Remote Sensing Image [J]. Chinese Agricultural Science Bulletin, 2012,28(12):273-278.
3. Dixon B, Candade N. International Journal of Remote Sensing, 2008, 29: 1185.
4. Wardlow B D, Egbert S L. Remote Sensing of Environment, 2008, 112: 1096.
5. TAO Chao, TAN Yihua, PENG Bifa1. A Probabilistic Latent Semantic Analysis Based Classification for High Resolution Remotely Sensed Imagery [J]. Acta Geodaetica et Cartographica Sinica, 2011, 40(2):155-161.
6. LI Gang, WAN You-chuan. Uncertainty classification method of remote sensing image based on high-dimensional cloud model and RBF neural network [J]. Science of Surveying and Mapping, 2012,37(1):115-118.
7. Ruru Qian. The Comparative Study of Remote Sensing Images Classification Methods [D].Chang' an University, 2007.
8. JIA Kun; 2; LI Qiangzi; TIAN Yichen. A Review of Classification Methods of Remote Sensing Imagery [J]. Spectroscopy and Spectral Analysis, 2011, 31(10):2618-2623.
9. GUO Yaqin, WANG Zhengqun,LE Xiaorong. Minimum distance classifier ensemble based on adaptive distance metric. Journal of Computer Applications, 2006, 26(7):1703-1706