

Research Progress of Computer Automatic Classification Technology and Methods Based on Remote Sensing Images

Zhang Hebing

School of Surveying and Land Information Engineering
Henan Polytechnic University
Jiaozuo, China, 454000

Wang Shidong

1. College of Resources Science & Technology, Beijing
Normal University, Beijing, China, 100875
2. School of Surveying and Land Information
Engineering Henan Polytechnic University, Jiaozuo,
China, 454000

Abstract—In this paper, the research progress of computer automatic classification technology and methods based on remote sensing images is introduced. Firstly, the traditional classification methods of remote sensing image are discussed, which include supervised classification and unsupervised classification. Finally, some new technologies and methods of remote sensing images classification at home and abroad.

Keywords—Remote sensing image, Classification methods, Supervised classification, Unsupervised classification

I. INTRODUCTION

For remote sensing technology, it is an important part of remote sensing technology to interpret and identify various ground objects from remote sensing images. Whether thematic information extraction, dynamic monitoring, thematic mapping, or remote database establishment, remote sensing image classification is very important to them. Remote sensing image classification is that put each pixel or region of image into a certain category or one of thematic elements; namely, select characteristic parameters through analyzing spectral signature of various types of ground objects, divide feature space into non-overlapping sub-space; then put each pixel of image into various sub-space which represent a type of actual ground objects, so classification is realized.

Remote sensing image classification is a very important application direction of digital image classification, which has many similarities as universal image classification. However, remote sensing images have some unique characteristics, if computer can be used to classify automatically remote sensing images according to certain meaning, it will be easier to extract and analyze vast amounts of remote sensing data. Thus, there have been researchers in remote sensing science to improve and innovate remote sensing image classification algorithmic.

The computer classification methods of remote sensing images are divided into two types: statistical pattern method and syntax pattern method. The common classification method is statistical identification pattern, such as maximum likelihood method and K minimum distance discrimination method. In recent years, most new classification methods

belong to syntax pattern method, including artificial neural network, fuzzy mathematics, expert systems and decision tree classification method and so on.

The statistical classification methods of remote sensing images are divided into two types: unsupervised classification and supervised classification. Unsupervised classification is the process that for remote sensing images without prior knowledge, only depends to the statistical difference of combination of different spectroscopic data, and then validates ground objects according to properties of various classified objects. Supervised classification is the process that for remote sensing images with prior knowledge, that is, its corresponding ground objects type is known, so can judge the types of non-sample data according to the type characters of samples.

II. TRADITIONAL CLASSIFICATION METHODS

A. Supervised classification

Supervised classification also can be called training field method that is classification technology according to typical sample training based on theoretical basis of establishing statistical identification function. Supervised classification is a method of pattern recognition that to classify remote sensing images by establishing discriminant function, according to known samples of training area, obtained characteristic parameters as the decision rules through selecting feature parameters. The training area must be typical and representative. If discriminant rule meet accuracy of classification, this rule is available; otherwise, need to re-establish the classification decision rules until meet accuracy of classification. The common methods of supervised classification include: minimum distance method, maximum likelihood method, cassette classification method, neural networks method and fuzzy classification method.

1) Minimum distance method

The method is to obtain first-order origin moment of each sample in feature space, then compare the Euclidean distance (or other distance such as Mahalanobis distance, taxi distance, etc.) of points to categories points and determine a category of minimum distance as result category. The method is similar to K-means method that do not need iterative and cluster centers have been identified.

2) Maximum likelihood method

Maximum likelihood method is a classification method that uses Bayes' rule and a classification method that minimum incorrect probability in terms of statistical rules. After calculating first-order origin moment and second-order central moment of samples, probability distribution is estimated, and then the points are classified into a category of maximum probability according to the posterior probability discriminant formula. After the membership in fuzzy theory is proposed, Bayes' rule can be thought as a selected mode of membership function and in statistical sense; it is consistent with overall maximum likelihood normal distribution requirements.

3) Cassette classification method

The basic idea is to first find location and shape of each category in the feature space through training area data, and then take a "box" with the cluster as discriminant function. Discriminant rule is that if an unknown vector X drop into the "box" and X is divided into the category, else it should be compared with the other box.

4) Neural networks method

Artificial neural network is the bionics results based on imitation of biological neural system signal transmission and an important algorithm in artificial intelligence. At present, artificial neural network has now developed into multi-layer from monolayer and has new some dynamic characteristics, such as BP algorithm and Hopfield algorithm. Artificial neural network algorithm needs prior knowledge, because its application of image classification is expressed as a supervised classification.

5) Fuzzy classification method

Because a lot of natural or semi-natural phenomena are difficult to clearly defined types, reflected in the remote sensing images, there are some mixed-pixel problems, and lots of phenomena different objects with same spectrum and same objects with different spectrum appear, which make pixels categories are difficult to clearly identify. For fuzzy classification, the training process ambiguity of supervised classification is ignored, using the traditional method, it is assumed that training samples is composed of a set of pixels that can be clearly defined and classified and is representative. The fuzzy classification of supervised classification can use characteristics of neural networks with good learning inductive mechanism, robust capabilities and ease of expansion into a dynamic system; and design a fuzzy classification method based on neural network technology. Fuzzy neural network model developed from ARTMAP to FasART to simplified FasART which make the functions of fuzzy neural network supervised classification are improved and its classification accuracy enhanced increasing.

Supervised classification has some certain advantages than unsupervised classification. But there is inevitably misclassification. The classification method need manual visual interpretation before classification and has a stronger dependence on priori knowledge of ground objects properties.

B. Unsupervised classification

Unsupervised classification refers to take "blind" classification with no any priori knowledge and only depend on data (distribution regularities of spectrum characteristics of remote sensing images) namely features of natural clustering. The classification results achieve distinction between different categories, but can not determine the attributes of categories; that is to say, the samples of unsupervised classification can be divided into several categories only, the attributes of classes can not be described. The classed attributes are defined by visual interpretation and field survey after classification. Unsupervised classification also is called as cluster analysis. General clustering algorithm is to first select several pattern points as the cluster centers. Each center represents a category, which put each pattern into category that each cluster center belongs to and realize initial classification. Then initial classification is determined according to clustering rule, if the result is not reasonable, the classification should be modified; and iteration is repeated until get a reasonable classification result. Unsupervised classification finds same categories by learning and then separates the class from other classes, which is different from supervised classification. Unsupervised classification and supervised classification take gray value of images as base which realizes classification by calculating some statistical parameters such as mean and covariance. So there are some commonalities between them. Common unsupervised classification methods include K-means method, ISODATA method, parallel pipeline clustering method, clustering systematically method and splitting method.

1) K-means method

The rule of K-means algorithm is that makes sum of squares of distance that multi-pattern point to center of category. The basic idea is that moves the centers of every category by iteration until get the best clustering results.

2) ISODATA method

The full name of ISODATA is "Iterative Self-Organizing Data Analysis Technique". ISODATA realizes clustering by using minimum spectral distance equation. The essence of ISODATA is the process that gets initial categories as "seeds" and cluster automatic iteration according to a discriminant rule. Between two iterations, the last iteration clustering result is analyzed statistically. Then the existing categories are canceled, split and merged based on statistical parameters, and continue to the next iteration until over maximum number of iterations or meet the classification parameters (threshold) and the classification process completes.

3) Parallel pipeline clustering method

This method is relatively simple, which takes the spectral curves as base and takes the similar of spectral curves of same category as discrimination criterion. After setting a similar threshold, the shape of same category objects in feature space is a "parallel pipeline" that takes

characteristic curve as center and takes threshold as radius. The essence of this clustering is a method of trial based on nearest neighbor rule.

4) *Clustering systematically method*

This approach is that regard all pixels of image as a class, calculate correlation coefficients matrix of mean between classes, choose the most relevant two classes to merge into two new class, re-calculate correlation coefficient matrix of each class, merge the most relevant two classes into a class, so that to continue, and merge classes according to the method of gradually integrating. Until all the correlation coefficient between new classes is less than a given threshold value so far.

5) *Splitting method*

Splitting method can also be called mixed distance classification that is inverse with clustering systematically classification, which at the beginning, all pixels are regarded as a class, and calculate mean and standard deviation of variables, and then calculate the center of two split categories by using a certain formula and calculate the clusters of each pixel to the two categories, and the pixels are merged to the category with minimum distance to form two new classes. Then each new class is classified, as long as the mean square deviation of any band is larger than specified threshold, the new class need to be split.

Unsupervised classification are widely used because it has many advantages that is affected less by human factors according to spectral characteristics of ground objects, and does not require detailed knowledge of ground information, what's more, its algorithm is mature and operation is simple. In the preliminary analysis of images, it is very valuable to use unsupervised classification method to study the natural cluster distribution of the data.

III. RESEARCH PROGRESS OF NEW METHODS

Both supervised classification and unsupervised classification realize classification based on point independent principles of spectrum characteristics and are statistical methods. These methods realize classification based on statistical characteristics of gray data of each band, because of limit of the resolution of satellite remote sensing data, many pixels are mixed pixels in a image with characteristics of mixed spectral information, resulting facing many fuzzy objects in the course of computer classification that can not determine these pixels belong to which class. Moreover, the phenomena that same objects with different spectrum and same spectrum belonging to different classes are widespread and can lead to misclassification, so people try new ways constantly to improve classification methods.

In recent years, many scholars in the field at home and abroad proposed a lot of new research methods such as decision tree classification method, comprehensive threshold method, expert system classification, multi-feature fusion methods, neural network classification method and classification method based on frequency spectrum.

A. *Research progress abroad*

In recent years, many foreign scholars made a large number of experimental studies on remote sensing image classification methods and proposed some new methods that has been applied in practice and achieved very good results.

Adel Shalaby monitored land cover and land use changes in the northwestern coastal zone of Egypt using remote sensing and GIS [1]. In his study, maximum likelihood supervised classification and post-classification change detection techniques were applied to Landsat images acquired in 1987 and 2001, respectively, to map land cover changes in the Northwestern coast of Egypt. A supervised classification was carried out on the six reflective bands for the two images individually with the aid of ground truth data. Ground truth information collected during six field trips conducted between 1998 and 2002 and land cover map of 1987 were used to assess the accuracy of the classification results. Using ancillary data, visual interpretation and expert knowledge of the area through GIS further refined the classification results. Post-classification change detection technique was used to produce change image through cross-tabulation.

VIDYA MANIAN 和 MIGUEL VELEZ-REYES proposed a novel wavelet and support vector machine (SVM) based method for hyperspectral image classification [2]. A 1-D wavelet transform is applied to the pixel spectra, followed by feature extraction and SVM classification. Contrary to the traditional method of using pixel spectra with SVM classifier, the approach not only reduces the dimension of the input pixel feature vector but also improves the classification accuracy. Texture energy features computed in the spectral dimension are mapped using polynomial kernels and used for training the SVM classifier. Results with AVIRIS and other hyperspectral images for land cover and benthic habitat classification are presented. The accuracy of the method with limited training sets and computational burden is assessed.

Peter F. Fisher proposed a classification method of remote sensing of land cover classes as type 2 fuzzy sets [4]. In recognition of the idea that land cover classes are vague phenomena, a number of researchers have advocated modeling of Land Cover as fuzzy sets, as opposed to Boolean sets. Once vagueness is admitted, it is not sufficient to stop at modeling land cover simply as a fuzzy set, because any threshold map (α -cut) of the fuzzy set is itself a Boolean set. Rather it may be appropriate to model it as a type 2 fuzzy set recognizing the existence of higher order vagueness. Using the example of estimating land cover extents in an area of the forest-savanna ecotone in Bolivia, he presents the results and possible advantages of a type 2 fuzzy analysis of land cover. The consideration of type 2 fuzzy sets reveals the depth of the uncertainty, and allows the analyst to be specific about minimum, maximum, and average extents of land cover types and even to report fuzzy area itself as a fuzzy number and to justify descriptive qualifications of the results. The nuanced analysis, which is possible with this approach,

seems an advance for the scientific community concerned with the conceptualization of land cover classes, but it may be a nuisance for those involved in policy matters.

B. Research progress in China

In recent years, the Chinese scholars keep up with international technological frontier, made a lot of experimental research work in the field of remote sensing image classification, and try to propose some new classification methods which achieved good results in practice. These new methods are as following.

Liu Li established a complex classification model based on comprehensive analysis of hierarchical classification and supervised classification^[4]. The method was tested by using SPOT images and it was proved that the result accuracy of this method was higher than single supervised classification.

Zhou Xingdong^[5] proposed a combined threshold value image classification method through combining different bands based on spectral features of various ground objects in Xuzhou and summarizing the method and approach of obtaining spectral information of various objects. The study showed that this method can distinguish classes that was not easy to distinguish such as town land and nivation land, can reduce impact of mixed pixels, and can improve the accuracy of land use classification.

Sun Xiubang thought that land cover classification based on expert system can solve complex problems in the process of classification with expert ability^[6]. However, in practice, the knowledge discovery and definition are often very difficult to be consistent with actual situation one-to-one.

Li Jinlian^[7] taken Anji region of Taihu drainage area as the test area, firstly, compressed and enhanced geometry information from SPOT5 image data using principal component analysis method, secondly, filtered images and processed noise using wavelet analysis method, analyzed texture information of high resolution images using gray symbiotic matrix, overlapped partition result and classification result based on NDVI threshold value method, finally, get ultimate classification results. The study show that can distinguish ground objects with spectral confusion by applying texture analysis method in image classification, and the classification accuracy based on combination spectrum feature with texture feature was higher than the method that only depends on spectrum feature or only depends on texture feature.

Sun Xiubang proposed a land cover remote sensing image classification method by transforming time domain to frequency domain from original time series. However this method is established based on variation of NDVI time series, it can not distinguish land classes without obvious changes and need should be complementary with other classification method in the application process. In addition, the classification accuracy also is affected by some factors

such as spatial resolution, lack of different features green degrees curve and biased factors^[8].

Generally speaking, in recent years, most of research in remote sensing image classification field combines traditional methods with new methods^[9, 10]. That is to say, in the basis of unsupervised classification and supervise classification, the new methods are used to improve traditional methods which can reduce misclassification, enhance accuracy of classification and achieved good effect in actual application.

IV. CONCLUSIONS

These traditional remote sensing image classification methods such as visual interpretation, supervised classification and unsupervised classification are widely used at present because these algorithms are mature and simple operation. In recent years, the new methods such as decision tree classification method, comprehensive threshold method, expert system classification, multi-feature fusion methods, neural network classification method and classification method based on frequency spectrum can extract more accurately target objects and can improve image classification in different degrees. In actual application, these new methods are usually combined with traditional classification methods. With continuous development of remote sensing technology, resolution of remote sensing images has greatly improved, but the image classification technology couldn't keep up with development of remote sensing technology. Although many scholars proposed many classification methods, no one is universal and efficient. With further development and mutual infiltration of remote sensing technology and computer technology, we can obtain more and more, more and more comprehensive information using remote sensing approaches, so research of new algorithms and new theories will become more and more thorough; if remote sensing combines with other information and features, the classification accuracy of remote sensing images will be improved remarkably.

ACKNOWLEDGMENT

The study is supported by Henan Provincial Education Office Research Plan (2010B420001) and Fund of Henan Polytechnic University Young Teachers (Q2010-42).

REFERENCES

- [1] Adel Shalaby. Remote sensing and GIS for mapping and monitoring land cover and land-use changes in the Northwestern coastal zone of Egypt [J]. Applied Geography, 2007(27): 28-41.
- [2] VIDYA MANIAN. SUPPORT VECTOR CLASSIFICATION OF LAND COVER AND BENTHIC HABITAT FROM HYPERSPECTRAL IMAGES [J]. International Journal of High Speed Electronics and Systems, 2008, 18(2): 337-348.
- [3] Peter F. Fisher. Remote sensing of land cover classes as type 2 fuzzy sets [J]. Remote Sensing of Environment, 2010(114): 309-321.
- [4] Liu Li, Yu Qiang. A Study on A Classification Method of Remote Sensing Combined Stratified Classification with Supervised Classification [J]. Forest Inventory and Planning, 2007, 32 (4): 37-39, 44.

- [5] Zhou Xingdong, Yu Shengwen, Zhao Changsheng et al. Study on land use classification base on remote sensing image [J]. Research of Soil and Water Conservation, 2007, 14 (4): 39-42.