Image Classification In Remote Sensing

P. Sathya[1], Dr. V. Baby deepa[2]
[1]Dept of computer science
[2] Assistant Professor, Dept of computer science
[1, 2] Government Arts College, Tamilnadu,

Abstract- *Classification is a particular case of Pattern Recognition. The important objective of the classification procedure is to automatically classify all pixels from an image into land cover classes based on the predefined classification model. In image classification pattern refers to the set of radiance measurements obtained in the various wavelength bands for each pixel. There are numerous classification algorithms. This paper gives most popular classifiers in the field of remote sensing. Classifiers are described under broad categories such as supervised and unsupervised classifiers, parametric and non-parametric, fuzzy classifiers and knowledge base classifiers. Only maximum likelihood and decision tree would be explained in this paper under parametric and non-parametric classifiers respectively.*

Keywords- Classifier, Pattern recognition, Measurement, Decision tree

I. INTRODUCTION

In remote sensing the supervised classification is one of the most important tasks image definitions, in which the image pixels are classified to different established land use/land cover classes based on the spectral reflectance values at various bands. In nature some classes have very close spectral reflectance values and therefore they overlap in the feature space. This makes spectral confusion among the classes and results in incorrect classified images. It is very difficult to classify some classes correctly using traditional parametric classifier like Maximum Likelihood Classifier. The other disadvantage of these traditional parametric classifiers is that they need sample data to be normally distributed. Due to this actuality it becomes difficult to add additional layers into classification procedures to increase accuracy of the classification.

Additional layers generally have bi- or multimodal distribution. Therefore maximum likelihood classifications don't produce quality results after adding of additional layers. To remove such spectral ignorance we need extra spectral and spatial information. Additional data can come from the additional information or from experience and knowledge of domain experts. Classification can done by using such knowledge is known as knowledge base classification. But some classification wants strong knowledge base, which sometimes become disadvantage of this process because of the knowledge acquisition process. Generally knowledge base is created with the help of knowledge acquired by interacting with the experts.

The traditional way of knowledge acquisition is that the knowledge engineer interacts with the corresponding domain expert; write up his /her experience and knowledge in a interpretable form and then feed the entire acquired knowledge in the computer in a symbolic form such as if-then rules.

This is usually long and repeated process takes too much of time. It is also not always possible that expert is available all the time. There is large amount of knowledge hidden in spatial databases that can be used in classification of satellite images. Some knowledge can be extracted by simple GIS query and other knowledge is so deep such as classification rules, spatial distribution rules, spatial association rules that are not stored explicitly in the database but we can be extracted them by computation and machine learning process. The question arises, from where and how we can extract this hidden knowledge automatically? This thesis explores a decision tree classifier based on machine learning to extract knowledge in the form of classification rules from the satellite and topographical.

II. SUPERVISED CLASSIFICATION

In this type of classification the image analyst supervises the pixel categorisation process by identifying to the algorithm particular information of the different land cover types available in a location. To do this, representative sample site of known cover type, called training areas, are applied to compose a numerical interpretation key that describes the spectral attributes for each feature type of interest. Reflectance value of each pixel in the image is then compared numerically to each class in the interpretation key mark with the name of the section it looks most like. Generally there are three major steps involved in the typical supervised classification procedures as follows:

Training Stage: The analyst identifies representative training areas and develops a numerical description of the spectral attributes of each land cover type of interest in the scene.
Classification Stage: Each pixel in the image is categorized into the land cover class it most resembles. The pixel is labelled as unknown when it was not matching to any predefined class.
Accuracy Assessment: To check the accuracy of the classification, the classified image is compared with some reference image or Ground truth

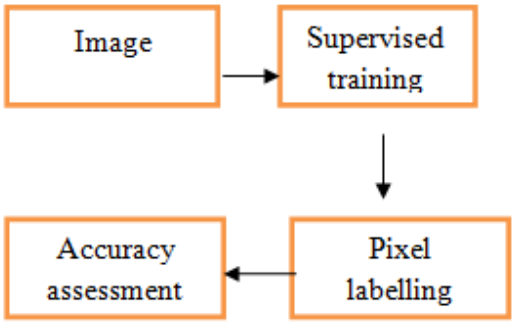

Figure1: Supervised classification process

III. UNSUPERVISED CLASSIFICATION

Unlike supervised classifiers, unsupervised ones do not utilize training data as the basis for classification. These classifiers try to aggregate reflectance value of pixels in the feature space into well separated clusters. Clusters are considered as classes. Once the spectral grouping has been done, the analyst identifies the obtained classes to some form of reference data. There are numerous clustering algorithms that can be used to determine the natural spectral clusters present in the image. The "*K*-means" is a most common algorithm. In this approach user has to define the number of clusters or classes to be located in the image. The algorithm automatically locates the centre means of various clusters present in the image and each pixel in the image is then assigned to the cluster whose mean is closest. After all pixels have been classified, revised mean vectors for each of the cluster is computed. The whole process is repeated again until there is no further change in the location of class means vectors.

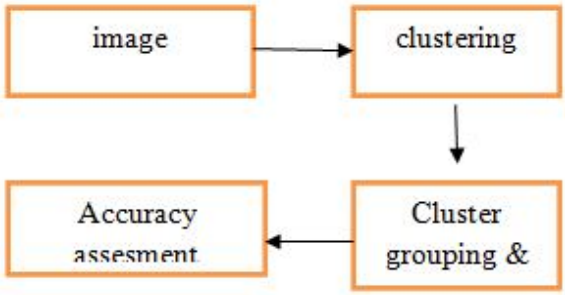

Figure2: Unsupervised classification process

IV. PARAMETRIC CLASSIFIER

Parametric classification algorithms assume that the observed measurement vectors *Xc* obtained during the training phase of the supervised classification for each class in each spectral band follow some statistical distribution such as Gaussian distribution. The major parametric classifiers under this category are minimum distance, Mahalanobis distance, and maximum likelihood classifier. Maximum likelihood gives better accuracy than others and frequently used in the remote sensing image classification. Therefore Maximum likelihood algorithm is described here as a representative of parametric classifiers.

4.1 Maximum Likelihood Classification

The Maximum Likelihood Classification quantitatively evaluates both the variance and covariance of the category spectral response pattern when classifying an unknown pattern. An assumption is made that the distribution of the training set is Gaussian. Under this assumption, the distribution of a training set of a class can be completely described by the mean vector and covariance matrix. Given these parameters, we may compute the statistical probability of a given pixel being a member of a particular class. Multivariate normal statistical theory describes the probability that an observation *X* will occur, given that it belongs to a class *k*, as the following function:

$$\Phi_k(X_i)=(\pi 2)^{-1/2p} \left| \sum_k \right|^{-1/2} \times e^{-1/2(X-\mu_k)'\sum_k^{-1}(X-\mu_k)} \quad (1)$$

The quadratic product

$$X^2=(X-\mu_k)\sum^{-1}{}_k(X-\mu_k) \quad (2)$$

can be thought of as a squared distance function between the observation and the class mean as scaled and corrected for covariance and variance of the class. As applied in a maximum likelihood decision rule, Equation (1) is used allows the calculation of the probability that an observation is a member of each of *k* classes. The individual is then assigned

to the class for which the probability value is greatest. In an operational context, observed means, variances, and co-variances substituted by the *log* form of the Equation (1).

$$In[\phi(X_i)]=-1/2pin(2л)-1/2in\lceil\sum_k\lceil-1/2(X_i-m_k)D^{-1}{}_k(X_i-m_k) \quad (3)$$

Since the *log* of the probability is a monotonic increasing function of the probability, the decision can be made by comparing values for each class as calculated from the right hand side of this equation. A simpler decision rule, *R1*, can be derived from Equation (3) by eliminating the constants

***R1*:** Select *k* which minimizes

$$F_{(1,k)}(X_i)=In\lceil D_k\lceil+(X_i-m_k)\ D^{-1}{}_k(X_i-m_k) \quad (4)$$

4.2 Bayes Rule

An extension of the maximum likelihood classification approach is Bayesian classifier. The maximum likelihood decision rule can be modified easily to take into account in the population of observations as a whole. The prior probability itself is simply an estimate of the objects which will fall into a particular class.

These prior probabilities are sometimes termed "weights" since the modified classification rule will tend to weigh more heavily those classes with higher prior probabilities. For example, when classifying a pixel, the probability of the rarely occurring “sand” category might be weighted lightly, and the more likely “urban” class weighted heavily. The analyst gets the a priori probabilities by evaluating historical summaries of the region. The prior probabilities have proved to be a useful way of separating classes with similar spectral reflectance values. Prior probabilities are incorporated into the classification through a manipulation of the law of Conditional Probability. To begin, two probabilities are defined: *P(w k),* the probability that an observation will be drawn from class wk; and *P(Xi*), the probability of occurrence of the measurement vector *Xi.* The law of Conditional Probability or Bayes theorem states that

$$P\{w_k\lceil X_i\}=\phi_k(X_i)P\{w_k\}\div\sum_{k\text{-}1}{}^k\ \phi_k(X_i)P\{w_k\}=\phi_k(X_i)\div\sum_{k\text{-}1}{}^k\ \phi_k(X_i) \quad (5)$$

This Equation provides the basis for the decision rule which includes prior probabilities. Since the denominator remains constant for all classes, the observation is simply assigned to the class for which F *k (Xi)* the product of F *k (Xi*) and *P*{*w k*}, is a maximum.

V. NON-PARAMETRIC CLASSIFIER

A non- parametric classifier is not based on statistics, therefore, it is independent of the properties of the data. Non-Parametric classification algorithm does not take into account the distribution of the training set. They do not require that that the observed measurement vectors *Xc* obtained for each class in each spectral band during the training phase of the supervised classification should follow Gaussian distribution. Best known classifiers in this category are parallelepiped, decision tree and neural network.

5.1 Decision Tree

Decision tree is nonparametric classifier. Decision tree is an example of machine learning algorithm. They involve a recursive partitioning of the feature space, based on a set of rules that are learned by an analysis of the training set. A tree structure is developed where at each branching a specific decision rule is implemented, which may involve one or more combinations of the attribute inputs. A new input vector then “travels” from the root node down through successive branches until it is placed in a specific class.

The thresholds used for each nodal decision are chosen using minimum entropy or minimum error measures. It is based on using the minimum number of bits to describe each decision at a node in the tree based on the frequency of each class at the node. With minimum entropy, the stopping criterion is based on the amount of information gained by a rule (the gain ratio). There are several well-established decision tree classifier implementations. Decision trees are not constrained by any lack of knowledge of the class distributions, as they do not try to model them in any way.

VI. FUZZY CLASSIFIER

Representing a geographical object is very difficult, as in most of the cases they do not have well defined boundaries, meaning that the boundaries between different phenomena are fuzzy, and/or there is heterogeneity within the class. If the class does not have sharp boundary then the assignment of the pixel to a class is uncertain and this uncertainty can be expressed by fuzzy class membership function. Fuzzy set theory provides useful concepts and methods to deal with uncertain information. It is achieved by applying a function called “membership function” on remotely sensed images. The set associated with a membership function and each element in this set has its own membership value towards that particular set. The membership values range between 0 and 1.

If the membership value of an element is 0, it means that, it does not belong to that set and if it is 1, then it belongs completely. But, in crisp sets, the membership value is 1 or 0.For fuzzy classification, this function takes values between 0 and 1. Therefore every pixel has certain membership values in every class. For example, a vegetation classification might include a pixel with grades of 0.68 for class "forest", 0.29 for class "urban" and 0.03 for "riverbed". We can see that pixel has higher membership value in class forest than other classes, and therefore it will be assigned to forest class.

VII. ACCURACY ASSESSMENTS

Classification process is not finished until its accuracy is assessed. Two sources of information are compared to perform accuracy assessment. (i) Remote sensing derived classification data and (ii) Reference test data. The relationship of above two sets is summarized in an error matrix, where columns represent the reference data while rows represent the classified data. An error matrix is a square array of numbers laid out in rows and columns that expresses the number of sample units assigns to a particular category relative to the actual category as verified in the field. From the error matrix various accuracies can be derived as explained in following sections.

7.1 Overall Accuracy

The overall accuracy is weighted by the number of samples (pixels) in each class, i.e. the sum of all samples on the diagonal divided by the total number of samples. However, as a single measure of accuracy, the overall accuracy (or percentage classified correctly) gives no insight into how well the classifier is performing for each of the different classes. In particular, a classifier might perform well for a class which accounts for a large proportion of the test data and this will bias the overall accuracy, despite low class accuracies for other classes. Therefore error matrix itself is not a sufficient way to predict the accuracy of the classified image.

7.2 User's and Producer's Accuracy

Other measures derived from the error matrix are 'error of omission' (or producer's accuracy) and 'error of commission' (or user's accuracy). Error of omission represents an error from including a pixel to a particular class, which is actually not a part of the class. Commission error represents that a pixel, which should be part of a particular class but is not included.

7.3 The Kappa Statistic

The Kappa statistic was derived to include measures of class accuracy within an overall measurement of classifier accuracy. It provides a better measure of the accuracy of a classifier than the overall accuracy, since it considers inter-class agreement. KAPPA analysis yields a *Khat* statistics that is a measure of agreement or accuracy. The *Khat* statistic is computed as:

$$K_{hat}=N\sum^{r}_{i-1}x_{ii}-\sum^{r}_{i-1(}X_{i+}xX_{+i})\div N^{2}-\sum^{r}_{i-1(}X_{i+}xX_{+i}) \quad (6)$$

Where *r* is the number of rows in the matrix, *xii* is the number of observation in row *i* and column *i*, and *xi+* and *x+i* are the marginal totals for row *i* and column *i*, respectively, and *N* is the total number of observations.

VIII. KNOWLEDGE BASE CLASSIFICATIONS

Besides the spectral data, expert's knowledge can also play an important role in improving accuracy of the classification of the satellite images. Human experience and knowledge about the topology, geology etc. of the study area can be embodied in the classification procedures to prepare accurate classified maps. Such classification is known as knowledge base classification. The most difficult part of knowledge base classifier is the creation of the knowledge base.

Generally, knowledge base is created with the help of knowledge acquired by interacting with experts. By acquiring such knowledge we can build knowledge based system that could help us in improving the classification accuracy. But Building a knowledge base is very difficult task, because of the knowledge acquisition process. It requires lot of time to acquire knowledge from domain experts. This acquired knowledge is used further for knowledge base image classification.

IX. CONCLUSION

The main objective of this study is to improve the accuracy of the classification of satellite images using extracted knowledge in the form of classification rules using a decision tree approach. Generally, accuracy can be improved by adding ancillary information or by incorporating expert's knowledge into classification process. The main emphasis is given here to automatically extract knowledge in the form of classification rules using decision tree classifier. Different methods have been tried to use extracted classification rules in classification of a satellite image.

REFERENCES

[1] Avelino, J.G. & D.D.Dankel (1993), The Engineering of Knowledge-Based Systems. UpperSaddle River, NJ, USA: Prentice Hall.

[2] Argialas, D. and C.Harlow (1990), Computational Image Interpretation Models: An Overviewabd Prespective, Photogrammetric Engineering and Remote Sensing, Vol. 56,6, pp. 871-886.

[3] Breiman, L.; J.H.Friedman; R.A.Olsen & C.J.Stone (1984), Classification and Regression Trees, Belmont, CA, USA: Wadsworth. In: Yohannes,Y and J. Hoddinott (1999), Classification and Regression Trees: An Introduction. Technical Guide 3. WWW site:http://www.ifpri.org/themes/mp18/techguid/tg03.pdf (accessed on 22-07-04).

[4] Blackmore, K. & T. R. J. Bossomaier, Comparison of See5 and J48.PART Algorithms for Missing Persons Profiling. WWW site: http://attend.it.uts.edu.au/icita05/CDROM-ICITA02/ICITA2002/papers/119-5.pdf (accessed on 21-04-04).

[5] Byungyong, K. & D.Landgrebe (1991), Hierarchical decision tree classifiers in high dimensional and large class data. IEEE Transactions on the Geosciences and Remote Sensing, Vol. 29, 4, pp. 518-528.

[6] Lees, B.G. & K.Ritman (1991), Decision Tree and Rule Induction Approach to Integration of Remotely Sensed and GIS Data in Mapping Vegetation in Disturbed or Hilly Environments, Environmental Management, Vol. 15, 6, pp. 823-831.

[7] Li, D.; K. Li & D.Li (2000), Land use classification of Remote Sensing Images with GIS Data Base on Spatial Data Dining Technique. International Archives of Photogrammetry and Remote Sensing, Vol. XXXIII, Part B, 3, pp. 666.

[8] Morgan, J.N. & R.C. Messenger (1973), THAID: A Sequential Search Program for the Analysis of Nominal Scale Dependent Variables. Survey Research Center, Institute for Social Research, University of Michigan.

[9] Morgan, J.N. & J.A. Sonquist (1963), Problems in the Analysis of Survey Data, and a Proposal. Journal of the American Statistical Association, Vol.58, pp. 415-434.

[10] Mather, P.M. (1985), A Computationally Efficient Maximum Likelihood Classifier Employing Prior Probabilities for Remotely Sensed Data, International Journal of Geographical Information System, Vol. 6, pp 369-376.