International Journal of Current Engineering and Technology
ISSN 2277 - 4106

Available at http://inpressco.com/category/ijcet

Research Article

A Comparative Study of Pixel Based Supervised Classification and Fuzzy-Supervised Classification over Area around Mysore District

Shivakumar.B.R[a*] and S. V. Rajashekararadhya[b]

[a]Department of E&C, G.M. Institute of Technology, Davanagere, Karnataka, India
[b]Department of E&C, K.I.T Tiptur, Karnataka, India

Accepted 05 September 2013, Available online 01 October 2013, **Vol.3, No.4 (October 2013)**

Abstract

Conventional image classification methods restricts each pixel of data set to exclusively just one cluster. As a consequence, with this approach the classification results are often very crispy, i.e., each pixel of the image belongs to exactly just one class. However, in many real situations, for images, issues such as limited spatial resolution, poor contrast, overlapping intensities, and noise and intensity in-homogeneities variation make this hard (crisp) segmentation a difficult task. In this paper, a comparative study is done over two classification methods: Supervised Classification method and Fuzzy Supervised Classification method. The fuzzy classifier makes use of spatial features extracted from a multispectral data, and a classification image, generated using maximum likelihood classification. Initially, the data is classified using Supervised classification method and accuracy assessment is done over it to find the overall classification accuracy (in %). Then Fuzzy Supervised classification analysis is performed by allowing gradual memberships, thus offering the opportunity to deal with data that belong to more than one cluster at the same time. The process of accuracy assessment is followed after classification. A case study is presented on the two classification methods considered and a comparison is done as to find which method results in higher accuracy and Kappa values. The Fuzzy Supervised classification method decreases the misclassifications between River bank cultivation and plain land vegetation as well as between dense forest region and light vegetations thereby raising the overall classification accuracy to above 75%.

Keywords: *Supervised Classification, Fuzzy Supervised Classification, Maximum Likelihood Classification and Accuracy Assessment.*

Introduction

Remote-sensing research focusing on image classification has long attracted the attention of the remote-sensing community because classification results are the basis for many environmental and socioeconomic applications.

Evaluation of classification results is an important process in the classification procedure. etTo evaluate the performance of a classification method, Cihlar *et al*. proposed six criteria: accuracy, reproducibility, and robustness, ability to fully use the information content of the data, uniform applicability, and objectiveness (Cihlar, J *et al*,1998). In reality, no classification algorithm can satisfy all these requirements nor be applicable to all studies, due to different environmental settings and datasets used. DeFries and Chan suggested the use of multiple criteria to evaluate the suitability of algorithms (Defries, R.S *et al*,1998). These criteria include classification accuracy, computational resources, stability of the algorithm, and robustness to noise in the Improving classification performance Classification accuracy assessment is, however, the most common approach for an evaluation of classification performance.

Image classification is a complex process that may be affected by many factors. Effective use of multiple features of remotely sensed data and selection of suitable classification method are significant for improving classification accuracy. Non parametric classifiers such as fuzzy logic, neural network, decision tree classifier and knowledge based classifiers have increasingly become important approaches for multisource data classification. In general image classification can be grouped into supervised and unsupervised, or parametric and non-parametric, or hard and soft (fuzzy) classification, or pixel, subpixel and perfield. In this paper, a Fuzzy clustering based method for image segmentation is considered.

Before implementing a classification accuracy assessment, one needs to know the sources of errors (Congalton, R.G *et al*,1999). In addition to errors from the classification itself, other sources of errors, such as position errors resulting from the registration, interpretation errors, and poor quality of training or test

*Corresponding author: **Shivakumar.B.R** is working as Asst. Prof and **Dr. S. V. Rajashekararadhya** is working as Prof and Head

samples, all affect classification accuracy. In the process of accuracy assessment, it is commonly assumed that the difference between an image classification result and the reference data is due to the classification error.

However, in order to provide a reliable report on classification accuracy, non-image classification errors should also be examined, especially when reference data are not obtained from a field survey.

A classification accuracy assessment generally includes three basic components: sampling design, response design, and estimation and analysis procedures (Stehman, S.V *et al*,1999). Selection of a suitable sampling strategy is a critical step (Congalton, R.G. *et al*,1991). The major components of a sampling strategy include sampling unit (pixels or polygons), sampling design, and sample size (Muller, S.V *et al*,1991). Possible sampling designs include random, stratified random, systematic, double, and cluster sampling. A detailed description of sampling techniques can be found in papers by Congalton and Green (Congalton, R.G *et al*,1999).

The error matrix approach is the one most widely used in accuracy assessment (Foody, G.M *et al*,2000). In order to properly generate an error matrix, one must consider the following factors: (1) reference data collection, (2) classification scheme, (3) sampling scheme, (4) spatial autocorrelation, and (5) sample size and sample unit. After generation of an error matrix, other important accuracy assessment elements, such as overall accuracy, omission error, commission error, and kappa coefficient, can be derived. Many papers have been published to define meanings and provide computation methods for these elements (Hudson, W.D. *et al*,1987) (Congalton, R.G. *et al*,1991) (Janssen, L.F.J *et al*,1994) (Kalkhan, M.A. *et al*,1997) (Stehman, S.V *et al*,1996). Meanwhile, many authors, such as Congalton, Janssen and van der Wel, Smith *et al*., and Foody, have conducted reviews on classification accuracy assessment (Congalton, R.G. *et al*,1991) (Janssen, L.F.J *et al*,1994) (Smith, G.M. *et al*,201) (Foody, G.M. *et al*,2002). They have assessed the status of accuracy assessment of image classification, and discussed relevant issues. Congalton and Green systematically reviewed the concept of basic accuracy assessment and some advanced topics involved in fuzzy-logic and multilayer assessments, and explained principles and practical considerations in designing and conducting accuracy assessment of remote-sensing data (Congalton, R.G *et al*,1999). The Kappa coefficient is a measure of overall statistical agreement of an error matrix, which takes non-diagonal elements into account.

Kappa analysis is recognized as a powerful method for analyzing a single error matrix and for comparing the differences between various error matrices (Congalton, R.G. *et al*,1991). Modified Kappa coefficient and tau coefficient have been developed as improved measures of classification accuracy (Foody, G.M. *et al*,1992) (MA, Z. *et al*,1995). Moreover, accuracy assessment based on a normalized error matrix has been conducted, which is regarded as a better presentation than the conventional error matrix (Congalton, R.G. *et al*,1991) (Hardin, P.J *et al*,1997).

The error matrix approach is only suitable for 'hard' classification, assuming that the map categories are mutually exclusive and exhaustive and that each location belongs to a single category. This assumption is often violated, especially for classifications with coarse spatial resolution imagery. 'Soft' classifications have been performed to minimize the mixed pixel problem using a fuzzy logic. The traditional error matrix approach is not appropriate for evaluating these soft classification results. Accordingly, many new measures, such as conditional entropy and mutual information, fuzzy-set approaches, symmetric index of information closeness, Renyi generalized entropy function, and parametric generalization of Morisita's index have been developed (Finn, J.T *et al*,1993) (Maselli, F *et al*,1994) (Gopal, S *et al*,1994) (Foody, G.M. *et al*,1996) [25] (Ricotta, C *et al*,2002). However, one critical issue in assessing fuzzy classifications is the difficulty of collecting reference data. More research is thus needed to find a suitable approach for evaluating fuzzy classification results.

Image classification approaches

There are many classifier algorithms. In this paper we mainly consider the following two classifier algorithms.

- Supervised Image Classification.
- Fuzzy Supervised Image Classification.

Supervised Image Classification: Supervised classification process is classified into two phases: (a) training phase, and (b) Decision making phase. In training phase, the analyst "trains" the computer by assigning a limited number of pixels to the respective classes they belong to in the particular image. In decision making phase, the computer assigns a class label to all (other) image pixels, by looking for each pixel the most similar trained class.

During the training phase, the classes to be used are previously defined. About each class some "ground truth" is needed. The ground truth can be obtained by assigning a number of places in the image area that are known to belong to that class. This knowledge must have been acquired beforehand, for instance as a result of fieldwork, or from an existing map (assuming that in some areas the class membership has not changed since the map was produced). If the ground truth is available, training samples (small areas or individual pixels) are indicated in the image and the corresponding class names are entered. These training samples are also termed as "Regions of Interest" (ROI).

Due to the large numbers of spectrally similar land cover types present in the urban environment, traditional classification approaches such as maximum likelihood often result in significant numbers of misclassifications, especially between the Road and Building classes, and the Grass and Tree classes. By utilizing spatial features in addition to the spectral information, the Fuzzy pixel-based classifier is able to more accurately classify high-resolution imagery of urban areas. This classifier uses the results of an initial maximum likelihood classification of the imagery to group the classes where significant misclassifications occur together into sets. Subsequent processing using spatial features are then performed to differentiate between the spectrally similar classes. This approach allows for different groups of classes to be classified using the features best suited for discrimination between those classes. This alleviates the problem of features simultaneously decreasing the confusion between one set of classes and increasing it for another set.

The Fuzzy pixel-based classification technique is significantly more accurate than maximum likelihood classification. However, more detail is needed to accurately represent the land cover types present in dense urban areas. A non-road, non-building Impervious Surface class is also needed to represent features such as parking lots, concrete plazas, etc. To distinguish between these urban land cover classes, an object based classification approach is used to examine features such as object shape and context (neighbourhood) and then classify the image objects using a Fuzzy logic rule base.

Accuracy assessment

Information derived from remotely sensed data is important for environmental models at local, regional, and global scales. The remote sensing–derived thematic information may be in the form of thematic maps or statistics derived from area-frame sampling techniques. The thematic information must be accurate because important decisions are made throughout the world using the information.

Unfortunately, the thematic information contains error. Scientists who create remote sensing–derived thematic information should recognize the sources of the error, minimize it as much as possible, and inform the user how much confidence he or she should have in the thematic information. Remote sensing–derived thematic maps should normally be subjected to a thorough accuracy assessment before being used in scientific investigations and policy decisions.

The ideal situation is to locate ground reference test pixels (or polygons if the classification is based on human visual interpretation) in the study area. These sites are not used to train the classification algorithm and therefore represent unbiased reference information. It is possible to collect some ground reference test information prior to the classification, perhaps at the same time as the training data. But the majority of test reference information is often collected after the classification has been performed using a random sample to collect the appropriate number of unbiased observations per category.

Fitzpatrick-Lins suggested that the sample size N to be used to assess the accuracy of a land-use classification map be determined from the formula for the binomial probability theory (Fitzpatrick-Lins, K *et al*,1981):

$$N = \frac{Z^2(p)(q)}{E^2} \quad (1)$$

Where p is the expected percent accuracy of the entire map, $q = 100 - p$, E is the allowable error, and $Z = 2$ from the standard normal deviate of 1.96 for the 95% two-sided confidence level.

Evaluation of Error Matrices

Once the ground reference test information has been collected from the randomly located sites, the test information is compared pixel by pixel with the information in the remote sensing–derived classification map. Agreement and disagreement are summarized in the cells of the error matrix. Information in the error matrix may be evaluated using simple descriptive statistics or multivariate analytical statistical techniques.

Overall Accuracy

Overall accuracy is the proportion of all reference pixels, which are classified correctly. It is computed by dividing the total number of correctly classified pixels (the sum of elements along the main diagonal) by the total number of reference pixels. According to the error matrix above, the overall accuracy can be calculated as:

$$OA = \frac{\sum_{K=1}^{N} a_{kk}}{\sum_{i,k=1}^{N} a_{ik}} = \frac{1}{n} \sum_{K=1}^{N} a_{kk} \quad (2)$$

Producer's Accuracy

Producer's accuracy estimates the probability that a pixel, which is of class I in the reference classification, is correctly classified. It is estimated with the reference pixels of class I divided by the pixels where classification and reference classification agree in class I. Given the error matrix above, the producer's accuracy can be calculated as:

$$PA(class\, I) = \frac{a_{ii}}{\sum_{i=1}^{N} a_{ki}} \quad (3)$$

Producer's accuracy tells how well the classification agrees with reference classification.

User's Accuracy

User's accuracy is estimated by dividing the number of pixels of the classification result for class I with the number of pixels that agree with the reference data in class I. It can be calculated as:

$$\text{UA (class I)} = \frac{a_{ii}}{\sum_{i=1}^{N} a_{ik}} \quad (4)$$

User's accuracy predicts the probability that a pixel classified as class I is actually belonging to class I.

Kappa Analysis

K_{hat} Coefficient of Agreement: Kappa analysis yields a statistic, $\hat{K}$, which is an estimate of Kappa. It is a measure of agreement or accuracy between the remote sensing–derived classification map and the reference data as indicated by a) the major diagonal, and b) the chance agreement, which is indicated by the row and column totals (referred to as marginal).

$$\hat{K} = \frac{N\sum_{i=1}^{k} x_{ii} - \sum_{i=1}^{k}(x_{i+} \times x_{+i})}{N^2 - \sum_{i=1}^{k}(x_{i+} \times x_{+i})} \quad (6)$$

Results and analysis

To validate the applicability of the selected methods, a case study is presented in this section, which is carried out on IRS-p6/LISS III sample image with 23m resolution. The area considered is a rectangular area between the points 12 10 09.82N 76 15 49.45E / 12 00 26.48N 76 46 46.68E as shown in Fig.1. The process of image to map registration is carried out on the study area so as to register the data correctly. After the data registration, the process of collecting the signatures is carried out. In signature collection process, the pixels belonging to similar group are framed together and given a specific class name. After signature collection, the data is classified using the above mentioned two image classification methods. At first, supervised image classification method is carried out on the data followed by Fuzzy supervised image classification. In Supervised image classification method, the data is classified such that one pixel belongs to only one class, whereas in Fuzzy Supervised image classification method, each pixel can belong to more than one pixel. In the current study, each pixel is allowed to belong to up to 5 classes.

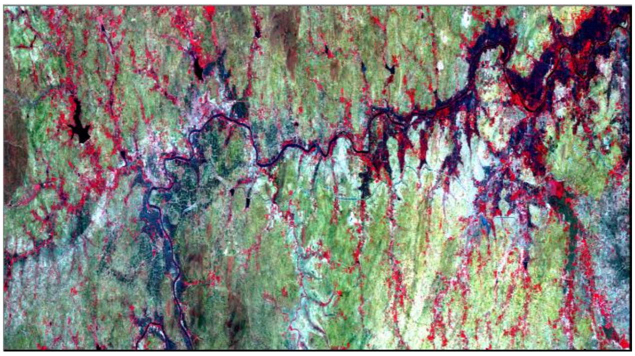

Fig. 1. 23m Spatial resolution study area considered.

The classified images are then subjected to the process of accuracy assessment to verify the accuracy assessment. A total of 100 validation points is considered in the study area and their class values are manually verified using the glabal maps in wikimapia. The resluts are tabulated as in Table.1 and Table.2 to compare the two classification methods considered. Table.1 represents the classification details for the Supervised classification method and Table.2 represents the classification details for Fuzzy Supervised classification method.

It can be seen from the qualitative analysis of the two classified images that Fuzzy Supervised image classification methods yields higher accuracy value when campared to the Supervised Classification method. The overall classification accuracy value can be improved by applying a appropriate parametric approach such as maximum likelihood, minimum distance, etc. The overall classification accuracy can also be improved by changing the value of classes-per-pixel in Fuzzy classification method.

Fig.2 shows the image of the study area considered after Supervised image classification process is carried over it. Fig.3 shows the image of the study area considered after Fuzzy Supervised Image classification process is carried over it. Table.3 presents the combined results of the two classification methods consideed for the case study.

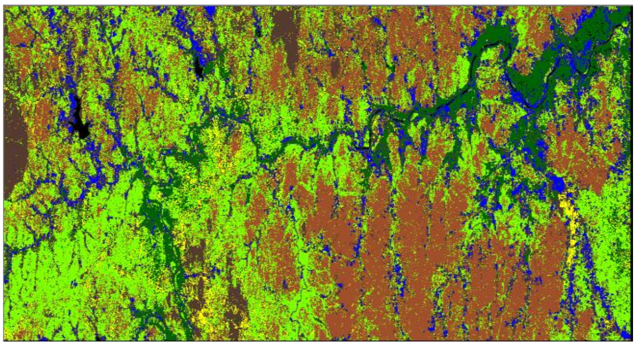

Fig.2. Study area after Supervised Classification process.

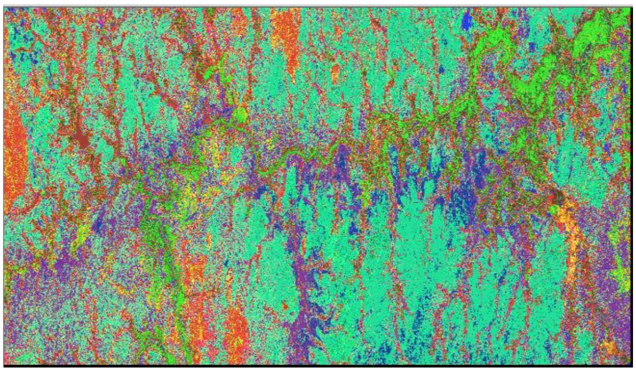

Fig.3. Study area after Fuzzy Supervised classification process.

Conclusion

The overall classification accuracy is dependent on various parameters like number of classes, spatial resolution of the data, classification methods used, number of classes per pixel and many more. Since we have used 23m spatial resolution data, both the methods are producing reasonably high classification accuracy values. It is hard to come to a conclusion by visually examining the classified images. Hence accuracy assessment is carried out for numerically finding which classification method results in highest accuracy value.

When supervised classification method was employed on the data, the overall classification accuracy value was 71.00%. After the same data was processed using Fuzzy Image classification method, the overall classification accuracy value was 71.00%. After the same

Table.1. Results for Supervised classification results

Class Name	Reference Totals	Classified Totals	Number Correct	Producers Accuracy	Users Accuracy
Unclassified	0	0	0	---	---
Uncultivated Land	19	25	14	73.68%	56.00%
River Bank Cultivation	7	7	5	71.43%	71.43%
Thick Vegetation	18	6	6	33.33%	100.00%
Light Vegetation	35	31	25	71.43%	80.65%
Water Bodies	0	0	0	---	---
River Water	0	0	0	---	---
Nagara Hole Forest	21	30	21	100.00%	70.00%
Black Soil Vegetation	0	1	0	---	---
Totals	100	100	71		
Overall Classification Accuracy = 71.00%					

Table.2 Results for Fuzzy Supervised Classification details

Class Name	Reference Totals	Classified Totals	Number Correct	Producers Accuracy	Users Accuracy
Unclassified	0	0	0	---	---
Uncultivated Land	16	29	12	68.75%	37.93%
River Bank Cultivation	4	5	4	100.00%	80.00%
Thick Vegetation	10	7	5	50.00%	71.43%
Light Vegetation	44	36	33	68.18%	83.33%
Water Bodies	1	1	1	100.00%	100.00%
River Water	0	0	0	---	---
Nagara Hole Forest	23	20	20	86.96%	100.00%
Black Soil Vegetation	2	2	2	100.00%	100.00%
Totals	100	100	77		
Overall Classification Accuracy=77.00%					

data was processed using Fuzzy Image classification method, the overall classification accuracy value increased to 77.00%. Hence, it can be concluded that for the data considered and the classification methods selected, Fuzzy supervised image classification method results in higher classification accuracy. The results can be improved by varying the number of classes per pixel or by repeating the same process over a higher resolution data.

It should be noted that these results are correct only for the data obtained. If thorough experimentation is done on the same data in different conditions like varied values of classes, classes per pixel and spatial resolutions, the above results may tend to vary.

References

Cihlar, J., Xiao, Q., Chen, J., Beaubien, J., Fung, K. And Latifovic, T (1998), Classification by progressive generalization: a new automated methodology for remote sensing multispectral data. *International Journal of Remote Sensing*, 19, pp. 2685–2704.

Defries, R.S. and Chan, J.C., (2000), Multiple criteria for evaluating machine learning algorithms for land cover classification from satellite data. *Remote Sensing of Environment*, 74, pp. 503–515.

Congalton, R.G. and Green, K (1999), Assessing the Accuracy of Remotely Sensed Data: Principles andpractices (Boca Raton, London, New York: *Lewis Publishers*).

Powell, R.L., Matzke, N., DE Souza JR, C., Clark, M., Numata, I., Hess, L.L. and Roberts, D.A. (2004), Sources of error in accuracy assessment of thematic land-cover maps in the Brazilian Amazon. *Remote Sensing of Environment,* 90, pp. 221–234.

Stehman, S.V. and Czaplewski, R.L. (1998), Design and analysis for thematic map accuracy assessment: fundamental principles. *Remote Sensing of Environment,* 64, pp. 331–344.

Congalton, R.G. (1991), A review of assessing the accuracy of classification of remotely sensed data. *Remote Sensing of Environment*, 37, pp. 35–46.

Muller, S.V., Walker, D.A., Nelson, F.E., Auerbach, N.A., Bockheim, J.G., Guyer, S. and Sherba, D (1998), Accuracy assessment of a land-cover map of the Kuparuk river basin, Alaska: considerations for remote regions. *Photogrammetric Engineering and Remote Sensing*, 64, pp. 619–628.

Congalton, R.G. and green, K (1999), Assessing the Accuracy of Remotely Sensed Data: Principles and practices (Boca Raton, London, New York: *Lewis Publishers*).

Foody, G.M (2002), Hard and soft classifications by a neural network with a nonexhaustively defined set of classes. *International Journal of Remote Sensing*, 23, pp. 3853–3864.

Congalton, R.G. and Green, K (1999), Assessing the Accuracy of Remotely Sensed Data: Principles and practices (Boca Raton, London, New York: *Lewis Publishers*).

Hudson, W.D. and Ramm, C.W (1987), Correct formulation of the Kappa coefficient of agreement. *Photogrammetric Engineering and Remote Sensing*, 53, pp. 421–422.

Janssen, L.F.J. and Van der Wel, F.J.M. (1994), Accuracy assessment of satellite derived land-cover data: a review. *Photogrammetric Engineering and Remote Sensing*, 60, pp. 419–426.

Kalkhan, M.A., Reich, R.M. and Czaplewski, R.L.(1997), Variance estimates and confidence intervals for the Kappa measure of classification accuracy. *Canadian Journal of Remote Sensing*, 23, pp. 210–216.

Stehman, S.V.(1996), Estimating the Kappa coefficient and its variance under stratified random sampling. *Photogrammetric Engineering and Remote Sensing*, 62, pp. 401–407.

Stehman, S.V. (1997), Selecting and interpreting measures of thematic classification accuracy. *Remote Sensing of Environment*, 62, pp. 77–89.

Smith, G.M. and Fuller, R.M. (2001), An integrated approach to land cover classification: an example in the Island of Jersey. *International Journal of Remote Sensing*, 22, pp. 3123–3142.

Foody, G.M., (2002), Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, 80, pp. 185–201.

Foody, G.M. (1992), On the compensation for chance agreement in image classification accuracy assessment. *Photogrammetric Engineering and Remote Sensing*, 58, pp. 1459–1460.

MA, Z. and Redmond, R.L.(1995), Tau coefficients for accuracy assessment of classification of remote sensing data. *Photogrammetric Engineering and Remote Sensing*, 61, pp. 435–439.

Hardin, P.J. and Shumway, J.M. (1997), Statistical significance and normalized confusion matrices. Photogrammetric Engineering and Remote Sensing, 63, pp. 735–740.

Finn, J.T. (1993), Use of the average mutual information index in evaluating classification error and consistency. *International Journal of Geographical Information Systems*, 7, pp. 349–366.

Maselli, F., Conese, C. and Petkov, L. (1994), Use of probability entropy for the estimation and graphical representation of the accuracy of maximum likelihood classifications. ISPRS Journal of Photogrammetry and Remote Sensing, 49, pp. 13–20.

Gopal, S. and Woodcock, C. (1994), Theory and methods for accuracy assessment of thematic maps using fuzzy sets. *Photogrammetric Engineering and Remote Sensing*, 60, pp. 181–188.

Binaghi, E., Brivio, P.A., Ghezzi, P. and Rampini, A.(1999), A fuzzy set accuracy assessment of soft classification. *Pattern Recognition Letters*, 20, pp. 935–948.

Foody, G.M. (1996), Approaches for the production and evaluation of fuzzy land cover classification from remotely-sensed data. *International Journal of Remote Sensing*, 17, pp. 1317–1340.

Ricotta, C. and Avena, G.C. (2002), Evaluating the degree of fuzziness of thematic maps with a generalized entropy function: a methodological outlook. *International Journal of Remote Sensing*, 23, pp. 4519–4523.

Fitzpatrick-Lins, K. (1981), Comparison of sampling procedures and data analysis for a land-use and land-cover map, *Photogramm. Eng. Remote Sens*. 47(3):343-351.