International Journal of Information and Computation Technology.
ISSN 0974-2239 Volume 3, Number 10 (2013), pp. 1059-1066

http://www. irphouse.com /ijict.htm

Deriving Uncertainty of Area Estimates from Satellite Imagery using Fuzzy Land-cover Classification

Rekha Gangula[1] and Jayakrishna Vasam[2]

[1]*Computer Science, Kamala Institute of Technology & Science, Karimnagar, India.*
[2]*AECOM India Private Limited, Gurgaon, India.*

Abstract

The use of remotely sensed data as input into geographical information systems has promoted new interest in issues related to the accuracy of multispectral classification. This paper investigates the impact of classification uncertainty on the estimation of area from satellite derived land-cover data. Applying four variant; of the maximum-likelihood classifier, it is shown that the estimated area for different land-cover classes is highly influenced by the methods which are used for classifier training. To evaluate the uncertainty of area estimates, a new error modeling strategy is proposed. Assuming that attribute uncertainty in image classification is field-based rather than pixel-based, the image is segmented in fields according to similarities in the probability vectors of adjacent pixels. In simulating uncertainty, this field structure is explicitly taken into account. Using different strategies for image segmentation, it is shown that the spatial correlation of classification uncertainty has a major impact on the assessment of the uncertainty of area estimates.

Fuzzy Land-Cover classification is relatively young theory. Major advantage of this theory is that it allows the natural description, in linguistic terms, of problems that should be solved rather than in terms of relationships between precise numerical values. This advantage, dealing with the complicated systems in simple way, is the main reason why fuzzy logic theory is widely applied in technique. It is easily possible to classify the remotely sensed image (as well as any other digital imagery), in such a way that certain land cover classes are clearly represented in the resulting image. Basic idea was to perform the classification procedure first in the supervised and then in fuzzy logic manner. Some information, needed for membership function definition, was taken from supervised maximum likelihood classification. Results of two procedures, both based on pixel-by-pixel technique and field based, were compared and certain encouraging conclusion remarks come out.

1. Introduction

Today much attention is given to error modeling and error propagation in GIS databases. The motivation behind this research is the general acknowledgment that a successful use of GIS as a decision support tool can only be achieved if it becomes possible to attach a quality label to the output of each GIS analysis. Most commercial GIS packages do not provide tools for the modeling of error in individual data layers and for the tracking of error when data layers are manipulated and combined in a GIS-based spatial analysis. Yet the issue of spatial data uncertainty is given high priority on the GIS research agenda and is one of the most frequently covered topics in recent scientific literature on GIS. With the increasing use of remotely sensed data as input into geographical information systems, the accuracy of multispectral classification also gained more attention and new approaches were taken to describe and model the uncertainty that characterizes the classification process.

Since the launch of high-resolution remote sensors, the use of satellite images as a major source of spatial information has been the subject of extensive research in a broad range of applications. As a result of these research efforts today, the benefits of high-resolution remote sensing technology are widely recognized. In particular, the extraction of land-cover information from remotely sensed data has received considerable attention over the last ten years and has become almost common practice in many application areas. A wide range of classification approaches for the extraction of land cover information has been developed, and many strategies and modifications have been proposed to improve classification accuracy obtained with these classifiers.

2. Satellite Image Classification

2.1 Fuzzy Logic and Algorithm

Fuzzy logic has been used in a wide range of problem domains. Although the fuzzy logic is relatively young theory, the areas of applications are very wide application like process control, management and decision making, operations research, economies and, for this paper the most important, field pattern recognition and classification. Dealing with simple ‘black’ and ‘white’ answers is no longer satisfactory enough; a degree of membership (suggested by Prof. Zadeh, 1965) became a new way of solving the problems. A fuzzy set is a set whose elements have degrees of membership. A element of a fuzzy set can be full member (100% membership) or a partial member (between 0% and 100% membership). That is, the membership value assigned to an element is no longer restricted to just two values, but can be 0, 1 or any value in-between. Mathematical function which defines the degree of an element's membership in a fuzzy set is called membership function. The natural description of problems, in linguistic terms, rather than in terms of relationships between precise numerical values is the major advantage of this theory. An idea to solve the problem of image classification in fuzzy logic manner as well as comparison of the results of supervised and fuzzy classification was the main motivation of this work. Behind this idea was also the question if the possible promising results can give the answer to the question of diminishing the influence of person dealing with supervised classification.

In this paper, in order to classify Linear Imaging Self-Scanner IV (LISS IV) spectral satellite image used in fuzzy logic manner as follows: input (image channels) and output variables (land classes) are introduced, a membership functions are defined using results from supervised classification, a ERDAS Fuzzy Logic Toolbox was used in definition of fuzzy logic inference rules, these rules are tested and verified through the simulation of classification procedure at random sample areas and at the end, a LISS IV image classification was conducted.

2.2 Supervised classification

As it was later used for fuzzy logic classification, the procedure of supervised image classification was conducted with ERDAS Imagine software. As the source for classification procedure, Resourcesat-2 LISS IV multispectral image was used. This image contains three spectral bands as follows: Ä band B2 covering 0.52-0.59 (green), a band B3 covering 0.62 to 0.68 µm (red) and a band 4 covering 0.77 to 0.86 µm (near infrared). In supervised image classification selected land cover classes are: sand dunes, deciduous trees, coniferous trees, urban area, agriculture land and fallow land. For these classes, training areas were pointed on the image (Figure 1.)

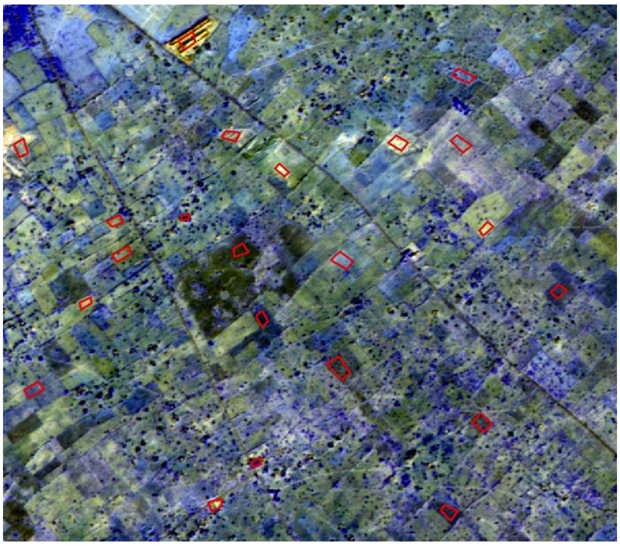

Figure 1: Liss IV Resourcesat-2 Satellite Image

Once spectral signatures are defined, supervised classification assigns each pixel in the image to exactly one spectral class using a discriminant function which may be based on the measurement of spectral distance (parametric methods) or on the frequency histograms of the classes considered (nonparametric methods). Although some of these methods (e.g., the maximum-likelihood classifier) produce class membership probabilities for each pixel, they are basically grounded on the concepts of classical set theory. While it may be tempting to consider the class membership probabilities of the traditional maximum-likelihood classifier as indicative of class mixture at the sub-pixel level, it remains questionable if reliable conclusions on fuzzy membership can be drawn from the output of a supervised classification procedure which generates spectral signatures from training pixels that are entirely assigned to a priori defined cover types through a hard classification procedure. Traditional unsupervised classification methods show the same inadequacy to deal with attribute uncertainty. By means of a clustering algorithm, spectral space is iteratively partitioned into regions, each of which corresponds to one spectral class. Pixels inside a region are considered as entirely belonging to that class. After classification no information on the probability of class membership is available.

In most classification studies, map accuracy is assessed by means of a confusion matrix which compares a sample of the classified pixels with reference data obtained from aerial photographs or ground surveys (Frank Canters, 1997). In recent years many authors have commented on the shortcomings of the confusion matrix as an estimator of classification accuracy. Much work has been published on optimal strategies for ground truth sampling, accuracy measures have been refined to deal with chance agreement, and methods have been presented to estimate statistically sound classification probabilities from the confusion matrix (van Genderen, 1978; Rosenfield et al., 1982; Congalton, 1988; Foody, 1992; Stehman, 1992; Green et al.,1993; Stehman, 1995). Yet the most fundamental drawback of the confusion matrix is its inability to provide information on the spatial structure of the uncertainty in a classified scene. Measures of uncertainty and class membership probabilities can only be derived for a land-cover class as a whole and are considered as representative for each pixel that is assigned to that class, thereby ignoring the complex spectral response that is present in a multispectral satellite image.

2.3 Fuzzy approaches to Image Classification

Fuzzy logic is increasingly used for the handling of uncertainty in geographical databases (Burrough and Frank, 1996). Building on concepts of fuzzy set theory, new methods for accuracy assessment have been presented (Gopal and Woodcock, 1994) and new classification approaches have been developed to properly deal with uncertainty in class allocation. Starting from the assumption that each training pixel can contribute to the spectral signature of each predefined class, (Wang 1990a; 1990b) modified the traditional maximum-likelihood classifier by calculating a fuzzy mean and a fuzzy covariance matrix for each class. The fuzzy mean is given by

$$\mu_i^F = \left(\sum_{j=0}^{n} f_i(x_j) x_j \right) / \left(\sum_{j=1}^{n} f_i(x_j) \right)$$

where n is the total number of training pixels, fi(x) is the membership function of class i, and xj is the jth pixel measurement vector, containing the brightness values of the pixel in each spectral band.

As can be seen, the contribution of each training pixel to the fuzzy mean and fuzzy covariance of a class i is determined by its membership value. For each training pixel, all membership values should be positive and add up to 1. A membership value close to 1 indicates that the pixel almost certainly belongs to that class; a value close to 0 indicates that the pixel is very unlikely to belong to that particular class. The membership values for each training pixel can be taken from a previously determined classification which is in a fuzzy representation (Wang, 1990a) or can be derived from the class membership probabilities of a traditional maximum-likelihood classification of the same image. Once fuzzy class signatures have been determined, the standard maximum-likelihood procedure is applied, except that the mean and covariance matrices in the definition of the probability density function for each class are replaced by their fuzzy counterparts.

The most obvious way of assessing classification accuracy is by comparing the class assigned to a pixel in the classification with the corresponding class observed in the field. The comparison of the classified land-cover to the actual land-cover is usually recorded in a confusion matrix. For each land-cover class the confusion matrix reports the number of pixels that are classified into class i but actually belong to class *j* (commission errors) as well as the number of pixels that belong to class i but are wrongly assigned to class *j* (ommission errors). From the confusion matrix global or class dependent measures of accuracy are derived by dividing the number of correctly classified pixels by the total number of pixels in the sample. Methods have been presented to estimate statistically sound class membership probabilities from the confusion matrix, taking into account the sampling strategy that is applied (Green et al., 1993). Estimated probabilities derived in this way are, however, identical for all pixels that have been assigned to the same class. Hence, no information about the spatial distribution of classification reliability within the classes is available.

3. Conclusion

The most important obstacle when handling uncertainty in geographical information systems is the lack of knowledge about the error present in the source data. When working with existing cartographic material, the information on data quality is usually sparse. Map making involves different data collection, generalization, and representation procedures, sometimes of a high level of abstraction, all leading to a degradation of data quality that is not easily described in a formal way. It is therefore clear that the key to a better understanding and modeling of source errors in GIS lies in the reduction of the level of abstraction in GIS databases (Goodchild and Wang, 1988). In this respect, the integrated use of digital satellite data in a GIs environment offers interesting challenges. Indeed, when using fuzzy classification approaches, class membership values indicating to what extent a pixel is likely to belong to each of the classes in the image can be derived. Being available for each pixel, these membership values provide valuable information with respect to the spatial structure of uncertainty in the classified image.

In this paper it has been demonstrated how class membership values can be used in the estimation of area for different land-cover classes as well as in the assessment of the uncertainty of these area estimates. Applying four variants of the maximum-likelihood classifier, it has been shown that the method which is used for the training of the classifier may have a considerable impact on the estimation of area, especially for classes with less distinctive signatures. A new method for the stochastic modeling of classification uncertainty has been proposed which makes optimal use of the class membership values derived from the classification. The method is based on the segmentation of the image in so called "fields," i.e., groups of adjacent pixels with similar class membership values. Instead of looking at the uncertainty for each individual pixel, the fields in the segmented image are considered as the elementary spatial units in the error simulation process. Using different strategies for image segmentation, it has been made clear that the spatial characteristics of classification uncertainty have a strong impact on the assessment of the uncertainty of area estimates. This stresses the importance, as also mentioned by other authors, of properly dealing with the spatial structure of uncertainty in stochastic error modeling. Although the assumption that classification uncertainty is field-based rather than pixel based is justified for agricultural areas, it is important to mention that it is probably not a useful strategy to apply in areas with semi-natural vegetation where classes tend to inter-grade gradually. Here, the use of a continuous model of spatial variation or a combination of both discrete and continuous modeling may be more appropriate.

References

[1] F Canters (1994). Simulating error in triangulated irregular network models, Proceedings of the Fifth European Conference and Exhibition on Geographical Information Systems, Paris, France, pp. 169- 178.

[2] J L van Genderen, B F Lock, and P A Vass (1978), Remote sensing: statistical testing of thematic map accuracy, Remote Sensing of Environment, 7:3-14.

[3] G H Rosenfield, K Fitzpatrick-Lins, and H S Ling (1982), Sampling for thematic map accuracy testing, Photogmmmetric Engineering & Remote Sensing, 48:131-137.

[4] R G and Congalton (1988), A comparison of sampling schemes used in generating error matrices for assessing the accuracy of maps generated from remotely sensed data, Photogrammetric Engineering & Remote Sensing, 54:593-600.

[5] G M and Foody (1992), On the compensation for chance agreement in image classification accuracy assessment, Photogrammetric Engineering & Remote Sensing, 58:1459-1460.

[6] S V Stehman,., (1992), Comparison of systematic and random sampling for estimating the accuracy of maps generated from remotely sensed data, Photogrammetric Engineering & Remote Sensing, 58:1343-1350.

[7] Green, E.J, W E Strawderman and T.M. Airola (1993), Assessing classification probabilities for thematic maps, Photogrammetric Engineering & Remote Sensing, 59635-639.

[8] Gopal, S., and C Woodcock (1994), Theory and methods for accuracy assessment of thematic maps using fuzzy sets, Photogrammetric Engineering & Remote Sensing, 60:181-188.

[9] Wang, F., (1990), Fuzzy supervised classification of remote sensing images, BEE Transactions on Geoscience and Remote Sensing, 28:194201.

[10] Wang, F., (1990), Improving remote sensing image analysis through fuzzy information representation, Photogrammetric Engineering & Remote Sensing, 56:1163-1169.

[11] Goodchild, M.F., and M H Wang (1988), Modeling error in raster-based spatial data, Proceedings of the Third International Symposium on Spatial Data Handling, Sydney, Australia, pp. 97-106.