STUDIA UNIV. BABEŞ–BOLYAI, INFORMATICA, Volume **LII**, Number 1, 2007

THE APPLICABILITY OF FUZZY THEORY IN REMOTE SENSING IMAGE CLASSIFICATION

GABRIELA DROJ

Abstract. In the recent years, the usage of Geographic Information Systems has been rapidly increasing and it became the main tool for analyzing spatial data in unprecedented number of fields of activities. The evolution of GIS led to the necessity of faster and better results. The processing time was reduced by using more and more advanced and applied mathematical and computer science knowledge. One of these mathematical theories is fuzzy logic. The fuzzy logic theory gives the possibility of enhancing spatial data management with the modeling of uncertainty. The usage of fuzzy theory has also applicability in processing remote sensing data. In this paper is presented the applicability of fuzzy set theory to the classification of raster images

1. Introduction

Geographical Information System (shortly GIS) represents a powerful set of tools for collecting, storing, retrieving at will, transforming and displaying spatial data from the real world. GIS extends the limits of Computer Aided Design (CAD) and Automated Mapping (AM) with the possibility of retrieving geospatial data at request and with the possibility of "what if" analysis and scenarios.

The difference between a CAD system and a GIS system is given by the possibility of spatial analysis, where new maps are computed from existing ones by applying either:

- Spatial Join operation between different geospatial data;
- Spatial aggregation;
- Buffer operations, which increase the size of an object by extending its boundary;
- Set operations, such as complement, union, and intersection.

Received by the editors: July 2007 .

2000 *Mathematics Subject Classification.* 68U10, 03E72, 03E75.

1998 *CR Categories and Descriptors.* I.4.6. [**Computing Methodologies**]: Image Processing and Computer Vision – *Segmentation*; I.5.3.[**Computing Methodologies**]: Pattern Recognition – *Clustering* F.4.1.[**Theory of Computation**]: Mathematical Logic and Formal Languages – *Mathematical Logic* .

In traditional GIS, these operations are exact quantitative operations. Humans, however, often prefer a qualitative operation over an exact quantitative one, which can be achieved by extending the standard map, overlaying operations to fuzzy maps and using fuzzy logic rather than crisp logic.

Fuzzy logic was implemented successfully in various GIS processes like:

- Data collection – analysis and processing remote sensing data for classification algorithms and object recognition;
- Spatial analysis - processing qualitative data, defining relationships between uncertain geospatial objects;
- Complex operations based on genetic algorithms or artificial intelligence; in this category we can include also the object recognition from airborne images.

Remote sensing from airborne and space borne platforms provides valuable data for mapping, environmental monitoring, disaster management and civil and military intelligence. However, to explore the full value of these data, the appropriate information has to be extracted and presented in standard format to import it into geo-information systems and thus allow efficient decision processes. The process usually used for extracting information from these images is based on classification.

Classification or clustering is the process of automatically grouping a given set of data into separate clusters such that data points with similar characteristics will belong to the same cluster. In this way, the number of clusters is reduced. The process of classification is usually based on object's attributes or characteristics than on its geometry but in the process of image classification is based on multi spectral analysis of the pixels. An image classification is acceptable if the distortion of the image is minim.

This paper will present the applicability of fuzzy set theory in classification of raster images. In the same time is presented the possibility to optimize the classification procedure by using fuzzy set theory to obtain a minimum distortion of the image.

2. On Classification Methods

In classical cluster analysis each pixel must be assigned to exactly one cluster. Fuzzy cluster analysis relaxes this requirement by allowing gradual memberships, thus offering the opportunity to deal with data that belong to more than one cluster at the same time.

The result of clustering using fuzzy classification consists of a multi-layer output file, one layer for each cluster. Each layer can be saved as an independent image. In the image layer, the black is representing the membership 0 and white is representing the membership 1. The pixels in different gray tones are representing the degree of membership to the cluster.

The theory of fuzzy sets has its main applicability in the process of raster classification

2.1. **Unsupervised Fuzzy Classification.** The unsupervised classification is a completely automatic process; it's eliminating the user and in the same time the influence of known information. This kind of clustering procedure is used where are no information about the site represented in the image to be classified.[Tur]

The most popular unsupervised classification methods are ISODATA and the k-means algorithm. [Eas01, Su03] There are many unsupervised classification algorithms based on fuzzy set theory, we mention fuzzy c-means, fuzzy Gustafson – Kessel algorithm, fuzzy c - shells and genetic algorithm and so on.[Ben03, Tur]. One of the most popular fuzzy based unsupervised classifications is the Fuzzy c-means algorithm, similar to K-means algorithm.

2.2. **Supervised Classification.** The results of classification can be optimized if there are known information about the image to be clustered. In this case the method used is called supervised classification. The supervised classification is made in two steps. The first step is to create signature files (training sites) and the second step is the classification itself. The fuzzy logic can be used in the creation of the signatures files and also in the process of the classification itself, but is not necessary to be used in both stages. [Ben03, Eas01, Liu03, Ste99, Su03]

The signatures files based on this training site can be created using unsupervised classification methods (like ISODATA), classical methods (like Maximum Likehood or Minim Distance) or by computing a fuzzy matrix filled with the values indicated by the membership grade of each training site [Eas01]. The membership degree is computed by using fuzzy membership functions like: piecewise linear functions, the Gaussian distribution function, the sigmoid curve and quadratic or cubic polynomial curve. The evaluation on the training can be done using statistical methods: minimum, maximum, mean, and standard deviation for each band independent and covariance matrix for all the three bands. The most relevant signature file evaluation is creating an error matrix as a matrix of percentages based on pixel counts that allows us to see how many pixels in each training sample were assigned to each class.

The second step of the supervised classification can also be processed with traditional methods (Minimum distance, Mahalanobis distance or Maximum Likehood) or by using fuzzy membership functions or genetic algorithm. This stage cannot be evaluated alone, it can be evaluated just the final result of the supervised classification.

In order to optimize the accuracy of the results we consider that is appropriate to use fuzzy logic in both stages of supervised classification. In the next chapter we are testing these hypotheses.

2.3. **Fuzzy convolution.** Independently of the classification method used, in order to optimize the results we propose usage of the fuzzy convolution operation. Fuzzy convolution creates a single classification band by calculating the total weighted inverse distance of all the classes in a window of pixels and assigning the center pixel the class with the largest total inverse distance summed over the entire set of fuzzy classification bands. This has the effect of creating a context-based classification to reduce the noise of the classification. Classes with a very small distance value remain unchanged while classes with higher distance values may change to a neighboring value if there are sufficient number of neighboring pixels with class values and small corresponding distance values.

3. Experiments with real-world data

3.1. **Input data.** For the procedures of image classification was used an orthorectified airborne image from the upper hills of Oradea municipality. This image contains three channels recorded in three bands: the first band for green, the second for red and the third for blue. In the figure below, we present a fragment of this image and some statistics for the whole image.

Figure 1. Image fragment and statistics

Statistics	Band1	Band 2	Band 3
Min value per cell	9	8	9
Max value per cell	255	255	227
Mean Deviation	127	138	109
Mean	127	137	108
Standard deviation	31.330	22.637	17.801
Correlation	2.7	0.169	0.061

3.2. **Definition and verification of the training areas.** Training is the first stage of a supervised classification. In this step the user must define training areas for each class interactively on the displayed image. The areas may be specified both on polygon and on pixel basis. The three classes of information defined are:

streets, houses and green areas (figure below). The signature files were created using a fuzzy membership function based on the sigmoid curve, in this case the accuracy of the signature files are acceptable, over 90 %(table below).

FIGURE 2. Training sites

Data	Streets	Houses	Green Areas
Streets	95.45	4.87	1.27
Houses	2.73	93.42	2.83
Green Areas	1.82	1.71	95.89

3.3. **Classification procedure.** In order to obtain a minimum distortion of the classified image it was studied the possibility of optimizing the classification procedure by using fuzzy theory. We analyzed the result of classification procedures in three experiments.

In the first experiment, the classification procedure is done using unsupervised classification, by using c-means algorithm. The result of fuzzy c-means algorithm on the input data, for three clusters is represented in the set of images presented in figure 3.

In order to optimize the accuracy of the results we consider that is appropriate to use fuzzy logic in both stages of supervised classification. In the following test cases, we are testing these hypotheses. In the second case we studied the result of classification by using fuzzy set theory just for making the signature file and in the third case we analyzed the results of using fuzzy theory in the both stage of supervised classification.

The actual classification process can be done by using fuzzy membership function or based on the standard distance of each pixel to the mean reflectance on each band for a signature.

On the second experiment we used a method based on the standard distance of each pixel till the relevant pixel. The result of the clustering on the chosen image,

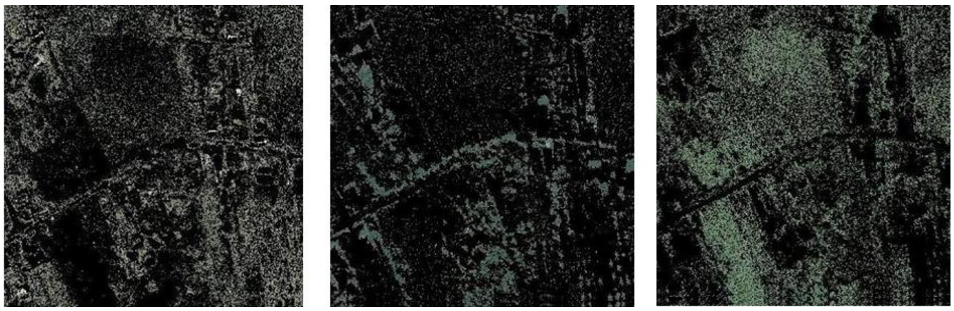

FIGURE 3. Unsupervised Fuzzy Classification - C-means

with the signature file defined before, is represented in figure 4, in three images, one for each cluster:

- The image in the left for the class representing the houses,
- The image in the middle for the class representing the green areas,
- The image in the right for the class representing the streets

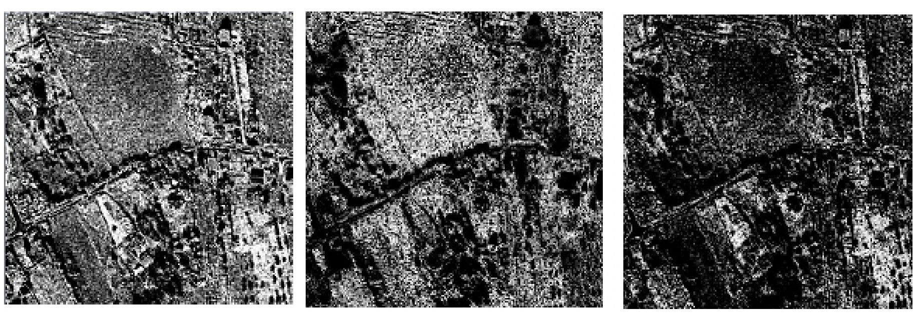

FIGURE 4. Supervised Fuzzy Classification - First Case

For the third experiment of classifying the input images, with the signature files defined before, we used a classifying method which combined the Maximum Likehood method with a fuzzy membership function (linear). The output images are represented in the figure 5, each image represent a cluster in the following order houses, grean areas, streets.

3.4. **Fuzzy convolution.** Classification results can be optimized through fuzzy convolution. The scope of this procedure is to combine the previously created bands in one band by interpolation. The interpolation method used is the weight inverse distance. Applying the fuzzy convolution on the results obtained after the

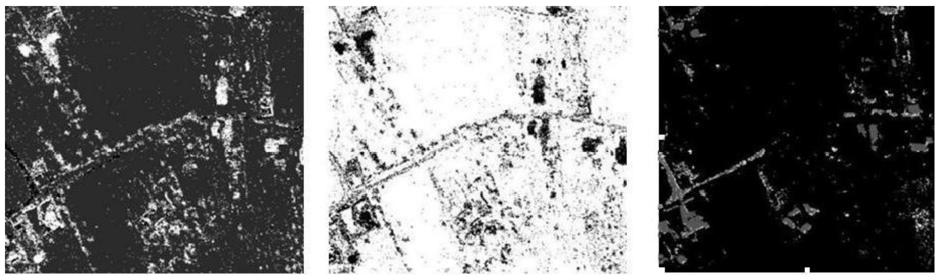

Figure 5. Supervised Fuzzy Classification - Second Case

second case of supervised classification, the image represented in Figure 6 was obtained.

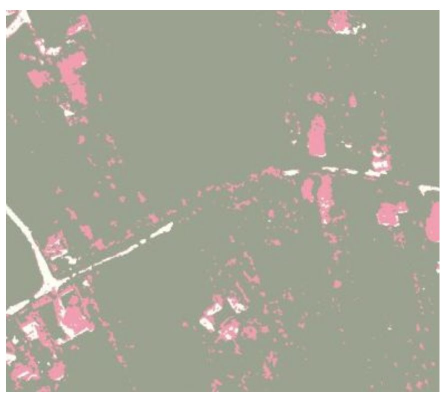

Figure 6. Fuzzy Convolution

4. Results of experiments

The results of classification based on fuzzy method were presented in the previous section, within the figures 1, 2 and 3. As the images show, the supervised classifications have higher quality than the unsupervised classification. For the evaluation of the fuzzy classifications presented before, three sets of accuracy measures have been considered:

- Overall accuracy – based on percent of correct identification;
- Overall kappa coefficient;
- Overall correlation coefficient

	Unsupervised	Supervised - first case	Supervised - second case
Overall accuracy	34%	72%	83,9%
Kappa	0.1151	0.3421	0.4476
Overall Agreement	0.31	0.69	0.89

5. Conclusion

The major advantage of Fuzzy logic theory is that it allows the natural description of data and problems, which should be solved, in terms of relationships between precise numerical values. Fuzzy sets make possible not only the definition of uncertain, vague or probabilistic spatial data, but also allow relationships and operations on them.

The usage of fuzzy theory has implications in improving the quality of the classification of airborne images and object recognition. The Fuzzy set theory offers instruments for supervised and unsupervised classification. The unsupervised fuzzy based classification allows clustering of data, where no a priori information known (c-means algorithms), but the supervised classification is offering higher quality.

The Fuzzy set theory represents a powerful instrument in designing efficient tools to process remote sensing images and also to support the spatial decision-making process.

References

[Ben03] Benz, U.; Hofmann, P.; Willhauck, G.; Lingenfelder, I.; Heynen, M. - Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information, Journal of Photogrammetry & Remote Sensing, 2004

[Eas01] Eastman, R – Idrisi 32 Release 2, Guide to GIS and Image Processing, Volume I- II, Clark University, May 2001

[Gue03] Guesgen, H.;Hertzberg, J.; Lobb, R.;Mantler, A. – Buffering Fuzzy Maps in GIS, Spatial Cognition and Computation, 2003

[Liu03] Liu, H., Li. J., Chapman, M. A.;- Automated Road Extraction from Satellite Imagery Using Hybrid Genetic Algorithms and Cluster Analysis, Journal of Environmental Informatics 1 (2) 40-47 (2003)

[Ste99] Stefanakis, E.; Sellis, T.- Enhancing a Database Management System for GIS with Fuzzy Set Methodologies, Proceedings of the 19th International Cartographic Conference, Ottawa, Canada, August 1999.

[Su03] Su-Yin Tan - A Comparison of the Classification of Vegetation Characteristics by Spectral Mixture Analysis and Standard Classifiers on Remotely Sensed Imagery within the Siberia Region, International Institute for Applied Systems Analysis Austria, 2003

[Tur] Turčan, A.; Ocelíková, E.; Madarász, L. – Fuzzy c-means Algorithms in Remote Sensing, Technical University of Kosice

City Hall of Oradea, GIS Department