

A COMPARISON OF THE PERFORMANCE OF FUZZY ALGORITHM VERSUS STATISTICAL ALGORITHM BASED SUB-PIXEL CLASSIFIER FOR REMOTE SENSING DATA

Anil Kumar^{a*}, S. K. Ghosh^b, V. K. Dadhwal^a

^aIndian Institute of Remote Sensing, Dehradun, India – (anil, dean)[@iirs.gov.in](mailto:iirs.gov.in)

^bIndian Institute of Technology Roorkee, India -scnagosh@datainfosys.net

KEYWORDS: Fuzzy c-Means (FCM), Possibilistic c-Means (PCM), Maximum Likelihood Classifier (MLC), Membership Function.

ABSTRACT:

It is found that sub-pixel classifiers for classification of multi-spectral remote sensing data yield a higher accuracy. With this objective, a study has been carried out, where fuzzy set theory based sub-pixel classifiers have been compared with statistical based sub-pixel classifier for classification of multi-spectral remote sensing data. Although, a number of Fuzzy set theory based classifiers may be adopted, but in this study only two classifiers are used like; Fuzzy c-Means (FCM) Clustering, Possibilistic c-Means (PCM) Clustering. FCM is an iterative clustering method that may be employed to partition pixels of remote sensing images into different class membership values. PCM clustering is similar to FCM but it does not have probabilistic constraint of FCM. Therefore, the formulation of PCM is based on modified FCM objective function whereby an additional term called as regularizing term is also included. FCM and PCM are essentially unsupervised classifiers, but in this study these classifiers are applied in supervised modes. Maximum Likelihood Classifier (MLC) as well as Possibilistic Maximum Likelihood Classifier (PMLC), the new proposed algorithm have been studied as statistical based classifier. All the algorithms in this work like; FCM, PCM, MLC and PMLC have been evaluated in sub-pixel classification mode and accuracy assessment has been done using Fuzzy Error Matrix (FERM) (Binaghi et al., 1999). It was observed that sub-pixel classification accuracy varies with different weighted norms.

1. INTRODUCTION:

Remote sensing images contain a mix of pure and mixed pixels. While digital image classification, however, a pixel is frequently considered as a unit belonging to a single land cover class. But due to limited image resolution, pixels often represent ground areas, which comprise two or more discrete land cover classes (Foody et al. 1996). For this reason it has been proposed that fuzziness should be accommodated in the classification procedure so that pixels may have multiple or partial class membership (Foody et al. 1996b, Foody and Lucas 1997). In this case a measure of the strength of membership for each class is output by the classifier, resulting in a sub-pixel classification technique (Yannis et al. 1999). Also Recent advances in supervised image classification have shown that conventional 'hard' classification techniques, which allocate each pixel to a specific class, are often inappropriate for applications where mixed pixels are abundant in the image (Foody et al. 1996a).

Mixed pixels are assigned to the class with the highest proportion of coverage to yield a hard classification. Due to which a considerable amount of information is lost. To overcome this loss, sub-pixel classification was introduced. A sub-pixel classification assigns a pixel to different classes according to the area it represents inside the pixel. This sub-pixel classification yields a number of fraction images equal to the number of land cover classes. Several researchers have addressed this sub-pixel mixture problem. Among the most popular techniques for sub-pixel classification are artificial neural networks (e.g. Kanellopoulos et al. (1992)), mixture modeling (e.g. Kerdlis and Grondona (1996)) and supervised fuzzy c-means classification (e.g. Foody (1994)) (Verhoeve et al. 2000). Other technique for sub-pixel classification is Possibilistic Fuzzy c-Means (PCM) (Raghu et al. 1993).

Possibilistic Fuzzy c-Means does not have probabilistic constraint that the membership of a data point across classes sum to one. In possibilistic fuzzy c-means case the resulting partition of the data can be interpreted as a possibilistic partition, and the membership values may be interpreted as degrees of possibility of the points belonging to the classes, i.e., the compatibilities of the points with the class prototypes.

In this work Fuzzy as well as Statistical sub-pixel classification algorithms like; FCM, PCM, MLC and PMLC have been compared with respect to overall classification of sub-pixel output using Fuzzy Error Matrix (FERM) (Binaghi et al., 1999). As commercially available image processing software's were not having sub-pixel classification algorithms used in this work. So in-house developed SMIC (Sub-Pixel Multi-Spectral Image Classifier) System (Kumar et al., 2005) having Fuzzy and Statistical classification algorithms with accuracy assessment module was used in this work.

2. SUPERVISED SUB-PIXEL TECHNIQUES:

The Fuzzy and Statistical based different techniques studied in this work are Fuzzy c-Means Approach (FCM), Possibilistic c-Means Approach (PCM), Maximum Likelihood Classifier (MLC) and Possibilistic Maximum Likelihood Classifier (PMLC). Possibilistic Maximum Likelihood Classifier (PMLC) is the new proposed sub-pixel classification algorithm, which is, modified form of Maximum Likelihood Classifier (MLC) algorithm. These algorithms are discussed in following sections;

2.1 FUZZY C-MEANS APPROACH (FCM):

Fuzzy c-Means (FCM) was originally introduced by Jim Bezdek in 1981. In this clustering technique each data point

* Corresponding author.

belongs to a cluster to some degree that is specified by a membership grade, and the sum of the memberships for each pixel must be unity. This can be achieved by minimizing the generalized least - square error objective function,

$$J_m(U, V) = \sum_{i=1}^N \sum_{j=1}^c (\mu_{ij})^m \|X_i - x_j\|_A^2$$

Subject to constraints,

$$\begin{aligned} \sum_{j=1}^c \mu_{ij} &= 1 \text{ for all } i \\ \sum_{i=1}^N \mu_{ij} &> 0 \text{ for all } j \\ 0 &\leq \mu_{ij} \leq 1 \text{ for all } i, j \end{aligned}$$

where X_i is the vector denoting spectral response of a pixel i , X is the collection of vector of cluster centers x_j , μ_{ij} are class membership values of a pixel, c and N are number of clusters and pixels respectively, m is a weighting exponent ($1 < m < \infty$), which controls the degree of fuzziness, $\|X_i - x_j\|_A^2$ is the squared distance (d_{ij}) between X_i and x_j , and is given by,

$$d_{ij}^2 = \|X_i - x_j\|_A^2 = (X_i - x_j)^T A (X_i - x_j)$$

where A is the weight matrix. Amongst a number of A -norms, three namely Euclidean, Diagonal and Mahalonobis norm, each induced by specific weight matrix, are widely used. The formulations of each norm are given as (Bezdek, 1981),

$A = I$	Euclidean Norm
$A = D_j^{-1}$	Diagonal Norm
$A = C_j^{-1}$	Mahalonobis Norm

where I is the identity matrix, D_j is the diagonal matrix having diagonal elements as the eigen values of the variance covariance matrix, C_j given by,

$$C_j = \sum_{i=1}^N (X_i - x_j)(X_i - x_j)^T$$

The class membership matrix μ_{ij} is obtained by;

$$\mu_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{d_{ij}^2}{d_{ik}^2} \right)^{1/(m-1)}}$$

$$\text{where, } d_{ik}^2 = \sum_{j=1}^c d_{ij}^2 ;$$

2.2 POSSIBILISTIC C-MEANS APPROACH (PCM):

The original FCM formulation minimizes the objective function given by

$$J_m(U, V) = \sum_{i=1}^N \sum_{j=1}^c (\mu_{ij})^m \|X_i - v_j\|_A^2 ; \text{ subject to } \sum_{j=1}^c \mu_{ij} = 1 \text{ for all } i$$

But in PCM one would like the memberships for representative feature points to be as high as possible, while unrepresentative points should have low membership in all clusters. The objective function, which satisfies this requirement, may be formulated as;

$$J_m(U, V) = \sum_{i=1}^N \sum_{j=1}^c (\mu_{ij})^m \|X_i - x_j\|_A^2 + \sum_{j=1}^c \eta_j \sum_{i=1}^N (1 - \mu_{ij})^m$$

Subject to constraints,

$$\begin{aligned} \max_j \mu_{ij} &> 0 \text{ for all } i \\ \sum_{i=1}^N \mu_{ij} &> 0 \text{ for all } j \\ 0 &\leq \mu_{ij} \leq 1 \text{ for all } i, j \end{aligned}$$

where η_j is the suitable positive number. The first term demands that the distances from the feature vectors to the prototypes be as low as possible, whereas the second term forces the μ_{ij} to be as large as possible, thus avoiding the trivial solution. Generally, η_j depends on the shape and average size of the cluster j and its value may be computed as;

$$\eta_j = K \frac{\sum_{i=1}^N \mu_{ij}^m d_{ij}^2}{\sum_{i=1}^N \mu_{ij}^m}$$

where K is a constant and is generally kept as 1. The class memberships, μ_{ij} are obtained from,

$$\mu_{ij} = \frac{1}{1 + \left(\frac{d_{ij}^2}{\eta_j} \right)^{1/(m-1)}}$$

2.3 MAXIMUM LIKELIHOOD CLASSIFIER (MLC):

Statistical methods are based on the assumption that the frequency distribution for each class is multivariate normal. Once this assumption is met, sets of parameters such as mean, standard deviation, variance covariance etc are calculated from the data. The accurate estimation of these parameters improves

the accuracy of the classification. That's why this method is called parametric method.

The decision rule assigns each pixel having pattern measurements or features X to the class c whose units are most probable or likely to have given rise to feature vector X . It assumes that training data statistics for each class in each band are normally distributed (i.e., Gaussian in nature). In other words, training data with bi- or tri-modal histograms in a single band are not ideal. In such cases the individual modes probably represent individual classes that should be trained upon individually and labeled as separate classes. This would then produce uni-modal, Gaussian training class statistics that would fulfill the normal distribution requirement.

Maximum likelihood classification makes use of the statistics, including the mean measurement vector, \bar{x}_j , for each class and the variance covariance matrix \sum_j and it allocates the pixel to a class having highest probability density. The probability density function can be written as (Polubinskas et al., 1995),

$$p(x/j) = \frac{1}{(2\pi)^{n/2} |\sum_j|^{1/2}} e^{(-1/2[(X - \bar{x}_j)^T \sum_j^{-1} (X - \bar{x}_j)]}$$

where $p(x/j)$ is the probability density function of a pixel X as a member of class j , n is the number of bands, X is the vector denoting spectral response of pixels, \bar{x}_j the mean vector and \sum_j variance covariance of a class are given by following equations respectively;

$$\bar{x}_j = \frac{\sum_{i=0}^{N_i} X_i}{N_i}$$

$$\sum_j = \frac{\sum_{i=0}^{N_i} (X_i - \bar{x}_j)(X_i - \bar{x}_j)}{N_i(N_i - 1)}$$

where N_i is the total number of the training pixels for class j . The soft classification output was derived from MLC using a posteriori probability, which was computed from;

$$p(x/j) = \frac{p(j)p(x/j)}{\sum_{j=1}^c p(j)p(x/j)}$$

where $p(x/j)$ is the a *posteriori* probability of a pixel belonging to class j , $p(j)$ is the a *priori* probability of the class j and c is the number of classes. The a *posteriori* probabilities were used for class proportion in a pixel and thus, reflect soft classified outputs.

2.4 POSSIBILISTIC MAXIMUM LIKELIHOOD CLASSIFIER (PMLC):

In the new modified Maximum Likelihood Classifier (MLC) known as Possibilistic Maximum Likelihood Classifier (PMLC) approach, $p(x/j)$ a *posteriori* probability is forced to be as large as possible, thus avoiding the trivial solution. This is done with the help of η_j factor, which depends on the shape and average size of the cluster j and its value may be computed as;

$$\eta_j = K \frac{\sum_{i=1}^N p(x/j) d_{ij}^2}{\sum_{i=1}^N \mu_{ij}^m}$$

where K is a constant and is generally kept as 1. The class memberships, $p(x/j)$ are obtained from,

$$p(x/j) = \frac{1}{1 + \left(\frac{d_{ij}^2}{\eta_j}\right)^{1/(m-1)}}$$

3. MULTI-SPECTRAL IMAGE:

A UTM rectified LISS-III image from Resource Sat -1, (IRS-P6) satellite acquired in four bands have been used. The image was acquired in 2003 and covers the rural area of Dehradun District, of Uttaranchal State, India. Approximately, 30% of the area in the image selected is covered by reserved forest, 30% by agriculture land, 15% by barn land, 15% sand, and 10% by river water. The size of the image is 250×296 pixels, with spatial resolution of 23.5 m.

The five classes of interest, namely forest, water, agriculture, barn, and sand have been used for this study work (Figure. 1).

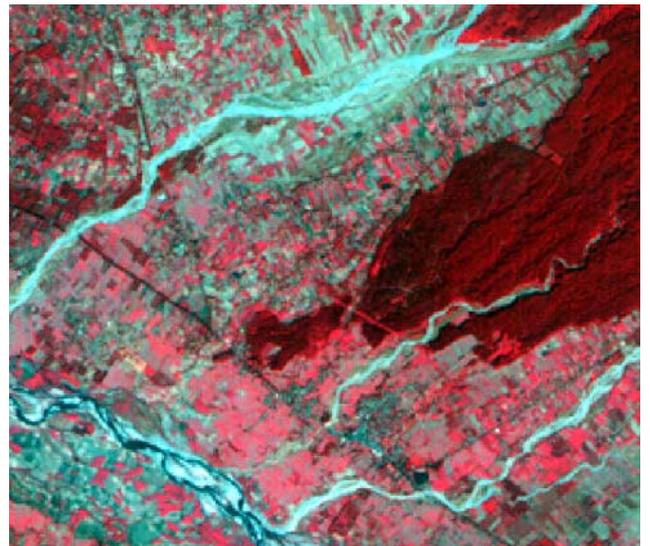


Figure 1. False Color composite from the Resource Sat-1 LISS-III multi-spectral image.

4. CLASSIFICATION RESULTS:

All the four classification algorithms mentioned in this paper were used as supervised sub-pixel classification approach. In the case of FCM, PCM, MLC and PMLC all the three norms of weighted matrix A were studied. Accuracy assessment of all the sub-pixel output was evaluated using FERM approach. The training as well as testing data used for supervised sub-pixel classification approach was greater than 10n, where n is dimension of data used. Separate data were used at training as well as at testing stage. At testing stage 500 samples were taken for overall accuracy assessment of sub-pixel output. Results of all the four classification approaches are mentioned in table 1.

Sl. No.	Different Sub-pixel Classification Algorithms	Overall Accuracy (%)
1	FCM with Euclidean Norm	97.90
2	FCM with Diagonal Norm	91.23
3	FCM with Mahalonobis Norm	90.87
4	PCM with Euclidean Norm	99.29
5	PCM with Diagonal Norm	88.85
6	PCM with Mahalonobis Norm	86.47
7	MLC with Euclidean Norm	98.12
8	MLC with Diagonal Norm	87.42
9	MLC with Mahalonobis Norm	84.47
10	Possibilistic MLC with Euclidean Norm	98.54
11	Possibilistic MLC with Diagonal Norm	94.38
12	Possibilistic MLC with Mahalonobis Norm	87.40

Table 1. Overall Accuracy while using different Sub-pixel classification approaches.

5. CONCLUSION:

Table: 1, shown in section 4.0, shows that overall sub-pixel classification accuracy of multi-spectral remote sensing data varies while using different classification approaches with different norms. While considering Euclidean Norm, PCM approach gives maximum sub-pixel classification accuracy of 99.29%. In case of diagonal norm, Possibilistic MLC gives maximum sub-pixel classification accuracy of 94.38%. While considering Mahalonobis Norm, FCM approach gives maximum sub-pixel classification accuracy of 90.87%. From this work it gives the level of confidence, which norm can give better accuracy while using Fuzzy as well as Statistical sub-pixel classification algorithm.

ACKNOWLEDGMENTS

We thank Dr. V. K. Dadhwal, Dean, Indian Institute of Remote Sensing, Dehradun and Dr. S. K. Ghosh, Associate Professors, Department of Civil Engineering, Indian Institute of Technology, Roorkee, for their valuable guidance and support.

REFERENCES:

1. Bezdek, J. C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum, New York, USA.
2. Binaghi, E., Brivio, P. A., Chessi, P. and Rampini, A. (1999). A fuzzy Set based Accuracy Assessment of

Soft Classification, *Pattern Recognition letters*, 20, 935-948.

3. Blanz, V., Schölkopf, B., Bühlhoff, H., Burges, C., Vapnik, V. and Vetter, T. (1996). Comparison of view-based object recognition algorithms using realistic 3d models. In C. von der Malsburg, W. von Seelen, J. C. Vorbrüggen, and B. Sendhoff, editors, *Artificial Neural Networks—ICANN'96*, pages 251 – 256, Berlin, Springer Lecture Notes in Computer Science, Vol. 1112.
4. Burges, C. J. C. (1996). Simplified support vector decision rules. In Lorenza Saitta, editor, *Proceedings of the Thirteenth International Conference on Machine Learning*, pages 71–77, Bari, Italy, Morgan Kaufman.
5. Burges, C. J. C. and Schölkopf, B. (1997). Improving the accuracy and speed of support vector learning machines. In M. Mozer, M. Jordan, and T. Petsche, editors, *Advances in Neural Information Processing Systems 9*, pages 375–381, Cambridge, MA, MIT Press.
6. Burges, C. J. C. (1998). Geometry and invariance in kernel based methods. In *Advances in Kernel Methods - Support Vector Learning*, Bernhard Schölkopf, Christopher J.C. Burges and Alexander J. Smola (eds.), MIT Press, Cambridge, MA.
7. Cortes, C. and Vapnik, V. (1995). Support vector networks. *Machine Learning*, 20:273–297.
8. Drucker, H., Burges, C. J. C., Kaufman, L., Smola, A. and Vapnik, V. (1997). Support vector regression machines. *Advances in Neural Information Processing Systems*, 9:155–161.
9. Foody G. M., Cox, D. P. (1994). Sub-pixel land cover composition estimation using a linear mixture model and fuzzy membership functions. *International Journal of Remote Sensing*, vol. 15, no. 3, pp. 619-631.
10. Foody, G. M. and Boyd D. S. (1996). Observation of temporal variations in environmental gradients and boundaries in tropical West Africa with NOAA AVHRR data, *Proceedings RSS: Science and Industry*, Remote Sensing Society, Nottingham, UK.
11. Foody, G. M. (1996). Land cover mapping from remotely sensed data with a neural network: accommodating fuzziness, *Proc. 1st COMPARES Workshop*, York, UK, July.
12. Foody, G. M. (1996). Approaches for the production and evaluation of fuzzy land cover classifications from remotely sensed data, *Int. J. Remote Sensing*, vol. 17, no. 7, pp. 1317-1340.
13. Foody, G. M., Lucas, R. M., Curran, P. J. and Honzak, M. (1997). Non-linear mixture modelling without end-members using an artificial neural network, *Int. J. Remote Sensing*, vol. 18, no. 4, pp. 937-953.

14. Girosi, F. (1998). An equivalence between sparse approximation and support vector machines. *Neural Computation*; CBCL AI Memo 1606, MIT.
15. Joachims, T. (1997). Text categorization with support vector machines. Technical report, LS VIII Number 23, University of Dortmund. <ftp://ftp-ai.informatik.uni-dortmund.de/pub/Reports/report23.ps.Z>.
16. Kanellopoulos, I., Varfis, A., Wilkinson, G. G. and Megier. (1992). Land cover discrimination in SPOT HRV imagery using an artificial neural network: a 20-class experiment. *International Journal of Remote Sensing*, vol. 13, no.5, pp. 917-924.
17. Kerdiles, H., and Grondona, M. O. (1996). NOAA-AVHRR NDVI decomposition and subpixel classification using linear mixing in the Argentinean Pampa. *International Journal of Remote Sensing*, vol. 16, no. 7, pp. 1303- 1325.
18. Kumar, A., Ghosh, S. K., Dadhwal, V. K. (2005). Advanced Supervised Sub-Pixel Multi-Spectral Image Classifier, Map Asia 2005 conference, Jakarta, Indonesia.
19. Mukherjee, S., Osuna, E. and Girosi, F. (1997). Nonlinear prediction of chaotic time series using a support vector machine. In *Proceedings of the IEEE Workshop on Neural Networks for Signal Processing 7*, pages 511–519, Amelia Island, FL.
20. Müller, K. R., Smola, A., Rätsch, G., Schölkopf, B., Kohlmorgen, J. and Vapnik, V. (1997). Predicting time series with support vector machines. In *Proceedings, International Conference on Artificial Neural Networks*, page 999. Springer Lecture Notes in Computer Science.
21. Osuna, E., Freund, R. and Girosi, F. (1997). An improved training algorithm for support vector machines. In *Proceedings of the 1997 IEEE Workshop on Neural Networks for Signal Processing*, Eds. J. Principe, L. Giles, N. Morgan, E. Wilson, pages 276 – 285, Amelia Island, FL.
22. Palubinskas, G., Lucas, R. M., Foody, G. M. and Curran, P. J. (1995). An evaluation of fuzzy and texture based classification approaches for mapping regenerating tropical forest classes from Landsat TM data, *International Journal of Remote Sensing*, 16, 747-759.
23. Krishnapuram, R., Keller, J. M. (1993). A Possibilistic Approach to Clustering, *IEEE Transactions on Fuzzy Systems*, Vol. 1, No. 2, May.
24. Refaat M. M. and Aly A. F. (2004). Mean Field Theory for Density Estimation Using Support Vector Machines, *Computer Vision and Image Processing Laboratory*, University of Louisville, Louisville, KY, 40292.
25. Schölkopf, B., Burges, C. and Vapnik, V. (1995). Extracting support data for a given task. In U. M. Fayyad and R. Uthurusamy, editors, *Proceedings, First International Conference on Knowledge Discovery & Data Mining*. AAAI Press, Menlo Park, CA.
26. Schölkopf, B., Burges, C. and Vapnik, V. (1996). Incorporating invariances in support vector learning machines. In C. von der Malsburg, W. von Seelen, J. C. Vorbrüggen, and B. Sendhoff, editors, *Artificial Neural Networks — ICANN'96*, pages 47 – 52, Berlin, Springer Lecture Notes in Computer Science, Vol. 1112.
27. Schölkopf, B., Smola, A., Müller, K. R., Burges, C. J. C. and Vapnik, V. (1998). Support vector methods in learning and feature extraction. In *Ninth Australian Congress on Neural Networks*.
28. Schmidt, M. (1996). Identifying speaker with support vector networks. In *Interface 1996 Proceedings*, Sydney.
29. Osuna, E. and Girosi, F. (1998). Reducing the runtime complexity of support vector machines. In *International Conference on Pattern Recognition* (submitted).
30. Smola, A. J., Schölkopf, B. and Müller, K. R. (1998). General cost functions for support vector regression. In *Ninth Australian Congress on Neural Networks*.
31. Smola, A. J. and Schölkopf, B. (1998). On a kernel-based method for pattern recognition, regression, approximation and operator inversion. *Algorithmica*.
32. Smola, A. J., Schölkopf, B. and Müller, K. R. (1998). The connection between regularization operators and support vector kernels. *Neural Networks*.
33. Schölkopf, B., Smola, A. and Müller, K. R. (1998). Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*. In press.
34. Vapnik, V., Golowich, S. and Smola. (1996). A Support vector method for function approximation, regression estimation, and signal processing. *Advances in Neural Information Processing Systems*, 9:281–287.
35. Verhoeve, Jan, Robert, D. W. (2000). Sub-pixel Mapping of Sahelian Wetlands using Multi-temporal SPOT Vegetation Images, *Laboratory of Forest Management and Spatial Information Techniques Faculty of Agricultural and Applied Biological Sciences, University of Gent*.
36. Wahba, G. (1998). Support vector machines, reproducing kernel hilbert spaces and the randomized gacv. In *Advances in Kernel Methods - Support Vector Learning*, Bernhard Schölkopf, Christopher J.C. Burges and Alexander J. Smola (eds.), MIT Press, Cambridge, MA.
37. Yannis, S. A., Stefanos, D. K. (1999). Fuzzy Image Classification Using Multi-resolution Neural with Applications to Remote Sensing, *Electrical Engineering Department, National Technical University of Athens, Zographou 15773, Greece*.