

# **LAND USE MAPPING OF MALAMPUZHA DAM AREA USING FUZZY C MEANS**

Josephina Paul<sup>1</sup>

Abstract: Decade by decade, the reserve of water on the earth is depleting at an alarming rate. Therefore the mapping and conservation of the water sheds and sources are the need of the hour for the survival of generations to come and the future of the universe. At this threat, the development of various methods and techniques to map the water logs are highly relevant. The signal processing and computational techniques can contribute for the study to a large extent. An important water reservoir, the Malambuzha dam, sprawling in the Bharathapuzha river basin of the state of Kerala, India has chosen for this study as it has been vested with the paramount importance in the irrigation and drinking water needs of the Palakkad district. We have used satellite imageries of the area chosen for study for the analysis and inference. Unlike images acquired using common cameras/devices, the noise integrated in the satellite images are mainly of speckle noise models. Wavelets are good tools in denoising the images, by thresholding or shrinking methods as they can represent the images in mathematical terms, in low frequency approximations and high frequency coefficients, which have been used for denoising images in this study. Conventional clustering methods create crisp clusters, i.e. one data object belongs to exactly one cluster, in reality, this may not be true always. Many problems related to nature may contain objects that cannot be considered into a definite class but has a variable degree of belonging to more than one class. This concept is exploited in fuzzy clustering where fuzzy membership is generated for each data member, to every cluster. Hence we have considered the Fuzzy C Means clustering in our study. Results showing the land cover, water body and barren areas are discussed with accuracy metric for estimating the quality of the clusters.

Keywords: Land Use, Watershed, Wavelets, Fuzzy C Means

## **I. INTRODUCTION**

Land use mapping and change detection have been emerged as an enthusiastic research area over the past decades. The importance of land use mapping is due to its widespread applications in the study of urbanization, growth of townships, industries and deforestation estimation etc. The development of signal processing techniques, efficient algorithms of segmenting the images, drastic development in the field of digitizing images and interpreting using software packages etc. have brought this research area into an investigative paradigm. The images are acquired by various means ranging from aerial photography to onboard satellite sensors, which have produced a large volume of data for the study also made the author to work towards this paper. Water is the essential medium for life. Decade by decade, the reserve of water on the earth is depleting at an alarming rate. Therefore the mapping and conservation of the water sheds and sources are the need of the hour for the survival of generations to come and the future of the universe. In this century, the development of various methods and techniques to map the water

---

<sup>1</sup> *KCAET Tavanur*

logs are highly relevant. The signal processing and computational techniques can contribute for the study in large extent. An important water reservoir, the Malambuzha sprawling in the Bharathapuzha river basin of the state of Kerala, India has chosen for this study as it has got the paramount importance in the irrigation and drinking water needs of the Palakkad district.

The land use types and area can be estimated with the help of the modern computing techniques and algorithms with less effort and more accuracy. When images acquired by the above mentioned methods, it may contain undesirable effects due to various intervening phenomena such as atmospheric haze, backscattering of radiations and CCD calibration errors. Such errors are mainly categorized into radiometric errors and geometric errors. The preprocessing of images are carried out to nullify the effect of these kinds of interruptions. Histogram equalization and geo-referencing are the common techniques used for the error correction. As the next step, the noise inherent in the images while capturing, has to be removed, as the noise inhibits the actual signals and signal to noise ratio, a quality measure of the signals will be dropped to very low value. Various filtering operations are done for removing noise from images [1] based on various noise models. Unlike images acquired using common cameras/devices, the noise integrated in the satellite images are mainly of speckle noise models. Speckle noise is inherent in images due to random scattering of objects while using imaging systems of coherent radiation, which creates a granular appearance all over the image and thus the quality of image is degraded severely [2]. Synthetic Aperture Radar (SAR) images are mainly affected with speckle noise, usually described as a multiplicative noise model. This may not be true always, especially at locations with built ups and man made structures[2]. In any case, the statistical dependence between reflectivity and scattering term may not hold [3]. The filtering can be done in spatial domain or in transform domain. Recently, several filtering algorithms have been formulated in transform domain, the most popular being the wavelet transform [4]. The wavelet shrinkage [5] wavelet thresholding and adaptive wavelet thresholding [6] are mainly employed for filtering in the wavelet domain. Wavelets are used to represent the images in mathematical terms, to low frequency approximations and high frequency coefficients. It is a mathematical model similar to Fourier Transform (FT), but the inefficiency of FFT to represent the time component is compensated in wavelets. The frequency as well as time localization is encoded in the approximation coefficients of the image. A number of wavelet families have been developed for representing signals in different scale [7]. Haar [1985] Meyer [1989] were the earliest steps in this and later Daubechies [1989] developed a family of wavelets called Daubechies wavelets in this respect.

The rest of the paper is divided into sections. Section 2 describes the state of the art Wavelet families that we have employed in this letter and the segmentation algorithm, the Fuzzy C Mean's algorithm. In section 3, the experimental set up and the details of the data set are given. In section 4, experimental results and its analysis are discussed in detail. The fifth and last section is provided with the conclusion and the direction of future improvements in this area.

## II. WAVELETS

The wavelets decompose the image into sub bands of multiple resolution – ie. the low frequency and high frequency components. The low frequency sub band is an approximation sub band, which contains the low frequency information. It can be decomposed into various scales subsequently. Thus, it can be said that a wavelet is a multiscale, multidirection signal decomposition tool. There are one dimensional wavelets which apply along a single dimension

that generates one low frequency sub band and one high frequency sub band; and two dimensional wavelets, that apply along row as well as column and thus produces four sub bands. Another categorization of wavelets is Continuous wavelets and Discrete wavelets. Discrete wavelets are the uniformly sampled functions of continuous signals and thus it is dyadic in nature. The two dimensional discrete wavelet decomposes the signal into four components, low frequency, and high frequency along the horizontal, vertical and diagonal direction [8,9]. While Morlet, Mexican hat and Shannon wavelets are continuous, the Daubechies family and Haar wavelet are a few examples of discrete wavelets[10]. The wide range application of wavelets made it so popular within a decade since its advent in late 1980's. In this paper, we have used discrete wavelet families. Wavelets have applications in denoising and image compression also. In this paper, wavelet denoising has been employed.

### III. FUZZY C MEAN'S ALGORITHM

In every image processing system, the segmentation has got a very prominent role in the image interpretation and analysis. A plenty of segmenting algorithms are available in the literature. Segmentation is the process of partitioning the image into homogenous areas, so that the image can be interpreted in a better manner. The segmentation can be in supervised model or unsupervised. While supervised models require a labeled training data set, the unsupervised does not require it and so the latter is recommended for situations where the ground truth is not available. Moreover, the prior knowledge of the number of clusters is not needed in this.

The unsupervised algorithms, are broadly classified into partitioning, hierarchical, density based and grid based, depending upon the working principle. In partitioning algorithm the dataset is divided into  $n$  partitions, in which each object belongs to exactly one cluster. In hierarchical algorithms, a bottom up agglomerative operation or top down divisive approach is performed[13]. In density based clustering, the density of neighborhood particles is taken into consideration as threshold value to form clusters and it can form any arbitrary shapes unlike the other clustering methods. A set of grids are formed in grid based algorithms and the dataset is collected in each grid to form clusters. While all the above methods create crisp clusters, ie. one data object belongs to exactly one cluster, in reality, this may not be true always. Many problems related to nature may contain objects that cannot be considered into a definite class but has a variable degree of belonging to more than one class. This concept is exploited in fuzzy clustering where fuzzy membership is generated for each data member, to every cluster. When the image resolution is much less than the size of the objects as in the case of satellite images, this kind of clustering is found to be nearer to the reality and hence we have considered the fuzzy clustering in our study.

The fuzzy partition algorithm was proposed by Zadeh in 1965, in which the dataset was divided into clusters by assigning a degree of membership to each cluster. Bezdek et al. have implemented the FORTRAN version of the Fuzzy C Means (FCM) algorithm in 1984 [11]. Since then, several monographs of FCM have been proposed by researchers implementing on large variety of applications. Fuzzy c-Means algorithm can be represented mathematically, at this point as follows. The fuzzy clustering that incorporates spatial information has been implemented in [14], which had used the neighborhood window of pixels,  $3 \times 3$  or  $5 \times 5$  in size, to exploit the nearby information for deciding the pixels membership in the cluster.

#### **Problem definition:**

Let  $X = \{x_1, x_2, \dots, x_n\}$  be a sample of  $n$  observations in  $R^n$  ( $n$  dimensional Euclidean space).

Let  $x_k$  is  $k^{th}$  feature vector. Let  $c$  is an integer  $1 \leq c \leq n$ , is the number of clusters or crisp partitions. Partitions  $P = \{p_1, p_2, \dots, p_c\}$  is subsets of  $X$ , which satisfies the following three conditions.

$$p_i \neq \phi$$

$$p_i \cap p_j = \phi$$

$$\cup p_i = X$$

In order to partition the data set  $X$  into  $c$  partitions, we have applied Fuzzy  $c$  means algorithm which generates  $c \times n$  membership values. Let  $U$  be a real  $c \times n$  matrix  $u = [u_{ik}]$ , it must satisfy the following rules:

$$u_{ik} = \begin{cases} 1, & p_k \in X_i \\ 0, & otherwise \end{cases}$$

$$\sum_{i=1}^c u_{ik} > 0 \quad \text{for all } i$$

$$\sum_{i=1}^c u_{ik} = 1 \quad \text{for all } k$$

It can be denoted the set of fuzzy partitions as

$$M_{fc} = \{U_{c \times n} | u_{ik} \in [0,1]\}$$

In order to identify an optimal set of fuzzy partitions, an objective function is used. The sum of the least square error function of all the clusters is to be minimum, for an optimal set. It is expressed as

$$J_m(U, v) = \sum_{k=1}^n \sum_{i=1}^c (u_{ik})^m \|X_k - v_i\|^2$$

where

$X = \{x_1, x_2, \dots, x_n\} \subset \mathbb{R}^n$  is the dataset.

$c$  is the number of clusters  $2 \leq c \leq n$

$m$  is the weighting factor  $1 \leq m \leq \infty$

$U$  is the fuzzy membership matrix, ie. fuzzy  $c$  partition  $U \in M_{fc}$ ,

$\|\cdot\|$  is norm of matrix. Here we have taken the Euclidean norm. The optimal clusters can be found out by minimizing the objective function  $J_m$ .

The cluster centers are calculated as

$$v_i = \frac{\sum_{k=1}^n (\hat{u}_{ik})^m X_{ik}}{\sum_{k=1}^n (\hat{u}_{ik})^m} \quad 1 \leq i \leq c$$

$$\hat{u}_{ik} = \left( \sum_{j=1}^c \left( \frac{d_{ik}}{d_{ij}} \right)^{2/(m-1)} \right)^{-1} \quad 1 \leq k \leq N,$$

$$1 \leq i \leq c$$

Where  $d_{ik}$  is the squared distance between  $X_{ik}$  and  $v_i$ , denoted as

$$d_{ik} = \|X_k - v_i\|^2$$

The Fuzzy C Means algorithm is described into steps as detailed below.

Input: Raw image, Output: segmented image

1. Fix the parameters  $m=2$ ,  $c$ = no of clusters to be formed. Initialize the fuzzy membership matrix with random numbers,  $U^{(0)} \in M_{fc}$
2. Compute centre of clusters using equation
3. Compute the updated membership matrix  
 $U^{k+1} = [\hat{u}_{ik}]$
4. If  $\|U^{k+1} - U^k\| < \text{epsilon}$  stop; Otherwise repeat step 2 to 4.

Finally, we get a membership matrix that indicates the membership of each data member to every cluster. If the membership is 1, then it becomes a crisp partition. When it is 0, the data member has no membership in  $k^{\text{th}}$  cluster.

A widely accepted method to map the pixel's membership to the segments of the image is to choose the cluster in which the data member is having maximum membership by applying the Max operator.

$$\text{Max } \{u_{ik}\}_{k=1,2,\dots,c}$$

#### IV. DATASETS AND STUDY AREA

The image was downloaded from earth explorer, a landsat image of the area covering 4 districts of Kerala from Palakkad, Malappuram, Thrissur and Ernakulam taken in the year 2006. The geometric co-ordinates ranges from  $10^{\circ}5'N, 76^{\circ}27'48''E$  to  $11^{\circ}35'20''N, 77^{\circ}17'33''E$ . A small portion of  $300 \times 300$  pixels were cut out from the georeferenced image which covers the Malampuzha reservoir that rests in the river basin of Bharathapuzha, in the Palakkad district area. The image was geo-referenced with ArcGIS software and registered with the co-ordinates. The Matlab software version 2010 working on a corei3 processor, 4GB RAM machine was used for image processing. The image was decomposed using discrete wavelets of various families and found out that the db5 of Daubechies wavelet family generates optimum result. In this work, the wavelet thresholding was applied for denoising the image using the Donoho & Johnstone formula[12],

Threshold =  $\sigma \sqrt{2 \log N}$  where  $\sigma$  is the standard deviation of the noise, N is the size of the noisy signals.

After denoising the image, the inverse wavelet transform was applied to bring back the image into the pixel domain. For the segmentation of the image, the FCM was executed as it is suitable for the land cover pixels.

#### V. RESULTS AND DISCUSSION

The source image and the resultant clusters are shown in Figure 1. The segmented image is displayed in figure 1.

The result obtained for various bands and the PCC are given in the Table 1.

Cluster	1	2	3	Total
1	6494	0	0	6494
2	0	47797	0	47797
3	0	2142	33567	35709
Total	6494	49939	33567	90000
	PCC	97.62%		

Table 1. Error Matrix with PCC of FCM, Band2



Fig. 1 (a) Image band 2 (b) Ground Truth (c) FCM cluster

The result obtained was verified with PCC metric. The overall accuracy obtained is 97.62%.

From the results, it is seen that three classes of land use types are found in the study area. The water body indicated in dark pixels, the barren area indicated in white pixels and the thick vegetation in gray pixels. The barren area may be exposed due to reduction of water in the reservoir over a long period. That means the water logged area in the reservoir has been reduced over the months or perhaps years. A thick vegetative land cover is seen in the nearby area of the dam.

## VI. CONCLUSION

The satellite images showing the area of the study on segmentation with a suitable algorithm, the Fuzzy C Means could generate a good quality segmented image that has shown the three classes of land cover, the one being the water reservoir, the second is the barren area appeared from the prior water logged area and the land cover with thick vegetation. This study could identify the depleting trend of water sheds, which can be used for proper measures to be taken to recharge and conserve the dam area. A better set of multitemporal images can be used to clearly delineate the changed areas, which is proposed as our future work.

## REFERENCES

- [1] Pei-Eng Ng and Kai-Kuang Ma, A Switching Median Filter With Boundary Discriminative Noise Detection for Extremely Corrupted Images, *IEEE TRANSACTIONS ON IMAGE PROCESSING*, VOL. 15, NO. 6, JUNE 2006.
- [2] Fabrizio Argenti and Luciano Alparone, Speckle Removal From SAR Images in the Undecimated Wavelet Domain, *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 40, No. 11, November 2002.
- [3] A. Lopès, E. Nezry, R. Touzi, and H. Laur, "Structure detection and statistical adaptive speckle filtering in SAR images," *Int. J. Remote Sens.*, vol. 14, no. 9, pp. 1735–1758, 1993.
- [4] S. S. O. Choy, Y.-H. Chan, and W. C. Siu, "Adaptive image noise filtering using transform domain local statistics," *J. Opt. Eng.*, vol. 37, no. 8, pp. 2290–2296, Aug. 1998.

- 
- [5] D. L. Donoho, "Denoising by soft-thresholding," *IEEE Trans. Inform.Theory*, vol. 41, pp. 613–627, May 1995.
- [6] J. R. Sveinsson and J. A. Benediktsson, "Speckle reduction and enhancement of SAR images using multiwavelets and adaptive thresholding," in *Proc. SPIE Conf. Image and Signal Processing for Remote Sensing V*, S. B. Serpico, Ed: EUROPTO Series, 1999, vol. 3871, pp. 239–250.
- [7] Ronald W. Lindsay, Donald B. Percival, and D. Andrew Rothrock, The Discrete Wavelet Transform and the Scale Analysis of the Surface Properties of Sea Ice, *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 40, No. 11, November 2002.
- [8] R. C. Gonzalez and R. E. Woods, *Digital Image Processing 2/E*. Upper Saddle River, NJ: Prentice-Hall, 2002, pp. 349-404.
- [9] S. Mallat, *Academic press - A Wavelet Tour of Signal Processing 2/E*. San Diego, Ca: Academic Press, 1999, pp. 2-121.
- [10] Pao-Yen Lin, An Introduction to Wavelet Transform.
- [11] James C. Bezdek, Robert Ehrlich, William Full C., Fcm: The Fuzzy C-Means Clustering Algorithm, *Computers & Geosciences* Vol. 10, No. 2-3, pp. 191-203, 1984.
- [12] Donoho, D.L. & Johnstone, I.M. (1992a). Ideal spatial adaptation via wavelet shrinkage. *Biometrika*, Vol. 81, No. 3 (Aug., 1994), pp. 425-455.
- [13] Jiawei Han and Micheline Kamber, *Data Mining Concepts and Techniques*, second Edition, Elsevier.
- [14] S. Bandyopadhyay (2005) Satellite image classification using genetically guided fuzzy clustering with spatial information, *International Journal of Remote Sensing*, 26:3, 579-593