

# A NOVEL APPROACH FOR GEOREFERENCED DATA ANALYSIS USING HARD CLUSTERING ALGORITHM

**Y.Sophiya Banu<sup>1</sup>, Y.Soniya Banu<sup>2</sup>, V.V. Karthikeyan<sup>3</sup>**

<sup>1,3</sup>SNS College of Engineering, Coimbatore.,<sup>2</sup>PSG College of Technology, Coimbatore  
*sophiyabanu05@gmail.com, soniyayusuffsai@gmail.com, karthi\_maharaja@yahoo.com*

### Abstract

The process of defining its existence in physical space is called as Georeferencing. That is establishing its terms of projections or coordinate systems. When data from different sources need to be combined and then used in a GIS application. In this work georeferenced data on soil map is clustered using a soft clustering algorithm. Most georeferencing tasks are undertaken to generate new map. Thus a map generated using GIS software is clustered for data analysis of soil type and vegetation possibilities. Remotely sensed data plays an important role in data collection, the platforms usually consist of aircraft and satellites. GIS is attached to many operations and has many applications related to engineering, planning, management, telecommunications and business.

**Keywords:** Soil map, K-Means Clustering Algorithm, Geographic Information System.

-----\*\*\*-----

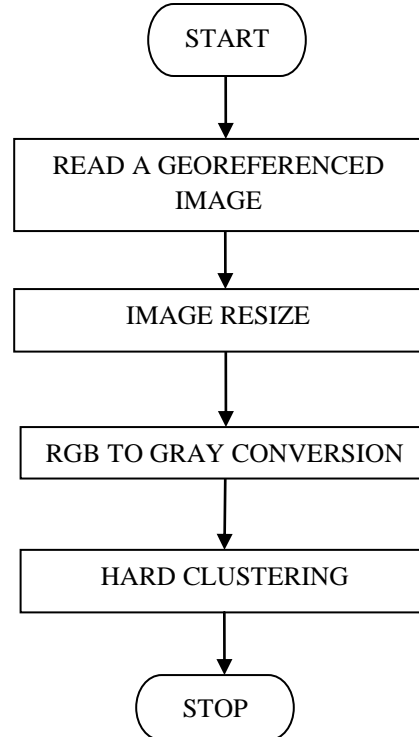
## 1. INTRODUCTION

A Geographic Information System (GIS) is a system designed to capture, store, manage, manipulate, analyze, and present all types of geographical data in a digital format. Greene et.al has given a detailed description. GIS is merging of cartography and computer science technology. Heywood et.al (2006) has denoted its applications. GIS and location intelligence applications can be the foundation for many location-enabled services that rely on analysis and dissemination of results for collaborative decision making. Remotely sensed data also plays an important role in data collection. It consists of sensors attached to a platform. J. Sun et.al (2010) has represented its process and operation. Sensors include cameras, digital scanners and LASER, while platforms usually consist of aircraft and satellites. More advanced data processing can occur with image processing techniques. It includes contrast enhancement, false color rendering and a variety of other techniques including use of two dimensional Fourier transforms. In this work spatial domain processing of the georeferenced data analysis has been attempted. The satellite data obtained by remote sensing is represented in Figure 2(a) and Georeferenced data are shown in Figure 2(b). The spatial location of other geographical features is determined. The hard clustering technique called K-means is adopted for the clustering of the georeferenced data.

## 2 METHODOLOGY

The Flow chart of the proposed method of georeferenced data analysis is shown in Figure 1. Preprocessing is considered as very important task in the hard clustering analysis of georeferenced data. Initially the image is converted into a gray scale image. In the second stage the image is resized to reduce

the computational complexity. At the third stage the image is clustered into individual regions. The georeferenced image is clustered into regions based on the flow chart shown in Figure 3.



**FIGURE 1** Flow chart of Georeferenced data clustering

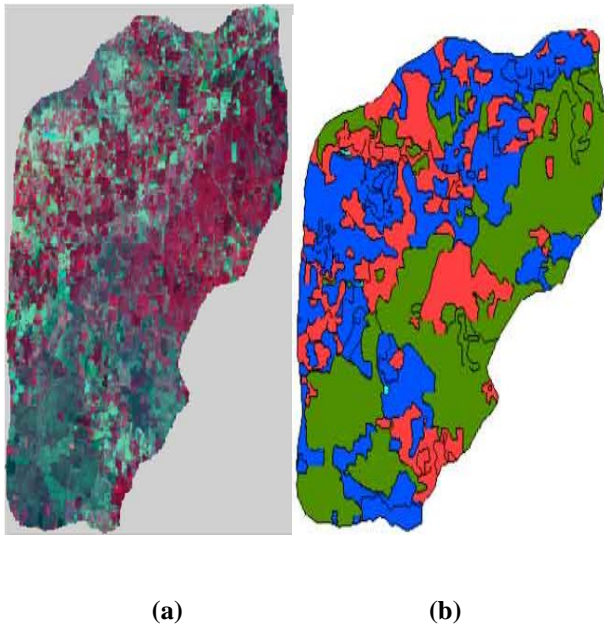


FIGURE 2 (a) Satellite image (b) Georeferenced data

## 2.1 Georeferenced Data

Cartography is a branch of science that deals with the design and production of maps. Calvert et al. (2013) has mentioned its importance. It is the visual representation of spatial data. The vast majority of modern cartography is done with the help of computers. Perry et al. (2006) has denoted a detailed procedure of georeferencing. GIS based maps are best in quality but production of quality cartography is also achieved by importing layers. A design program to refine it is applied. Fonseca and Egenhofer (1999) has given a brief introduction to soil maps. Most GIS software gives the user friendly control over the appearance of the data. A map produced using a GIS software is considered as the input to the soft clustering analysis of georeferenced data analysis process. A soil map is a visual representation of an area. Soil map is a map that is a geographical representation. Sarma and Jean-Paul Legros (2005) has denoted its developments. It shows the diversity of soil types and soil properties in the area of interest. It is typically the end result of a soil survey investigations. Chang, K. T. (1989) has denoted its importance. Soil maps are most commonly used for land evaluation. Spatial planning, agricultural extension, environmental protection and similar process can also be applications of soil map. In this work the soil map is read as an input through the Matlab command. Many maps are static two-dimensional, geometrically accurate representation of three dimensional spaces. Clarke K. C. (1986) has given its process and applications. Soil maps are typically richer in context and show higher spatial details than the traditional soil maps.

## 2.2 Image Resize

Image resize is the second task of the process flow as depicted by the Figure 1. The `imresize` command in matlab without altering the pixel values it resizes the image and also computes the number of rows or columns automatically to preserve the image aspect ratio. Bentoutou et al. (2002) has given detailed procedure. Hence the georeferenced Image is resized to (256 x 256).

## 2.3 RGB to Gray Conversion

There are several established color models used in computer graphics, but the two most common are the RGB model (Red-Green-Blue) for computer display and the CMYK model (Cyan-Magenta-Yellow-Black) for printing. Gonzalez and Woods (2008) has presented a detailed view about digital image processing. When red, green and blue light is combined it forms white. Thus to reduce the computational complexity the georeferenced data that exists in RGB color model is converted into a gray scale image. The range of gray scale image from black to white values can be calculated by equation (1). Where L is Luminance, R is RED, G is Green and B is Blue.

$$L = 0.2989 * R + 0.5870 * G + 0.1140 * B \quad (1)$$

## 2.4 Hard Clustering Simulation Results

K-Means Clustering algorithm is also called as Hard clustering technique. Clustering based image segmentation is a pixel by pixel segmentation method, and it stops the process of clustering when the image is segmented into predefined number of clusters. The flow chart of hard clustering technique is denoted in the Figure 3. A cluster is therefore a collection of objects which are "similar" between them and are "dissimilar" to the objects belonging to other clusters. J. Wood (1996) has denoted its results. The segmentation results of georeferenced data are shown in Figure 3. The input image is preprocessed and clustered into individual regions. Tian and Guo (2011) has presented the advantages of clustering the soil maps and georeferenced data. In this work georeferenced soil map is taken as input is represented in Figure 2(b). The Hard clustering technique is initiated with cluster value 5. In this work Image segmentation is carried using K-Means Clustering algorithm. K-Means is a method of clustering which allows one piece of data to belong to two or more clusters. It is based on minimization of the distance between the clusters. Partitioning is carried out through an iterative optimization of the mean distance between the cluster centers.

Segmentation is the process of partitioning a digital image into multiple segments. Choi et al. (2011) has denoted the segmentation process. The goal of segmentation is to simplify and change the representation of an image into something that is more meaningful and easier to analyze. In this project clustering technique is used for image segmentation of chromosomal images. The segmentation results are represented in Figure 4. The region by region segmented

image are denoted in Figure 4((a),(b),(c),(d)). Figure 4((e)and(f)) are the segmented results of FCM. The segmented image is coded with specific color for understanding using Matlab commands.

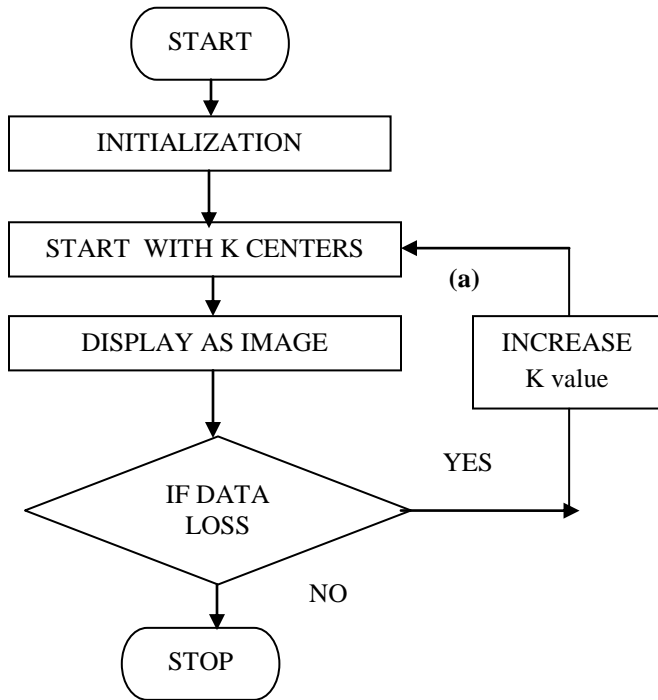


FIGURE 3 Flow chart of hard clustering

**ALGORITHM**

- STEP 1: Initialize the **K** with 5 ( $K=1, \dots, N$ ).
- STEP 2: Initialize cluster center  $c_n(n=1, \dots, K)$
- STEP 3: Data grouping based on defined **K** values.
- STEP4: Reshaping the clustered data and display the segmented image.
- STEP 5: Determine the presence and absence of data remaining.
- STEP 6: Increment the **K** value for further segmentation if the remaining data exist. If else stop the process

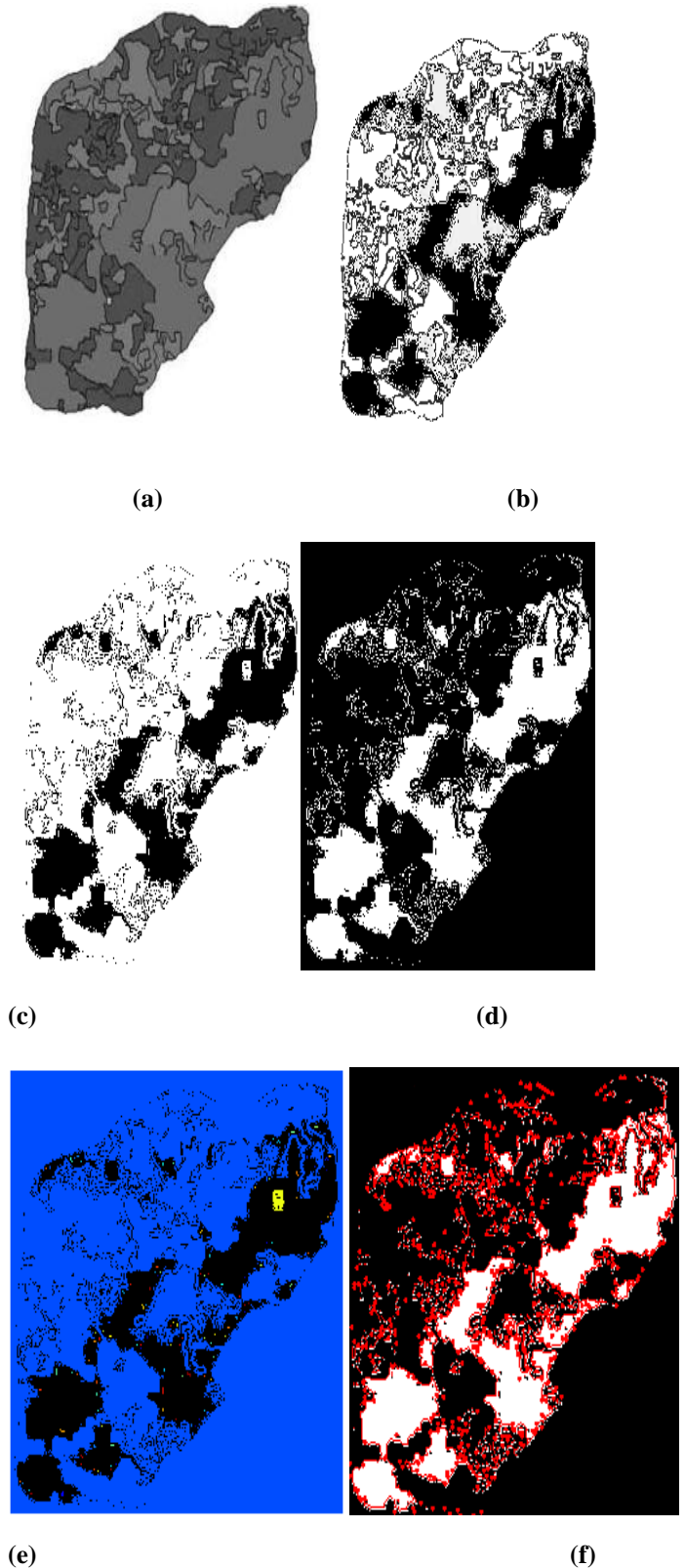


Figure 4 Simulation results (a) Gray scale image, ((b),(c),(d),(e),(f))Regions segmented

### 3. CONCLUSION AND FUTURE WORK

Georeferenced data analysis is one of the important tasks in the field of Geospatial Information Studies. It denotes the academic discipline or career of working with geographical information. During the analyzing process complex image retrieved from the satellite is georeferenced using GIS software. In the similar manner the soil map thus received using a resource satellite is the input to this proposed method of Hard clustering technique. A Hard clustering technique which is the other name of the K- Means Clustering algorithm has been attempted with a soil map image. The segmented results produced by K- Means are denoted in Figure 4. When the complexity of the image increases the efficiency in computing decreases. The proposed method of georeferenced data analysis uses a preprocessing stage before clustering. Thus it reduces the segmentation error. Georeferenced data which are read as an input has been segmented into region by region. The pixel by pixel clustering for georeferenced data analysis is an ideal method even in the case of a noisy image because the input image can be denoised using any of the spatial domain or frequency domain filters. It would be less complex to analyze the possibility of vegetation and the types of soil with the help of a soft clustering technique by a Geoinformatics scholar when compared with the hard clustering results represented in Figure 4((c) and (d)). The K-Means clustering algorithm techniques has been evolving with its adaptive nature. Hence if such Adaptive nature of K-Means clustering algorithm is used it would produce even better quality of segmented result. A comparative study is possible with the emerging algorithms of clustering. Thus when compared with the soft clustering segmented result shown in Figure 4((e) and (f)) the K- Means segmented results are needed to be improved for further analysis. Hence for georeferenced data analysis process using clustering technique the soft clustering technique based on Fuzzy clustering is advantageous than the hard clustering technique.

### REFERENCES

- [1] Greene, R., Devillers, R. and Luther, J. "GIS-based multi-criteria analysis". *Geography Compass* 5/6, 412–432.
- [2] Heywood, I., Cornelius, S., & Carver, S. "An Introduction to Geographical Information Systems", Third Edition. Essex, England: Prentice Hall, 2006.
- [3] K. Calvert, J. M. Pearce, W.E. Mabee, "Toward renewable energy geo-information infrastructures: Applications of GIScience and remote sensing that can build institutional capacity" *Renewable and Sustainable Energy Reviews* 18, pp. 416–429, 2013.
- [4] Chang, K. T. "A comparison of techniques for calculating gradient and aspect from a gridded digital elevation model". *International Journal of Geographical Information Science* 3 (4): 323–334, 1989.
- [5] Chang, K. T. "Introduction to Geographical Information Systems". New York: McGraw Hill. pp. 184, 2008.
- [6] Fu, P., and J. Sun, "Web GIS: Principles and Applications" ESRI Press. Redlands, CA, 2010.
- [7] Clarke, K. C., "Advances in geographic information systems, computers, environment and urban systems", Vol. 10, pp. 175–184, 1986.
- [8] Sarma and Jean-Paul Legros, "Mapping of the soil cover", Enfield, NH, USA, Science Publishers, 2005.
- [9] Cao H. and Wang Y. "Segmentation of M-FISH images for improved classification of chromosomes with an adaptive Fuzzy C-Means Clustering Algorithm". *IEEE Transaction on Biomedical Imaging*, 20(1), 1442–1445, 2011.
- [10] Rafael C. Gonzalez, and Richard E. Woods. "Digital Image Processing". Pearson Education Third ed. 2008.
- [11] Perry, Matthew, Hakimpour, Farshad, Sheth and Amit, "Analyzing Theme, Space and Time", *proc. ACM International Symposium on GIS*. pp. 147–154, 2006.
- [12] Fonseca Frederico, and Egenhofer, "Ontology-Driven Geographic Information Systems". *Proc. ACM International Symposium on Geographic Information Systems*. pp. 14–19, 1999.
- [13] Ma, Y., Guo, Y., and Tian, X., "Distributed Clustering-Based Aggregation Algorithm for Spatial Correlated Sensor Networks". *IEEE Sensors Journal* 11, 2011.
- [14] J. Wood, "Invariant pattern recognition: A review", *Pattern Recognit.*, vol. 29, no. 1, pp. 1–17, Jan. 1996.
- [15] Y. Bentoutou, N. Taleb, M. C. E. Mezouar, M. Taleb, and L. Jetto, "An invariant approach for image registration in digital subtraction angiography", *Pattern Recognit.*, vol. 35, no. 12, pp. 2853–2865, 2002.
- [16] B. Mahdian and S. Saic, "Detection of copy-move forgery using a method based on blur moment invariants", *Forensic Sci. Int.*, vol. 171, no. 2/3, pp. 180–189, Sep. 2007.
- [17] V. Ojansivu and J. Heikkilä, "A method for blur and similarity transform invariant object recognition", *in Proc. ICIAP*, R. Cucchiara, Ed., 2007, pp. 583–588.