ISSN : 2347 - 8446 (Online)
ISSN : 2347 - 9817 (Print)

Vol. 2, Issue 2, Ver. 2 (April - June 2014)

International Journal of Advanced Research in
Computer Science & Technology (IJARCST 2014)

Color Based Image Segmentation Using Data Mining Functionalities

[I]Nishi Sharma, [II]Mr. Anurag Joshi
[I]CSE, Northern India Engineeing College, Shastri Park, Delhi, India
[II]CSE, BIT Mesra, Noida Campus, Noida, India

Abstract

Image classification the process of classifying image into n number of classes .classification can be categorized into two types supervised classification and unsupervised classification.

In this work Remote sensing image classification has been done using unsupervised technique which is known as clustering.There are so many methods used for clustering such as model based method,grid based method ,hierarchical methods ,partitioning methods. Partition methods and Model based method has been used to classify the images.

Image Classification plays an important role in image retrieval system. Image classification refers to the process of partitioning a digital image into multiple regions. The goal of image classification is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze. There are several approaches to image classification. Existing approach which is partitioning methods "K-Means Algorithm" has been discussed with the solution Model based method "Neural Network approach (SOFM) algorithm".

Keywords

Clustering, k-means, som, Remote sensing

I. Introduction

Image classification can be defined as the process of reducing an image to information classes. The categorization of image pixels is based on their digital numbers/grey values in one or more spectral bands. Image classification is perhaps the most important part of digital image . Clustering is the unsupervised classification of patterns (observations, data items, or feature vectors) into groups (clusters). It is important to understand the difference between clustering (unsupervised classification) and discriminate analysis (supervised classification). In supervised classification, we are provided with a collection of labeled (pre- classified) patterns; the problem is to label a newly encountered, yet unlabeled, pattern. Typically, the given labeled (training) patterns are used to learn the descriptions of classes which in turn are used to label a new pattern. In the case of clustering, the problem is to group a given collection of unlabeled patterns into meaningful clusters. In a sense, labels are associated with clusters also, but these category labels are data driven;that is, they are obtained solely from the data.

II. Related Work

Image segmentation can be done using clustering methods.model based methods,grid based methods,partitioning methods are the methods of clustering.Partitioning methods which are based on the distance function are mostly used to form cluters.

A. k-means algorithm

k-means algorithm are used for color based image segmentation which is a type of partitioning method. Simply speaking it is an algorithm to classify or to group your objects based on attributes/ features into K number of group. K is positive integer number. The grouping is done by minimizing the sum of squares of distances between data and the corresponding cluster centroid. Thus, the purpose of K-mean clustering is to classify the data.The distance function is used to calculate the distance between the object and centroid.

$$J = \sum_{j=1}^{k}\sum_{i=1}^{n}\left\| x_i^{(j)} - c_j \right\|^2$$

K-means algorithm
Take a satellite image as input.
Assume k clusters.
Randomly guess k cluster center locations.
Assign r, g, and b values to the centroid location.
Repeat until convergence is obtained.
For counter_height= 1 to heigh
For counter_width= 1 to width
For counter_cluster= 1 to cluster
(Calculate Euclidean distance between input and clusters to find out which center it is closest to)
Distance=((rgb[i][j][0]-CentroidPixel[k][0])^2+(rgb[i][j][1]-CentroidPixel[k][1])^2 + (rgb[i][j][2]-CentroidPixel[k][2])^2) 1/2
DistanceMatrix [counter_cluster] =Distance;
Endloop
Find minimum index
For counter_cluster= 1 to cluster
Recompute the cluster center by averaging all the pixels in the clusters.
Endloop
Endloop
Endloop

B. Self organize map

Artificial neural networks are processing systems analogous to biological neural networks, presenting neurons, axons, dendrites, neural layers, transfer functions, and so on. Their paradigms fall in three main categories: supervised, reinforced and self-organized. This classification takes into account the amount of data needed for the training phase. Supervised networks use a

ISSN : 2347 - 8446 (Online)
ISSN : 2347 - 9817 (Print)

previous knowledge about the desired outputs, in such a way that the error between the actual input and expected output is a suitable parameter. Reinforced networks rely on the measure of the overall error but do not need their exact output. Self-organizing networks determine by themselves the internal weight representation for the presented input data and do not need supervision. The self-organizing neural networks, also known as Kohonen networks, are networks which incorporate a topology scheme, i.e., take into account the topological structure among units. The input signals are n-tuples and there is a set of m cluster units. Each input is fully connected to all units, which respond differently to the input pattern. At each step in the training phase, the cluster unit with weights that best match the input pattern is elected the winner (usually in a minimum Euclidean distance sense). This winning unit and a neighborhood around it are then updated in such a way that their internal weights be closer to the presented input. The adopted updating factor is not equal for all neurons, but stronger near the winning unit, decreasing for more distant units. Figure 1 shows the basic structure of self-organizing maps. There is input components (white circles) connected to all cluster units (shaded circles). The cluster units can assume any spatial distribution, which are usually linear or planar arrays. Weights are associated to each connection. With time, the gain factor must be reduced and also the neighborhood decreases in size. During the learning phase the node weights are changed in an ordered manner, in such a way that the main image features tend to be organized according to a topological distribution in the network. Adjacent nodes respond similarly, while distant nodes respond diversely. The convergence of the features in the self-organizing map occurs considering some limitations on the gain factor while updating the weights.

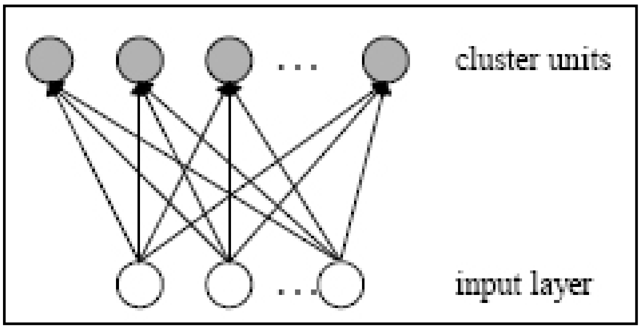
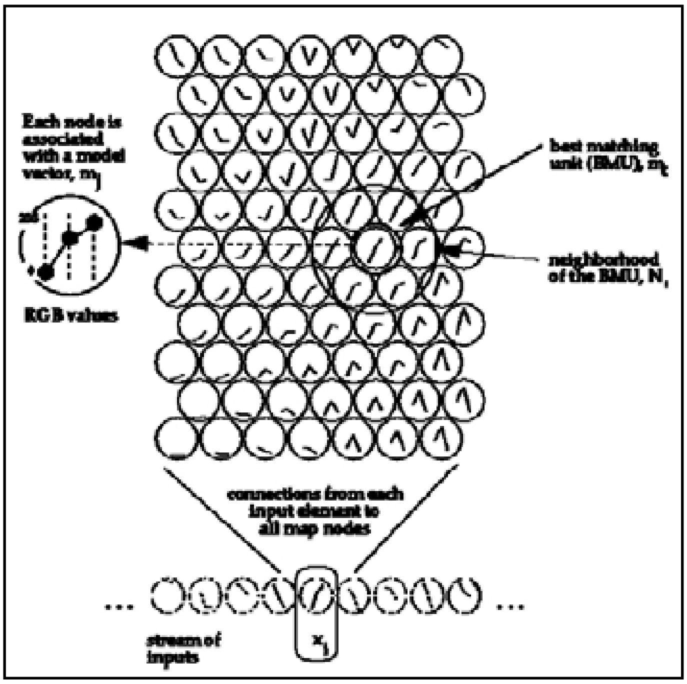


Fig. 1

Image features are represented by their chromaticity values. Chromaticity of a color is the classifications of color with reference to its hue and its saturation i.e. its departure from white light. Hue is the property that distinguishes red from orange, blue from blue-green and so forth. Saturation is the property that distinguishes red from pink. It describes the purity of the color. Chromaticity is obtained by normalizing the RGB components of the image i.e.
r=R/(R+G+B) (Chromaticity of R)
g=G/(R+G+B) (Chromaticity of G)
b=B/(R+G+B) (Chromaticity of B)



C. SOM algorithm

1). Take a satellite image as input
2). Assume k clusters in the beginning.
3). Choose small random weights for cluster nodes.
4). Repeat for N iterations

For counter_height= 1 to height
For counter_width= 1 to width
Express R, G, and B components in terms of their chromaticity values.
r=R/(R+G+B) (Chromaticity of R)
g=G/(R+G+B) (Chromaticity of G)
b=B/(R+G+B) (Chromaticity of B)
For counter_cluster= 1 to cluster
(Every cluster node is examined to determine which one's weights are most like the input vector. This is done by computing the Euclidean distance between the input vector and the cluster node.)
Distance=((r-WeightMatrix[k][0])^2 + (g-WeightMatrix[k][1])^2+(bWeightMatrix[k][2])^2) 1/2
DistanceMatrix [counter_cluster] =Distance;
Endloop
Find best Matching unit (BMU).
(BMU is the node with minimum index).
Next, determine BMU's neighborhood.
Neighborhood Radius=Map Radius * e (-Iteration count / Time Constant)Where, Map Radius= Number of Clusters and Time Constant=Total number of iterations / log (Map Radius)
Determine the nodes that lie within this neighborhood radius.
Update the weights of BMU and its neighbors
W (t+1) =W (t) + L (t) * A (t) * (I (t) – W (t)) Where,Learning rate L (t) = Learning rate parameter * e (-Iteration count / Total no. of Iterations)
Learning rate parameter = 0.1
Amount of Influence A (t) = e (-distance * distance / 2 * Neighborhood radius * Neighborhood radius)
distance = distance between node and BMU
Endloop
Endloop

III. Experimental Results

Satellite image has been used to perform the experiment. Result has been generated using java code for n number of classes. maximum number of clusters can be 10.classified image has been generated using k-means as well as self organized feature map algorithm.Color based image segmentation has been done using 5,8 & 10 classes.

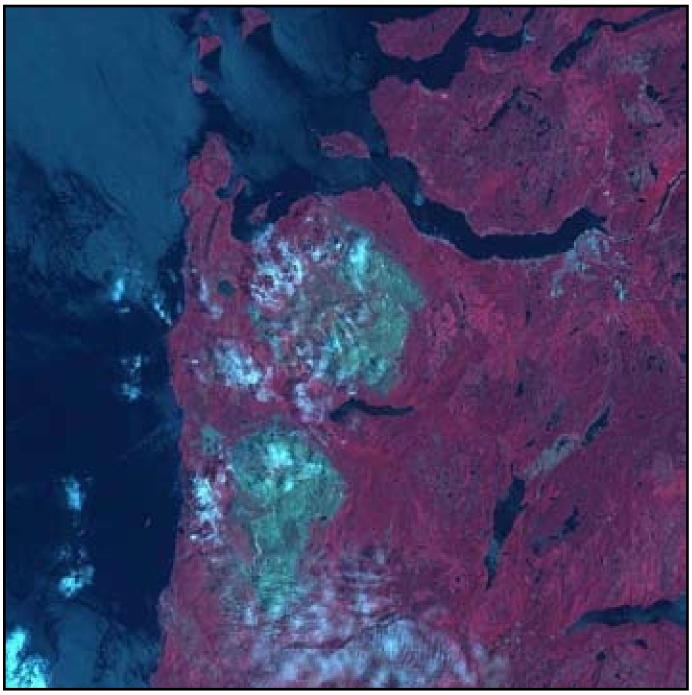
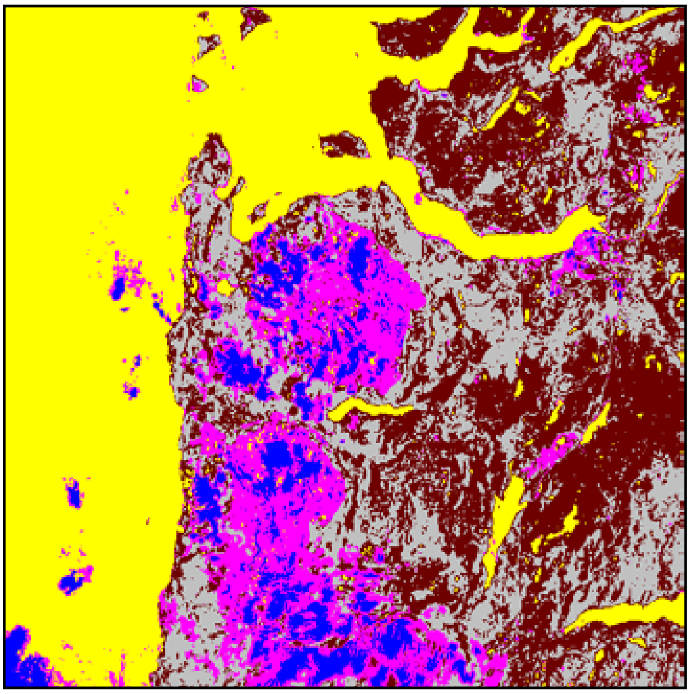
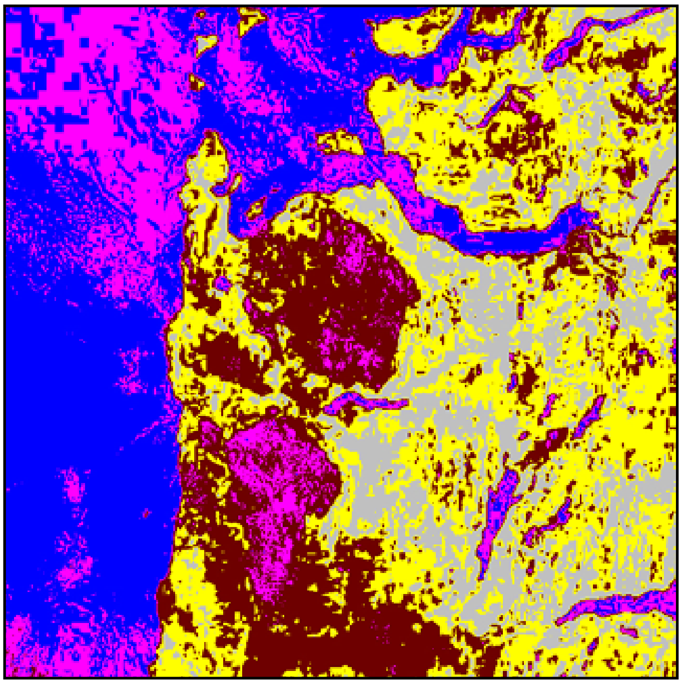


Fig. 2 : Input image bay of island



Color based segmentation using K-means algorithm having 5 classes



Color based image segmentation using SOFM having 5 classes

Algorithms	Classification Accuracy	Time	Iterations
K-Means	Less for higher density	Less	Less
SOFM	High for higher density	More	More

Results of experiment

Result analysis has been shown in the above table. Image classification has been done using 5,8,10 number of classes. As number of classes will increase classification accuracy will decrease.SOFM gives more accuracy in image classification as compare to K-mean algorithm.Image classification time will increase in case of SOFM & number of iterations will also increase in case of SOFM algorithm

IV. Future Work & Conclusion

In this paper Image classification approaches has been discussed. Unsupervised Satellite Image classification has been done with the help of K-means and Self organized feature map algorithm. Both algorithms are based on unsupervised learning. Higher Classification accuracy has been achieved in case of SOFM as compare to K-means algorithm.K-means is based on only distance function so it is sensitive to outliers while SOFM is based on the concept of artificial intelligence which is more accurate as compare to k-means.

There is no single method that clearly outperforms all methods in all problem situations in research papers. Moreover no thorough paper has been written to compare all the methods for all data characteristics on same datasets.

This paper too is confined to analyze impact of change in number of clusters and number of observations.There are other methods also can be used for classified images such as soft-computing methods, Hierarchical methods. A detailed study to compare different methods for different data characteristics and to examine possible interactions among these characteristics can throw more light on this aspect. Each method needs to be analyzed separately

Vol. 2, Issue 2, Ver. 2 (April - June 2014)
ISSN : 2347 - 8446 (Online)
ISSN : 2347 - 9817 (Print)

in detail to examine its performance to change in various data characteristics.

References

[1] Image Segmentation Method based of self organizing map and K-means algorithm, Dragan M. Ristic, Milan Pavlovic, Irini Reljin, Member, IEEE, 9th Symposium on Neural Network Applications in Electrical Engineering, NEUREL-2008

[2] Self Organizing Maps as substitue for K-Means Clustering, Fernando Bacao, Vector Lobo, Computers and Geosciences, Vol 31, pp155-163 Elsevier 2005.

[3] Kohnen Self-Organizing maps, December 2005, Shyam M. Guthikonda, Wittenberg University

[4] "The Self-organizing Map", Teuvo Kohonen, PROCEEDINGS OF THE IEEE, VOL. 78, NO 9, SEPTEMBER 1990

[5] Clustering Methods, Lior Rokach Department of Industrial Engineering Tel-Aviv University, Oded Maimon Department of Industrial Engineering,, Tel-Aviv University.

[6] Comparison between data clustering algoritms,Osama Abu Abbas,Computer science Department, Jordan, The International Arab journal of Information Technology,Vol 5,No. 3,July 2008.