Performance Evaluation of Image Segmentation Using Clustering

Suchi Maheshwary
PG Student SCET,Rajpur,Kadi

Abstract : Segmentation of satellite images is an important issue in various applications. Though clustering techniques have been in vogue for many years, they have not been too effective because of several problems such as selection of the number of clusters. This proposed work tackles this problem by having a validity measure coupled with the new clustering technique. This method treats each point in the data set, which is the map of all possible color combinations in the given image, as a potential cluster center and estimates its potential with respect to the other data elements. The point with the maximum value of potential is considered to be a cluster center and then its effect is removed from the other points of the data set. This procedure is repeated to determine the different cluster centers. At the same time we compute the compactness and the minimum separation among all the cluster centers, also the validity function as the ratio of these quantities. The validity function can be used in making a choice of the number of clusters. This technique has been compared to the Fuzzy C-means technique and the results have been shown for a sample color image of satellite data.

1. Introduction

Segmentation of satellite images is an important task required in many fields. It is especially of a great significance in the area of Geographical Information Systems (GIS) as this helps in planning the activities in the development of resources, study of changing environment and observing the impact of disasters. The basis of segmentation may be mainly made on the image properties such as color or texture [1] or both. Mostly, color can be used in the segmentation of such images, but textural features may also prove to be useful. The perfect segmentation has eluded the researchers still forcing them to try alternative approaches. However, we will make an attempt to use the color property in this paper. In view of wide acceptability and facility of fuzzy approach, we mainly devote our attention on these approaches for the segmentation of color images. Some of the important contributions are the fuzzy C- means approach and robust clustering. However, we will follow the mountain clustering of Yager and Filev but modify the same for increased efficiency and adaptability to the color imagery in the lines of Azeem et all. Techniques available.

II K-means clustering

It is an iterative technique that is used to partition an image into K clusters'-means is one of the simplest unsupervised learning algorithms that solve the well-known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed a priori. The main idea is to define k centroids,[7] one for each cluster. These centroids should be placed in a cunning way because of different location causes different result. So, the better choice is to place them as much as possible far away from each other. The next step is to take each point belonging to a given data set and associate it to the nearest centroid. When no point is pending, the first step is completed and an early group age is done. At this point we need to re-calculate k new centroids as barycenter of the clusters resulting from the previous step. After we have these k new centroids, a new binding has to be done between the same data set points and the nearest new centroid. A loop has been generated. As a result of this loop we may notice that the k centroids change their location step by step until no more changes are done. In other words centroids do not move anymore. K-Means clustering generates a specific number of disjoint, flat (non-hierarchical) clusters. It is well suited to generating global clusters. The K-Means method is numerical, unsupervised, non-deterministic and iterative.

This algorithm aims at minimizing an objective function, e.g. a squared error function. The objective function

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

Where‖xi(j)-cj‖is a chosen distance measure between a xi(j) data point and the cluster center cj, is an indicator of the distance of the n data points from their respective cluster centers.

III Fuzzy c-means clustering

In hard clustering, data is divided into distinct clusters, where each data element belongs to exactly one cluster. In fuzzy clustering (also referred to as soft clustering), data elements can belong to more than one cluster, and associated with each element is a set of membership levels. These indicate the strength of the association between that data element and a particular cluster. Fuzzy clustering is a process of assigning these membership levels, and then using them to assign data elements to one or more clusters.

One of the most widely used fuzzy clustering algorithms is the Fuzzy C-Means (FCM) Algorithm. The FCM algorithm attempts to partition a finite collection of n elements X = {x1,...,xn} into a collection of c fuzzy clusters with respect to some given criterion. Given a finite set of data, the algorithm returns a list of c cluster centers C = {c1,...,cc} and a partition matrix

J U = u i, j ϵ [0,1], i = 1,….,n, j = 1,…,c

III. Mountain Clustering

The mountain clustering approach is a simple way to find cluster centers based on a density measure called the mountain function. This method is a simple way to find approximate cluster centers, and can be used as a preprocessor for other sophisticated clustering methods. The first step in mountain clustering involves forming a grid on the data space, where the intersections of the grid lines constitute the potential cluster centers, denoted as a set V . The second step entails constructing a mountain function representing a data density measure. The height of the mountain function at a point v€ V

$$m(\mathbf{v}) = \sum_{i=1}^{N} \exp\left(-\frac{\|\mathbf{v}-\mathbf{x}_i\|^2}{2\sigma^2} \right),$$

X is data point ,σ is an application specific constant. This equation statesthat the data density measure at a point v is affected by all the points i x in the data set, and this density measure is inversely proportional to the distance between the data points i x and the point under consideration v. The constant σ determines the height as well as the smoothness of the resultant mountain function. The third step involves selecting the cluster centers by sequentially destructing the mountain function. The first cluster center 1 c is determined by selecting the point with the greatest density measure. Obtaining the next cluster center requires eliminating the effect of the first cluster. This is done by revising the mountain function: a new mountain function is formed by subtracting a scaled Gaussian function centered at 1 c

$$m_{\text{new}}(\mathbf{v}) = m(\mathbf{v}) - m(\mathbf{c}_1)\exp\left(-\frac{\|\mathbf{v}-\mathbf{c}_1\|^2}{2\beta^2} \right)$$

The subtracted amount eliminates the effect of the first cluster. Note that after subtraction, the new mountain function new( ) m v reduces to zero at 1=v c . After subtraction, the second cluster center is selected as the point having the greatest value for the new mountain function. This process continues until a sufficient number of cluster centers is attained.

D. Subtractive Clustering

The problem with the previous clustering method, mountain clustering, is that its computation grows exponentially with the dimension of the problem; that is because the mountain function has to be evaluated at each grid point. Subtractive clustering solves this problem by using data points as the candidates for cluster centers, instead of grid points as in mountain clustering. This means that the computation is now proportional to the problem size instead of the problem dimension. However, the actual cluster centers are not necessarily located at one of the data points, but in most cases it is a good approximation, especially with the reduced computation this approach introduces.

Since each data point is a candidate for cluster centers, a density measure at data point i x is defined as

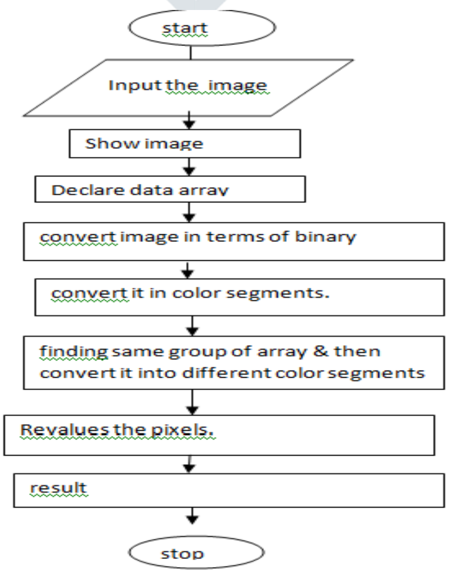
$$D_i = \sum_{j=1}^{n} \exp\left(-\frac{\|\mathbf{x}_i-\mathbf{x}_j\|^2}{(r_a/2)^2} \right),$$

Where are is a positive constant representing a neighborhood radius. Hence, a data point will have a high density value if it has many neighboring data points. The first cluster center 1 cx is chosen as the point having the largest density value 1 cD . Next, the density measure of each data point i x is revised as follows:

$$D_i = D_i - D_{c_1}\exp\left(-\frac{\|\mathbf{x}_i-\mathbf{x}_{c_1}\|^2}{(r_b/2)^2} \right)$$

Where b r is a positive constant which defines a neighborhood that has measurable reductions in density measure. Therefore, the data points near the first cluster center 1 cx will have significantly reduced density measure. After revising the density function, the next cluster center is selected as the point having the greatest density value. This process continues until a sufficient number of clusters is attainted.

FLOW CHART of proposed work



In the implementation work step..
1 Input image reads an underwater image from the file specified by the string filename. If the file is not in the current folder, or in a folder on the MATLAB path, specify the full pathname.
2 Show Image displays the underwater image in a Handle Graphics figure, where image is a grayscale, RGB (true color), or binary image. For binary images, displays pixels with the value 0 (zero) as black and 1 as white.
3 Declare data array A data **array** is a systematic arrangement of objects, usually in rows and column. Declaration of array in this step is there in the algorithm.
4 Convert image in terms of binary & find the minimum distance.
5 After finding minimum distance convert it in color segments.
6 Finding same group of array & then convert it into different color segments.
7 Revalues the pixels.

Conclusion

After observation it is concluded that fuzzy c means cluster which gives less value of Mutual Information (MI) which is best suited for underwater image segmentation.Theuse of clustering is to identify natural groupings of data from a large data set to produce a concise representation of a system's behavior. Fuzzy c-means (fcm) is a data clustering method in which a dataset is grouped into n clusters. The fuzzy c-means (FCM) algorithm is a clustering algorithm developed by Dunn, and later on modified by Bezdek.FCMalgorithm takes more time.but obtain much information form an image.in future work FCM algorithm takes high time, but obtain much information form an image. future work is to decrease the FCM time.

References:

[1] IssamDagherandKamalElTom,WaterBalloons: A Hybrid Watershed Balloon Snake Segmentation, Proceedings of International Joint Conference on Neural Networks, Orlando, Florida, USA, in 2009 IEEE SOCIETY.

[2] YI-WEI YU and JUNG-WA WANG, Image Segmentation Based on Region Growing and Edge Detection, 2011 IEEE VI – 798.

[3] WANG Shi-Iong,WANLei,TANGXu-dong, A Modified Fast Fuzzy C-means Algorithm based on the Spatial Information for Underwater Image Segmentation, 2010 International Conference On Computer Design And Applications (ICCDA 2010), 2010 IEEE VI-524.

[4] Dr.G.Padmavathi, Mr.M.Muthukumar, Mr. Suresh Kumar Thakur, Implementation and Comparison of Different Segmentation Algorithms used for Underwater Images Based on Nonlinear Objective Assessments, 2010 3rd International Conference on Advanced Computer Theory and Engineering(ICACTE), 2010 IEEE(V2-393)

[5] Peizhezhang,canpingli,region based image segmentation of fishes with complex bacground in water,2011 IEEE

[6] Guangyu Liu and HongyuBian, XiufenYe,Hong Shi, An Improved Spectral Clustering Sonar Image Segmentation Method, Proceedings of the 2011 IEEE\ICME International Conference on Complex Medical Engineering May 22 - 25, Harbin, China, 20 11 IEEE.

[7] Ding Caihong, Jiang Lin, Du Fenghai,*Segmentation of Color Image Contour Using Fuzzy Cluster Analysis,IEEE 2010.*

[8] R. Adollah, M.Y Mashor, E. U. Francis, N. H. Harun, Bone Marrow Image Segmentation Based on Multilevel Thresholding,2011 IEEE.

[9] DongPing Ming1, Qun Wang1, Jiancheng LUO2, Zhanfeng SHEN2,Evaluation of High Spatial Resolution Remote Sensing Image Segmentation Algorithms, 2009 IEEE.