International Journal of Computer Science and Engineering Research and Development (IJCSERD), ISSN 2248-9363 (Print), ISSN 2248-9371 (Online), Volume 1, Number 2, May-October (2011)
pp. 70-74 © PRJ Publication
http://www.prjpublication.com/IJCSERD.asp

IJCSERD

© PRJ PUBLICATION

A COMPREHENSIVE SURVEY OF CONTEMPORARY RESEARCHES ON IMAGE SEGMENTATION THROUGH CLUSTERING

Ritu Agrawal[1] , Prof. Manisha Sharma[2]
1 Electronics and Telecommunication , C.I.T, Rajnandgaon, C.G, India
2 Electronics and Telecommunication , B.I.T, Durg, C.G, India
ritube_03@yahoo.co.in , manishasharma1@rediffmail.com

ABSTRACT

This paper presents an analysis on different clustering techniques for image segmentation. Clustering is the unsupervised classification of patterns (observations, data items, or feature vectors) into groups (clusters). Clustering is most widely spread approach in Image segmentation because of its robust characteristics for data classification. Clustering is done on different attributes of an image such as size, color, texture etc. The purpose of clustering is to get meaningful result, effective storage and fast retrieval in various areas.

Keywords: Clustering, Image segmentation, Exclusive Clustering ,Overlapping Clustering , Hierarchical clustering , Probabilistic D-Clustering

1. INTRODUCTION

Clustering in Image segmentation is defined as the process of identifying groups of similar objects. A cluster is a collection of data points that are similar to one another within the same cluster and dissimilar to data points in other clusters [1]. Clustering techniques can broadly be classified as Unsupervised Clustering (calculated by software) and Supervised (human-guided) clustering. Unsupervised clustering is defined as groupings of pixels with common characteristics which are based on the software analysis of an image. Supervised clustering is based on the idea that a user can select sample pixels in an image that are representative of specific classes. Segmentation is a process of subdividing an image into its constituent regions or objects. The level of details, to which the subdivision is carried on, depends on the problem being solved So, the segmentation should stop when the objects or regions of interest in an application have been detected.

Image segmentation is a technique that partitions an image into uniform and non-overlapping regions. This technique has a variety of applications including computer vision, image analysis, medical image processing, remote sensing and geographical information system.

2. CLUSTERING

Clustering is a common technique for data analysis, which is used in many fields, including machine learning, data mining, pattern recognition, image analysis and bioinformatics. Clustering is the method of classification of similar objects into different groups or more precisely the partitioning of a data set into subsets (clusters), so that the data in each subset (ideally) share some common trait . The goal of clustering algorithm is to maximize the intra-cluster similarity and minimize the inter-cluster similarity.

A variety of clustering technique has been introduced to make the segmentation more effective. The clustering technique can be broadly classified as: 1. Exclusive Clustering 2. Overlapping Clustering 3. Hierarchical Clustering 4. Probabilistic D-Clustering.

3. SEGMENTATION

Segmentation is a process of partitioning a digital image into multiple segments or a sets of pixels. The goal of segmentation is to simplify an image into some more meaningful and easier to analyze. Image segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in images. More precisely, image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain visual characteristics.

Segmentation has been used in a wide range of applications. Different applications require different types of images. The most commonly used images are light intensity (LI), range (depth) image(RI), computerized tomography(CT), magnetic resonance images(MRI). Image segmentation is highly dependent on the image type, hence there is no single generalized technique that is suitable for all images.

There are numerous image segmentation techniques in the literature , which can be broadly classified into two categories, namely i) classical ii) fuzzy mathematical . Fuzzy mathematical techniques are widely used in computer vision applications as they are far better able to handle and segment images, particularly noisy images.

4. CLUSTERING TECHNIQUES

An image may contain more than one objects and to segment an image in a meaningful feature is a very difficult job.

This paper is a review and summarizes different clustering technique.

4.1 Exclusive Clustering

In this case data are grouped in an exclusive way, so that if a certain datum belongs to a definite cluster then it could not be included in another cluster. K-means clustering is one of the type of exclusive clustering and is one of the simplest unsupervised learning algorithms.

In case of K means clustering , k centroid must be defined for each cluster. The algorithm is composed of the following steps:

Step 1: Place K points into the space represented by the objects that are being clustered . These points represent initial group centroids.

Step2: Assign each object to the group that has the closest centroid.
Step3: When all objects have been assigned, recalculate the positions of the K centroids.
Step 4: Repeat Step2 and 3 until the centriods no longer moves. This produces a separation of the objects into groups which the metric to be minimized can be calculated.

4.2 Overlapping Clustering

The overlapping clustering, uses fuzzy sets to cluster data, so that each point may belong to two or more clusters with different degrees of membership. In this case, data will be associated to an appropriate membership value. Fuzzy C means is one of the type of overlapping clustering algorithm. Fuzzy c-means (FCM) is a method of clustering which allows one piece of data to belong to two or more clusters. This method is frequently used in pattern recognition.

The algorithm is composed of the following steps:
Step1: Initialize prototype
$V=\{v_1,v_2,\ldots v_c\}$
Repeat $V^{previous} \leftarrow V$
Compute membership function using equation

$$\mu_{ci}(x) = 1 / \sum_{i=1}^{k} [(\| x-v_i \|^2) / (\| x-v_j \|^2)]^{1/m-1} \quad 1\le i\le k,\ x\in X$$

update the prototype , v_i in V using equation

$$v_i = \sum_{x\in X} (\mu_{ci}(x))^m \times x / \sum_{x\in X}^{n} (\mu_{ci}(x))^m \quad 1\le i\le k$$

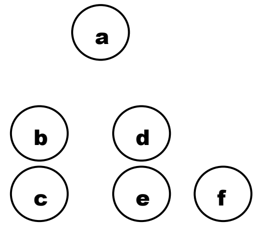
$$\text{until} \sum_{i=1}^{c} \| v_i^{previous} - v_i \| \le \varepsilon$$

where, X : an unlabeled data set
c: the number of clusters to form
m: the parameter in the objective function.
ε: a threshold for the convergence criteria.

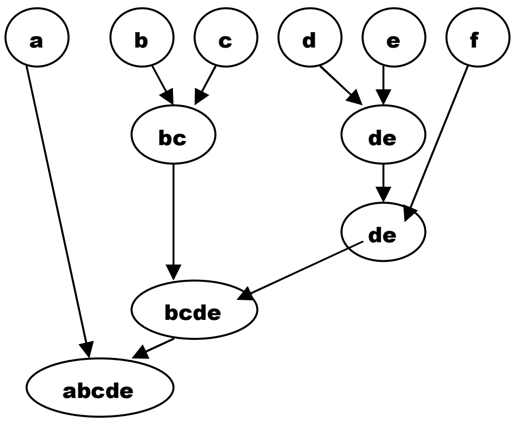
4.3 Hierarchical clustering

Hierarchical clustering creates a hierarchy of clusters which may be represented in a tree structure. The root of the tree consists of a single cluster containing all observations, and the leaves correspond to individual observations.

For example, suppose this data is to be clustered



In this example, we have six elements {a} {b} {c} {d} {e} and {f}. The first step is to determine which elements is to merge in a cluster. Usually, we want to take the two closest elements, according to the chosen distance by using Euclidian distance.
In this example, cutting the row yield clusters {a} {b c} {d e} {f}. Then again Cutting another row yields clusters {a} {b c} {d e f}.The combination of clusters depend on the distace between the clusters. As clustering progresses, rows and columns are merged and as the clusters are merged, the distances updated. This is a common way to implement this type of clustering.



Hierarchical representation

To stop clustering either when the clusters are too far apart to be merged or when there is a sufficiently small number of clusters.

The algorithm used are Given a set of N items to be clustered, and an N*N distance matrix then.

The basic algorithm of hierarchical clustering is the N*N matrix is D = [d(i,j)]. The clustering's are assigned sequence numbers 0,1,......, (n-1) and L(k) is the level of the k^{th} clustering. A cluster with sequence number "m" is denoted (m) and the proximity between clusters (r) and (s) is denoted d [(r),(s)].
The algorithm is composed of the following steps:
Step1: Begin with the disjoint clustering having level L(0)= 0 and the sequence number m=0 .
Step2: Find the least dissimilar pair of clusters in the current clustering , say pair (r), (s) according to d[(r),(s)]= min
d[(i),(j)] where the minimum is over all pairs of clusters in the current clustering. **Step 3:** Increment the sequence number m= m+1 . Merge clusters (r)and (s) into single clusters to form the next clustering m. Set the level of this clustering to L(m)= d[(r),(s)].
Step 4: Update the proximity matrix D , by deleting the rows and columns corresponding to clusters (r) and (s) and adding a row and columns corresponding to the newly formed cluster. The proximity between the new cluster, denoted(r,s) and the old cluster (k) is defined in this way
d[(k),(r,s)] = min d[(k),(r)],d[(k),(s)]
if all objects are in one cluster , stop else go to Step 2 .

4.4 Probabilistic – D Clustering

Here the word D means distance (Euclidean/ Exponential). The probability of cluster membership at any point is assumed to be inversely proportional to the distance from the centre of cluster.

If, $P_k(x)$ = probability that the point x belongs to cluster C_k.
$d_k(x)$ = distance of point x from cluster C_k.
Then: $P_k(x) \cdot d_k(x)$ = constant , depending on (x).
The clustering criterion used here is Euclidean distance

Mathematically

$$P_k(x) = \prod_{j \neq k} d_j(x) \;/\; \sum_{i=1}^{k} \prod_{j \neq i} di(x)$$

If we consider the distance as Exponential then, Probability equation will be changed as $d_j(x)$ will be replaced by e $d_j(x)$.

5 CONCLUSION

The paper presents an analysis on different clustering techniques used for image segmentation. Through clustering algorithms, image segmentation can be done in an effective way. Fuzzy C-means algorithm proved to be superior over other clustering approaches in terms of segmentation efficiency. The major drawback of FCM is the huge computational time required for convergence. To avoid the computational time of FCM , probabilistic – D clustering is proposed .

REFERENCES

[1] www.wikipedia.com
[2] Andrew Moore: “K-means and Hierarchical Clustering - Tutorial Slides”
[3] Brian T. Luke: “K-Means Clustering”
[4] J. C. Dunn (1973): "A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters", *Journal of Cybernetics* 3: 32-5
[5] J. C. Bezdek (1981): "Pattern Recognition with Fuzzy Objective Function Algorithms", Plenum Press, New York.
[6]Data Clustering A Review ACM Computing Surveys, Vol. 31, No. 3, September 1999
[7] Paper 193-2011
Comparison of Probabilistic-D and k-Means Clustering in Segment Profiles for B2B Markets SAS Global Forum 2011
[8] A Survey on Image Segmentation Through Clustering International Journal of Research and Reviews in Information Sciences Vol. 1, No. 1, March 2011
[9] Image Segmentation using Fuzzy Clustering: A Survey. 6th International Conference on Emerging Technologies (ICET) 2010