ISSN: 2321-7782 (Online)
Special Issue, December 2013
International Journal of Advance Research in Computer Science and Management Studies
Research Paper
Available online at: www.ijarcsms.com

A Survey on Image Segmentation Techniques and Clustering

P. Sravani[1]
M.E Computer Science and Engineering
Velammal Engineering College
Chennai – India

S. Deepa[2]
Asst Prof, Computer Science and Engineering
Velammal Engineering College
Chennai - India

Abstract: ***Image is information which has to be processed effectively. Segmentation is a branch of image analysis, which partitions the image into multiple segments. Image segmentation assigns label to every pixel in an image such that pixels with the same label share certain visual characteristic and these segmented is analyzed for classification. Several segmentation algorithms are used for this purpose, but all are not suitable for all types of images. In this paper, Threshold method, Region based method and Clustering are discussed. Types of clustering are also studied. Fuzzy clustering is a class of cluster analysis and is powerfully growing approach. Clustering techniques are been widely used for effective automated segmentation.***

Keywords: Segmentation, Clustering, Fuzzy C-Means Clustering (FCM).

I. INTRODUCTION

The images are considered one of the most important means of information transmission; therefore the image processing is an important tool in a variety of fields such as video coding, computer vision and medical imaging. Image processing in general encompasses an enormous range of techniques which are used in very wide range of applications. Many image analysis processes is based on image segmentation. It is a process of partitioning an image into different regions having same features [1] and is often used to extract region of interests (ROI). Segmentation forms a set of homogeneous and meaningful regions, such that the pixels in each partitioned region possess an identical set of properties or attributes. The sets of properties of the image may include gray levels, contrast, spectral values, or texture properties, etc. The result of segmentation is a number of homogeneous regions, each having unique label. Image segmentation can be classified into three categories:

- *Supervised*- In this method require the interactivity in which the pixels belonging to the same intensity range pointed out manually and segmented.
- *Automatic*- This is also known as unsupervised methods, where the algorithms need some priori information, so these methods are more complex.
- *Semi-automatic*- This is the combination of manual and automatic segmentation.

The most promising in segmentation of image in general based on clustering. Clustering is unsupervised learning method. Here a collection of "points" are examined and grouped into "clusters". Similarities between the points are measured. Similarity can be distance measure. Pixels are grouped with their closeness.

Some of practical applications of image segmentation are: the medical imaging tasks that consist of location of tumors and other pathologies, recognition of the objects in images of remote sensing obtained, automated-recognition systems to inspect the electronic assemblies, biometrics, automatic traffic controlling systems, machine vision, separating and tracking the regions appearing in consequent frames of an sequence, and finally, the real time mobile robot applications employing vision systems.

The rest of the paper is organized as follows: Section II contains different segmentation methods, in Section III clustering types and Section IV conclusion and finally Section V with references.

II. SEGMENTATION METHOD

This section describes about various image segmentation methods. In this paper, techniques like thresholding approaches, region growing approaches, clustering approaches, wavelet segmentation are discussed. All these methods can used together to solve a complex problem.

Many of the segmentation algorithms have common steps for image segmentation can be found in Fig. 2.1. Noise in image degrades the segmentation process. Initially the noise must be removed by filtering process. Noises are of different types like salt-pepper noise, Poison noise, etc. Various filters like Mean filter, Median filter, Wiener filter, etc can be used for noise removal.

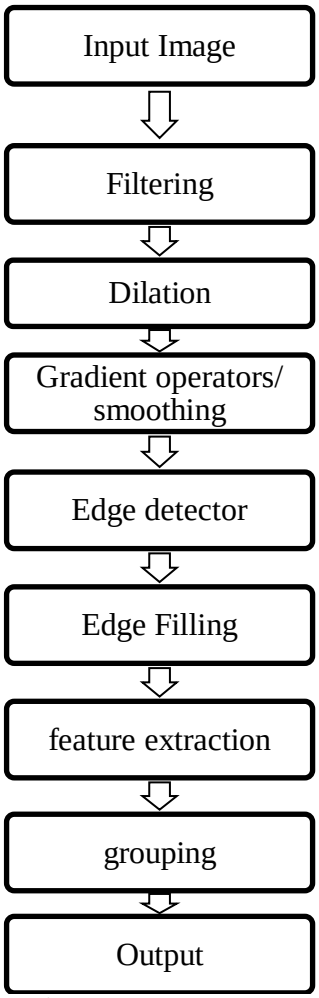

Fig. 2.1 Example of a typical image segmentation algorithm

The choice of filter depends on the type of noise in the image.

A. *Thresholding*

This is the simplest method of image segmentation and is extensively used. Thresholding is used to create binary image based on intensity of the image. This method attempts to find an intensity called threshold. The range function is applied to the intensity values of the image pixels. This technique is to partition an input image into two or more sub-images by comparing with the predefined threshold value T [1]. Let I (i, j) be an image,

$$I(i,j) = \begin{cases} 0, & p(i,j) < T \\ 1, & p(i,j) \geq T \end{cases}$$

(1)Where p (i, j) refers to the pixel value at position (i, j). Thresholding can be either locally or globally. Global thresholding partitions the image into two based on the above equation. In local thresholding image is divided into sub images and thresholding properties are derived from the local properties of its pixels [3]. The disadvantages of this method are difficulty in finding the threshold value, they often lack the sensitivity and specificity needed for accurate classification, thresholding does not take into account the spatial characteristics of the image.

B. Region Growing

Region growing is simple segmentation technique. Regions of an image are connected based on some criteria. These criteria can be intensity information or edges in the image. Region based segmentation is partitioning of an image into similar areas of connected pixels based on some criteria [4]. This technique starts by choosing an arbitrary seed pixel and compare it with neighboring pixels. Neighboring pixels that are similar are grouped with the seed points, which increases cluster size. Its main disadvantage is that manual interaction is needed to obtain the seed point. Initial selection of seed point affects the final result. This method can be sensitive to noise and also sensitive if seed lies on edge. Seeded Region Growing (SRG) is one of the simplest region growing algorithms. In this algorithm,

(i) A number of seed points are selected as initial clusters. Threshold value is assigned.

(ii) Difference between neighboring pixels and these set of points are taken.

$$\delta(x) = |g(x) - mean_{y \epsilon A_{i(x)}}[g(y)]| \quad (2)$$

where g(x) is the gray value of the image pixel.

(iv) If a pixel can be added to more than one seed point clusters the one with minimum δ(x) is chosen.

(v) The mean value of the clusters are recalculated and the boundary.

(vi) At each point, one pixel is added to the existing cluster.

(v) This procedure is repeated until all the pixels are allocated.

Threshold is user-defined parameter and is based on intensity, color, etc. The seed selection process van also be automated, which reduces the improper seed selection. Another problem in SRG is, it is time-consuming, as the computational time is more.

C. Wavelet segmentation

Wavelet technique divides the image into detailed sub signals. A mathematical tool for hierarchically decomposing functions in the frequency domain by preserving the spatial domain is the wavelet [5]. Wavelets represent the image in time-frequency domain. Wavelets are used as basic functions ψ_k (t) in representing other functions f (t).

Wavelets are functions generated from a single basis function by its operations dilations and translations. The Haar transform is commonly applied tool used in the wavelet transform for the image decomposition and feature extraction. The mathematical analysis of Haar transform is used for image compression and image pixels features extraction using decomposition and reconstruction matrices. This method provides perfect image reconstruction. Haar transform is the simplest compression process of this type. The wavelet transform offers great design flexibility. To enable real time processing capability fast implementation of wavelet transforms using a filter-bank framework is used. Multi-resolution representation is one of the most important features of wavelet transform [6]. Wavelets are used in applications such as computer graphics, image compression and digital image processing and feature detection.

D. Clustering

Clustering method is unsupervised method because it does not use the training data. Clustering method train themselves using the available data. Commonly used clustering algorithms are K-means clustering, fuzzy c-means algorithm and

expectation minimization (EM) algorithm. K-means clustering algorithm iteratively calculates mean of each cluster and assigns the pixels according the distance measured. The Expectation Maximization (EM) algorithm extends k-means algorithm. Fuzzy c-means (FCM) algorithm is soft segmentation based on fuzzy set theory. Clustering method requires an initial segmentation. It is sensitive to noise and intensity in-homogeneities.

III. CLUSTERING TYPE

A. Hard Clustering

Hard clustering assumes sharp boundaries between clusters [2]; a pixel belongs to one and only one cluster. A popular and well known hard clustering algorithm is K-means clustering algorithm [7]. K-means algorithm is a clustering technique to partition n pixels into k clusters, where k < n. K-means algorithm Developed by Mac Queen in 1965 and then refined by Hartigan and Wong in 1979. K-means algorithm is a clustering technique, which classify pixels in an image into K number of clusters, where K is a positive integer, according to some similarity feature like grey level intensity of pixels and distance of pixel intensities, from centroid pixel intensity. The main advantages of this algorithm are its simplicity and low computational cost, which allow it to run efficiently on large data sets. The main drawback is that: K the number of clusters must be determined, it does not yield the same result each time the algorithm is executed and the resulting clusters depend on the initial assignments of centroids. They assume a Euclidean space, and they also assume the number of clusters, k, is known in advance. It is, however, possible to deduce k by trial and error. The process is as follow:

(i) Randomly choose number of clusters K.

(ii) Randomly chooses K pixels of different intensities as Centroids.

(iii) Centroids are calculated out by finding mean of the region and place Centroids as much far away from each other as possible.

(iv) Now, compare a pixel to every centroid and assign pixel to closest centroid to form a cluster.

(v) Re-estimate the mean of the cluster as new centroid.

(vi) Repeat step 5 & 6, until Centroids no longer move.

(vii) Image is now separated into K clusters.

The complexity of the k-means algorithm is O (n*K*I*d) where, n is number of points; K is cluster number: I is number of iterations; d is number of attributes.

B. Soft Clustering

In real time applications, one of the most difficult task in image analysis & computer vision is to classify the pixel in an image correctly [8], when there is no crisp boundaries between objects in an image thus in order to address this difficulty , fuzzy clustering techniques are used.

Soft or Fuzzy clustering technique classify pixel values with great extent of accuracy & it is basically suitable for decision oriented applications like tissue classification & tumor detection etc. Fuzzy clustering divides the input pixels into clusters or groups on the basis of some similarity criterion, such that similar pixels belong to same cluster. Similarity criterion used can be distance, connectivity, intensity. The resulting partition improves the understanding of human beings & helps in a more informed decision making. The advantage of fuzzy system is that they are easy to understand, as the membership function partition the data-space properly [8]. Fuzzy clustering algorithms include FCM (fuzzy C means) algorithm, GK (Gustafson-Kessel), GMD (Gaussian mixture decomposition), FCV (Fuzzy C varieties), AFC (Adaptive Fuzzy Clustering) algorithm and etc. The FCM is the most accepted method since it can preserve much more information than other approaches.

Fuzzy C-Means Clustering

Fuzzy c-means algorithm (FCM) is most popular objective function based fuzzy clustering algorithm. It is first developed by Dunn and improved by Bezdek [10]. The objective (cost) function used in FCM presented in [10] is:

$$J_m(U,V,X) = \sum_{i=1}^{c}\sum_{j=1}^{N} u_{ik}^{m} d_{ik}^{2}$$

$$= \sum_{i=1}^{c}\sum_{j=1}^{N} u_{ik}^{m} ||x_j - v_i||^2$$

where, $U \in M_{fc}$ is a fuzzy partition matrix, $V = [V_1, V2, V3... Vc] \in R^n$ is a vector of cluster prototypes (centers), which has to be determined.

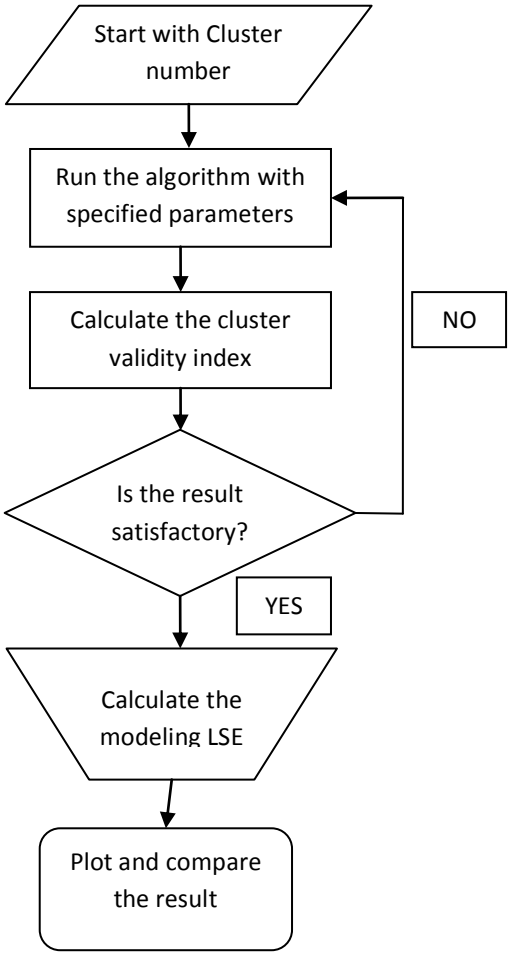

Fig. 3.1 Fuzzy C-Means clustering algorithm Procedure [11]

Fig. 3.1 shows the steps of FCM algorithm. The procedure is as follows:

(i) Initially the number of clusters c ($2 \leq c < n$), m, U are specified by user.

(ii) All the user-defined parameters are defined, run the algorithm.

(iii) The validity index measures the compactness of the output cluster is measured.

Validity index measures like Partition Coefficient, Partition Entropy, etc ate used to measure the fitness of the outcome.

(iv) If the calculated index is satisfactory, then it moves to next step or the parameters are changed and process is repeated.

(v) Least Square Estimation (LSE) is calculated. This error should be minimum.

(vi) Finally the outputs obtained for various inputs are compared and plotted.

The value of the cost function can be seen as a measure of the total variance.Several parameters must be specified to run the FCM algorithm; the number of clusters c, the fuzziness exponent m, the termination tolerance ε, the cluster center and also the fuzzy partition matrix U must be initialized. The number of clusters c is the most important parameter. When clustering real data without any priori information, the number of underlying clusters must be selected carefully. This value ranges from (0, n]. Different values of c results different results. The selection of initial cluster center is directly related to the performance of the algorithm. If cluster center is close to actual center, then the algorithm converges quickly. This also reduces the processing time. The weighting exponent m, on fuzzy membership controls the fuzziness of the clusters. The FCM algorithm stops iterating when the norm of the difference between U_{in} two successive iterations is smaller than the termination parameter.

Disadvantage of FCM [9] is that for noisy images it does not take into account spatial information, which makes it sensitive to noise & other image artifacts. To overcome these drawbacks of FCM, several other algorithms are introduced as modified FCM, GSFCM (Generalized spatial FCM), mean shift based FCM, FLICM (fuzzy logic information C-means clustering algorithm), NFCM (novel FCM), and ISFCM (improved spatial FCM).

Distance Metrics

Distance function is the distance between objects in given space. The performance of many algorithms depends on choosing a good metric over input data. A distance metrics is, d (X, Y) a function or algorithm for calculating a distance between two things, X and Y having following properties [12]:

- It is always positive or zero.
- The distance from a point to itself is zero.
- It obeys inequality property of a triangle.
- Similarity axiom.

Anything that obeys these 3 properties is a distance metric. Euclidean distance measure is one of the commonly used metric, which is used to find the distance between two objects. This metric takes the square root difference of two objects. The least square measure reduces the error rate. There are many other metrics like Chebyshev measure, Chi-Square measure, and so on.

IV. CONCLUSION

In this survey, an overview of different segmentation methods and clustering are studied. Though many techniques are developed, not all types are useful for all types of images. Segmentation segments the image and clusters according to some similarity. Distance metric is a similarity measure and has direct impact on the clusters formed. In this, Fuzzy is powerful unsupervised clustering method which is widely used for robust segmentation of real time images. Traditional FCM and many other algorithms use Euclidean Distance metric. This review provides a platform for the development of the novel techniques in this area as future work.

References

1. Rafael C. Gonzalez, Richard E. Woods, “Digital Image Processing”, Pearson Education, 2002
2. K. Jain, M. N. Murty, and P. J. Flynn, “Data clustering: a review,” ACM Computing Surveys, vol. 31, issue 3, pp. 264-323, Sep. 1999
3. P.K. Sahoo, S. Soltani, and A.K.C.Wong, “A survey of thresholding techniques”,Comput. Vis. Graph. Im. Proc., 41:233–260, 1988
4. J. Freixeenet, ET. Al, “Yet Another Survey on Image Segmentation: Region and Boundary Information Integration”, Springer: 408 – 422, 2002
5. E.A. Ashton, M.J. Berg, K.J. Parker, J. Weisberg, C.W. Chen, and L. Ketonen, “Segmentation and feature extraction techniques, with applicationstoMRIheadstudies”, Mag. Res. Med., 33:670–677, 1995
6. S. Mallat, “A compact multiresolution representation: the wavelet model.” Proc. IEEE Computer Society Workshop on ComputerVision, IEEE ComputerSociety Press, Washington, D.C., p.2-7, 1987

7. K. Dehariya, S. K. Shrivastava, R. C. Jain, "Clustering of Image Data Set Using K-Means and Fuzzy K-Means Algorithms", International conference on CICN, pp. 386- 391, 2010
8. F .Z. Kettaf, D. BI, J. P.,"A Comparison Study of Image Segmentation by Clustering Technique", Proceedings of ICSP, pp. 1280-1282, 1996
9. S. Naz, H. Majeed, H. Irshad, "Image Segmentation using Fuzzy Clustering: A Survey", International Conference on ICET, pp.181-186, 2010
10. J. Bezdek, "Pattern Recognition with Fuzzy Objective Function Algorithms", New York: Plenum, 1981
11. K. M. Bataineh, M. Naji, M. Saqer, "A Comparison Study between Various Fuzzy Clustering Algorithms", Jordan Journal of Mechanical and Industrial Engineering, Vol. 5, Number 4, Aug. 2011
12. O. A. Mohamed Jafar and R. Sivakumar, "A Comparative Study of Hard and Fuzzy Data Clustering Algorithms with Cluster Validity Indices", Elsevier Publications, 2013

AUTHOR(S) PROFILE

P. Sravani, received B.E degree in Computer Science and Engineering from Velammal Institute of Technology, Anna University, Chennai, Tamilnadu, India and doing M.E Computer Science and Engineering in Velammal Engineering College, Anna University, Chennai, Tamilnadu, India 2012-2014.

S. Deepa, completed by Bachelor of computer science and engineering in Sri Venkateswara college of engineering and technology and Master of computer science and engineering in Jaya Engineering college. I have 7.5 years of experience in teaching and I am currently working in Velammal engineering college for the past 2.5 years.

