A survey of K means Clustering with modified gradient magnitude region growing technique for lesion segmentation

Navneet Kaur
Department of Computer Science Engineering
Chandigarh Engineering College, Landran, Punjab, India

Gagan Jindal
Department of Computer Science Engineering
Chandigarh Engineering College, Landran, Punjab, India

Abstract- Image segmentation is very useful in lesion segmentation which is needed for monitoring and quantifying Lesion. Automated lesion segmentation in CT/MRI poses many challenges with regard to characteristics of an image. Many techniques are involved to detect the lesion in the medical images such watershed, wavelet transform etc. In this paper, we are defining the modified k means cluster to segment the lesion and combining the technique with Modified gradient magnitude region growing technique. Integrated techniques are analyzed as better for lesion segmentation.

Keywords- segmentation, K mean clustering, region growing.

I. INRODUCTION

In computer vision, Segmentation refers to the process of partitioning a digital image into multiple regions (sets of pixels). The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze. Image segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in images. Partitioning of an image into several constituent components is called segmentation. Segmentation is an important part of practically any automated image recognition system, because it is at this moment that one extracts the interesting objects, for further processing such as description or recognition. Segmentation of an image is in practice the classification of each image pixel to one of the image parts. Segmentation subdivides an image into its constituent regions or objects. The level to which the segmentation is carried depends on the problem being solved. That is, segmentation should stop when the objects of interest in an application have been isolated.
For example in the automated inspection of electronic assemblies the interest lies in analyzing images of products with objective of determining the presence or absence of specific anomalies, such as missing components or broken connection paths

II. IMAGE SEGMENTATION TECHNIQUES

The most important step in Image Segmentation is to generate a compact description of an Image. In order to do that following two approaches could be used a) Contour Segmentation (edge detection and contour tracing), b) Region Segmentation (Grouping of connected pixels in to regions of uniform properties)We propose, in this paper, Region Segmentation Algorithm based on a edge preserving smoothing filter, the symmetric nearest neighbor mean and a fast anisotropic diffusion. There are two techniques of Image Segmentation –

2.1. Edge Preserving Smoothing Filter [4]

Edge preserving smoothing filters cog nominate edge preserving noise-cleaning filters. Systematic Nearest Neighbor (SNN) filters, in many of the correlative studies of edge-preserving noise smoothing techniques, is considered to give the best results in smoothing and preserving edges. Therefore, SNN is most widely used as an Image Improvement Technique. Westman et al has used a SNN filter to preprocess color images before segmenting them.

The principle on which the SNN filter works is stated below The SNN filter makes use of both the Spatial and Gray value information in the neighborhood of pixel to be processed. In a squre window, half the number of Pixels is selected by choosing one pixel nearest in Gray value to the center pixel from each pair of pixels located symmetrically opposite the center. Only the selected pixels are used to compute a new value for the center pixel.

2.2. Anisotropic diffusion [4]

In Image processing, anisotropic diffusion is also called Perona–Malik diffusion. It is used to remove the Image Noise without losing the significant parts of the image content and other details that are important for the Image interpretation. This technique enhances the contrast by using a modified heat diffusion equation. This technique is a discontinuity preserving smoothing approach and is closely related to the adaptive smoothing proposed by Chen et al. The principle is that a pixel should become weighted average of its neighbors. The weight resembles the continuity measures of these pixels. Repetitive Implementation of anisotropic diffusion is Adaptive smoothing in which the unwanted edges will disappear along with repetition. However, this scheme is considered to be slow, and to avoid this slowness we use Toboggan Contrast enhancement as proposed by Fairfield.

2.3. Application of Edge Preserving Smoothing and Anisotropic diffusion [4]

SNN filters are considered good for Clearing the noises and preserving the edge, but they cannot make potential regions Uniform. Whereas, anisotropic diffusion can make potential regions uniform but it is impressionable to noises. So, we propose to make a collective use of both the techniques to enhance an image before segmenting it. To make Fairfield's diffusion algorithm less impressionable to noises, Canny-Deriche detector can be used by us.

III. CLUSTERING

Clustering is a collection of objects which are "similar" between them and are "dissimilar" to the objects belonging to other clusters. Therefore, it deals with finding a structure in a collection of unlabeled data. The goal of clustering is to determine the intrinsic grouping in a set of unlabeled data. Clustering Techniques can be divided into two categories-

3.1. Hierarchical Clustering [12]

Hierarchical clustering techniques are based on the use of a proximity matrix indicating the similarity between every pair of data points to be clustered. The end result is a tree of clusters called a dendrogram representing the nested grouping of patterns and similarity levels at which groupings change [A.K. Jain and P.J.Flynn, 1999]. It proceeds successively by either merging smaller clusters into larger ones (agglomerative, bottom-up), or by splitting larger clusters (divisive, top-down). The clustering methods differ in regards to the rules by which two small clusters are merged or a large cluster is split. Some of the Hierarchical [R.Bhowmik, 2003] algorithms include COBWEB, CURE and CHAMEL.

3.2. Partitional Clustering

Partitional algorithms are categorized into Partitioning Relocation Algorithms and Density-based Partitioning. Partitional clustering techniques such as K-means clustering have an advantage over the hierarchical clustering techniques, where a partition of the data points which optimizes some criterion functions. In hierarchical clustering once a "thing" is as- signed to a particular cluster, it cannot be altered. Therefore if a "thing" is incorrectly assigned to a particular cluster at an early stage there is no way to correct the error. There is however a disadvantage of the partitional clustering techniques and this is the determination of the number of clusters K.

IV. K-MEANS CLUSTERING ALGORITHM

There are two existing basic versions of k-means clustering, a non-adaptive version introduced by Lloyd and an adaptive version [14] introduced by MacQueen. The most commonly used k-means clustering is the adaptive k-means clustering based on the Euclidean distance. Adaptive k-means clustering can be considered as a special case of the gradient descent algorithm where only the winning cluster is adjusted at each learning step. This paper concentrates only on adaptive k-means clustering as the algorithm can be used for on-line training of RBF network. Adaptive k-means clustering tries to minimize the cost function in equation (1) by searching for the centre c_j on-line as the data are presented. As the data sample is presented, the Euclidean distances between the data sample and all the centers are calculated and the nearest centre is updated according to:

$$\Delta c_z(t) = \eta(t)\left[v(t) - c_z(t-1)\right] \quad (1)$$

Where z indicates the nearest centre to the data v(t). Notice that, the centres and the data are written in terms of time t where $c_z(t-1)$ represents the centre location at the previous clustering step. The adaptation rate, $\eta(t)$, can be selected in a number of ways. MacQueen [13] set $\eta(t) = 1/n_z(t)$, where $n_z(t)$ is the number of data samples that have been assigned to the centre up to the time t. Darken and Moody used a constant adaptation rate and a square root method $\left(\eta(t) = 1/\sqrt{n_z(t)}\right)$. Another method called search-then-converge has been introduced by Darken and Moody. According to this method $\eta(t)$ is updated using:

$$\eta(t) = \eta(0)\frac{1 + \frac{\alpha}{\eta(0)}\frac{t}{\tau}}{1 + \frac{\alpha}{\eta(0)}\frac{t}{\tau} + \tau\frac{t^2}{\tau^2}} \quad (2)$$

The basic idea is to keep $\eta(t)$ approximately constant at times small compared to τ and decrease $\eta(t)$ at the rate of α/t as time t becomes large compared to τ. This method yields optimally fast asymptotic convergence if $\alpha > 1/2\beta$, where β is the smallest eigenvalue of the Hessian matrix of the cost function defined in equation (1) . Chen et al. used an adaptation rate that is updated at each step according to:

$$\eta(t) = \eta(t-1)\Big/\sqrt{1 + int\left(t / n_c\right)} \quad (3)$$

where int(.) denotes the integer part of the argument and n_c is the number of centres.

The problem of assigning the adaptation rate to adaptive k-means clustering is very similar to the problem of assigning the learning rate to the back propagation algorithm. Both algorithms are based on the gradient descent method except that in back propagation all the parameters are updated at the same time. Therefore, all the methods that are used to choose the learning rate for the back propagation algorithm may also be applied for the adaptation rate in k-means clustering. The usual approach is to update $\eta(t)$ according to the variation of the cost function during the clustering process, such as:

$$\Delta\eta(t) = \begin{cases} +a & \text{if } \Delta E < 0 \textit{ consistently} \\ -b\eta(t-1) & \text{if } \Delta E > 0 \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

Where ΔE is the change in the cost function and, a and b are parameter constants. The term consistently in equation (5) means a constantly decrease of E for the last few clustering steps. Cater suggested that this kind of adaptive scheme can be made more effective if each parameter (the centre in this case) has a different adaptation rate.

Another method to improve the back propagation algorithm that may be adapted to k-means clustering is the method of momentum that has been introduced by Plaut et al. For k-means clustering, a momentum term can be included as follows:

$$\Delta c_k(t) = \eta(t)\left[v(t) - c_k(t-1)\right] + \alpha \Delta c_k(t-1) \quad (5)$$

The momentum constant α is between 0 and 1, and is often chosen to be close to 1. In this case, $\eta(t)$ can be a constant or adaptive. Other updating methods such as Newton method, stochastic method and conjugate gradient method may also be adapted to improve the k-means clustering algorithm at the expense of computational time.

In the current study, two updating methods are proposed as alternatives to update $\eta(t)$. The first method update $\eta(t)$ according to:

$$\eta(t) = \eta(t-1) \big/ e^{\left(1/r\right)} \quad (6)$$

where $r = n_c + t$ for off-line clustering and $r = \sqrt{n_c + t}$ for on-line clustering. The updating method uses different r for on-line and off-line clustering because in on-line clustering problems, $\eta(t)$ should be decreased rapidly so that the weights of the network can converge properly. This will not be a problem with the off-line clustering since the weights are estimated after the centers are located. The second proposed updating method updates $\eta(t)$ according to:

$$\eta(t) = \eta(0)\left[e^{-p\left(t^2/n_c^3\right)} + b e^{-\left[n_c \eta(t-1)\right)}\right] \quad (7)$$

Where p is a constant, 0 < p £ 1 and $b = 1/\left(n_c + n_z(t)\right)$. n_c and $n_z(t)$ are the number of centers and the number of data assigned to centre c_z up to time t respectively. This method involves two terms in the bracket on the right hand side. At the beginning, $\eta(t)$ will be dominated by the first term but as time t becomes large, $\eta(t)$ will converge to the value of b in the second term. The constant term p will determine how long $\eta(t)$ will be dominated by the first term.

In the present study, methods of updating $\eta(t)$ are selected such that the computational time will be minimized, which is beneficial for on-line clustering problems. For this reason the two proposed updating methods (described by equations (6) and (7)) together with the three methods that have been used by Chen et al. and Darken and Moody are studied:

1. $\eta(t) = 1/n_z(t)$, the MacQueen method.
2. $\eta(t) = 1/\sqrt{n_z(t)}$, the square root method
3. $\eta(t) = a/\sqrt{1 + int(t / n_c)}$, Chen�s method, where $a = \eta(0)$ for off-line clustering and $a = \eta(t-1)$ for on-line clustering.

4 $\eta(t)=\eta(t-1)/e^{[1/r]}$, where r = n_c + t for off-line clustering and $r=\sqrt{n_c+t}$ for on-line clustering, the first proposed method.

5 $\eta(t)=\eta(0)\left[e^{-p[t^2/n_c^3]}+be^{-[n_c\eta(t-1)]}\right]$, p is a constant, 0 < p £ 1 and $b=1/(n_c+n_z(t))$

the second proposed method.

Where n_c, $n_z(t)$ are the number of centers and the number of data assigned to centre c_z up to time t respectively. Notice that all these updating methods update the centers based on equation (1).

V. REGION GROWING SEGMENTATION

The region based segmentation [8] is partitioning of an image into similar/homogenous areas of connected pixels through the application of homogeneity/similarity criteria among candidate sets of pixels. Each of the pixels in a region is similar with respect to some characteristics or computed property such as color, intensity and/or texture. Failure to adjust the homogeneity/similarity criteria accordingly will produce undesirable results. The following are some of them:

- The segmented region might be smaller or larger than the actual
- Over or under-segmentation of the image (arising of pseudo objects or missing objects)
- Fragmentation

Region growing is a simple region-based image segmentation method. It is also classified as a pixel-based image segmentation method since it involves the selection of initial seed points. This approach to segmentation examines neighboring pixels of initial “seed points” and determines whether the pixel neighbors should be added to the region. The process is iterated on, in the same manner as general data clustering algorithms. The fundamental drawback of histogram-based region detection is that histograms provide no spatial information (only the distribution of gray levels).

The main goal of segmentation is to partition an image into regions. Some segmentation methods such as "Thresholding" achieve this goal by looking for the boundaries between regions based on discontinuities in gray levels or color properties. Region-based segmentation is a technique for determining the region directly.

The first region-growing method was the seeded region growing method. This method takes a set of seeds as input along with the image. The seeds mark each of the objects to be segmented. The regions are iteratively grown by comparing all unallocated neighboring pixels to the regions. The difference between a pixel's intensity value and the region's mean, δ, is used as a measure of similarity. The pixel with the smallest difference measured this way is allocated to the respective region. This process continues until all pixels are allocated to a region.

VI. CONCLUSIONS

In the paper, we have presented the different techniques that automatically segment and locate the lesion in body organs, especially brain and lung CT/MRI images. These techniques overcome the accuracy and sensitivity limitations of the current solutions. Recently attention is being paid to the semi-automatic segmentation methods on Lesion measurements in order to avoid the observer variability and therefore to increase the accuracy. The region-based labeling drastically reduces the segmentation time.

REFERENCES

[1] Huawu Deng, David A. Clausi,” Unsupervised image segmentation using a simpleMRFmodel with a newimplementation scheme”, *April 2004 Department of Systems Design Engineering, University of Waterloo*, 200 University Avenue West, Waterloo, Ont., Canada N2L 3G.

[2] Wenbing Tao, Hai Jin and Yimin Zhang,” Color Image Segmentation Based on Mean Shift and Normalized Cuts”, *OCTOBER 2007 IEEE Transactions on systems,man and cybernetics*.

[3] T. G´eraud, P.-Y. Strub, J. Darbon,”Color Image Segmentation based on automatic morphological clustering”,*October 2001 IEEE International Conference on Image Processing*.

[4] Tuan DAND,” An image segmentation technique based on Edge-preserving smoothening filter and anisotropic Diffusion”, *http://www.multimania.com/dangt*.

[5] Ye Hou, Bao-long Guo, Jeng-Shyang Pan,Chin-Shiuh Shieh,” Image Segmentation Based on GC-CV”, *2009 Ninth International Conference on Hybrid Intelligent Systems*.

[6] Yi-hua Lan, Cun-hua , Yong Zhang, Xue-feng Zhao,” A Novel Image Segmentation Method Based on Random Walk” ,*2009 Second Asia-Pacific Conference on Computational Intelligence and Industrial Applications.*

[7] Clifton Forlines, Ravin Balakrishnan,”Improving Visual Search with Image Segmentation”, *April 7th, 2009 ~ Boston, MA, USA.*

[8] Shilpa Kamdi1, R.K.Krishna,” Image Segmentation and Region Growing Algorithm”, *International Journal of Computer Technology and Electronics Engineering (IJCTEE) Volume 2, Issue 1.*

[9] Laura Bradbury and Justin W.L. Wan,” A Spectral K-Means Approach to Bright-field Cell Image Segmentation”, *32nd Annual International Conference of the IEEE EMBS Buenos Aires, Argentina, August 31 - September 4, 2010.*

[10] Tse-Theyi Chen, Hsiao-Hang Su, Yi-Ling Chen, and Shao-Yi Chien,” Edge-adaptive image segmentation based on seem processing and k-means clustering”,*SEPTEMBER 2010 IEEE 17th International Conference on Image Processing.*

[11] R. V. Patil, K. C. Jondhale,” Edge based technique to estimate Number of Clusters in k-means Color Image Segmentation”,*2010 IEEE Transactions on image processing.*

[12] S. Nandagopalan, Dr. B. S. Adiga, C. Dhanalakshmi, N. Deepak,” Automatic Segmentation and Ventricular Border Detection of 2D Echocardiographic Images Combining K-Means Clustering and Active Contour Model”, *2010 IEEE Transactions on image processing*

[13] Jos´e Antonio Mart´ın H, Javier Montero , Javier Y´a˜nez and Daniel G´omez,” A Divisive Hierarchical k-means based Algorithm for Image Segmentation*”, IEEE Transactions on image processing 2010.*

[14] Siti Noraini Sulaiman and Nor Ashidi Mat Isa,” Adaptive Fuzzy-K-means Clustering Algorithm for Image Segmentation”,*IEEE 2010 Adaptive Fuzzy-K-means Clustering Algorithm for Image Segmentation.*

[15] Jun Tang,” A Color Image Segmentation algorithm Based on Region Growing”,*IEEE 2010 School of Electronic Engineering Xi'an Shiyou University Xi'an China.*

[16] Juyong Zhang Jianmin Zheng Jianfei Cai,” **A** Diffusion Approach to Seeded Image Segmentation”, *IEEE 2010 Nanyang Technological University, Singapore.*

[17] Fasahat Ullah Siddiqui and Nor Ashidi Mat Isa,” Enhanced Moving K-Means (EMKM) Algorithm for Image Segmentation”,*JUNE 2011 IEEE Transaction on image processing.*

[18] Zhigen Fei, Junjie Guo,” A New Hybrid Image Segmentation Method for Fingerprint Identification”,*2011 Mechanical & Electrical Engineering Department Zhengzhou University of Light Industry, Zhengzhou 45002, China.*

[19] Lei He, L. Rodney Long, Sameer Antani, George R. Thoma,” Multiphase Level Set Model with Local K-Means Energy for Histology Image Segmentation”, *2011 First IEEE International Conference on Healthcare Informatics, Imaging and Systems Biology.*

[20] Vijayakalyani Natarajan,” Indexing and retrieval of images using wavelets”,*2011 First International Conference on Informatics and Computational Intelligence.*

[21] Shaohua Zhu, Zhaohua Wu,” Study on Solder Joint Image Segmentation Techniques Based on Matlab”,*2011 International Conference on Electronic Packaging Technology & High Density Packaging.*

[22] H C Vijay Lakshmi and S. PatilKulakarni,” Segmentation Algorithm for Multiple Face Detection in Color Images with Skin Tone Regions using Color Spaces and Edge Detection Techniques”, *International Journal of Computer Theory and Engineering, Vol. 2, No. 4, August, 2011.*

[23] A.K. Jain, M. and P.J.Flynn [1999]. Data clustering: A review, ACM Computing Survey vol.31(3): 264–323.