A Survey of Image Segmentation Algorithms Based on Expectation-Maximization

[1]R.Ravindraiah, [2]K.Tejaswini

[1]*Assistant Professor, ECE Department, Madanapalle Institute of Technology and Science/JNTU University, India,*
[2]*Student Electronics and Communication Engineering, Madanapalle Institute of Technology and Science/JNTU University, India,*

ABSTRACT: *Medical image segmentation plays an important role in one of the most challenging fields of engineering. Imaging modality provides detailed information about anatomy. It is also helpful in the finding of the disease and its progressive treatment. More research and work on it has enhanced more effectiveness as far as the subject is concerned. Different methods are used for medical image segmentation such as Clustering methods, Thresholding method, Classifier, Region Growing, Deformable Model, Markov Random Model etc. The main purpose of this survey is to provide a comprehensive reference source for the researchers involved in Expectation-Maximization based medical image processing. There are different types of Expectation-Maximization algorithms for medical image. Their advantages and disadvantages are discussed.*
Keywords: Image segmentation, Medical Image Processing, Expectation-Maximization

I. Introduction

Segmentation is the process of separating a digital image into different regions which have similar properties such as gray level, colour, texture, brightness etc. So that the image can be more understandable and helpful to analyzing. On the basis of pixel intensity we can differentiate the boundaries of different objects.

Segmentation identifies separate object within an image and also find boundary between different regions. Segmentation can be classified into two types: local segmentation and global segmentation. Local segmentation is small windows on a whole image and deal with segmenting sub image. Global segmentation deals with segmenting whole image. Global segmentation mostly deals with relatively large no of pixel. But local segmentation deal with lower no of pixel as compare to global segmentation. Image segmentation is one of the classical problems in image processing and computer vision. Using of Image segmentation we can able to understand the fundamental of digital image processing. Image segmentation is used to enhancement of image and also useful to different medical application. Image segmentation can also use for analysis of the image and further pre-processing of the image. After a segmentation process each phase of image treated differently. Now we are going through about medical images like Ultrasound Images (US) which is widely used today.

Medical images play a vital role in assisting health care providers to access patients for diagnosis and treatment. Studying medical images depends mainly on the visual interpretation of the radiologists. However, this consumes time and usually subjective, depending on the experience of the radiologist. Consequently the use of computer-aided systems becomes very necessary to overcome these limitations. Artificial Intelligence methods such as digital image processing when combined with others like machine learning, fuzzy logic and pattern recognition are so valuable in Image techniques. The computerization of medical image segmentation plays an important role in medical imaging applications. It has found wide application in different areas such as diagnosis, localization of pathology, study of anatomical structure, treatment planning, and computer-integrated surgery. However, the variability and the complexity of the anatomical structures in the human body have resulted in medical image segmentation remaining a hard problem.

Atherosclerosis is a specific form of arteriosclerosis (thickening & hardening of arterial walls) affecting primarily the intima of large and medium-sized muscular arteries and is characterized by the presence of fibrofatty plaques. The term atherosclerosis is derived from "athero" referring to the soft lipid-rich material in the centre of atheroma, and "sclerosis" referring to connective tissue in the plaques. Atherosclerosis is often called as arteriosclerosis. Arteriosclerosis can occur in several forms, including atherosclerosis. Atherosclerosis is a heart disease, and is a type of arteriosclerosis or hardening of the arteries. Coronary artery mainly consists of three layers. The inner layer is intima, middle layer is media and the outer one is adventitia. An artery is made up of several layers: an inner lining called the endothelium, an elastic membrane that allows the artery to expand and contract, a layer of smooth muscle, and a layer of connective tissue. Arteriosclerosis is a broad term that includes a hardening of the inner and middle layers of the artery. It can be caused by normal aging, high blood pressure, mental stress, lack of inactivity and by diseases such as diabetes. Atherosclerosis is a type of

arteriosclerosis that affects only the inner lining of an artery. It is characterized by plaque deposits that block the flow of blood. Plaque is made of fatty substances, cholesterol, waste products from the cells, calcium, and fibrin, a stringy material that helps clot blood. The plaque formation process stimulates the cells of the artery wall to produce substances that accumulate in the inner layer. Fat builds up within these cells and around them, and they form connective tissue and calcium. The inner layer of the artery wall thickens, the artery's diameter is reduced, and blood flow and oxygen delivery are decreased. Plaques can rupture, causing the sudden formation of a blood clot (thrombosis). Atherosclerosis can cause a heart attack if it completely blocks the blood flow in the heart (coronary) arteries. It can cause a stroke if it completely blocks the brain (carotid) arteries.

Atherosclerosis can also occur in the arteries of the neck, kidneys, thighs, and arms, causing kidney failure or gangrene and amputation.
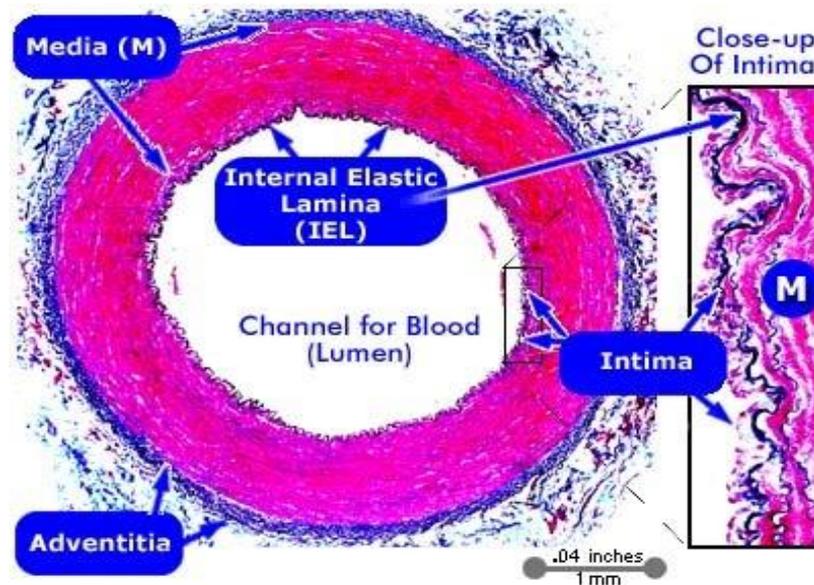

Figure: Cross section of coronary artery

Different types of Expectation-Maximization Algorithms
1. Expectation Maximization
2. Gaussian Mixer Model Expectation maximization
3. Diffused Expectation Maximization
4. Adaptive Distance Expectation Maximization
5. Multiresolution Diffused Expectation Maximization
6. Neighborhood Expectation Maximization)
7. Variational Expectation Maximization

II.   Expectation-Maximization (EM)

The EM (expectation-maximization) algorithm is an iterative approach to compute maximum-likelihood estimates when the observations are incomplete. Expectation Maximization (EM) is one of the most common algorithms used for density estimation of data points in an unsupervised setting. The expectation maximization (EM) algorithm introduced by Dempster (1977) for maximization likelihood functions with missing data. This algorithm is a popular tool for simplifying difficult maximum likelihood problems. It has two steps; in E-step we compute the expectation and in M-step the maximization of the last step is done and iteration EM-steps continue until convergence occur. The EM algorithm is used for the research of the parameter achieving the maximum likelihood. The criteria of stop of the algorithm, is either a maximum number of iterations to limit the time of calculation, either a lower mistake. It is put easily in application because it leans on the calculation of the complete data.

1. Initialize

$$\theta^{(0)} = (p_1^{(0)}, \ldots, p_k^{(0)}, m_1^{(0)}, \ldots, m_k^{(0)}, D_1^{(0)}, \ldots, D_k^{(0)}) \qquad (1)$$

2. (E-step) Compute

$$P^{(r+1)}(i|x_s) = \frac{p_i^{(r)} N(x_s | m_i^{(r)}, D_i^{(r)})}{\sum_{i=1}^{k} p_i^{(r)} N(x_s | m_i^{(r)}, D_i^{(r)})} \tag{2}$$

3. (M-step) Compute

$$m_i^{\wedge(r+1)} = \frac{\sum_{s=1}^{n} P^{(r+1)}(i|x_s) x_s}{\sum_{s=1}^{n} p^{(r+1)}(i|x_s)} \tag{3}$$

$$D_i^{\wedge 2(r+1)} = \frac{\sum_{s=1}^{n} P^{(r+1)}(i|x_s)(x_s - m_i^{\wedge(r+1)})^2}{\sum_{s=1}^{n} P^{(r+1)}(i|x_s)} \tag{4}$$

$$p_i^{\wedge(r+1)} = \frac{1}{n} \sum_{s=1}^{n} p^{(r+1)}(i|x_s) \tag{5}$$

4. Iterate steps 2 and 3 until convergence.
The drawback of EM is, it does not directly incorporate spatial modeling and can therefore be sensitive to noise and intensity inhomogeneities.

2.1 Gaussian Mixer Model Expectation Maximization (GMM-EM)
The application of standard GMM (using a grey level as a single feature of pixel) yields a good segmentation of pixels inside regions and pixels of contours. However, it includes a bad clustering of noisy pixels of white region. This clustering drawback is essentially due to the only utilization of the intrinsic feature of pixel to be classified (gray level) without take into account the information relative to spatial position in the image. This information turns out to be important in the segmentation context. Image is a matrix within which each element is a pixel. The value of the pixel is a number that shows intensity or color of the image. Let X be a random variable that takes these values. For a probability model determination, we can suppose to have mixture of Gaussian distribution as the following form:

$$f(x) = \sum_{i=1}^{k} p_i N(x | \mu_i, \sigma_i^2) \tag{6}$$

Where k is the number of components or regions and $p_i > 0$ are weights such that $\sum_{i=1}^{k} p_i = 1$,

$$N(\mu_i, \sigma_i^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(x-\mu_i)^2}{2\sigma^2}\right) \tag{7}$$

Where μ_i, σ_i^2 are mean and standard deviation of class i. For a given image X, the lattice data are the values of pixels. However, the parameters are $= (p_1, \dots., p_k, \mu_1, \dots, \mu_k, \sigma_1^2, \dots, \sigma_k^2)$.
The application of the GMM segmentation algorithm can generate a good classification of the internal pixels to the regions as well as the noisy pixels, but produces a deterioration of the contours between the two regions. Of this study the advantage and the inconvenient of the use of the grey level and the spatial feature for the noise and the contours classification.

2.2 Diffused Expectation Maximization (DEM)
The DEM algorithm is different from previously proposed methods. Several methods have tried to incorporate within the EM algorithm a prior term in order to maximise a log posterior probability instead of log-likelihood, thus leading to quite complex EM steps [1]. Alternatively, Haker et al. [2] compute an initial posterior probability map, through some kind of classification (e.g., clustering), followed by anisotropic diffusion on such map in order to account for spatial constraints among sites; clearly, in this way final results strictly depend upon the goodness of the initial labelling. Here, we follow a different approach: we operate on the maximisation of the log-likelihood function, and spatial context is implicitly accounted for via label diffusion along maximisation. As a result we obtain a quite simple but effective segmentation algorithm, which can be easily interpreted in terms of a competition/cooperation scheme on the *k* label probability planes: the E and M steps implement a competition among the different labels at site *i*, while the D-step can be considered as a cooperation step among sites on the same plane. Its flexibility makes it suitable for any type of application.
The DEM algorithm generally uses to measure the similarity between an individual and a class given by its center, a distance that grants the same importance to the different attributes took in account in the process of classification. To introduce the adaptive effect as for the selection of the attributes, we propose to use a dynamic distance and weighted derivative of the Euclidian distance. This new distance is given by the equation:

$$D(xj, vj) = (1 - pj)(xj^{NG} - vj^{NG})^2 + pj(xj^{spatial} - vj^{spatial})^2 \qquad (8)$$

D is a bidimensional distance based on the two NG and Spatial attributes. In this equation, the weight pj permits to control the importance of every attribute for the classification of the pixel xj. So if the pj is thigh we privileges the spatial attribute otherwise one privileges the gray level. The term pj must be calculated for every pixel to classify according to its spatial configuration in the image. , we can deduct that the weight p_j must be maximized (tender verse 1) when the pixel to classify is a PR or a PB, because the decision of its adherence to the different classes must be taken only on the basis of the spatial attribute. However this p_j must be minimized (tender verse 0) in case of a PC or a NNP because the gray level in these cases is going to constitute a good criteria of classification. Advantage of the DEM is the regions to be segmented usually resides at different scales. The drawback of DEM is for e.g. points in almost uniform regions have longer interaction range than those in weak texture, and, moreover such range depends on the scale on which the analysis is performed [3]. Then the problem arises on which is the optimal scale at which segmentation must be carried out: a low resolution might result in a too coarse segmentation, with different regions merged together, whereas a high resolution, on the contrary could result in many fragmented images and involves a heavy computational load.

2.3 Adaptive Distance Expectation Maximization (ADEM)
The complementarity of the level of gray and the average of the gray levels with regard to the classification can let consider an use joined of these two describers in the image segmentation with the help of the EM method. In this section

We present a new version of the EM baptized Adaptive Distance EM (ADEM) that enrolls in this direction. This version tried to take advantage of the advantages of the two aforesaid features while avoiding their inconveniences and this while using an adaptive manner one or the other according to the spatial configuration of the pixel to classify. In the setting of our work, we distinguished four possible spatial configurations for the pixels asking for each a specific choice of the criteria of classification. These configurations are the following:
• PR : Pixel which belongs to a region
• PC : Pixel which belongs to a contour,
• NP : Noisy pixel,
• NNP : Neighbour of Noisy Pixel.

Formally, the different spatial configurations are characterized by the two statistical descriptors presented as follows:

The standard deviation (σ) characterizes the dynamics of distribution around the pixel to be classified. This attribute is calculated as follows:

$$\sigma(xj) = \sqrt{\frac{1}{N}\sum_{k=1}^{N}\left(xk - \mu(xj)\right)^2} \qquad (9)$$

The NCN: represents the closest neighbours number in term of grey level regarding the considered pixel. It's defined as follows:

$$\mathrm{NCN}(x_j) = \mathrm{card}\{x_p \in neighborhood(x_j)/\,|\,x_p - x_j\,|\,<S\} \qquad (10)$$

S designates a threshold which is generally chosen in year empiric manner. From thesis two descriptors we can characterize the different spatial configurations possible of the pixels. In case of a PR the standard deviation is generally Low, it is null for the constant regions. However the becomes high for gthe PC, NP and NNP. The distinction between these three configurations is made by using the NCN feature, which is generally low for a NP, moderate for a PC and high for a NNP. The classification of a PR or a PB must privilege the spatial attribute because the decision must be taken on the basis of the information of its neighborhood. On the other hand the classification of a PC or a NCN must privilege the level of gray of the pixel (NG) respectively to preserve the contours and to avoid the influence of the noise. The choice of the criteria of classification as well as the characterizations of the spatial configurations are summarized .The segmentation consists in delimiting the three cerebral structures: gray matter (MG), white matter (WM) and the cerebrospinal fluid (CSF).

Table: Badly classified pixels number contours and inside regions of MRI cerebral image

		EM	DEM	ADEM
Region	CSF	200	10	4
	GM	500	60	30
	WM	15	5	0
Contour	CSF	500	450	100
	GM	1120	1100	190
	WM	450	70	15

2.4 Multiresolution Diffused Expectation Maximization (MDEM)

MDEM is a novel scheme that results in a simple but effective segmentation algorithm for color images that: 1) retains the appealing characteristics of a feature clustering based approach; 2) takes into account spatial constraints while avoiding complex schemes such as MRFs; 3) it operates within a multiresolution framework, in order to reliably define regions of interest and efficiently perform required computations.

The MDEM algorithm is different from previously proposed methods. For instance, different approaches have tried to incorporate within the EM algorithm a prior term in order to maximize a log-posterior probability instead of log-likelihood, thus leading to quite complex EM steps [3],[4]. On the other hand, Haker et al. [5], have suggested to compute an initial posterior probability map, through some kind of preliminary classification (e.g., clustering), followed by anisotropic diffusion performed on such initial map in order to diffuse spatial constraints among probability sites; clearly, in this way final results strictly depend upon the goodness of the initial labeling. Here, we follow a different approach: we operate on the maximization of the log-likelihood function, and spatial context is implicitly accounted for along maximization via diffusion. Further, in order to take advantage of the structure of the image as represented at different scales, this methods has been carried out in a multiscale framework, thus yielding an accurate segmentation, without unduly increasing the computational load.

Here it as been assumed K fixed, in that we are not concerned with the problem of model selection. In general, this problem could be tackled by resorting to BIC or Akaike's information criteria [6]. However, this may not be necessary for biomedical images, since, depending on the application, the value of K is often assumed to be provided by prior knowledge of the anatomy being considered.

As a result we obtain a simple iterative segmentation algorithm which can be easily interpreted in terms of a multiresolution competition/cooperation scheme: at each resolution level of the pyramid, the E and M steps can be seen as an individual site competition between the k different label planes, while the interleaved diffusion step can be considered as a cooperation step within sites on the same plane with respect to the computational load, the whole algorithm is slightly slower than the EM procedure. Currently, it takes 30 seconds for a 256 x 256 2-D image, using an Intel P*IV* 3*.*4 GHz processor, equipped with 2 GHz RAM, under Windows XP operating system. Finally, simulations show that the method performs quite well on a variety of medical images either with respect to more standard methods or also techniques specifically designed for an application (for instance, our skin lesion segmentation results can be compared with examples provided by Xu et al. [7] and it is flexible enough to be used in a wide range of applications.

2.5 Neighborhood Expectation Maximization (NEM)

In order to incorporate the spatial dependence into the objects, a modified version of the conventional expectation maximization (EM) [8] algorithm has been proposed in [9]. In this approach, the maximum likelihood criterion is penalized by a term that quantifies the degree of spatial contiguity of the pixels supporting the respective components of the probability density function (pdf) model. The spatial structure of a given data set is defined by using matrix $W=\left(w_{jk}\right)$:

$$\left(w_{jk}\right) = \begin{cases} 1 \text{ if } x_j \text{ and} x_k \text{ are neighbours and } j \neq k, \\ \quad\quad 0 \text{ otherwise.} \end{cases} \tag{11}$$

The following term is then used for regularizing the maximum likelihood criterion

$$G(c) = \frac{1}{2}\sum_{j=1}^{n}\sum_{i=1}^{n}\sum_{k=1}^{c} c_{ij}\, c_{ik}\, w_{jk} \tag{12}$$

Where c is the number of classes and c_{ij} is the probability that x_j belongs to class i.This term characterizes the homogeneity level of the partition. The more the classes contain adjacent elements, the greater this term is. The new criterion of the NEM algorithm is obtained by optimizing the weighted sum of two terms

$$U(c, \varphi) = D(c, \varphi) + \beta G(c) \tag{13}$$

Where $D(c, \varphi)$ is the log-likelihood function of EM algorithm, $\beta > 0$ is a fixed coefficient. Details about NEM can be found in [9]. This algorithm is maximized to get the optimum results just as the same structure as the PersonNameProductIDEM algorithm. Successful EM algorithm. Successful results have been reported for image segmentation using this algorithm.

2.6 Variational Expectation-Maximization (VEM)

In this paper we use a deterministic Bayesian approach such as the VEM algorithm. The number of mixture components is decided by using the Bayesian Information Criterion (BIC), which corresponds to the Minimum Description Length (MDL) [10].Integrating over the entire parameter space would amount to a very heavy computational task, involving multidimensional integration. Variational Bayes algorithm has been used for estimating hyperparameters of mixture models [2, 9]. In our approach we use the initialisation provided by the maximum log-likelihood as described in previous Section. The proposed algorithm is called variational expectation-maximization (VEM) algorithm.

The VEM algorithm is iterative and consists of two steps at each iteration: variational expectation (V-E) and variational maximization (V-M).

In the first step we compute the a posteriori probabilities, given the hidden variable distributions and their hyperparameters.

In the V-M step we find the hyperparameters that maximise the log-likelihood, given the observed data and their a posteriori probabilities.

In the V-E step we calculate the a posteriori probabilities for each data sample $\mathbf{x}_j$, depending on the hyperparameters:

$$\hat{p}(i|x_j) = \exp\left[\frac{-1}{2}\log|S_i| + \frac{1}{2}d\log 2 + \frac{1}{2}\sum_{k=1}^{d}\Psi\left(\frac{v_i+1-k}{2}\right) + \Psi(\lambda_i) - \Psi(\sum_{k=1}^{N}\lambda_k) - \frac{v_i}{2}(X_j - m_i)^T \beta_i s_i^{-1}(X_j - m_i) - d2\beta i\right] \tag{14}$$

In the V-M step we perform an intermediary calculation of the mean parameter as in the EM algorithm, but considering the a posteriori probabilities from (14):

$$v_k = \left\{ X_j \mid k = \operatorname*{argmax}_{i=1}^{N} \hat{P}(i \mid X_j) \right\} \tag{15}$$

Where $\hat{P}(i \mid X_j)$ are the a posteriori probabilities obtained at the convergence of VEM algorithm. An important issue in colour image segmentation is the selection of an appropriate colour space and that of a representative set of colours. Image respresentation in the L*u*v* colour space has been found as appropriate to be modelled by Gaussian mixtures [11, 12].

References

[1]   SANJAY-GOPAL, S. and HEBERT, T.J.: 'Bayesian pixel classification using spatially variant finite mixtures and the Generalized EM algorithm', IEEE Trans. Image Proc., 1998, 7, pp.1014-1028

[2]   HAKER, S., SAPIRO, G. and TANNENBAUM, A.: 'Knowledge-based segmentation of SAR data with learned priors', IEEE Trans. Image Proc., 2000, 9, pp.299-301

[3]   G. Boccignone, M. Ferraro, An Information-theoretic Approach to Interactions in Images, Spatial Vision 12 (1999) 345-362.

[4]   S. Sanjay-Gopal, T.J. Hebert, Bayesian pixel classification using spatially variant finite mixtures and the generalized EM algorithm, IEEE Trans. Image Proc. 7 (1998) 1014-1028.

[5]   Y. Zhang, M. Brady, and S. Smith, Segmentation of Brain MR Images Through a Hidden Markov Random Field Model and the Expectation-Maximization Algorithm, IEEE Transactions on Medical Imaging 20(1) (2001) 45-57.

[6]   S. Haker, G. Sapiro, A. Tannenbaum, Knowledge-based segmentation of SAR data with learned priors, IEEE Trans Image Proc. 9 (2000) 299-301.

[7]   D.J.C. MacKay, Information Theory, Inference, and Learning Algorithms, Cambridge University Press, UK, 2003.

[8]   L. Xu, M. Jackowski, A. Goshtasby, D. Roseman, S. Bines ,C. Yu, A. Dhawan, A. Huntley, Segmentation of skin cancer images, Image and Vision Computing 17(1999) 65-74.

[9]   Dempster, A.P.—Laird, N.M.—Rubin, D.B.: Maximum-Likelihood from Incomplete Data Via the EM Algorithm. J.Roy. Statist. Soc., Vol. B39, 1077, pp. 1–38.

[10] Ambroise, C.—Govaert, G.: Convergence of an EM-Type Algorithm for Spatial Clustering. Pattern Recognition Letters, Vol. 19, 1998, pp. 919–927.
[11] J. Rissanen. Stochastic Complexity in Statistical Inquiry. World Scientific, 1989.
[12] D. Comaniciu. An algorithm for data-driven bandwidth selection. IEEE Trans. On Pattern Analysis and Machine Intelligence, 25(2):281–288, 2003.
[13] Z.H. Zhang, C.B. Chen, J. Sun, and K.L. Chan. EM algorithms for Gaussian mixtures with split-and-merge operation Pattern Recognition, 36(9):1973–1983, 2003.

Mr.R.Ravindraiah received his B.Tech degree from G.Pullareddy Engineering College, Kurnool, India in 2008 and M.Tech from AITS, Rajampet, India in 2011. He is doing research work in Bio-Medical Image Processing and presented some papers in reputed journals, IEEE & ELSEVIER Conferences. His research work includes analysis and processing of images of pathologies like Atherosclerosis, Retinopathy and Leukemia. Currently he is the Assistant Professor in department of Electronics and Communication Engineering, Madanapalle institute of technology and science, Madanapalle.   Under his guidance many UG and PG students have completed their projects.
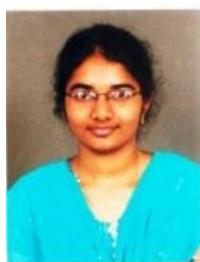
Ms.K.Tejaswini received her B.Tech degree in Electronics and Communication Engineering from Mekapati Rajamohan Reddy Institute of Technolgy and Science, Udayagiri, India in 2011.Currently she is pursuing M.Tech degree Digital Electronics and Communication Engineering in Madanapalle institute of technology and science, Madanapalle, India.