“1st International Symposium on Computing in Informatics and Mathematics (ISCIM 2011)”
in Collabaration between EPOKA University and “Aleksandër Moisiu” University of Durrës
on June 2-4 2011, Tirana-Durres, ALBANIA.

Expectation Maximization And Gaussian Model Based Segmentation on Histology Slides

Evgin GÖÇERİ
Department of Computer Engineering, Pamukkale University, Denizli-TURKEY
Email: egoceri@pau.edu.tr

ABSTRACT

The importance of the Expectation Maximization (EM) algorithm is increasing day by day in order to solve Maximum A Posteriori (MAP) estimation problems and Gaussian Mixture Models (GMMs), which are parametric probability density functions, have become more popular in computerized applications due to the EM algorithm. This article explains an automatic GMM based image segmentation method for histology cell images. For this purpose, the GMM parameters, which are recomputed iteratively starting with initial values, are calculated by using the EM algorithm which classifies each pixel into the class with the largest probability distribution using maximum likelihood. The accuracy of this segmentation algorithm depends on how much close the probabilistic model to the gray level distributions of the input images.

INTRODUCTION

Medical image segmentation is generally performed manually by experts, which is cumbersome, time consuming, tedious task due to the complex and unclear edge structures. Therefore such a segmentation leads to high outcome variability. Also, not only variations in image shape and size, but also the usage of different electronic devices for imaging, overlapping signal intensities and several imaging artifacts cause the segmentation process to be a more challenging task.

Although the advantages of semi-automatic segmentation algorithms, they are still remain labor intensive and so the need for fully automatic segmentation techniques which means without any user interaction is becoming more and more important. Automated and accurate segmentation of medical images to separate anatomically meaningful partitions is needed. For instance, caudate nucleus segmentation in patients with specific neurologic diseases is of great interest because it can help in improving patient treatment and/or diagnose.

There are several image segmentation approaches in the literature such as region growing method, watershed method, level set method, morphological operations, active shape models, deformable models, artificial neural network and atlas based segmentation techniques. Also, different combinations of these techniques are applied to obtain accurate results on different images. For instance, in [1-3] k-means based segmentation methods are applied for cDNA microarray spots, brain and fingerprints. K-means method, which is an iterative and unsupervised classification technique, based on minimization of the sum of distance between data points and their cluster centers. Data points are assigned to their nearest centers and each center is re-computed. Usually Euclidean distance is used to measure the distance. Region-based segmentation techniques [4-6] attempt to segment an image by identifying the various homogeneous regions that correspond to different objects in an image. Unlike clustering methods, region-based methods explicitly consider spatial interactions between neighboring voxels. In its simplest form, region growing methods usually start by locating some seeds representing distinct regions in the image. The seeds are then grown until they eventually cover the entire image. The region growing process is therefore governed by a rule that describe the growth mechanism and a rule that check the homogeneity of the regions at each growth step. In the split-and-merge technique [7], an image is first split into many small regions during the splitting stage according to a rule, and then the regions are merged if they are similar enough to produce the final segmentation. In addition to these generally used techniques, a neural network based approach is proposed [8] for liver segmentation. Another neural network based method [9] uses radial basis functions for segmentation. A deformable model based approach is proposed in [10] for kidney segmentation. Probabilistic atlases are constructed and used for medical image segmentation [11]. Also, an active shape model based segmentation technique by using non-rigid image registration is explained [12].

In this study, EM based image segmentation is explained and an application result by using a histological image is presented.

IMAGE SEGMENTATION USING EXPECTATION MAXIMIZATION AND GAUSSIAN MODEL

GMMs are known as a weighted sum of Gaussians. When we know the mean, covariance and prior probability values then we can define a multivariate Gaussian probability density function. Because, the probability distribution of each voxel can be represented as a mixture of probability distributions with coefficients, which can be taken as prior probabilities. The parameters of GMMs can be estimated by using EM algorithm [13,14] that interleaves probability density function estimation for each tissue class.

Assume that the intensities in an image be denoted as a one-dimensional array. $Y = \{y_j, j = 1,2,3,.....J\}$ that have J pixels, where y_j is the intensity of voxel *j*. Assume that we have *K* tissue types in the image, and voxel *j* can be belong to a tissue type, which can be represented as $l_j \in \{1,2,3.....K\}$. In the GMM, it is assumed that each tissue type *k* has an intensity «_k in the image. In other words, the probability density function (1) is represented with voxel *j* of tissue type l_j, which has intensity y_j, can be written as,

$$p(y_j \mid l_j, \Phi) = G_{\Sigma_{l_j}}\left(y_j - \text{«}_{l_j}\right) \tag{1}$$

where $G_\Sigma(.)$ refers to a zero-mean Gaussian distribution with covariance Σ and $\Phi = \left\{\text{«}_k, \Sigma_k, k = 1,2,3....K\right\}$ represents the set of model parameters.

To simplify the explanation, let all the pixel labels l_j represent by grouping in a label $L = \{l_j, j = 1,2,3,.....J\}$. It is assumed that the label l_j of each pixel is drawn independently from the labels of the other pixels by using a priori known probability ®_k, which can be represented as,

$$p(L) = \prod_j \text{®}_{l_j} \tag{2}$$

The probability density function for image *Y* with the given the model parameters Φ is then written as,

$$p(Y \mid \Phi) = \sum_L p(Y \mid L, \Phi) p(L)$$

$$= \prod_j p(y_j \mid \Phi)$$

with $$p(y_j \mid \Phi) = \sum_k p(y_j \mid l = k, \Phi) \circledR_k \qquad (3)$$

(3) shows a mixture model that models the histogram of gray values of an image as a sum of Gaussian distributions, each distribution is weighted with its given prior probability $\circledR_k$

The goal of an image segmentation process is to reconstruct the tissue labels *L* using the image *Y*. When we have the estimated parameters of the model, then each pixel can be assigned to the tissue type which best fits to its gray level value. However, the obtained results mainly depend on the model parameters. The iterative EM algorithm is used to estimate the maximum likelihood parameters $\hat{\Phi}$

$$\hat{\Phi} = \arg\max_{\Phi} \log p(Y \mid \Phi) \qquad (4)$$

by filling in the unknown tissue labels *L* and using the current estimated parameter values, This algorithm recalculates the parameters that maximizes the likelihood. Therefore, the algorithm consists of two steps:

Expectation step calculates that

$$\mathbb{Q}(\Phi \mid \Phi^{(m-1)}) = E_L \left[\log p(Y, L \mid \Phi) \mid Y, \Phi^{(m-1)} \right] \qquad (5)$$

Maximization step performs the recalculation by

$$\Phi^{(m)} = \arg\max_{\Phi} \mathbb{Q}(\Phi \mid \Phi^{(m-1)}) \qquad (6)$$

and by using *m* the iteration number. It has been proved that the likelihood $\log p(Y \mid \Phi)$ is increased at each iteration for EM algorithms [15].

The expectation step statistically classifies the voxels by using the estimation of the Gaussian distribution parameters that

$$p(l_j \mid Y, \Phi^{(m-1)}) = \frac{p(y_j \mid l_j, \Phi^{(m-1)})®_{l_j}}{\sum_k p(y_j \mid l_j = k, \Phi^{(m-1)})®_{l_j}} \tag{7}$$

and the maximization step estimates the Gaussian distribution parameters based on the classification as,

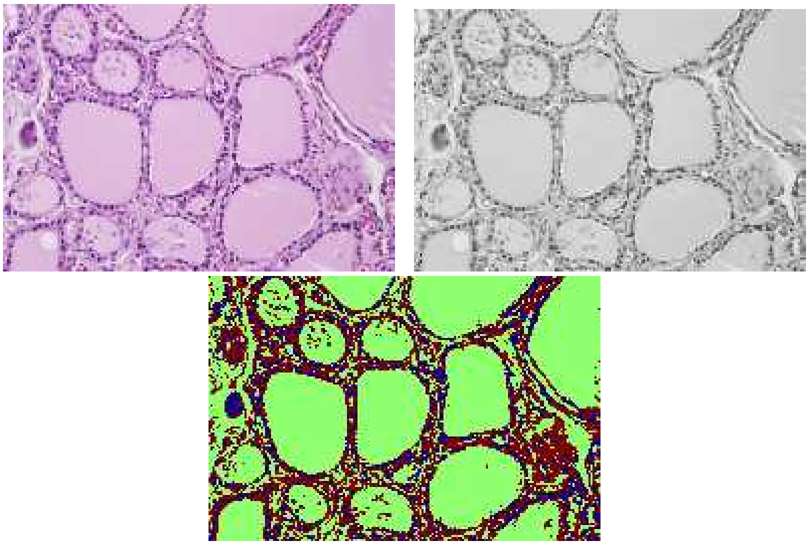
$$«_k^{(m)} = \frac{\sum_j p(l_j = k \mid Y, \Phi^{(m-1)}) y_j^{(m-1)}}{\sum_j p(l_j = k \mid Y, \Phi^{(m-1)})} \tag{8}$$

$$\Sigma_k^{(m)} = \frac{\sum_j p(l_j = k \mid Y, \Phi^{(m-1)})(y_j^{(m-1)} - «_k^{(m)})(y_j^{(m-1)} - «_k^{(m)})^t}{\sum_j p(l_j = k \mid Y, \Phi^{(m-1)})} \tag{9}$$

According to a given convergence criteria, equation (7) gives the final classification results.

APPLICATION ON HISTOLOGICAL IMAGES

An example application of the EM based segmentation method is performed by using a thyroid image as shown in Figure 1.a. The grayscale image is obtained (Figure 1.b) before application of the EM algorithm. The result image is shown (Figure 1.c) using different colors for each region in the original image.



(a) Original image image (b) Grayscale image (c) Segmented

Figure 1. EM based histological image segmentation

CONCLUSIONS

The advantage of the automated segmentation method is to achieve the process without any user supervision. Semi-automated or automated segmentation methods generate more agreeable results.

EM based probabilistic segmentation approaches for histological images give acceptable results. The accuracy of this segmentation algorithm depends on how much close the probabilistic model to the gray level distributions of the input images.

REFERENCES

[1] Yijun, H. and Guirong, W. (2009) Segmentation of cDNA microarray spots using k-means clustering algorithm and mathematical morphology. *WASE International Conference on Information Engineering (ICIE 2009),* Shanxi, China

[2] Juang, L. H. and Wu, M. N. (2010) MRI brain lesion image detection based on color-converted k-means clustering segmentation. *Measurement* **43**, 941–949

[3] Yang, G., Zhou, G. T., Yin, Y. and Yang, X. (2010) K-means based fingerprint segmentation with sensor interoperability. *EURASIP Journal on Advances in Signal Processing*, **2010**,1-12

[4] Chang Y.L and Li X. (1994) Adaptive image region-growing. *IEEE Trans.on Image Processing* **3**(6), 868-72.

[5] Adams R. and Bischof L. (1994) Seeded region growing. *IEEE Trans. Pattern Analysis Machine Intelligence* **16**(6),641-7.

[6] Pohle R. and Toennies K.D. (2001) Segmentation of medical images using adaptive region growing. *Proceedings of SPIE (Medical Imaging 2001)* **46**(1),4322-1337.

[7] Manousakas IN, Undrill PE, Cameron G. G, Redpath T. W. (1998) Split-and-merge segmentation of magnetic resonance medical images: Performance evaluation and extension to three dimensions. *Computers and Biomedical Research* **31**(6), 393–412.

[8] Rafiee, A., Masoumi, H. and Roosta, A. (2009) Using neural network for liver detection in abdominal MRI images. *IEEE Int. Conf. on Signal and Image Processing*, Kuala Lumpur

[9] Yang, P., Zhu, Q. and Zhong, X. (2009). Subtractive clustering based RBF neural network model for outlier detection. *Journal of Computers* **4**(8),755-783

[10] Tsagaan, B., Shimizu, A., Kobatake, H., Kunihisa, M., Hanzawa, Y. (2001) Segmentation of kidney by using a deformable model. *International Conference on Image Processing* (ICIP), Thessaloniki, Greece, **3**(1),1059–1062

[11] Park, H., Bland, P. and Meyer, C. (2003). Construction of an abdominal probabilistic atlas and its application in segmentation. *IEEE Trans. Med. Imaging* **22**(4),483–492.

[12] Spiegel, M., Hahn, D. A., Daum, V., Wasza, J., Hornegger, J. (2009). Segmentation of kidneys using a new active shape model generation technique based on non-rigid image registration. *Computerized Medical Imaging and Graphics* **33**(1),29–39

[13] Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977) Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society*, **39**(1),1-38.

[14] Wu, C. F. J. (1983) On the convergence properties of the EM algorithm. *The Annals of Statistics*, vol. **11**(1),95-103.

[15] Wu, Z., Chung, H. W. and Wehrli, F.W. (1994) A bayesian approach to subvoxel tissue classification in NMR microscopic images of trabecular bone. *Journal of the Society of Magnetic Resonance in Medicine* **31(1),**302–308.