DETECTION OF CANCEROUS LESION BY UTERINE CERVIX IMAGE SEGMENTATION

P. Priya

Department of Electronics and Communication Engineering, Vivekanandha Institute of Engineering and Technology for Women, India
E-mail: pleazantmail@gmail.com

Abstract
This paper works at segmentation of lesion observed in cervical cancer, which is the second most common cancer among women worldwide. The purpose of segmentation is to determine the location for a biopsy to be taken for diagnosis. Cervix cancer is a disease in which cancer cells are found in the tissues of the cervix. The acetowhite region is a major indicator of abnormality in the cervix image. This project addresses the problem of segmenting uterine cervix image into different regions. We analyze two algorithms namely Watershed, K-means clustering algorithm, Expectation Maximization (EM) Image Segmentation algorithm. These segmentations methods are carried over for the colposcopic uterine cervix image.

Keywords:
Segmentation, Uterine Cervix, Cervical Cancer, Colposcopy, Acetowhite, Watershed, Clustering

1. INTRODUCTION

CERVICAL cancer, one of the most common cancers affecting women worldwide and the most common in developing countries that can be cured in almost all patients, if detected and treated in time. Segmentation is a tool that has been widely used in many applications involving image processing and several techniques have been developed. One such application is in diagnosis using medical image analysis. For effective and correct diagnosis it is necessary to segment and find the different types of tissues and organs in the image. The cervix region, which is the main region of interest within the cervigram, is located in the central part of the image. Typical colposcopic uterine cervix image is shown in Fig.1. The main tissues of interest as defined are as follows. 1) The squamous epithelium (SE) the normal cervix tissue. It appears as a homogenous pinkish region and consists of multiple layers of cells [1]. 2) The columnar epithelium (CE), which exerts out of the Os (the opening of the cervix) when the cervix grows rapidly. This tissue is characterized by its bright red color and a rough textured appearance. The CE region is not always visible within a cervigram image. 3) The AW (Aceto White) region, the main focus of this paper, is epithelium that turns white and visible for a period of time following the application of acetic acid, and is a major visual indicator of cervical cancer.

This paper focuses on the segmentation of a specific tissue within the cervix, known as the AW tissue. The AW is a major indicator of cervical cancer.

Colposcopy is a medical procedure used in the detection of cervical cancer. A large microscope, or lens, is placed outside the vagina and focused on the cervix. This microscope is used to look for changes on the cervix that may signal the early signs of cervical cancer. The colposcope is the most important piece of equipment used during the procedure. The colposcope basically functions as a lighted microscope, which magnifies the view of the cervix, vagina and vulvar surface and helps to identify abnormal tissue.
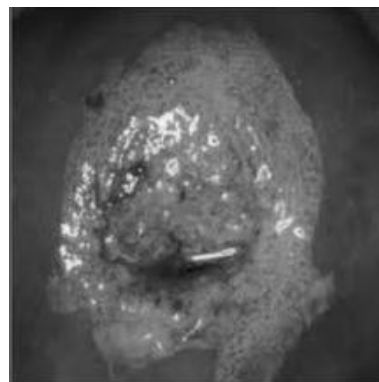

Fig.1. Colposcopic Uterine Cervix image

Segmentation is one of the first steps in image analysis. It refers to the process of partitioning a digital image into multiple regions (sets of pixels). Each of the pixels in a region is similar with respect to some characteristic or computed property, such as color, intensity, or texture. We can segment color images by reducing the number of colors in the images, and then recognizing contiguous pixels of the same color as a region. One very efficient technique for image segmentation that is being used for high quality segmentation in many complex images is the watershed algorithm. K-means clustering is a very popular clustering technique, which is used in numerous applications. With a k-means clustering algorithm, we can reduce the number of colors in a given image to K while maintaining the quality of the image.

2. METHODOLOGY

2.1 WATERSHED SEGMENTATION

The watershed transformation was widely and successfully applied in different domains such are biomedicine, industry and generally in computer vision application as a powerful segmentation tool. Watershed algorithm is an image segmentation method based on mathematical morphology, it can realize parallel regional partition and get complete segmentation regions. Watershed algorithm is a classical and effective image segmentation method [2].

Watershed algorithm [2]-[3] is an image segmentation method based on mathematical morphology, it was firstly introduced to image processing by Vincent and Soille. The gray value of every point in the image represents this point's altitude, for example, Fig.2 can be considered as a three dimensional surface. If the rainwater falls on the three dimensional surface, it will inflow the two regions which are marked as the slope basin, else if the rainwater fall just on the ridge line of watershed, it will inflow the two slope basins with the same probability, the watershed

transform will find out the slope basin and the ridge line of watershed in gray level image. Every image after transformation, the slope basins are the identifying objects or regions.

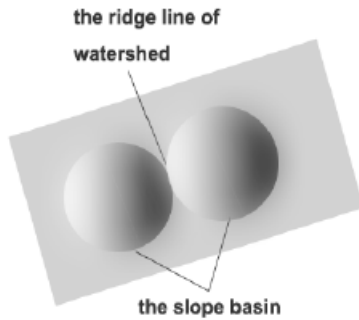the ridge line of watershed

the slope basin

Fig.2. The principle of watershed algorithm

Two different watersheds transform techniques, namely rain falling and water immersion, can be used to find watershed within an image. The rain falling algorithm [4] is a straightforward way to find watershed but requires extensive computations. On the other hand, the water immersion algorithm [5], [6], [7] can find watershed of an image quickly. In this algorithm, all pixels in the gradient image are sorted in order of increasing gray-level values. Flooding is performed beginning from the global minimum. Suppose that flooding has reached level h, where h represents the gradient magnitude of the pixel. Then in the next step the pixels at level h +1 need to be divided between the catchment basins of level h and the new basins corresponding to the local minima at level h + 1. For this, first the pixels at level h + 1 that are neighbors to pixels from a catchment basin at level h are put into a FIFO queue. Under the constraint of considering only pixels at level h +1, the catchment basins are extended by propagation. When several catchment basins of level h are connected at level h +1, the resulting basins are separated along the geodesic skeleton by influence zone. The pixels at level h +1 that are not reached by one of the catchment basins must be local minima, and they become the seeds of new basins. Once the image is completely flooded, thus obtained will correspond to watershed lines. But this method easily produces the over segmentation which results in the edge line buried in disorderly watershed lines. The watershed algorithm is illustrated in Fig.2
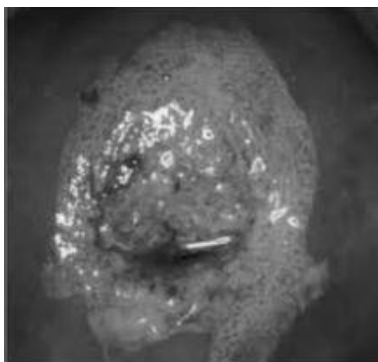
(a)
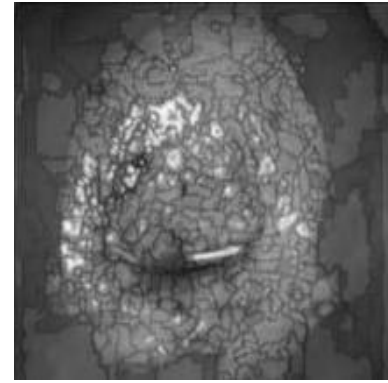
(b)

Fig.3. Segmented results for Colposcopic images
(a). Colposcopic uterine cervix input image, (b). Lesion segmented by Watershed Method

2.2 K-MEANS CLUSTERING ALGORITHM

K-means clustering algorithm [8] is one of the most influencing methods in many clustering algorithms. It is a most universal algorithm to adjust centroid of clustered region by continuous iterative. A non-hierarchical approach to forming good clusters is to specify a desired number of clusters, say, k, then assign each case to one of k clusters so as to minimize a measure of dispersion within the clusters. A very common measure is the sum of distances or sum of squared Euclidean distances from the mean of each cluster. The problem can be set up as an integer programming problem but because solving integer programs with a large number of variables is time consuming, clusters are often computed using a fast, heuristic method that generally produces good solutions. The k-means algorithm is one such method.
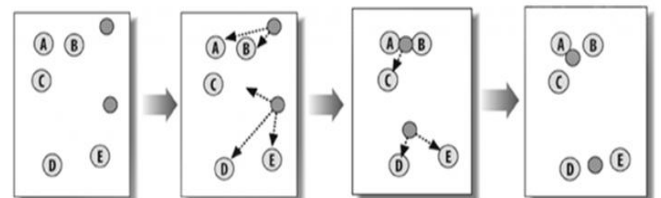

Fig.4. Demonstration of k-means clustering algorithm

The procedure follows a simple and easy way to classify a given data set through a certain number of clusters, assume k clusters. The main idea is to define k centroids, one for each cluster. These centroids should be placed in a cunning way because of different location causes different result. So the better choice is to place them as much as possible far away from each other. The next step is to take each point belonging to a given data set and associate it to the nearest centroid. When no point is pending, the first step is completed and an early groupage is done. The algorithm is also significantly sensitive to the initial randomly selected cluster centers. The k-means algorithm can be run multiple times to reduce this effect. These steps are demonstrated in Table.1.

K-Means method [9] starts with a single cluster with its center as the mean of the data. This cluster is split into two and the means of the new clusters are iteratively trained. These two clusters are again split and the process continues until the

specified number of clusters is obtained. If the specified number of clusters is not a power of two, then the nearest power of two above the number specified is chosen and then the least important clusters are removed and the remaining clusters are again iteratively trained to get the final clusters.

Firstly, select K points as initial clustering center, and calculate the distance between each data object and clustering center, the data object will be assigned to the cluster which has the closest mean. Then calculate the new clustering center of adjusted cluster. If there is no change of pre-and post clustering centers, it shows that clustering criteria has been converged. K-Means algorithm [10] is an unsupervised clustering algorithm that classifies the input data points into multiple classes based on their inherent distance from each other. The algorithm assumes that the data features form a vector space and tries to find natural clustering in them. The points are clustered around centroids which are obtained by minimizing the objective by

$$V = \sum_{i=1}^{k} \sum_{x_i \in s_i} \left(x_i - \mu_j \right)^2 \tag{1}$$

where ,there are k clusters S_i, $i = 1,2\dots k$ and μ_i is the centroid or mean point of all the points $x_j \in S_i$

As a part of this project, an iterative version of the algorithm was implemented. Various steps in the algorithm are as follows:

Step 1: Compute the intensity distribution of the intensities.

Step 2: Initialize the centroids with k random intensities.

Step 3: Repeat the following steps until the cluster labels of the image do not change anymore.

Step 4: Cluster the points based on distance of their intensities from the centroid intensities.

$$c^{(i)} = \arg \min{}_j \left\| x^{(i)} - \mu_j \right\|^2 \tag{2}$$
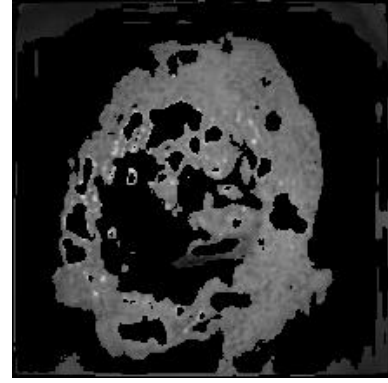
Step 5: Compute the new centroid for each of the clusters.

$$\mu_i = \frac{\sum_{i=1}^{m} \left\{ c_{(i)} = j \right\} x^{(i)}}{\sum_{i=1}^{m} \left\{ c_{(i)} = j \right\}} \tag{3}$$

where, k is a parameter of the algorithm (the number of clusters to be found), i iterates over all the intensities, j iterates over all the centroids and μ_i are the centroid intensities.
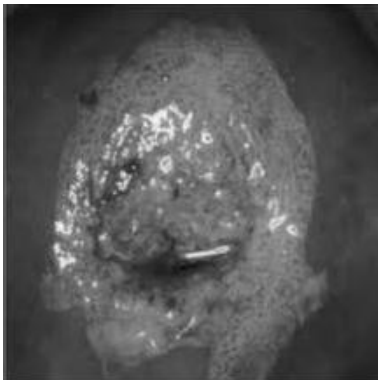
(a)

(b)

Fig.5. Segmented results for Colposcopic images (a). Colposcopic uterine cervix input image, (b). Lesion segmented by K-means Clustering method

2.3 EXPECTATION MAXIMIZATION (EM) IMAGE SEGMENTATION

The EM [11] algorithm is the algorithm for calculating the maximum-likelihood estimates when the observations can be viewed as incomplete data. Each iteration of the algorithm consists of an expectation step followed by a maximization step. Expectation-Maximization relies on the assignment of data to a given set of partitions. Every data value is associated with every partition through a system of weights based upon how strongly the data value should be associated with a particular partition. EM algorithms give very heavily on the assumption that the initial partition values are close to the natural clusters of the given data. The EM cycle begins with an Expectation step which is defined by the following equation:

$$E\left[Z_{ij}\right] = \frac{p\left(x = x_i \mid \mu = \mu_j\right)}{\sum_{n=1}^{k} p\left(x = x_i \mid \mu = \mu_n\right)} \tag{4}$$

$$= \frac{e^{-\frac{1}{2\sigma^2}\left(x_i - \mu_j\right)^2}}{\sum_{n=1}^{k} e^{-\frac{1}{2\sigma^2}\left(x_i - \mu_j\right)^2}} \tag{5}$$

This Eq.(4) states that the expectations or weight for pixel z with respect to partition j equals the probability that x is pixel x_i given that μ is partition μ_i divided by the sum over all partitions k of the same previously described probability. This leads to the lower expression for the weights. The sigma squared seen Eq.(5) represents the covariance of the pixel data. Once the E step has been performed and every pixel has a weight or expectation for each partition, the M step or maximization step begins. This step is defined by the following equation:

$$\mu_j \leftarrow \frac{1}{m} \sum_{i=1}^{m} E\left[Z_{ij}\right] x_i \tag{6}$$

This Eq.(6) states that the partition value j is changed to the weighted average of the pixel values where the weights are the weights from the E step for this particular partition. This EM cycle is repeated for each new set of partitions until, as in the K-Means algorithm, the partition values no longer change by a significant amount.

The EM algorithm is introduced to improve the initial estimates and possibly reduce the unlikely segmentation. One advantage of the proposed algorithm is that it gives a stable solution for the color image segmentation problem, while the results of the segmentation methods[12] that use a random initialization, e.g., the k-means algorithm[13], may differ according to the selection of initial parameters.
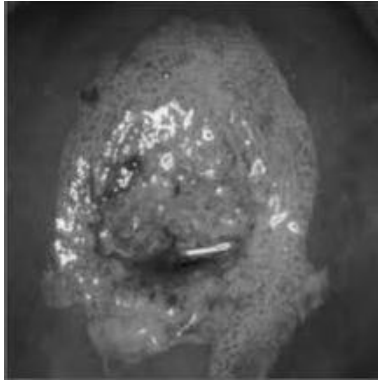
(a)

(b)

Fig.6. Segmented results for Colposcopic images.
(a). Colposcopic uterine cervix input image, (b). Lesion segmented by Expectation Maximization (EM) method
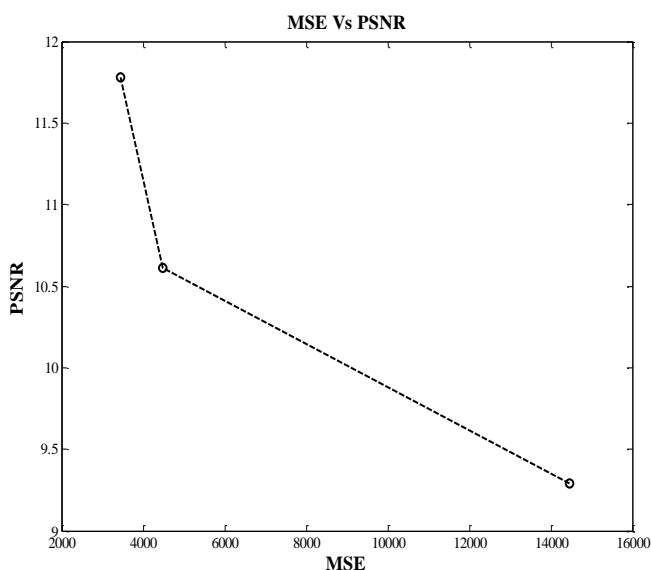

Fig.7. Graph plots for MSE vs PSNR for three segmentation algorithms

Table.1. PSNR for three segmentation algorithms

S. No	ALGORITHMS	PSNR
1	Watershed Segmentation Algorithm	9.2897
2	K-Means Clustering segmentation Algorithm	10.6116
3	Expectation Maximization Segmentation Algorithm	11.7787

The experimental results shows that the Watershed method gives the over segmentation problem shown in Fig.3. By segmentation using k-means clustering the results are shown in Fig.4. The segmentation results using Expectation Maximization (EM) Image Segmentation are shown in Fig.5. The Fig.6 shows the graph drawn for various values of MSE (Mean Squared Error) and PSNR (Peak Signal to Noise Ratio). The Table.2 shows the comparison of PSNR value for three different algorithms.

3. CONCLUSION

The purpose for the segmentation is to locate the abnormal region in a cervix for taking a biopsy for diagnosis of cervical cancer. We have presented a cervical image analysis system to extract acetowhite epithelium in cervical images. This algorithmic ability of segmenting the abnormal regions is a significant step in the process of developing a fully automatic diagnostic tool for cervical cancer. A lower value for MSE means lesser error and as seen from the inverse relation between the MSE and PSNR this translates to a high value of PSNR. Logically, a higher value of PSNR is good because it means that the ratio of Signal to Noise is higher. So by scheme having a lower MSE and a high PSNR is the better segmentation. Among these segmentation algorithm, EM image segmentation having the higher PSNR value and it gives the better segmentation. While the present project performed the Expectation Maximization method and hence found the lesion part effectively than the other segmentation methods. Future work deals with merging these small regions with larger ones efficiently to reduce the number of regions, and to make it easy to recognize objects in the images. And also more improved and separated regions or objects can be obtained using the other higher level segmentation techniques such as level set methods, graph partitioning methods, neural networks segmentation etc.

REFERENCES

[1] Amir Alush, Hayit Greenspan and Jacob Goldberger, "Automated and Interactive Lesion Detection and Segmentation in Uterine Cervix Images", *IEEE Transactions on Medical Imaging*, Vol. 29, No. 2, pp. 488 - 501, 2010.

[2] R.C. Gonzales, R.E. Woods and S.L. Eddins, "*Digital Image Processing Using MATLAB*", Publishing House of Electronics Industry, 2005.

[3] Kostas Haris, Serafim N. Efstratiadis, N. Maglaveras and Aggelos K. Katsaggelos, "Hybrid Image Segmentation Using Watersheds and Fast Region Merging", *IEEE Transactions on Image Processing*, Vol. 7, No. 12, pp. 1684 - 1699, 1998.

[4] L. Vincent and P. Solille, "Watershed in digital spaces: an efficient algorithm based immersion simulations", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 13, No. 6, pp. 583 - 598, 1991.

[5] A. Moga, B. Cramariuc and M. Gabbouj, "An efficient watershed segmentation algorithm suitable for parallel implementation", *Proceedings of IEEE International Conference on Image processing*, Vol. II, pp. 101 - 104, 1995.

[6] P. Salembier and M. Pardas, "Hierarchical morphological segmentation for image sequence coding," *IEEE Transactions on Image Processing*, Vol. 3, No. 5, pp. 639 - 651, 1994.

[7] Krishnan Nallaperumal et al., "A Novel Multiscale Morphological Watershed Segmentation Algorithm", *International Journal on Imaging Science and Engineering*, Vol. 1, No. 2, pp. 60 - 64, 2007.

[8] A. Goswami, R. Jin and G. Agrawal, "Fast and Exact Out-of- Core K-Means Clustering", *Fourth IEEE International Conference on Data Mining*, pp. 83 - 90, 2004.

[9] Weili Yang, Lei Guo, Tianyun Zhao and Guchu Xiao, "Image segmentation method based on watersheds and clustering", *Chinese Journal of Quantum Electronics*, Vol. 25, No. 1, pp. 19 - 24, 2008.

[10] T. Kanungo et al., "An Efficient k-Means Clustering Algorithm: Analysis and Implementation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 7, pp. 881 - 892, 2002.

[11] K. Revathy, "Applying EM algorithm for segmentation of textured images", *Proceedings of the World Congress on Engineering*, Vol. I, 2007.

[12] Fei gao, "An Efficient Approach to Automated Segmentation in Medical Image Analysis" A thesis in Electrical Engineering, 2005.

[13] Ying Wu, "Segmentation: Clustering, Graph cut and EM", *EECS 432-Advanced Computer Vision Notes Series 5*, Vol. II, pp. 1-10, 2000.

[14] J.W. Sellors and R. Sankaranarayanan, "Colposcopy and Treatment of Cervical Intraepithelial Neoplasia: A Beginner's Manual", *International Agency for Research on Cancer*, 2003.