INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact factor: 4.295
(Volume3, Issue3)
Available online at www.ijariit.com

Segmentation of Breast Images Using Gaussian Mixture Models

Miss. Ankita Satyendra Singh
Department of Electronics and Telecommunication Engineering
SVERI's COE Pandharpur, Solapur, Maharashtra, India

Prof. Mrs. M. M. Pawar
Department of Electronics and Telecommunication Engineering
SVERI's COE Pandharpur, Solapur, Maharashtra, India

***Abstract:** Breast cancer originates when cells grow uncontrollably in the breast resulting in a tumour that can be felt as a lump or observed on x-ray. The tumour is malignant that is cancerous if the cells invade surrounding tissues or spread to distant areas of the body. Breast Cancer can be observed both in women as well as men. Image processing aims at divide one picture into different types of classes or regions, recognition of objects and detecting of edges, etc that is done after the image is segmented. The main aim of this paper is to detect and separate background and foreground by using Gaussians Mixture Model, the parameters of the model and training data are learned by EM-algorithm. Pixel labeling corresponding to each pixel of the true image is done by Bayesian rule. This hi labeled image is constructed during of running EM-algorithm.Numerical method of finding maximum a posterior estimation is done by using EM-algorithm and Gaussians mixture model which we called EM-MAP algorithm.*

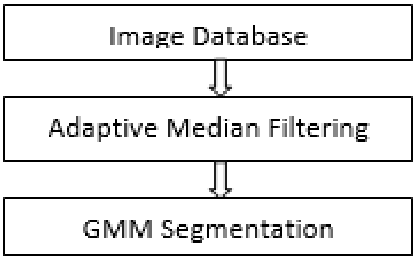
***Keywords:** Image segmentation, Expectation-Maximization (EM) Algorithm, Maximum a Posterior (MAP), Bayesian Rule, Gaussian Mixture Model (GMM).*

1. INTRODUCTION

Breast cancer is the second leading cause of death in women next to lung cancer. About 1 in 8 (12%) women in the will develop invasive breast cancer during their lifetime. The American Cancer Society's estimates for breast cancer in the United States for 2017 are: About 252,710 new cases of invasive breast cancer will be diagnosed in women. About 63,410 new cases of carcinoma in situ (CIS) will be diagnosed (CIS is non-invasive and is the earliest form of breast cancer). About 40,610 women will die from breast cancer. Breast cancer is found to be the most common

Cancer occurring in women all over India and accounts for 25% to 31% of all cancers in Indian women. Breast cancers can start from different parts of the breast. Most breast cancers begin in the ducts that carry milk to the nipple (ductal cancers). Some start in the glands that make breast milk (lobular cancers).Masses can be either benign i.e. non-cancerous or malignant i.e. cancerous. There are many mathematical and statistical methods for image problems, but this paper talks about GMM as a generally Gaussian distribution, EM-algorithm, and Bayesian rule. Markov Chain Monte Carlo algorithms or Variational methods with high computing to find MAP estimation is used as Bayesian framework usually has many difficulties because the posterior probability has a complex form. In this paper, the preprocessing technique is applied on the breast images and further, the preprocessed images are segmented using GMM model. The aim of this paper is to separate the region of interest by using GMM technique.

2. METHODOLOGY



3. IMAGE DATABASE

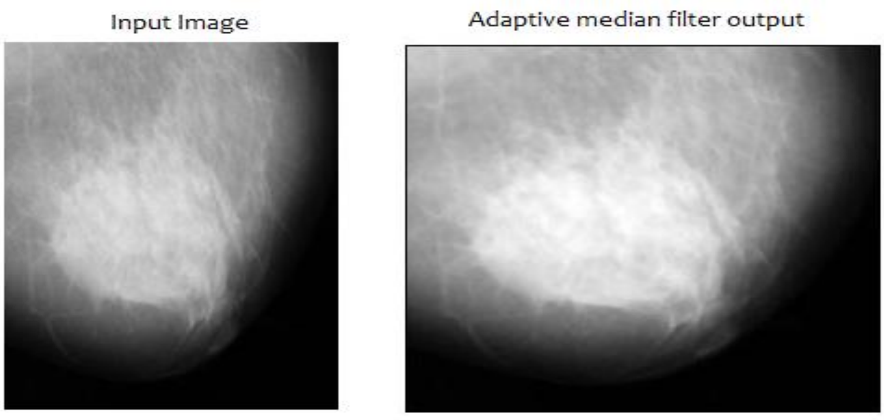
The mammogram images are taken from mini mammography database of MIAS. The Mammographic Image Analysis Society database contains all total 322 mammographic images in MLO out of which 207 are normal, 63 benign and 52 are malignant cases.Originally the images are available in .pgm (Portable Gray Map) format. The Mammographic Image Analysis Society has provided the image database for the purpose of research.

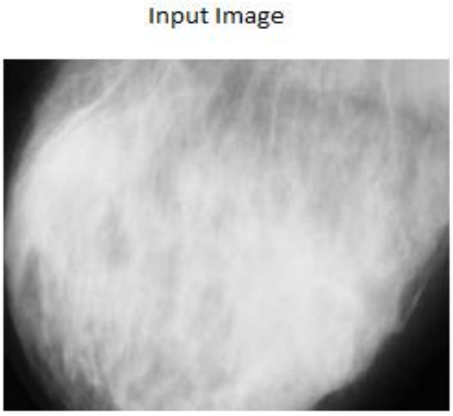
4. ADAPTIVE MEDIAN FILTERING

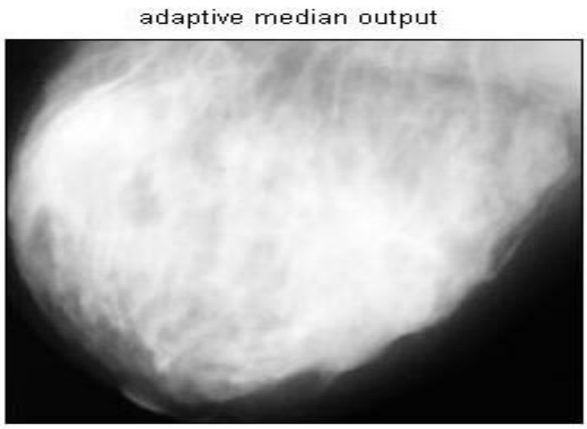
The aim of the pre-processing technique is to enhance the image quality and make ready to further processing by removing the unrelated and surplus parts in the Background of the mammogram images. The Adaptive Median Filter performs spatial processing to determine which pixels in an image have been affected by impulse noise. The Adaptive Median Filter classifies pixels as noise by comparing each pixel in the image to its surrounding neighbor pixels. The size of the neighborhood is adjustable, as well as the threshold for the comparison. A pixel that is different from a majority of its neighbors, as well as being not structurally aligned with those pixels to which it is similar, is labeled as impulse noise. These noise pixels are then replaced by the median pixel value of the pixels in the neighborhood that have passed the noise labeling test. The main purpose of this preprocessing technique is to remove the impulse noise while smoothing other noises present in the mammograms and reduce the distortions like excessive thinning or thickening of object boundaries. Adaptive median filter changes size of Sxy (the size of the neighborhood) during operation. The size of Sxy changes during the filtering operation depending on certain conditions. Each output pixel contains the median value in the 3-by-3 neighborhood around the corresponding pixel in the input images. Zeros replace the edges of the images. The output of the filter is a single value, which replaces the current pixel value at (x, y), the point on which S is centered at the time. The notations used are

Zmin = minimum pixel value in Sxy
Zmax = maximum pixel value in Sxy
Zmed = median pixel value in Sxy
Zxy= pixel value at coordinates (x, y)
Smax = maximum allowed size of Sxy

Adaptive Median filtering used to smooth the non-repulsive noise from two-dimensional signals without blurring edges and preserved images. This makes, it particularly suitable for enhancing mammogram images. The preprocessing techniques used in a mammogram, orientation, label, artifact removal, enhancement, and segmentations. The preprocessing involved in creating masks for pixels with the highest intensity, to reduce resolutions and to segment the breast.







a) Input Image **b)Adaptive Median Filter Output**

5. K-MEANS CLUSTERING

K-means clustering is a technique of vector quantization which is widely used for cluster analysis in data mining. The aim of K-means clustering is to partition the n no of observations into k clusters in which each observation belongs to the particular cluster with the nearest mean, serving as a prototype of the cluster. If a set of observations is given as (x1, x2, x3,…,xn) and each observation is a real vector having d-dimensions. This clustering technique aims to cluster aims to divide this n observations into k (≤ n) sets S = {S1, S2, Sk} in order to minimize the variance within-cluster sum of squares (WCSS).

$$\underset{\mathbf{S}}{\arg\min} \sum_{i=1}^{k} \sum_{\mathbf{x} \in S_i} \|\mathbf{x} - \boldsymbol{\mu}_i\|^2 = \underset{\mathbf{S}}{\arg\min} \sum_{i=1}^{k} |S_i| \operatorname{Var} S_i$$

μi is the mean of points in Si which is equivalent to minimizing the pair wise squared deviations of points in the same cluster.

$$\sum_{\text{Cluster } C_i} \sum_{\text{Dimension } d} \sum_{x,y \in C_i} (x_d - y_d)^2$$

As the total variance is constant and is equal to maximizing the squared derivations between the points from the different clusters. K-means clustering is widely used in widely used as a pre-processing step for different algorithms. The K-means clustering is used as a pre-processing step for the GMM segmentation where the breast images are partitioned into three clusters as the background region, breast tissues and the lumpy part which is the main area of focus.

6. GAUSSIAN BLUR

The Gaussian blur is an image-blurring filter which uses a Gaussian function to calculate the transformation that is applied to transform each of the pixels from the image. For the two dimensions, we consider multiplication of two Gaussians one in both the dimensions stated as

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$$

x and y are the distances from the origin in the horizontal axis vertical axis respectively, and σ is the standard deviation of the Gaussian distribution.

7. GAUSSIAN MIXTURE MODELS

A GMM (Gaussian mixture model) is used for modeling data which comes from one of the numerous groups, the groups might be different from each other, but data points within the same group can be modeled by a Gaussian distribution. The image is a matrix in which each element is a pixel. The value of the pixel is nothing but a number that shows intensity or color of the input image. Consider X to be a random variable that takes values. For Probability model determination, we can assume to have a mixture of the Gaussian distribution.

$$f(x/\theta) = \sum_{i=1}^{k} \alpha_i f_i(x/\theta_i)$$

4. (M-step)

•

$$\hat{p}_i^{(r+1)} = \frac{1}{n}\sum_{j=1}^{n} p_{ij}^{(r)}$$

•

$$\hat{\mu}_i^{(r+1)} = \frac{\sum_{j=1}^{n} p_{ij}^{(r+1)} x_j}{n\hat{p}_i^{(r+1)}}$$

•

$$\hat{\sigma}_i^{2(r+1)} = \frac{\sum_{j=1}^{n} p_{ij}^{(r+1)}(x_j - \hat{\mu}_i^{(r+1)})}{n\hat{p}_i^{(r+1)}}$$

5. Iterate steps 3 and 4 until an specify error i.e. $\sum_i e_i^2 < \varepsilon$
6. Compute

$$p_{lj} = ArgMax_i p_{ij}^{(final)}, j = 1, 2, \ldots, n$$

7. Construct labeled image corresponding of each true image pixel.

Where k is the number of regions and pi _ 0 weight such that

$$\sum_{i=1}^{k} pi = 1$$

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

Where µi,σi are mean, the standard deviation of class i.The parameters are

$$\theta = (p_1, \ldots, p_k, \mu_1, \ldots, \mu_k, \sigma_1^2, \ldots, \sigma_k^2)$$

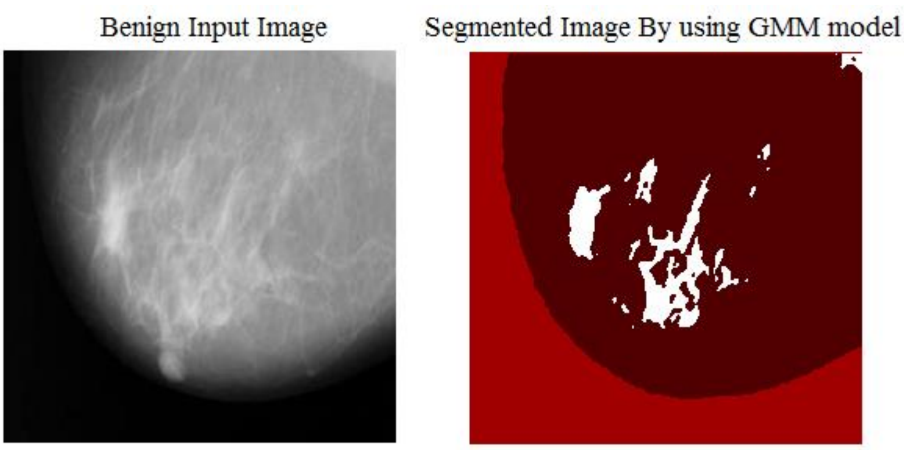
8. EM-MAP ALGORITHM

EM algorithm is stated as below:

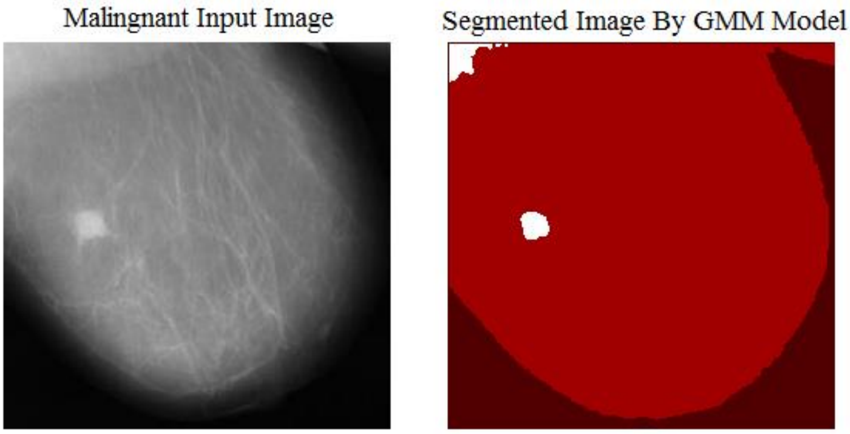
1. Input:Observed Image in a vector $x_j, j = 1, 2, \ldots, n$ and $i \in \{1, 2, \ldots, k\}$]
2. Initialize: $\theta^{(0)} = (p_1^{(0)}, \ldots, p_k^{(0)}, \mu_1^{(0)}, \ldots, \mu_k^{(0)}, \sigma_1^{2(0)}, \ldots, \sigma_k^{2(0)})$
3. (E-step)

$$p_{ij}^{(r+1)} = P^{(r+1)}(i|x_j) = \frac{p_i^{(r)} N(x_j|\mu_i^{(r)}, \sigma_i^{2(r)})}{f(x_j)}$$

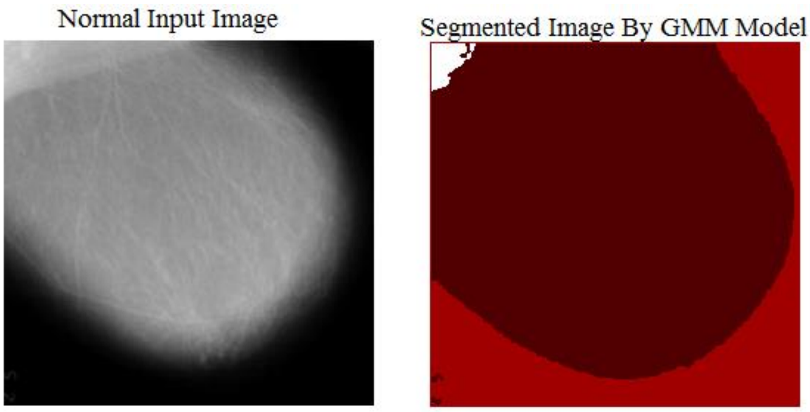
EM-MAP algorithm is a pixel labeling method such that the labelled image shows each segment or object by a different type of labels.



a) Segmentation of Benign Breast Images



b) Segmentation of Malignant Breast Images



c) Segmentation of Normal Breast Images

CONCLUSION

In this paper, we have formed a new numerical EM-GMM-Map algorithm for noise reduction and image segmentation. The adaptive median filter is the most appropriate pre-processing method used for mammographic images. The images are accurately segmented to separate out the tumour region from Benign, Malignant and Normal Images.

REFERENCES

1. H. D. Cheng and Muyi Cu, ―Mass Lesion Detection with a Fuzzy Neural Network‖, J. Pattern Recognition, 37, pp.1189-1200, 2004.
2. Belsare AD, Mushrif MM. Histopathological image analysis using image processing techniques: An overview. Signal Image Proc Int J 2012;3(4):23–36.
3. A.B. Tosun and C. Gunduz-Demir, C. Graph run-length matrices for histopathological image segmentation, IEEE Transactions on Medical Imaging 30 (3), 2011,732-566
4. Ali I, Wani WA, Saleem K. Cancer scenario in India with future perspectives. 2011;8(1):56–70.
5. Ho,S.,Bullitt,E.,&Gerig,G.(2002).Level-setevolutionwithregioncompetition:Au-tomatic3-Dsegmentationofbraintumors.InProceedingsofthe16thInternationalConferenceonPatternRecognition,2002:1(pp.532–535).IEEE.
6. Zucker, S.W.(1976).Regiongrowing:childhoodandadolescence.ComputerGraphicsandImageProcessing,5(3),382–399.
7. Tahmasbi,A.,Saki,F.,&Shokouhi,S.B.(2011a).ClassificationofbenignandmalignantmassesbasedonZernikemoments.ComputersinBiologyandMedicine,41(8),726–735.
8. Rouhi,R.,Jafari,M.,Kasaei,S.,&Keshavarzian,P.(2015).BenignandmalignantbreasttumorsclassificationbasedonregiongrowingandCNNsegmentation.Expert SystemswithApplications,42(3),990–1002.
9. Oliver,A.,Freixenet,J.,Marti,J.,Perez,E.,Pont,J.,Denton,E.R.,&Zwiggelaar,R.(2010).Areviewofautomaticmassdetectionandsegmentationinmammographicimages.MedicalImageAnalysis,14(2),87–110
10. Li,C.,Xu,C.,Gui,C.,&Fox,M.D.(2005).Level set evolutionwithreinitialization:anewvariationalformulation.InProceedingsoftheIEEEComputerSocietyConferenceonComputerVisionandPatternRecognition,2005.CVPR2005:1(pp.430–436).IEEE.
11. U,T.,Nandi,A.K.,&Rangayyan,R.M.(2008).Classificationofbreastmassesusingselectedshape,edge-sharpness,andtexturefeatureswithlinearandkernelbasedclassifiers.JournalofDigitalImaging,21(2),153–169
12. Anitha,J.,&Peter,J.D.(2015).Mammogramsegmentationusingmaximalcellstrengthupdationincellularautomata.Medical&BiologicalEngineering&Computing,53(8),737–749.

