Computer-Aided Breast Tumor Segmentation

Shweta Bhanushali
Computer Department
K.J. Somaiya College of Engineering
Mumbai, India

Sonia Lad
Computer Department
K.J. Somaiya College of Engineering
Mumbai, India

Vaibhavi Haria
Computer Department
K.J. Somaiya College of Engineering
Mumbai, India

Poonam Bhogale
Assistant Professor
Computer Department
K.J. Somaiya College of Engineering
Mumbai, India

ABSTRACT
Breast cancer is one of the most prevalent cancers diagnosed among the middle aged women. The rate of curacy depends on how well and early the tumor has been detected. One of the most effective methods of breast tumor segmentation is by using x-ray mammography. The accuracy of the results varies with the experience of the radiologists and the quality of the mammograms. In order to overcome these drawbacks a computer aided system has been developed that can accurately identify, position and segment the tumor.

General Terms
Tumor, Segmentation, Computer-aided.

Keywords
Computer-aided, Tumor segmentation, Thresholding, LBG, KPE, KMeans.

1. INTRODUCTION
Breast cancer is one of the most frequently diagnosed forms of cancer in women. Image segmentation is a fundamental issue in biomedical imaging area. The system aims at studying and comparing segmentation algorithms, based on standard parameters like efficiency and area.

Our objective is to develop a system incorporating image processing, and computer vision techniques for enhancement, segmentation of breast tumors. Our study aims at enhancing the current accuracy (diagnostic) of digital mammograms using industry standard simulation software tool, MATLAB and the online dataset. These techniques involve pre-processing of digital mammograms by resizing them and then apply the proposed algorithms for segmentation. Computer-aided breast tumor segmentation system can be used by radiologists and health-care specialists. The system is expected to improve the efficiency of breast cancer screening, and possibly reduce health care costs by decreasing the need for follow-up procedures such as biopsy. Several processing steps are required for the accurate characterization and analysis of biomedical image data. In our system, we have implemented four algorithms and have made a brief comparison about the working and efficiency of these algorithms based on parameters like area, segmentation accuracy, mean absolute error, etc. The algorithms are Global Thresholding, Linde-Buzo-Gray algorithm, Kekre's Proportionate Error algorithm and Kmeans Clustering. All except Global Thresholding algorithm are based on concepts of Vector Quantization. The algorithms are explained in detail and comparison is performed.

2. PROPOSED SYSTEM
2.1 Global Thresholding
Thresholding is the simplest method of image segmentation. From a gray scale image, thresholding can be used to create binary images. The simplest property that pixels in a region can share is intensity. So, a natural way to segment such regions is through thresholding, the separation of light and dark regions. Thresholding creates binary images from grey-level ones by turning all pixels below some threshold to zero and all pixels about that threshold to one.

The Algorithm for Global Thresholding is as follows:

Step1: The input image is first selected from the database.

Step2: Read the selected image.

Step3: Resize the file to 256*256 size

Step4: Convert original image to Gray Scale image.

Step5: Plot the histogram of this image and select threshold value.

Step6: Perform erosion and dilation on the obtained image.

Step7: Perform superimposition of images.

Step8: Obtain the desired tumor region as output.

2.2 LBG Algorithm [3]
Image vector quantization (VQ) includes four stages: vector formation, Training set selection, codebook generation and quantization. We first divide the input image into set of vectors. Later subset of vectors in the set is chosen as a training sequence. The codebook of codewords is obtained by an iterative clustering algorithm.

In 1980, Linde et al. proposed a Generalized Lloyd Algorithm (GLA) which is also called Linde-Buzo-Gary (LBG) algorithm. A mapping function to partition training vectors into N clusters. The mapping function is defined as $R^k \rightarrow CB$. Let $X = (x_1, x_2, \ldots, x_k)$ be a training vector and $d(X; Y)$ be the Euclidean distance between any two vectors.

The iteration for a codebook generation is given as follows:

Step1: Read the image, convert it into grayscale and resize image to 256*256.

Step 2: Generate an initial codebook CB_0 using mean.

Step 3: $i = 0$.

Step 4: Perform the following process for each training vector.

Compute the Euclidean distances between the training vector and the code words in CB_i. The Euclidean distance is defined as

$$d(X,C)=\sqrt{\sum_{t=1}^{k}(x_t - c_t)^2}$$

Equation 1[3]

• Search the nearest code word among CB_i.

Step 5: Partition the codebook into N cells.

Step 6: Compute the mean of each cell to obtain the new codebook CB_{i+1}.

Step 7: $i = i + 1$ and go to Step 4.

Repeat till the codebook of desired size is obtained.

2.3 KPE Algorithm [1]

Kekre's Proportionate Error algorithm is also based on Vector quantization method.

This algorithm is same as the LBG algorithm, just in place of adding a constant error, we are adding proportionate error.

2.4 K-means Clustering [8]

K-means clustering is a simple way to classify the data set into clusters on the basis of some pre-defined criteria.

Step 1: There are K clusters with at least one item in each cluster.

Step 2: The choice for the centroid of the clusters is done randomly.

Step 3: The distance calculation is done between the centroids and the pixels of the image. Clusters are formed based on minimum distance.

Step 4: Now the centroid of all the pixels in any cluster is calculated which will give us a new centroid and this process of clustering is repeated till we obtain a constant centroid or constant data set of a cluster.

3. EXPERIMENTAL RESULTS:

We have implemented these algorithms for 20 mammogram images and the screenshots for one of the tumor mammogram is shown below. The algorithm is tested for efficiency on basis of parameters like Area, Entropy, Segmentation Accuracy, Peak Signal to Noise Ratio and Mean Absolute Error.

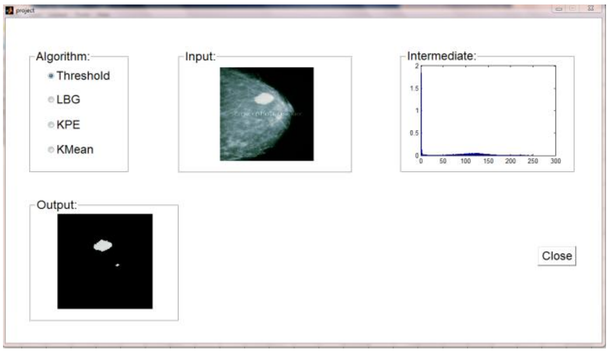

Figure 1: Global Thresholding

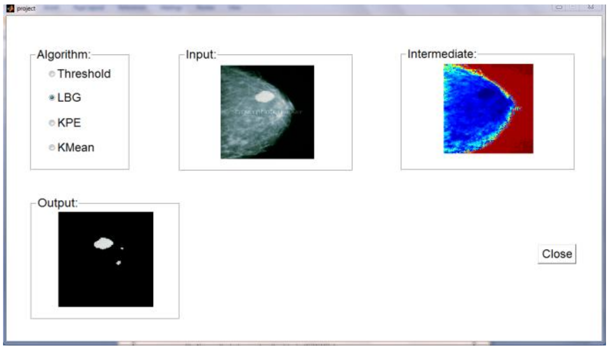

Figure 2: LBG Algorithm

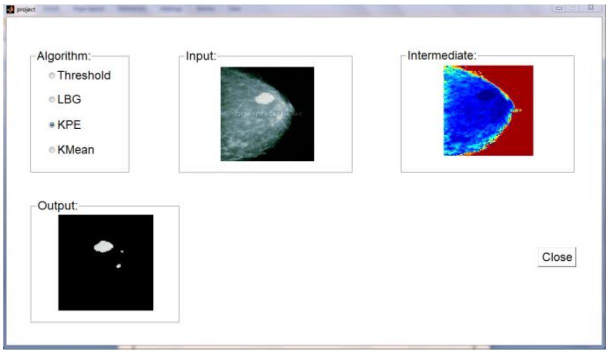

Figure 3: KPE Algorithm

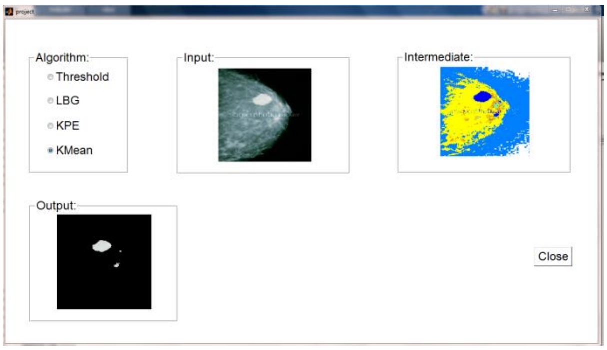

Figure 4: KMeans Algorithm

	Threshold	LBG	KPE	KMeans
Before Extraction Area	2195	2195	2195	2195
After Extraction Area	1093	1260	1260	1263
Entropy	0.198074	0.233242	0.233242	0.232132
Mean Absolute Error	551	467.5	467.5	466
Peak Signal to Noise Ratio	0.174703	0.205907	0.205907	0.20657
Segmentation Accuracy	49.79499	57.40319	57.40319	57.53986

Figure 5: Efficiency Parameters

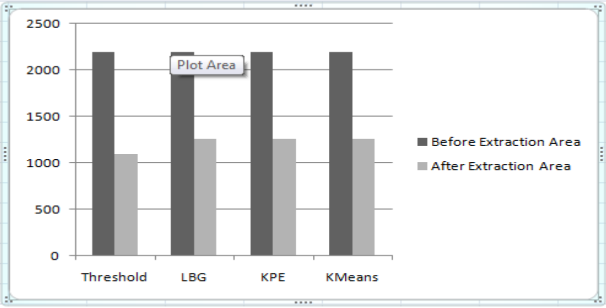

Figure 6: Area

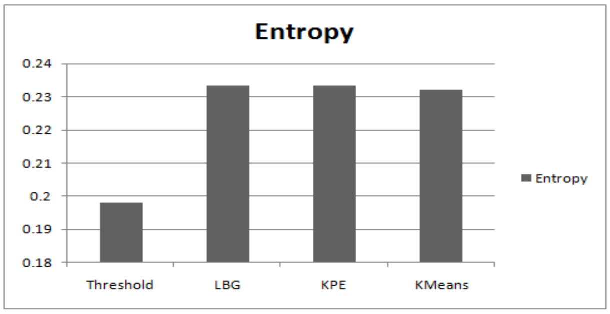

Figure 7: Entropy

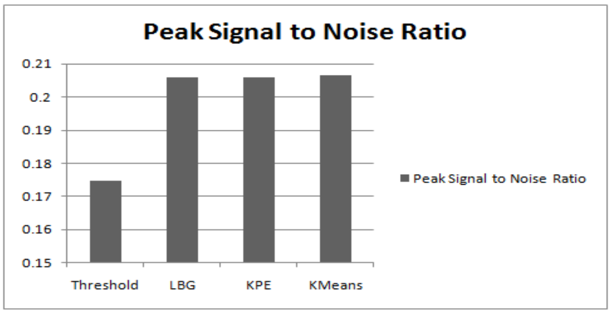

Figure 8: Peak Signal to Noise Ratio

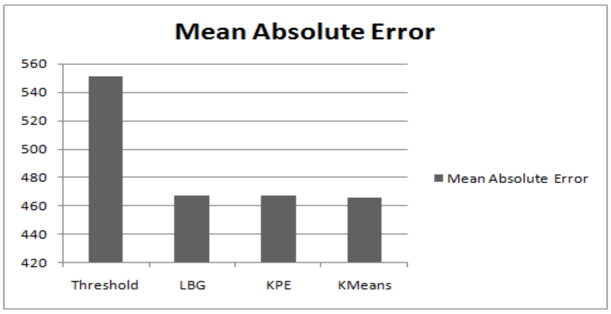

Figure 9: Mean Absolute Error

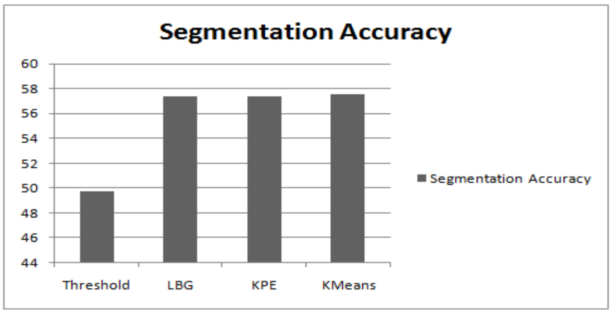

Figure 10: Segmentation Accuracy

Table 1: Comparison of Algorithms [5]

Sr. No.	Parameters	Global Thresholding	LBG	KPE	KMeans
1.	Human Intervention	Yes. To enter the threshold value	No intervention	No intervention	No intervention
2.	Area	Poor Extraction	Better extraction than thresholding	Better extraction than thresholding	Best extraction
3.	Entropy	Low information retrieved	Better information retrieved	Better information retrieved	Good information retrieved
4.	Peak Signal to Noise Ratio	Segmented image is of low quality	Segmented image is of better quality	Segmented image is of better quality	Segmented image is of good quality
5.	Mean Absolute Error	Higher value, so poor quality of image	Better quality of the image	Better quality of the image	Good quality of the image

4. CONCLUSION

The four algorithms, Global Thresholding, LBG, KPE and KMeans Clustering were successfully implemented using MATLAB 13.0. These algorithms were tested on 20 digital mammogram images. It is observed that the results obtained by KPE algorithm are more accurate as compared to other three algorithms. Segmentation of breast tumors from the mammogram images using KPE is much faster than the LBG method. So it can be concluded that the KPE algorithm is a better segmentation algorithm for breast tumors. The segmented tumor can be useful for further procedures including biopsy. This segmented tumor can be used further for classification of cancerous tumors, whether malignant or benign.

5. ACKNOWLEDGEMENTS

Regards to Prof. Poonam Bhogle for her valuable guidance and suggestions and K J Somaiya College of Engineering for providing us with the required platform.

6. REFERENCES

[1] Dr.H.B.Kekre,Tanuja K.Sarode. 2012," Two-level vector-quantization method for Codebook generation, using Kekre's Proportionate error algorithm, International Journal of Image Processing, Volume(4):Issue(1)

[2] Theodosios Goudas and Ilias Maglogiannis, "Cancer Cells Detection and Pathology Quantification Utilizing Image Analysis Techniques", 34th Annual International Conference of the IEEE EMBS San Diego, California USA, 2012.

[3] Tzu-Chuen Lu and Ching-Yun Chang, "A Survey of VQ Codebook Generation", Journal of Information Hiding and Multimedia Signal Processing, Volume 1, Number 3, 2010.

[4] Dr.H.B.Kekre, TK.Sarode, S.M.Gharge.Kekre's," Fast Codebook Generation Algorithm for Tumor Detection in Mammography Images", ICWET 2010.

[5] Manisha Sharma, Vandana Chouhan, "Objective Evaluation parameters of Image Segmentation Algorithms" , IJEAT 2012.

[6] Manoj Kumar, Poonam Saini, "Image Compression With Efficient Codebook Initialization Using LBG Algorithm", *International Journal of Advanced Computer Technology*, Volume 2, Number 2.

[7] Mukesh Mittal, RuchikaLamba, "Image Compression Using Vector Quantization Algorithms: A Review", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 6, 2013.

[8] Huanyi Yang, Lauren A. Christopher, Nebojsa Duric, Erik West and Predrag Bakic, "Performance Analysis of EM-MPM and K-means Clustering in 3D Ultrasound Image Segmentation", IEEE2012.

[9] P.Spandana, Dr.Kunda M.M Rao.SMIEEE, Prof. B.V.V.S.N.Prabhakar rao, Dr.Jwalasrikala, "Novel Image Processing Techniques for Early Detection of Breast Cancer, Mat lab and Lab view implementation", 2013 IEEE Point-of-Care Healthcare Technologies(PHT) Bangalore, India, 2013.