

# Tumour Delineation using Statistical Properties of The Breast US Images and Vector Quantization based Clustering Algorithms

Dr. H. B. Kekre

Senior Professor, Department of Computer Engineering  
MPSTME, SVKM's NMIMS University, Mumbai, India  
hbkekre@gmail.com

Pravin Shrinath

Associate Professor and Ph.D. Scholar, Department of Computer Engineering  
MPSTME, SVKM's NMIMS University, Mumbai, India  
shrinath.pm@gmail.com

**Abstract** — Breast cancer is most common and leading cause of death among women. With improvement in the imaging modalities it is possible to diagnose the cancer at an early stage moreover treatment at an early stage reduces the mortality rate. B-mode ultrasound (US) imaging is very illustrious and reliable technique in early detection of masses in the breast. Though it is complimentary to the mammography, dense breast tissues can be examined more efficiently and detects the small nodules that are usually not observed in mammography. Segmentation of US images gives the clear understanding of nature and growth of the tumor. But some inherent artifact of US images makes this process difficult and computationally inefficient. Many methods are discussed in the literature for US image segmentation, each method has its pros and cons. In this paper, initially region merging based watershed and marker-controlled watershed transforms are discussed and implemented. In the subsequent sections we proposed a method for segmentation, based on clustering. Proposed method consists of three stages, in first stage probability images and its equalized histogram images are obtained from the original US images without any preprocessing. In the next stage, we used VQ based clustering technique with LBG, KPE and KEVR codebook generation algorithm followed by sequential cluster merging. Last stage is the post processing, where we removed unwanted regions from the selected cluster image by labeling the connected components and moreover used morphological operation for closing the holes in the final segmented image. Finally, results by our method are compared with initially discussed methods.

**Index Terms** — Probability image, vector quantization, codebook generation, connected components

## I. INTRODUCTION

Breast cancer is the most commonly diagnosed cancer

among women and leading cause of death in developed and developing countries. In the westernized countries, it is observed that mortality rates have been stable or decreasing from the past few years. This reduction is credited to early detection through mammography, ultrasound imaging and subsequently with improved treatment [1, 2]. Nowadays, ultrasound imaging becoming more popular due to its real-time image availability, low health risk to patient and low cost. B-mode ultrasound imaging is complimentary to mammogram and commonly used modality to examine and diagnose breast cancer [3, 4]. In the dense and thick breasts, mammography has limited sensitivity, therefore omnipresent breast cancer often visible as normal masses and difficult to detect. Ultrasound imaging is useful to examine dense breast tissues effectively and detects the presence of small malignant nodule moreover it is routinely used for remedial diagnosis, to overcome the limitations of mammogram. In case of solid malignant masses, only ultrasound gives better tissue characterization and differentiation from benign masses [5, 6, 7]. Malignant tumor has typical characteristics, such as irregular shape, ill defined boundaries, lobulated, microlobulated margins and heterogeneous echo texture whereas benign tumor has round or oval shape with well defined margins and homogeneous echo texture. Sometime heterogeneous texture in the tumor region may present in both the cases, therefore it cannot be considered as clinically valuable information for diagnosis. However shape and boundary characteristic of a tumor be the most valuable information in the detection and differentiation of tumor [8].

Image segmentation process can differentiate between a region of interest (object) and other regions (background). Accurate segmentation of US images usually provides clinically important information for radiologists, such as shape irregularity, boundary characteristics and quantitative measurement of tumor and also helps in treatment and planning for surgery [9].

In ultrasound imaging, quality of image is highly subjective (dependence on expert radiologist, operator and technology), therefore detection, quantification (shape and size) of a tumor manually is difficult and time consuming process. Though, there is enormous improvement in the scanning devices (transducers), and technology, presence of some inherent artifacts degrades the quality of images. Furthermore they are major concerns in achieving accurate segmentation. Speckle (sometime called texture) is common artifacts found in US images and most of the segmentation algorithm has an additional step to remove it [10]. The artifact such as attenuation, causes by the gradual loss in the intensity of the ultrasound waves, generates intensity patterns within the boundary region. Therefore existence of blurred boundary and tissue intensity variation in the region of interest is major hurdles to achieve accuracy in the automated intensity based segmentation [11]. Various methods are discussed in the literature which handles this issue with the help of multiple images of the same region (sequence of images), but the processing of multiple images together is computationally inefficient [12, 13]. Some segmentation methods are based on GLCM texture feature, watershed transform [14], self organizing map neural network, etc, but they use speckle removal techniques for de-noising the real US images. Noise removal techniques such as Median filter, Gaussian filter and Wiener filter smear the margins of the region of interest and may cause of loss of important image information. Furthermore texture feature calculation based on GLCM is computationally inefficient. [15, 16, 17] Due to high sensitivity to the noise traditional methods such as histogram thresholding, region growing and merging are not suitable for US image segmentation. Adaptive thresholding faces the serious problem of selecting optimum block size to divide image and automatic selection of threshold, moreover require certain condition to stop the algorithm, since it is iterative. Model based method such as Active contour may perform better for US images, but it require manually delineated images and manual delineation is time consuming, further accuracy is subject to human expertise [10, 18]. K-means, Fuzzy C-means clustering based methods are also studied and used for US image segmentation, but they are very sensitive to the noise [19, 20]. Here, in this paper, we are delineating the boundary of tumor in high degree noisy (speckle) and attenuated US images without involving any preprocessing step (i.e. image enhancement). We are proposing three stages algorithm for segmentation, in first stage probability and its histogram equalized images are obtained from the original US images. In second stage, Vector Quantization (VQ) based clustering is done by LBG, KPE and KEVR codebook generation algorithms along with sequential cluster merging. Post processing is done at the third stage.

The other sections of this paper are organized as follows, in section II, vector quantization is discussed with encoding technique and its usability in

segmentation. In section III, watershed and marker controlled watershed algorithms are discussed. Section IV describes the flow of proposed algorithm. In section V and VI probability images formation and clustering algorithms are discussed respectively along with sequential merging of clusters. Section VII, discussed post processing techniques i.e. connected component labeling subsequently morphological operation for closing the hole in the cluster images. In section VIII results are discussed followed by conclusion in section IX.

## II. WATERSHED AND MARKER CONTROLLED WATERSHED TRANSFORMS

### A. *Watershed transform and region mearging (WT)*

Immersion based Watershed Transform shown its usefulness in digital image processing and was introduced by Luc Vincent and Pierre Soille [21]. Nowadays this region based algorithm extensively used in medical image segmentation like Ultrasound, MRI, CT and mammography, but it has a problem of high sensitivity towards noise. Therefore, due to inherent speckle noise in US images, it always involves additional noise removal techniques, such as Multi-scale Gradient filtering, Wiener filter or Anisotropic Diffusion filter [22, 23]. In this algorithm, firstly image can be viewed topographically using its gradient map, which is made up of ridges and catchment basins of various sizes and shapes depends on local maxima and minima. Hole is there in each catchment basin, which is represented by the minimum gradient value in the region (catchment basin). The water will start penetrating after immersion from the catchment basin which has lowest bottom. As water level increases it starts merging the adjacent regions, to avoid this, dam is constructed and eventually regions are separated. But this method has major problem of over-segmentation, therefore improvement such as region merging is suggested to overcome this problem [24], still it has constraints to select the criteria for merging of regions. Results of this method is shown in Figure 24 (a), Figure 25 (a), Figure 26 (a) and Figure 27 (a), In the following section a newly developed marker controlled watershed transform is discussed.

### B. *Marker-Controlled Watershed transform (MCWT)*

Over segmentation in the watershed transform is due to formation of large number of catchment basins with trivial regional minima, which may not be the part of region of interest. To overcome this problem a concept of markers has been introduced, it is nothing but the connected component belongings to an image. In this technique two types of markers are located as foreground and background markers. Since region of interest in the US images having lower intensity values as compare to other region, foreground markers are located at regional minima and background markers are located at regional maxima. To calculate these markers many methods are discussed in the literature, such as

gray level thresholding, linear filtering, non linear filtering and morphological operators [25, 26]. Further gradient magnitude image is modified by mapping these markers and minima imposition. Modified gradient magnitude image has regional minima at marked location. Eventually watershed transform is applied on this modified gradient magnitude image [21, 26]. This technique overcomes the problem of over segmentation up to some extent, but not completely. As shown in Figure 24 (b), Figure 25 (b), Figure 26 (b) and Figure 27 (b), tumor is located with some false positive regions (over segmentation). Here in this paper we introduced textural property and clustering based technique for segmentation and our results are compared with these techniques described in the literature.

### III. VECTOR QUANTIZATION

Vector Quantization (VQ) was formerly introduced as image compression technique and proved to be efficient. Many codebook generation and quantization algorithms were developed and tested for compression [27, 28, 29, 30]. Nowadays it has shown its value in other applications as well, such as pattern recognition and face detection [31, 32], image segmentation [33], tumor demarcation in MRI and Mammogram images [34, 35], content based image retrieval [36] etc. In this paper, this method have been used as part of algorithm to make clusters and delineate the area of interest in ultrasound images of the breast.

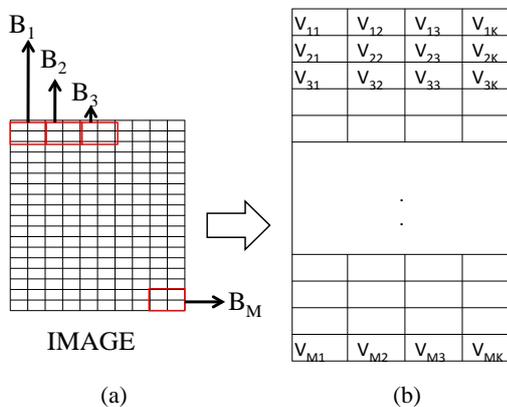


Figure.1 (a) Original image, divided horizontally into M nonoverlapping blocks shown by red boxes. i.e. B<sub>1</sub>, B<sub>2</sub>, ..., B<sub>M</sub>. (b) Training set of dimension K = 4, generated by storing gray levels in the vectors.

As shown in Figure.1, a two dimensional image I(X, Y) is converted into K dimensional vector space of size M,  $V = \{V_1, V_2, V_3, \dots, V_M\}$  (training set). VQ is used as a mapping function to convert this K dimensional vector space to finite set  $CB = \{C_1, C_2, C_3, C_4, \dots, C_N\}$ . CB is a codebook of size N and each code vector from C<sub>1</sub> to C<sub>N</sub> represents the specific set of vectors of the entire training set of dimensions K and size M. The codebook size is much smaller than size of the training set and it can represent entire training set.

Here, in this paper, the work has been done in spatial domain and size of the codebook is limited to only eight codevectors, which are further used to forms eight clusters. As discussed in the section VI A, B and C, LBG, KPE and KEVR codebook generation algorithms are used for clustering. All cluster obtained by using these algorithms represents the differet regions of the images and each region has different codevector. Therefore codebook can represents entire image. This phenomenon is very important information and further used in segmentation process.

### IV. PROPOSED ALGORITHM

In this paper, a new method has been proposed for the segmentation of the ultrasound images. This method is fundamentally based on principle of random distribution of the gray levels and texture of the image. Since texture is the dominant characteristics of the ultrasound image, proposed method focuses on homogeneous and heterogeneous texture regions and tries to separate them using probability of distribution of gray levels. This method has three major steps. In first step we obtained the probability image from the original ultrasound image and also get the histogram equalized image of the probability image. Further these two images are handled separately and used as input to the next step. In second step VQ based clustering techniques as discussed in the section VI, has been used separately on these images and set of clusters are obtained. Here, we restrict the set up to eight clusters and further they merged sequentially and one cluster is selected from this merged clusters set for post processing, which is last step of algorithm. Figure.2 shows the flow of proposed segmentation algorithm and steps are explain sequentially in the subsequent sections.

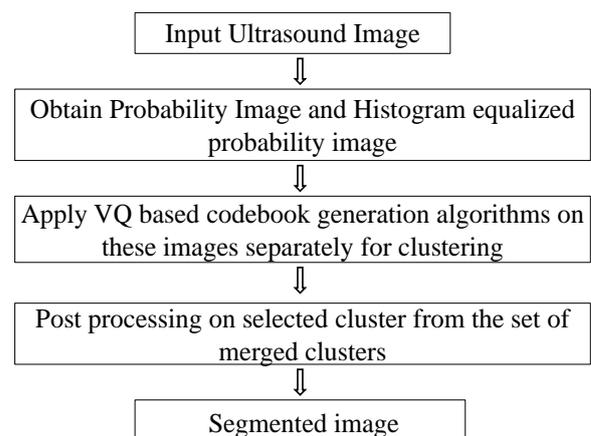


Figure.2 Architecture of proposed algorithm

### V. PROBABILITY IMAGES

Texture is the most dominant characteristics in ultrasound images, therefore acquisition of texture patterns provides useful information required in understanding and detection of region of interest. Since

image is the distribution of random values, statistical properties such as variance, probability, entropy reveals clear-cut understanding of homogeneous and heterogeneous regions present in the image. Here in this paper, probability of the pixel in image is measured as vital information in analyzing the texture patterns and extraction of the features. Probability of every pixel is calculated for entire image using equation 1.

$$P(i) = \frac{X_i}{M \times N} \quad (1)$$

Where,  $P(i)$  is probability of the  $i^{\text{th}}$  gray level,  $X_i$  is number of occurrence of  $i^{\text{th}}$  gray level in the image,  $M$  and  $N$  are number of rows and column of the image respectively.

After calculation of probability of each pixel, respective pixel value is replaced by its probability value in the image and probability image is obtained from original ultrasound image shown in Figure.6(b). The values of probability images are too small and some texture patterns may not clearly visible, therefore histogram equalization method has been used over these probability images. New set of equalized histogram probability images are obtained from the probability images shown in Figure.6(c). In the next step probability and its histogram equalized images are handled separately by clustering algorithms and obtained different set of results.

## VI. CODEBOOK GENERATION ALGORITHMS

### A. Linde Buzo Gray (LBG) Algorithm [37, 38, 39]

This VQ based algorithm is mainly used to divide images into clusters and further these clusters are use in segmentation process. In this algorithm, average of entire training set is calculated, this average vector is considered as centroid (first codevector). As shown in the Figure 1, a certain constant error is added to (+1) and subtracted from (-1) first codevector and two new codevectors  $C_1$  and  $C_2$  are obtained respectively. Euclidean distance between vectors of entire training set with respect to  $C_1$  and  $C_2$  is calculated and training set is divided into two clusters based on the closest of  $C_1$  or  $C_2$ . This process is repeated till we get desired number of clusters (eight clusters). For two dimensional vectors, as shown in Figure.3, this technique gives inefficient clustering, since it generates elongated clusters which has constant angle of  $45^\circ$  with x-axis.

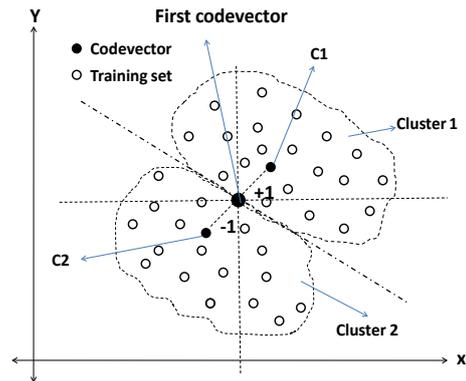


Figure.3 Clustering using LBG for two dimensional vector,  $K=2$ .

### B. Kekre's Proportionate Error (KPE) Vector

In this technique, proportionate error is calculated [40]. Thus unlike LBG, this proportionate error is added to and subtracted from first codevector which is generated by obtaining average of entire training set. Rest of the procedure is same as that of LBG. Care is taken to keep codevector  $C_1$  and  $C_2$  within the limit of vector space while adding proportionate error. In this clustering process clusters are not elongated since angles of codevectors are not constant with the x-axis like LBG, as shown in Figure.4. These clusters are formed in different directions, thus they contain vectors which have close relation and it gives efficient clustering than LBG.

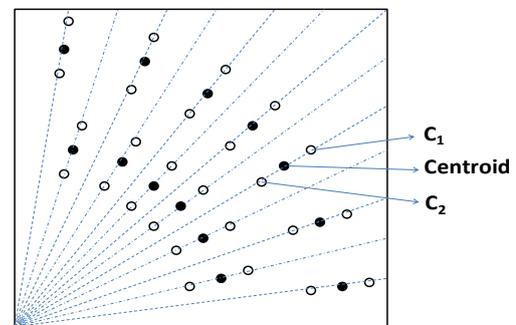


Figure.4 Orientation of the lines joining two codevectors  $C_1$  and  $C_2$  after addition of proportionate error to the centroid

### C. Kekre's Error Vector Rotation (KEVR) [41, 42]

This algorithm is based on generation of matrix of error vectors. As shown in Figure.5, error vector matrix  $E$  is generated for dimension  $K$  and  $i^{\text{th}}$  row of the error matrix is considered as error vector  $e_i$ . Binary sequence of numbers from 0 to  $K-1$  is acquired and each error vector is generated, replacing 0 by 1, 1 by -1. First error vector is added to and subtracted from first code vector and codevector  $C_1$  and  $C_2$  are obtained respectively. This process is repeated till desired number of clusters has been formed with each entry of the error matrix. With the addition and subtraction of error vector the cluster formation is rotated in different direction and elongated clusters are not formed, thus cluster contains

more appropriate vectors. Cluster formation using this method is efficient than LBG and KPE.

$$\begin{matrix}
 e_1 & 1 & 1 & 1 & 1 & \dots & 1 & 1 & 1 & 1 \\
 e_2 & 1 & 1 & 1 & 1 & \dots & 1 & 1 & 1 & -1 \\
 e_3 & 1 & 1 & 1 & 1 & \dots & 1 & 1 & -1 & 1 \\
 E = e_4 & = & 1 & 1 & 1 & 1 & \dots & 1 & 1 & -1 & -1 \\
 \cdot & & & & & & & & & & \\
 \cdot & & & & & & & & & & \\
 e_k & & & & & & & & & &
 \end{matrix}$$

Figure.5 Error Matrix generated for K dimensions [35]

VII. POST PROCESSING

Is the last step of algorithm, we tried to remove the unwanted regions from the selected image from the set of merged clusters. To remove the false positive regions from the cluster image, number of independent patches (regions) are identified and labeled them with different integer values, but all the pixels of same region has same integer value. Then number of pixels is counted for each region and region with largest number of pixel is selected as foreground region (tumor) and other regions are merged with background. Furthermore we used morphological operators for closing the holes. We use disk structuring element with radius of 3 pixels.

VIII. EXPERIMENTAL RESULTS AND ANALYSIS

To validate the performance of the proposed segmentation algorithm, we tested it on set of real US images. This method is executed using MATLAB 7.0 on Intel Core2 Duo 2.10GHz processor with 2 GB RAM. Here, in this paper results of four different images are shown, in which, Figure 6(a), 16(a) has malignant mass and Figure 12(a), 20(a) has benign mass. In the second stage of proposed algorithm we used three different clustering techniques; therefore results of different images are shown accordingly. Eight cluster images and its sequentially merged cluster images are shown only for one image shown in Figure 6(b). Segmentation results along with the selected merged cluster image and image after post processing are directly shown for others. Best results of each clustering algorithm over original US image, probability image and histogram equalized probability image are shown by red box. Figure.9, Figure.10 and Figure.11 shows results for LBG, KPE and KEVR respectively for original US Image 1. Similarly all the results are shown in same order for other three original US images. Result analysis and opinion of expert radiologist is shown in Table.1.

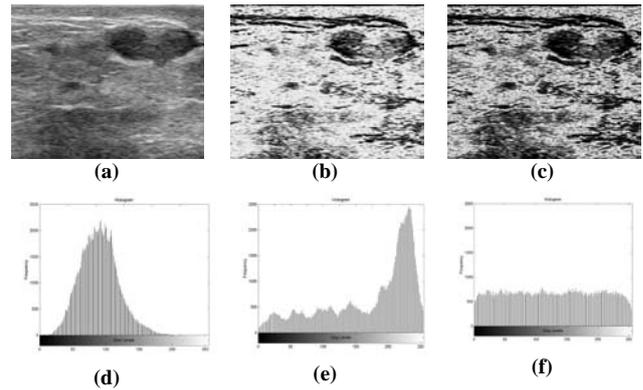


Figure.6 Ultrasound Image 1 with malignant tumor

(a) Original Image (b) Probability image obtained from original image (c) Image obtained after Histogram equalization on Probability image. (d) Histogram of original image (e) Histogram of probability image (f) Histogram after equalization

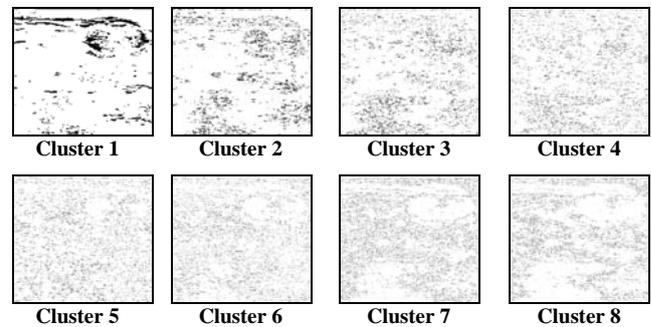


Figure.7 Eight cluster images obtained from Probability image shown in Figure. 6 (b) using LBG algorithm for dimension K= 4.

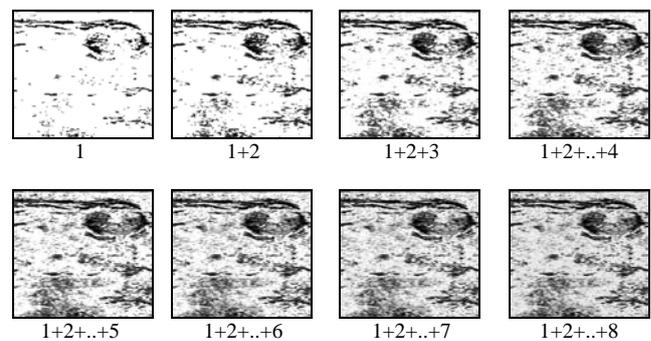
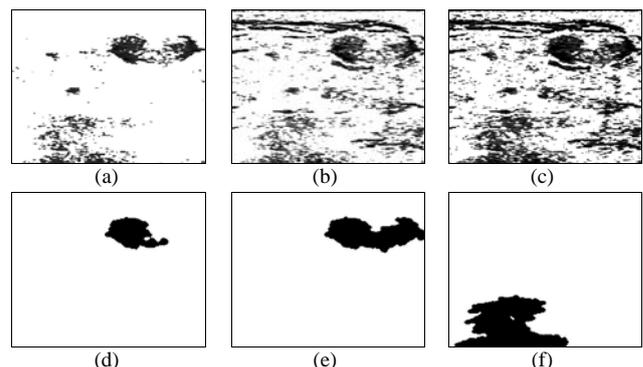


Figure.8 Set of eight sequentially merged cluster images obtained from clusters shown in Figure. 7



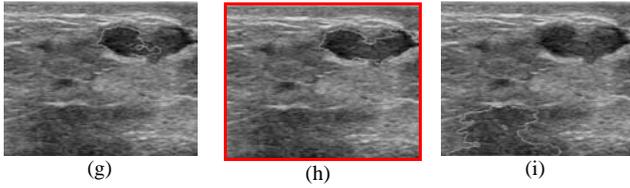


Figure.9 Segmentation results by LBG codebook generation algorithm.

(a), (b) and (c) Selected cluster images of Original US, Probability and Histogram equalized probability images respectively. (d), (e) and (f) Images obtained after post processing on (a), (b) and (c) respectively. (g), (h) and (i) Superimposed image.

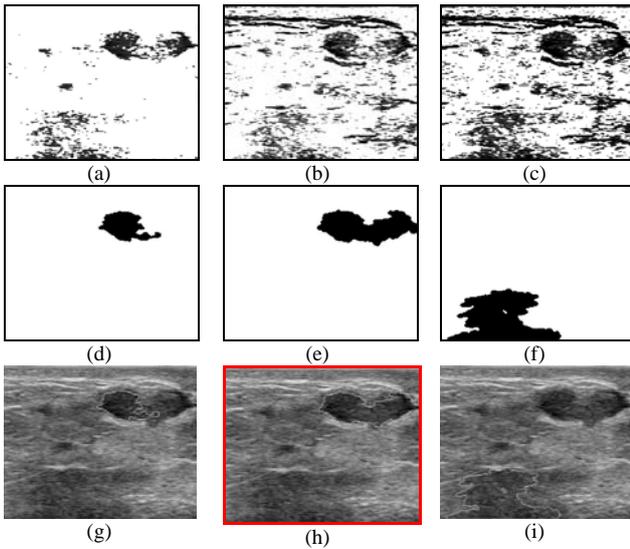


Figure.10 Segmentation results by KPE codebook generation algorithm

(a), (b) and (c) Selected cluster images of Original US, Probability and Histogram equalized probability images respectively. (d), (e) and (f) Images obtained after post processing on (a), (b) and (c) respectively. (g), (h) and (i) Superimposed image.

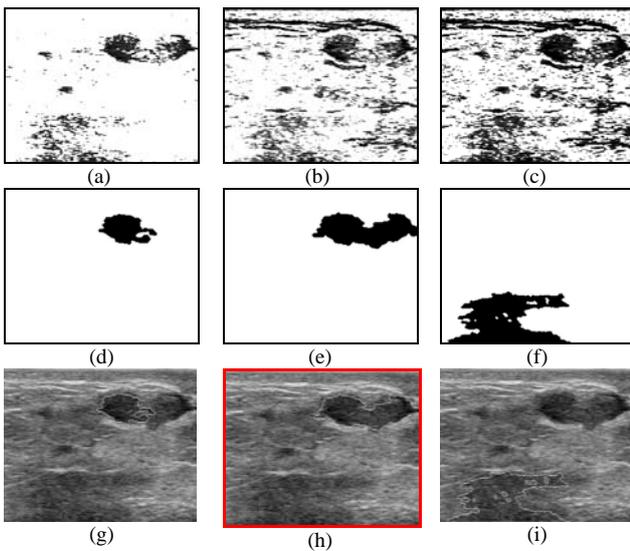


Figure.11 Segmentation results by KEVR codebook generation algorithm

(a), (b) and (c) Selected cluster images of Original US, Probability and Histogram equalized probability images respectively. (d), (e) and (f) Images obtained after post processing on (a), (b) and (c) respectively. (g), (h) and (i) Superimposed image.

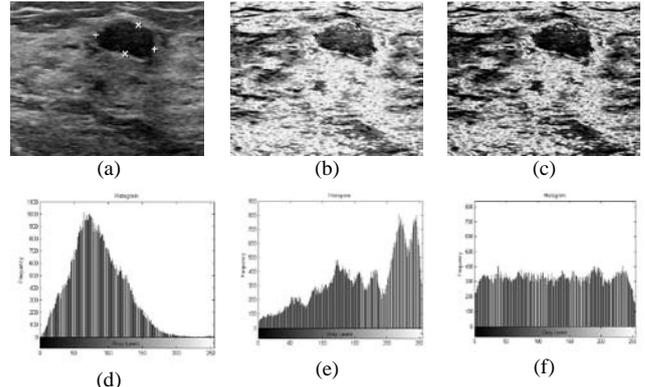


Figure.12 Ultrasound Image 2 with benign tumor

(a) Original Image (b) Probability image obtained from original image (c) Image obtained after Histogram equalization on probability image. (d) Histogram of original image (e) Histogram of probability image (f) Histogram after equalization

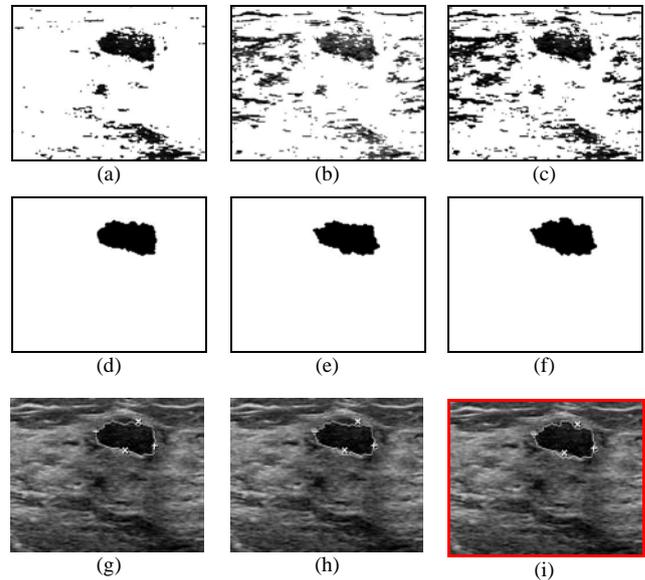
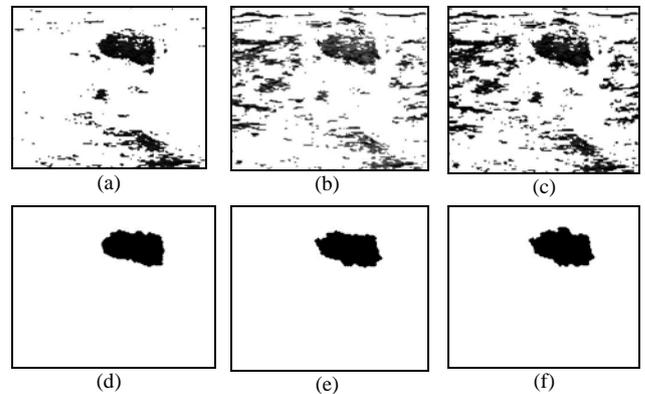


Figure.13 Segmentation results by LBG codebook generation algorithm.

(a), (b) and (c) Selected cluster images of Original US, Probability and Histogram equalized probability images respectively. (d), (e) and (f) Images obtained after post processing on (a), (b) and (c) respectively. (g), (h) and (i) Superimposed image.



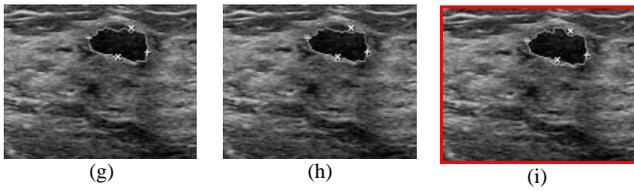


Figure.14 Segmentation results by KPE codebook generation algorithm.

(a), (b) and (c) Selected cluster images of Original US, Probability and Histogram equalized probability images respectively. (d), (e) and (f) Images obtained after post processing on (a), (b) and (c) respectively. (g), (h) and (i) Superimposed image.

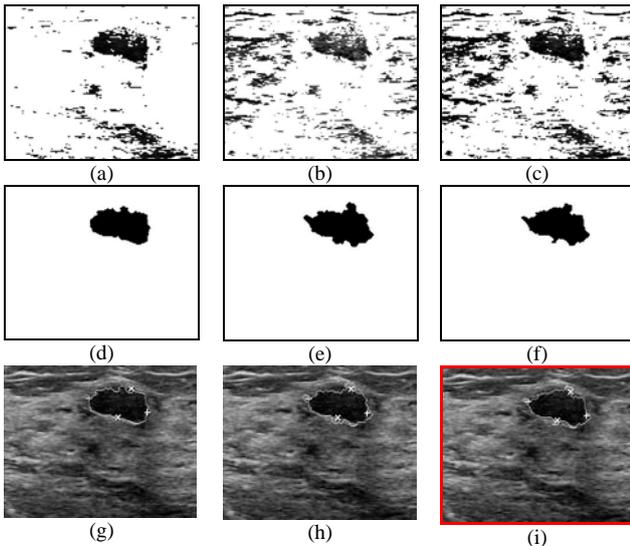


Figure.15 Segmentation results by KEVR codebook generation algorithm.

(a), (b) and (c) Selected cluster images of Original US, Probability and Histogram equalized probability images respectively. (d), (e) and (f) Images obtained after post processing on (a), (b) and (c) respectively. (g), (h) and (i) Superimposed image.

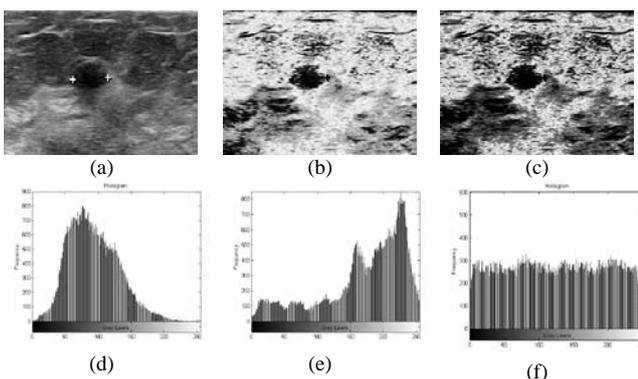


Figure.16 Ultrasound Image 3 with malignant tumor

(a) Original Image (b) Probability image obtained from original image (c) Image obtained after histogram equalization on probability image. (d) Histogram of original image (e) Histogram of probability image (f) Histogram after equalization

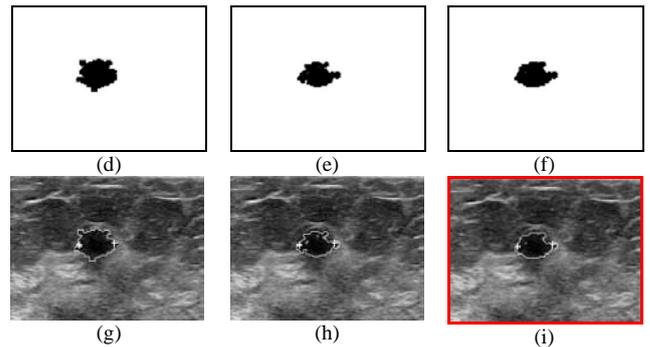
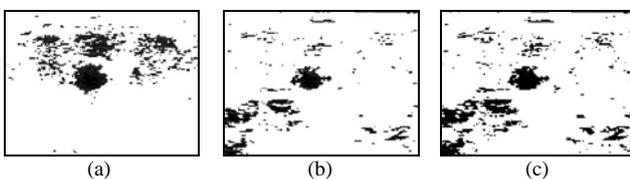


Figure.17 Segmentation results by LBG codebook generation algorithm.

(a), (b) and (c) Selected cluster images of Original US, Probability and Histogram equalized probability images respectively. (d), (e) and (f) Images obtained after post processing on (a), (b) and (c) respectively. (g), (h) and (i) Superimposed image.

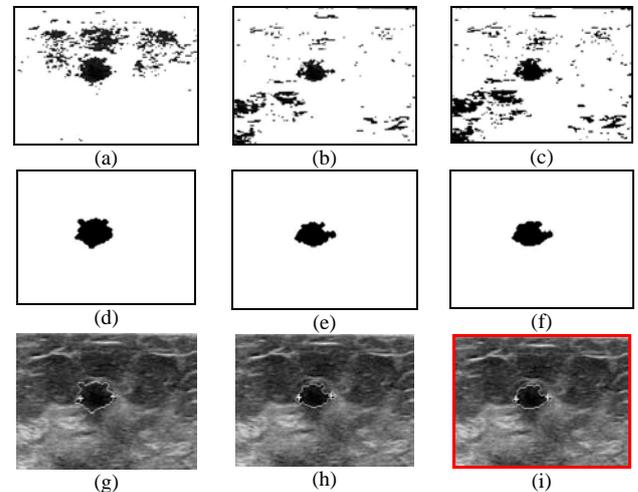


Figure.18 Segmentation results by KPE codebook generation algorithm.

(a), (b) and (c) Selected cluster images of Original US, Probability and Histogram equalized probability images respectively. (d), (e) and (f) Images obtained after post processing on (a), (b) and (c) respectively. (g), (h) and (i) Superimposed image.

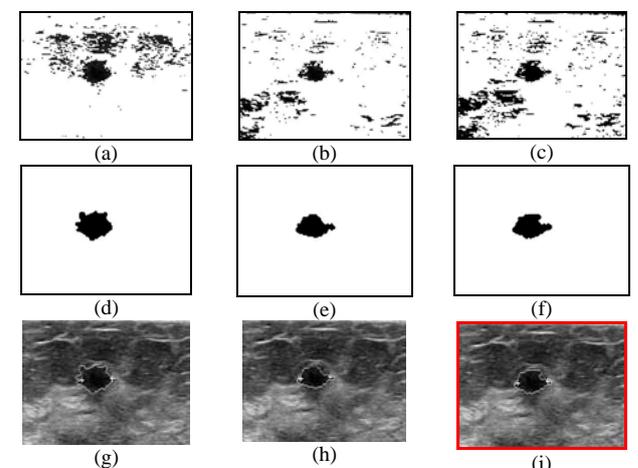


Figure.19 Segmentation results by KEVR codebook generation algorithm.

(a), (b) and (c) Selected cluster images of Original US, Probability and Histogram equalized probability images respectively. (d), (e) and (f) Images obtained after post processing on (a), (b) and (c) respectively. (g), (h) and (i) Superimposed image.

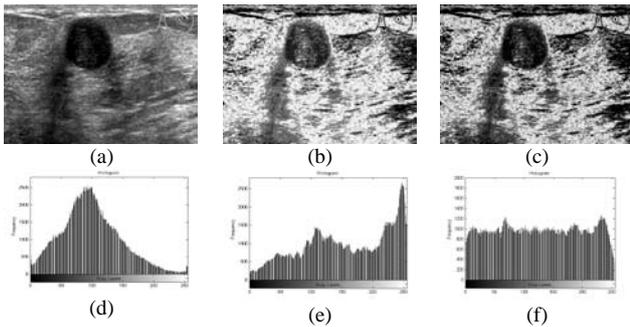


Figure.20 Ultrasound Image 4 with benign tumor

(a) Original image (b) Probability image obtained from original image  
 (c) Image obtained after histogram equalization on probability image.  
 (d) Histogram of original image (e) Histogram of probability image (f) Histogram after equalization

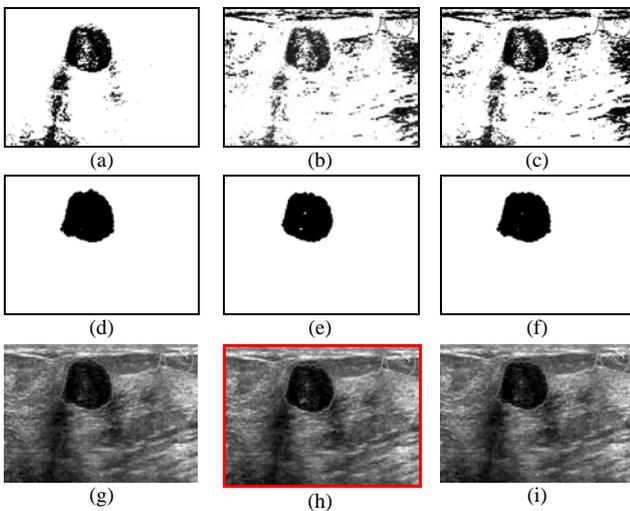


Figure.21 Segmentation results by LBG codebook generation algorithm.

(a), (b) and (c) Selected cluster images of Original US, Probability and Histogram equalized probability images respectively. (d), (e) and (f) Images obtained after post processing on (a), (b) and (c) respectively. (g), (h) and (i) Superimposed image.

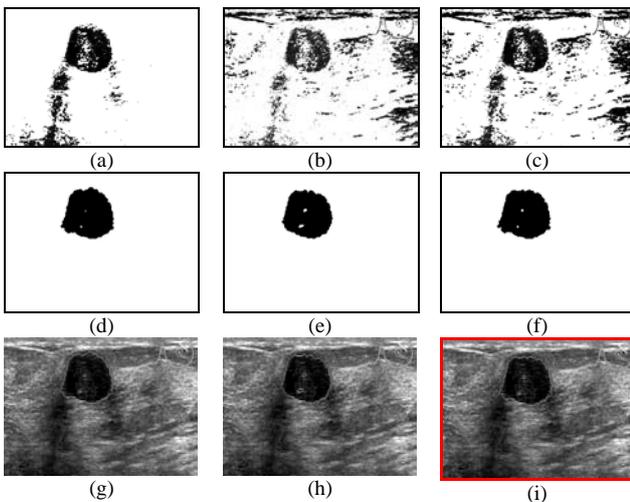


Figure.22 Segmentation results by KPE codebook generation algorithm.

(a), (b) and (c) Selected cluster images of Original US, Probability and Histogram equalized probability images respectively. (d), (e) and (f) Images obtained after post processing on (a), (b) and (c) respectively. (g), (h) and (i) Superimposed image.

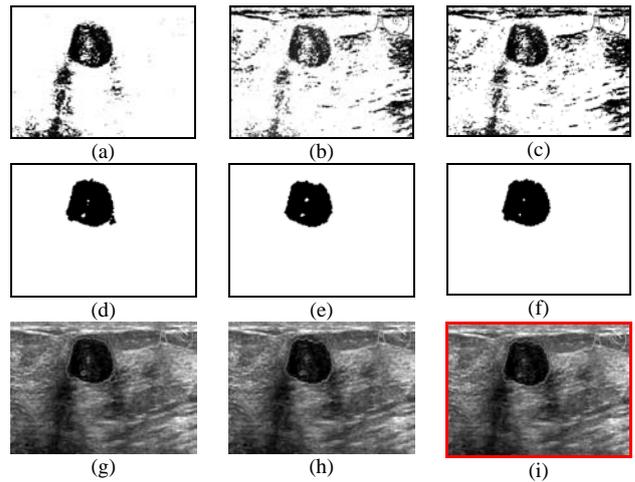


Figure.23 Segmentation results by KEVR codebook generation algorithm.

(a), (b) and (c) Selected cluster images of Original US, Probability and Histogram equalized probability images respectively. (d), (e) and (f) Images obtained after post processing on (a), (b) and (c) respectively. (g), (h) and (i) Superimposed image.

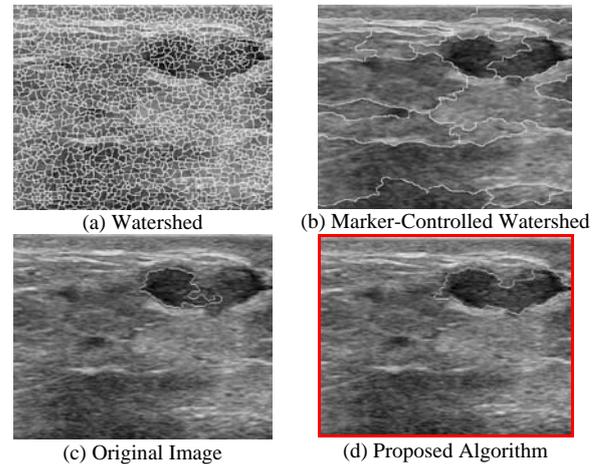


Figure.24 Result comparison of US Image 1 shown in Figure. 6

(a) Watershed transforms (b) Marker-Controlled Watershed transform (c) Clustering on original image (d) Our proposed method with probability image

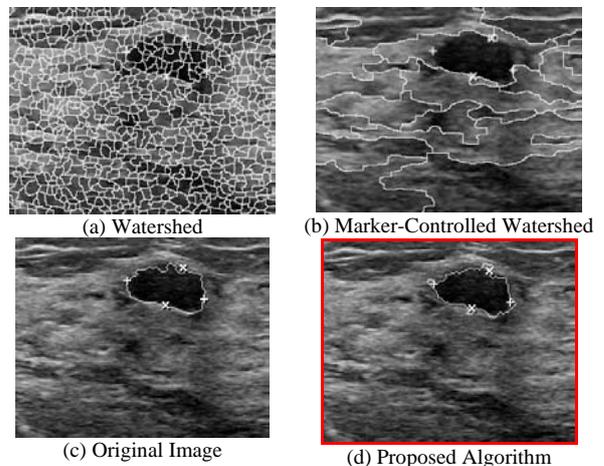


Figure.25 Result comparison of US Image 2 shown in Figure. 12

(a) Watershed transform on original image (b) Marker-Controlled Watershed transform on original image (c) Clustering on original image (d) Our proposed method with Histogram equalized probability image

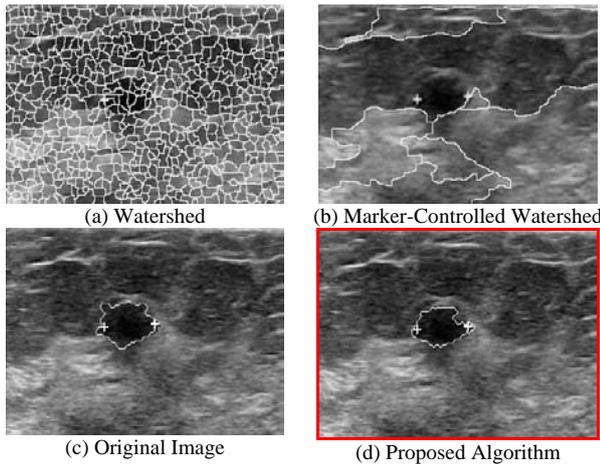


Figure.26 Result comparison of US Image 3 shown in Figure. 16

(a) Watershed transform on original image (b) Marker-Controlled Watershed transform on original image (c) Clustering on original image (d) Our proposed method with Histogram equalized probability image

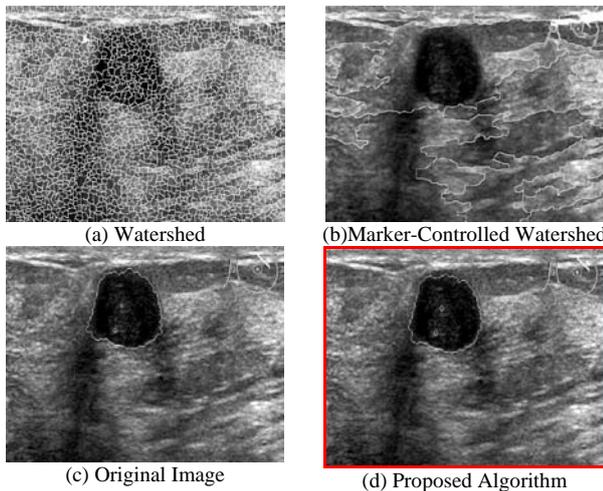


Figure.27 Result comparison of US Image 4 shown in Figure. 20

(a) Watershed transform on original image (b) Marker-Controlled Watershed transform on original image (c) Clustering on original image (d) Our proposed method with Histogram equalized probability image

TABLE.1 RESULT COMPARISION AND OPINION OF CONSULTANT RADIOLOGIEST

Images	Images obtained at first stage of algorithm	Tumor Delineation using				
		WT	MCWT	LBG	KPE	KEVR
Image 1 with Malignant tumor	Original	NA	NA	NA	NA	NA
	Probability			A	G	BEST
	Histogram equalized probability			NA	NA	NA
Image 2 with Benign Tumor	Original	NA	NA	A	A	A
	Probability			NA	NA	A
	Histogram equalized probability			G	G	BEST
Image 3 with	Original	NA	NA	NA	NA	NA
	Probability			NA	NA	NA

Malignant tumor	Histogram equalized probability			A	G	BEST
Image 4 with Benign tumor	Original	NA	NA	A	A	NA
	Probability			G	A	NA
	Histogram equalized probability			A	G	BEST
NA – Not acceptable, A – Acceptable, G – Good WT- Watershed Transform, MCWT- Marker-Controlled Watershed Transform						

IX. CONCLUSION

This paper, presents three stages segmentation algorithm for US breast images. In first stage it focused on statistical properties of image i.e. pixel’s probability. Due to certain artifacts such as speckle, attenuation, tissues of the tumor region may fall in the class of background tissue regions resultantly there could be occurrence of blurred boundary. Probability image makes blurred boundary clear by rearranging tissue classes and provides useful information such as tumor shape. This probability images are enhance further by histogram equalization. Further along with the post processing step, VQ based clustering algorithms are used to detect and delineate tumor from these probability and histogram equalized probability images. We used three different clustering algorithms and results of them are compared with each other.

LBG algorithm, use the constant error, therefore unidirectional elongated clusters are formed leading to inefficient clustering. To overcome this drawback, in KPE, proportionate error has been used to improve the formation of clusters but proportionate error for ultrasound image would have small magnitude, so results may be similar to LBG. In KEVR, this limitation is overcome by using rotation of error vector and produced clusters with new orientation every time. Proposed algorithm is tested on different US images, for first image KEVR with probability image gives best result. On the contrary for other images KEVR with histogram equalized probability gives better results amongst all combinations. Therefore results obtained using our proposed algorithm is image specific and depends on characteristics of original US image. We also tested Watershed and Marker-Controlled Watershed algorithm on the same images, but these results are not accepted by the expert radiologists and moreover compared with best results of our proposed method. All the results are validated by the consultant radiologists through visual inspection and their opinion is provided in Table.1. According to opinion of consultant radiologist KEVR give best results amongst all clustering algorithms.

ACKNOWLEDGMENT

The authors would like to thank Dr. Wrushali More, and Dr. Mushtak Chaudhari for valuable inputs to

understand the nature, shape of the tumor and their opinion to validate the results.

## REFERENCES

- [1] American Cancer Society, "Global Cancer Facts & Figures 2<sup>nd</sup> Edition, Atlanta: American Cancer Society; 2011.
- [2] Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM, "GLOBOCAN 2008: Cancer Incidence, Mortality and Prevalence Worldwide IARC", CancerBase No. 10, Lyon, France: International Agency for Research on Cancer, 2010.
- [3] Thomas L Szabo, "Diagnostic Ultrasound Imaging: Inside out", Elsevier Inc., pp 19-23, Sept 2004.
- [4] K. J. W. Taylor, C. Merritt, C. Piccoli, R. Schmidt, G. Rouse, B. Fornage, E. Rubin, D. Georgian-Smith, F. Winsberg, B. Goldberg, and E. Mendelson, "Ultrasound as a complement to mammography and breast examination to characterize breast masses," *Ultrasound Med. Biol.*, vol. 28, pp. 19–26, Jan. 2002.
- [5] Stavros, A., Thickman, D., Rapp, C., Dennis, M., Parker, S., and Sisney, G., "Solid breast nodules: Use of sonography to distinguish between benign and malignant lesions," *Radiology* 196, 122-134, 1995.
- [6] Bosch, A., Kessels, A., Beets, G., Vrancken, K., Borstlap, A., von Meyenfeldt, M., and van Engelshoven, J., "Interexamination variation of whole breast ultrasound", *British Journal of Radiology* 76(905), 328-331 May 2003.
- [7] Helmut Madjar, "Role of Breast Ultrasound for the Detection and Differentiation of Breast Lesions", *Breast Care*, 5: pp 109–114, April 2010.
- [8] Guita Rahbar, Angela C. Sai, Gail C Hansen, Jeffrey S. Prince, "Benign Versus Malignant Solid Breast Masses: US Differentiation", 213: 889-894, *Radiology*, December 1999.
- [9] K. Horsch, M.L. Giger, L.A. Venta, C.J. Vyborny, "Computerized diagnosis of breast lesions on ultrasound", *Medical Physics*, vol. 29, no.2, pp 157–164, Feb 2002.
- [10] J. Alison Noble, Djamel Boukerroui, "Ultrasound Image Segmentation: A Survey", *IEEE Transactions on Medical Imaging*, Vol. 25, No. 8, pp 987-1010, Aug 2006.
- [11] Guofang Xiao, Michael Brady, J. Alison Noble, Yongyue Zhang, "Segmentation of Ultrasound B-Mode Images With Intensity Inhomogeneity Correction, *IEEE TRANSACTIONS ON MEDICAL IMAGING*, VOL. 21, NO. 1, pp 48-57, JANUARY 2002.
- [12] D. Boukerroui, O. Basset, A. Noble, and A. Baskurt, Gérard Gimenez, "A Multiparametric and Multiresolution Segmentation Algorithm of 3-D Ultrasonic Data", *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, Vol 48 , No1 pp 64-77, 2002.
- [13] D. Boukerroui, O. Basset, A. Noble, and A. Baskurt, "Segmentation of ultrasound images—multiresolution 2D and 3D algorithm based on global and local statistics", *Pattern Recognition Letters Vol 24* , pp 779–790, 2003, Elsevier, Science Direct, 2003.
- [14] Y.L. Huang, D.R. Chen, "Watershed segmentation for breast tumor in 2-D sonography", *Ultrasound Med. Biol. Volume 30, Issue 5*, pp 625-632, May 2004.
- [15] Yanni Su, Yuanyuan Wang, Jing Jiao, Yi Guo, "Automatic Detection and Classification of Breast Tumours in Ultrasonic Images Using Texture and Morphological Features", Volume 5, pp 26-37, *The Open Medical Informatics Journal*, 2011.
- [16] Dar-Ren Chen, Yu-Len Huang, Sheng-Hsiung Lin "Computer-aided diagnosis with textural features for breast lesions in sonograms" *Computerized Medical Imaging and Graphic*, Elsevier, vol. 35, pp- 220–226, 2011.
- [17] D. R. Chen, R. F. Chang, W. J. Kuo, M. C. Chen, and Y. L. Huang, "Diagnosis of breast tumors with sonographic texture analysis using wavelet transform and neural networks," *Ultrasound Med. Biol.*, vol. 28, no. 10, pp. 1301–1310, Oct. 2002.
- [18] Bo Liu, H.D.Cheng, JianhuaHuang, JiaweiTian, XianglongTang, JiafengLiu, "Probability density difference-based active contour for ultrasound image segmentation", 43, 2028–2042, *Pattern Recognition, Science Direct, Elsevier*, 2010.
- [19] Jin-Hua Yu, Yuan-Yuan Wang, Ping Chen, Hui-Ying Xu, "Two-dimensional Fuzzy Clustering for Ultrasound Image Segmentation", published in the proceeding of IEEE International Conference on Bioinformatics and Biomedical Engineering, pp 599-603, 1-4244-1120-3, July 2007.
- [20] Chang Wen Chen, "Image Segmentation via Adaptive k-Mean Clustering and Knowledge-Based Morphological Operations with Biomedical Applications", *IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 7, NO. 12, PP 1673-1683, DECEMBER 1998*.
- [21] Luc Vincent, Pierre Soille, "Watershed in Digital Space: An Efficient Algorithm Based on Immersion Simulations", *IEEE Transaction on Pattern Analysis and Machine Intelligence*, pp 583-598, Vol 13, No 6, 1991.
- [22] Chung-Ming-Cheng, Henry Horning-Shing-Lu, Bin-Shin Su, "Cell-Based Region Competition for Ultrasound Image Segmentation", *Journal of Medical and Biomedical Engineering*, pp 59-66, 22(2), 2002.
- [23] Amr R. Abdel-Dayem, Mahmoud R. EI-Sakka1, Aaron Fenster2, "Watershed Segmentation for Carotid Artery Ultrasound Images", 3rd ACS /IEEE International Conference, 2005.
- [24] Deka B, Ghosh D, "Ultrasound Image Segmentation using Watershed and Region Merging", *IET International Conference on Visual Information Engineering, IEEE xplore*, pp 110-115, 2006.

- [25] W. Gomez, L. Leija, W.C.A.Pereira, A.F.C. Infantosi, "Segmentation of Breast Nodules on Ultrasonographic Images Based On Marker-Controlled Watershed Transform", *Computation Systems*, Vol 14, No 2, 2010.
- [26] Samual H. Lewis, Aijuan Dong, "Detection of Breast Tumor Candidates using Marker-Controlled Watershed and Morphological Analysis", *Southwest Symposium (SSIAI)*, pp 1-4, IEEE International Conference, 2012.
- [27] R. M. Gray, "Vector quantization", *IEEE ASSP Magazine.*, pp. 4-29, Apr.1984.
- [28] Pamela C. Cosman, Karen L. Oehler, Eve A. Riskin, and Robert M. Gray, "Using Vector Quantization for Image Processing", *Proceedings of the IEEE*, pp- 1326-1341, Vol. 81, No. 9, September 1993.
- [29] W. H. Equitz, "A New Vector Quantization Clustering Algorithm," *IEEE Trans. on Acoustics, Speech, Signal Proc.*, pp 1568-1575. Vol-37, No-10, Oct-1989.
- [30] Huang, C. M., Harris R.W., "A comparison of several vector quantization codebook generation approaches", *IEEE Transactions on Image Processing*, pp 108 – 112, Vol-2, No-1, January 1993.
- [31] Qiu Chen, Kotani, K., Feifei Lee, Ohmi, T., "VQ-based face recognition algorithm using code pattern classification and Self-Organizing Maps", *9th International Conference on Signal Processing*, pp 2059 – 2064, October 2008.
- [32] C. Garcia and G. Tziritas, "Face detection using quantized skin color regions merging and wavelet packet analysis," *IEEE Trans. Multimedia*, vol. 1, no. 3, pp. 264–277, Sep. 1999.
- [33] H. B. Kekre, Tanuja K. Sarode, Bhakti Raul, "Color Image Segmentation using Kekre's Algorithm for Vector Quantization *International Journal of Computer Science (IJCS)*, Vol. 3, No. 4, pp. 287-292, Fall 2008. Available at: <http://www.waset.org/ijcs>.
- [34] H. B. Kekre, Tanuja K. Sarode, Saylee Gharge, "Detection and Demarcation of Tumor using Vector Quantization in MRI images", *International Journal of Engineering Science and Technology*, Vol.1, Number (2), pp.: 59-66, 2009. Available online at: <http://arxiv.org/ftp/arxiv/papers/1001/1001.4189.pdf>.
- [35] H. B. Kekre, Dr.Tanuja Sarode, Ms.Saylee Gharge, Ms.Kavita Raut, "Detection of Cancer Using Vector Quantization for Segmentation", Volume 4, No. 9, *International Journal of Computer Applications* (0975 – 8887), August 2010.
- [36] H.B.Kekre, Tanuja K. Sarode, Sudeep D. Thepade, "Color Texture Feature based Image Retrieval using DCT applied on Kekre's Median Codebook", *International Journal on Imaging (IJI)*, Available online at [www.ceser.res.in/iji.html](http://www.ceser.res.in/iji.html).
- [37] Pamela C. Cosman, Karen L. Oehler, Eve A. Riskin, and Robert M. Gray, "Using Vector Quantization for Image Processing", *Proceedings of the IEEE*, pp- 1326-1341, Vol. 81, No. 9, September 1993.
- [38] R. M. Gray, "Vector quantization", *IEEE ASSP Magazine.*, pp. 4-29, Apr.1984.
- [39] Yoseph Linde, Andres Buzo, Robert M.Gray, "An Algorithm for Vector Quantizer Design", *IEEE Transaction On Communication*, pp 84-95, Vol. Com-28, No. 1, January 1980.
- [40] H.B.Kekre, Tanuja K. Sarode, "Two-level Vector Quantization Method for Codebook Generation using Kekre's Proportionate Error Algorithm", *International Journal of Image Processing*, Volume (4): Issue (1).
- [41] H. B. Kekre, Tanuja K. Sarode, "An Efficient Fast Algorithm to Generate Codebook for Vector Quantization," *First International Conference on Emerging Trends in Engineering and Technology, ICETET-2008*, held at Rasoni College of Engineering, Nagpur, India, 16-18 July 2008, Available at online IEEE Xplore.
- [42] H. B. Kekre, Tanuja K. Sarode, "New Clustering Algorithm for Vector Quantization using Rotation of Error Vector", *(IJCSIS) International Journal of Computer Science and Information Security*, Vol. 7, No. 3, 2010.



**Dr. H. B. Kekre** has received B.E. (Hons.) in Telecomm. Engineering. from Jabalpur University in 1958, M.Tech (Industrial Electronics) from IIT Bombay in 1960, M.S.Engg. (Electrical Engg.) from University of Ottawa in 1965 and Ph.D. (System Identification) from

IIT Bombay in 1970 He has worked as Faculty of Electrical Engg. and then HOD Computer Science and Engg. at IIT Bombay. For 13 years he was working as a professor and head in the Department of Computer Engg. at Thadomal Shahani Engineering. College, Mumbai. Now he is Senior Professor at MPSTME, SVKM's NMIMS University. He has guided 17 Ph.Ds, more than 100 M.E./M.Tech and several B.E./ B.Tech projects. His areas of interest are Digital Signal processing, Image Processing and Computer Networking. He has more than 450 papers in National / International Conferences and Journals to his credit. He was Senior Member of IEEE. Presently He is Fellow of IETE and Life Member of ISTE. 13 Research Papers published under his guidance have received best paper awards. Recently 5 research scholars have been conferred Ph. D. by NMIMS University. Currently 07 research scholars are pursuing Ph.D. program under his guidance.



**Mr. Pravin Shrinath** has received B.E. (Computer science and Engineering) degree from Amravati University in 2000. He has done Masters in computer Engineering in 2008. Currently pursuing Ph.D. from Mukesh Patel School of Technology

Management & Engineering, NMIMS University, Vile Parle (w), Mumbai. He has more than 10 years of teaching experience and currently working as Associate Professor in Computer Engineering Department, MPSTME.