

Tumor Demarcation of Breast US Images using VQ Based Clustering Algorithms on Probability and its Histogram Equalized Images

H. B Kekre and Pravin Shrinath

*Department of Computer Engineering, MPSTME, NMIMS University,
Mumbai, India*

hbkekre@yahoo.com, pravin.srinath@nmims.edu

Abstract

Breast cancer is the leading cause of death among the women worldwide. Early detection and proper treatment may reduce the mortality rate and ultrasound imaging is used as a complimentary with the mammography to detect and diagnose the tumor. During the inspection of thick and dense breast tissue ultrasound imaging gives better tissue characterization than mammography. Detection and demarcation of tumor is acquired through segmentation process, but it is challenging due to some inherent artifacts (speckle, attenuation) of US image. Here in this paper, new segmentation algorithm has been proposed with three stages. It has also discarded the overhead of preprocessing (image enhancement, and noise removal techniques) used by traditional and other algorithms discussed in the literature. In the first stage of the proposed algorithm, probability image and its histogram equalized image are obtained from the original US image. In the second stage, clustering process is implemented on probability and histogram equalized image separately using KMCG, KFCG and augmented KMCG/KFCG codebook generation algorithms. Further, these clusters are merged sequentially one-by-one. In the third stage, post processing has been implemented on the selected merged cluster to obtain resultant segmented image. Complexity analysis is done for all codebook generation algorithms used for clustering and the results are compared. KFCG is found to be the fastest algorithm amongst all. Eventually in consultation with the expert radiologist all segmentation results are compared with each other and best results are displayed with red border.

Keywords: *Vector Quantization, Probability Image, Codebook, Merged Clusters, Histogram Equalized Image*

1. Introduction

Among the women worldwide breast cancer is the frequently diagnosed disease and leading cause of cancer death. In the developed countries, mortality rate have been decreasing from the past few years. This reduction in the mortality rate is due to early detection through improved imaging modalities such as mammography, ultrasound, MRI and subsequently with proper treatment given in Ferlay J. [1] and American Cancer Society [32]. From the past few years, it is observed that, B-mode ultrasound imaging becoming more popular because of its non invasive nature, low health risk and cost. In fact ultrasound imaging is complimentary to mammogram and commonly used modality to examine and diagnose breast cancer given in Thomas Szabo [2] and K. J. W. Taylor [3]. Mammography has limitations over the dense and thick breasts tissues, where common malignant region often visible as normal mass and

difficult to understand. Ultrasound imaging is useful to examine dense breast tissues effectively and detect the presence of small malignant nodules. Moreover it is routinely used for remedial diagnosis to overcome the limitations of mammography. Ultrasound imaging provides better tissue and boundary characterization and differentiation between malignant and benign masses than mammography given by Stavros [4], Bosch [5], and Helmut [6]. Malignant tumor has typical characteristics, such as irregular shape, ill defined boundary, micro lobulation and heterogeneous echo texture whereas benign tumor has round or oval shape with well defined boundaries and homogeneous echo texture. In many instances tumor region exhibits heterogeneous texture in both the cases, therefore it cannot be considered as clinically valuable information for identification. However shape and boundary characteristic of a tumor is vital information in the detection and classification of tumor discussed by Guita Rahbar [7].

Image segmentation commonly used to detect and quantify the shape of a region of interest. Therefore segmentation of US image is widely open field for the researchers. Accurate segmentation of US images usually provides clinically valuable information, such as shape and boundary characteristics, further it can also helps in treatment and planning for surgery, but the accuracy is strongly influenced by quality of US images given in K. Horsh [8]. In US modality, quality of image depends on the professionals and technology used. There are recent advances in transducer design, spatial resolution, portability, *etc.*, as a result quality of data has improved significantly but presence of some inherent artifacts makes segmentation process challenging. Speckle is common artifact found in US images and degrades the quality of information. Therefore most of the segmentation algorithms apply an additional speckle removal step, which is an overhead on segmentation process discussed by J. A. Noble [9]. The artifact such as attenuation, causes by the gradual loss in the intensity of the ultrasound waves, generates different intensity patterns around the tumor. Therefore blurred boundary and tissue intensity variation exists in the region of interest. This occurrence becomes the major concern to achieve accuracy in the automated intensity based segmentation discussed in Guofang Xiao [10]. Various methods are discussed in the literature which handles this issue with the help of multiple images of the same region (sequence of images) discussed in D. Boukerroui [11, 12]. But processing of multiple images together to achieve accurate segmentation is computationally inefficient. Some segmentation methods are based on texture feature, watershed transform given by Y.L. Huang [13], self organizing map neural network, etc, but these methods are very sensitive to noise. Watershed transform has a serious drawback of over segmentation. Therefore these algorithms are having an overhead of noise removal technique. Noise removal techniques such as Median filter, Gaussian filter and Wiener filter smear the margins of the regions of interest and may cause loss of image information discussed in Yanni Su [14], Dar-Ren Chen [15], and D. R. Chen [16]. Due to high sensitivity to the noise, traditional algorithms such as global thresholding, edge based segmentation algorithms, region growing and merging are not suitable for US image. Although an adaptive thresholding gives good results in some cases but selection of optimum block size and accurate threshold is major challenge. Moreover stopping condition required to be set manually to stop the iteration of the algorithm given by H. B. Kekre [17]. 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 given by J. A. Nobel [9], Bo Liu [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 and require preprocessed images, which is an

overhead on segmentation algorithm given in Jin-Hua Yu [33], and Chang Wen [19]. Here, in this paper, we are demarking the boundary of tumor in original US images (which is poor in quality and has high degree of noise) without using any image enhancement technique. We are proposing three stages in the algorithm for segmentation, in first stage probability and its histogram equalized images are obtained. Subsequently, KMCG, KFCG, augmented KMCG and augmented KFCG algorithms are used for clustering separately on these images. Further these clusters are merged sequentially and post processing is done at the last stage on selected merged cluster.

The other sections of this paper are organized as follows, in Section 2, training set formation from the image has been discussed with the help of Vector Quantization. In section 3, various stages of the proposed algorithm are explained with flow chart. Section 4 describes formation of probability and its histogram equalized image from the original image. Various codebook generation algorithm used for clustering are discussed in Section 5. Section 6, elaborate post processing techniques where we remove noise by using connected component labeling and subsequently morphological closing operation is done on the selected cluster image. In Section 7 complexity analysis and its comparison is discussed. Section 8 discusses the results followed by conclusion in Section 9.

2. Vector Quantization

Vector Quantization (VQ) was initially proposed and used as efficient image compression algorithm. Many codebook generation and quantization algorithms were developed and tested successfully to achieve good compression ratio discussed in R. M. Gray [20], Pamela C. [21], and Huang C. [22]. Currently researcher focusing on the usefulness of this technique in other image processing domains such as content based image retrieval by H. B. Kekre [23], pattern recognition and face identification by Qiu Chen [34], and C. Garcia [24], segmentation of MRI and Mammogram images By H. B. Kekre [25, 26], iris and finger print recognition etc. In this paper, we used vector quantization based codebook generation as an essential step in our proposed segmentation algorithm. This step divides the image into clusters and these clusters are used in subsequent steps to delineate the area of interest in ultrasound image of the breast.

In VQ based algorithms, image is converted into M vectors, *i.e.*, $V = \{V_1, V_2, V_3, \dots, V_M\}$. Further it uses 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 codevector from C_1 to C_N represents the specific set of training vectors. The codebook size is much smaller than size of the training set and used to represents entire training set. Here, in this paper, the work has been done in spatial domain and size of the codebook is limited to eight codevectors only, which are further used to forms eight clusters. As discussed in the section 5.1, 5.2 and 5.3, these codebook generation algorithms are used for clustering. Each cluster with its codevector represents separate region of the image. Therefore a single codebook can represent entire image. This phenomenon provides clinically vital information and further used to identify tumor in segmentation process.

As shown in Figure 1, image $I(X, Y)$ is converted in to training set (matrix) of size $M \times K$. Where image I is divided in to 2×2 non overlapping blocks (*i.e.*, B_1, B_2, \dots, B_M) shown by red boxes. These blocks are further changed to individual vectors of dimension 4 (*i.e.* V_{11}, V_{12}, V_{13} and V_{14} for Block B_1 and same process is repeated up to block B_M).

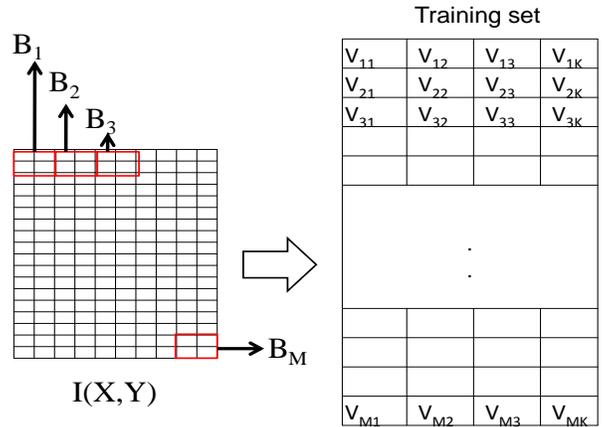


Figure 1. Conversion of Image I in to training set of size M with K dimension vectors (K=4)

3. Proposed method

In this paper, multistage algorithm has been proposed for detection and demarcation of tumor from the ultrasound images of the breast. A mixture of homogeneous and heterogeneous texture in the tumor and around the tumor makes segmentation process difficult. Therefore at first stage, our algorithm explores this phenomenon of random distribution of the gray levels and also handles the speckle. Here, we tried to acquire the information of heterogeneous and homogeneous regions using probability of individual gray levels. So probability image and its histogram equalized image are obtained from the original US image. Subsequently these two images are handled separately and used as input to the next stage. In second stage four VQ based clustering algorithms as discussed in the Section 5, are used on these images separately and sets of clusters are obtained. Further these clusters are merged sequentially and one cluster is selected for post processing, which is last step of the proposed algorithm. Figure 2 shows the flow of proposed segmentation algorithm and stages are explained sequentially in the subsequent sections.

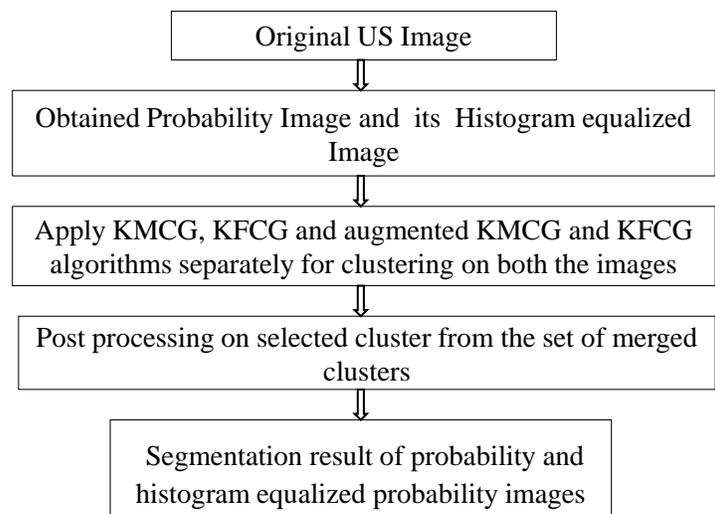


Figure 2. Flow of proposed algorithm

4. Probability Image

In US images, due to random distribution of the gray levels and enormous amount of speckle, coarse texture is the dominant characteristic and these texture patterns hold vital information about the images. This pattern exhibits three different characteristics, such as echo texture (coarse texture) in the tumor region, heterogeneous texture at boundary of the tumor and typically homogeneous texture at normal region. Therefore this variation in the patterns produced a challenging phenomenon to achieve accurate segmentation. To overcome this difficulty, here we proposed to focus on the statistical property in terms of probability of distribution of the gray levels in an image. To obtain the probability of a gray level we have calculated the number of occurrence of gray level in the image and divided it by total number of gray levels of the image, shown in equation 1. The probability of the gray levels contains vital information and is further used to differentiate between the region of interest and other regions including noise. Probability image reveals complete change in the representation of gray levels than the original US image without affecting the region of interest.

$$P(i) = \frac{Z_i}{X \times Y} \quad (1)$$

Where, $P(i)$ is probability of the i^{th} gray level, Z_i is number of occurrence of i^{th} gray level in the image, X and Y are number of rows and column of the image respectively.

After calculation of probability of a pixel, respective pixel value is replaced by its probability value. Eventually a probability image is obtained from original ultrasound image shown in Figure 3(ii). The values of probability images are too small and some texture patterns may not clearly visible, therefore histogram equalization method has been used over this probability image. Histogram equalized probability image is shown in Figure 3(iii). In the next step probability, its histogram equalized probability images are handled separately and used as input to the codebook generation algorithms for clustering.

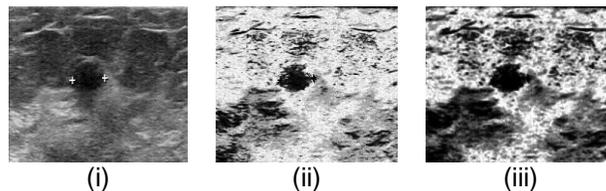


Figure 3. (i) Original US image with malignant tumor (ii) Probability image (iii) Histogram equalized probability image

5. Codebook Generation Algorithm

5.1. Kekre's Median Codebook Generation Algorithm (KMCG)

This VQ based codebook generation algorithm was proposed in data compression initially by H. B. Kekre [27, 28, 29]. Nowadays this algorithm has been extensively used in various applications, such as segmentation of mammogram images by H. B. Kekre [35], content based image retrieval, face recognition etc. In some application it gives better results than the traditional application specific algorithms. Here, in this paper, this iterative algorithm has been used in the second stage of the algorithm to make clusters from the training set. Initially, the input image is divided into M non-overlapping blocks of size 2×2 and these

blocks are further converted into vectors of dimension $1 \times K$, ($K=4$). These individual vectors are further clubbed together to form a training set of size M .

Let, V be the training set, *i.e.*, $V = \{V_1, V_2, V_3, \dots, V_M\}$, where V_1 to V_M are individual vectors of the training set. This entire training set (matrix) of dimension $M \times 4$ is considered as first cluster and become the input to the KMCG algorithm. Further in the clustering process, during the first iteration entire training set has been sorted with respect to first column (*i.e.*, first value of all the vectors) and obtained the median. The vector holding median is consider as first codevector, and then divides training set into two clusters with respect to this codevector. (*i.e.*, upper part of training set including codevector is considered as first cluster and lower part is considered as second cluster). In second iteration, these two clusters are sorted separately with respect to second column (*i.e.*, second value of all the vectors) of the training set and obtain the respective medians. Further, these two clusters are divided into four clusters by using their respective median values. Same procedure has been repeated till we obtained the eight clusters. After acquiring eight clusters, they are merged sequentially one-by-one and produced a new set of merged clusters.

5.2. Kekre's Fast Codebook Generation Algorithm (KFCG)

This algorithm is proposed by H. B. Kekre [27, 28, 36] to generate the codebook used in data compression. Euclidean Distance calculation is removed in KFCG, but it used in other VQ based algorithms such as LBG, KPE, and KEVR given by H. B. Kekre [30]. KFCG algorithm also eliminates sorting process, which is required in KMCG. Therefore this iterative algorithm is computationally efficient than the KMCG and other VQ based algorithms discussed in the literature by H. B. Kekre [31]. In this algorithm, similar to KMCG, training set of size $M \times K$ has been obtained from the image. During first iteration, the centroid is calculated by taking the average of first column of training set and considered as the first codevector. Using this codevector, entire training set is further divided into two clusters CL_1 and CL_2 with respect to first column as shown in the Figure 4(a). Cluster CL_1 is obtained by putting vectors which has lower value than the codevector and cluster CL_2 obtained with larger values than the codevector. In second iteration, repeat the same procedure separately on CL_1 and CL_2 with respect to second column and four clusters are obtained as shown in Figure 4 (b). Repeat this procedure further on these four clusters with respect to third column to divides the training sets into eight clusters. These clusters are displayed as eight cluster images by putting vector values at appropriate location and other vector values are considered as maximum gray level (*i.e.*, 255).

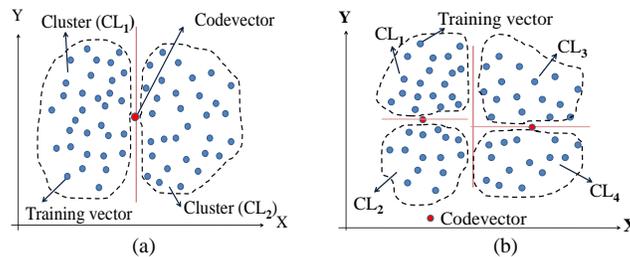


Figure 4. Clustering using KFCG algorithm for $K=2$: (a) cluster formation after first iteration (b) cluster formation after second iteration

5.3. Augmented KMCG and KFCG algorithms

Here, in this paper, augmented method for formation of vectors has been used on original KMCG and KFCG algorithms proposed by H. B. Kekre [37]. In this method, image I is divided into blocks of size 2×2 , addition of all gray levels of each of these blocks are done separately and stored at the first column in the respective vectors. Original gray levels of the blocks are stored at the subsequent columns, *i.e.*, from second column to fifth column. Unlike original KMCG and KFCG algorithms, all clusters are formed with respect to only first column. In augmented KMCG sorting of training set is done only once with respect to first column, here we obtained the median of first column and divide the training set into two clusters. These two clusters are handled separately and obtained median value from each of them. Using two median values these clusters are further divided into four clusters. This process is repeated until desired number of cluster has been formed. Augmented KFCG is similar to augmented KMCG up to the formation of training set. In the next step average of the first column of the training set is calculated and compared with the other values of same column and divides training set into two clusters, vectors with small values than average goes to first cluster, otherwise goes to second cluster. Similarly, these two clusters are handled separately and further divided into four clusters with respect to first column only. This process is repeated three times to obtained eight clusters.

6. Post Processing on Selected Cluster

Is the last stage 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 present in the desired region. Disk structuring element has been used with radius 3.

7. Complexity Analysis of Codebook Generation Algorithms

As shown in the Figure 1, M is the total number of vectors in the training set, K is the size of a vector (dimension of the vector) and N is the size of codebook. In this paper codebook size is limited up to eight codevectors, therefore eight clusters are created using codebook generation algorithm. By considering this fact merely 3 iterations are required for the algorithm to make eight clusters from the training set of size M.

To compute time complexity of the algorithms used in this paper in terms of number of vectors in the training set, let us consider 1 CPU unit require for one addition / subtraction of two 8 bit numbers, 8 CPU units are required for multiplication /division of two 8 bits numbers and 1 CPU unit required to compare two 8 bits numbers. In this hypothesis, we are not considering the CPU unit required in computation of probability and its histogram equalized probability images, they are used as a input to these clustering algorithms. As explained in Section 5.1, KMCG algorithm sorts the training set and sorting algorithm requires M^2 comparisons, therefore it requires M^2 CPU units for sorting the training set once. Here we have neglected one CPU unit required for single division which computes median, since it is negligible as compare to M^2 units. Since the algorithms is of iterative nature, same sorting process repeated 3 times to get eight clusters on different columns of the training set. As a

result, M^2 CPU units, $M^2/2$ CPU units and $M^2/4$ CPU units required during first iteration, second iteration and third iteration respectively. Same hypothetical consideration is applied for other algorithms used in this paper and Table 1 illustrates total number of CPU units required for all the codebook generation algorithms in terms of size of the training set. Table 2 illustrates actual number of CPU units required for different images (values of M) and Table 2 shows that KFCG is the fastest algorithm among all the codebook generation algorithms for the given set of images.

Table 1. Complexity analysis of clustering algorithms in terms of total number of CPU units required based on number of training vector M and vector dimension K

Analysis Parameters	Iterations required for clustering	KMCG	KFCG	Augmented KMCG	Augmented KFCG
Number of Comparison Required	Iteration 1	M^2	M	M^2	M
	Iteration 2	$M^2/2$	M	0	M
	Iteration 3	$M^2/4$	M	0	M
Number of Additions Required	Iteration 1	0	M-1	$M(K-1)$	$(M-1) + M(K-1)$
	Iteration 2	0	M-2	0	M-2
	Iteration 3	0	M-4	0	M-4
Total CPU units required		$M^2 + M^2/2 + M^2/4 = 7M^2/4$	$3M + 3M - 7 = 6M - 7$	$M^2 + M(K-1)$	$6M - 7 + M(K-1)$

Table 2. Number of CPU units required for different values of M and K=4 for images used in this paper

Images used in the paper	Number of vectors in training set	KMCG	KFCG	Augmented KMCG	Augmented KFCG
Image 1 (206x336)	M=17304	52.39×10^7	10.38×10^4	29.94×10^7	15.75×10^4
Image 2 (346x336)	M=29064	14.78×10^8	17.43×10^4	84.48×10^7	26.15×10^4
Image 2 (242x322)	M=19481	66.41×10^7	11.68×10^4	37.98×10^7	17.52×10^4
Image 3 (402x620)	M=62310	67.94×10^8	37.38×10^4	38.82×10^8	56.07×10^4

8. Result analysis and comparison

To authenticate our proposed algorithm, shown in Figure 2 we implemented and tested it using set of real US images. This method is executed using MATLAB 7.0 on Intel Core2 Duo 2.10 GHz processor with 2 GB RAM. Here, in this paper results of four different images are shown, in which, according to expert radiologist image 1, image 3 exhibits malignant tumor and image 2, image 4 exhibits benign tumor. In this paper, all the results are verified and validated by the expert radiologists through visual inspection. In first stage of the algorithm

probability and its histogram equalized images are obtained and used as input to the second stage. In the second stage of proposed algorithm we used four different codebook generation algorithms for clustering, therefore results of different images are shown accordingly.

Figure 6 shows eight cluster images and Figure 7 shows its sequentially merged cluster images obtained from the probability image shown in Figure 5 (iii). Segmentation results are shown directly for all other images. Best results amongst original US image, probability image and histogram equalized probability images are shown by red boxes. Figure 8 shows the segmentation results of KMCG, KFCG and augmented KMCG / KFCG algorithms applied on images shown in Figure 5(i), (iii) and (v). Similarly Figure 10, Figure 12 and Figure 14 shows segmentation results of other three sets of images (original, probability and its histogram equalized probability image). All the results are displayed in pairs first image from the pair is segmentation result and other is result superimposed on original image. Analysis of the results and opinion of the expert radiologist is shown in Table 3.

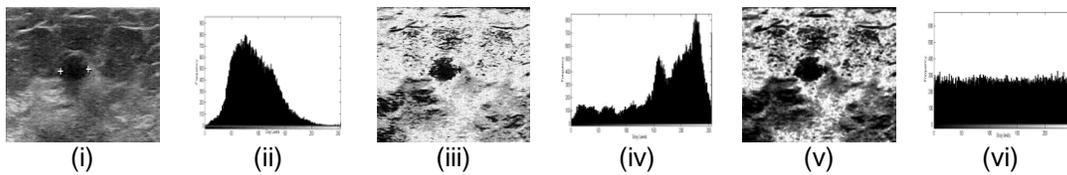


Figure 5. (i) Original US image (Image 1) (iii) Probability image (v) Histogram equalized probability image, with their respective histogram in (ii), (iv) and (vi)

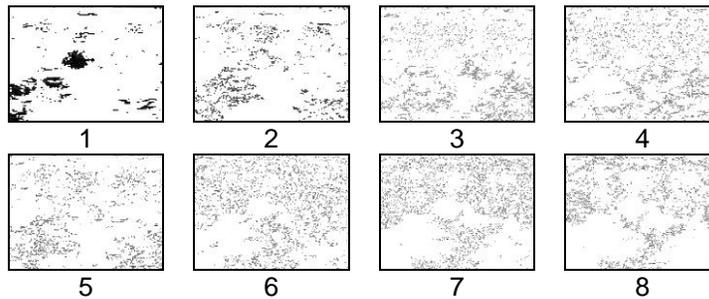


Figure 6. Eight cluster images obtained using KMCG codebook generation algorithm on probability image shown in Figure 4 (iii).

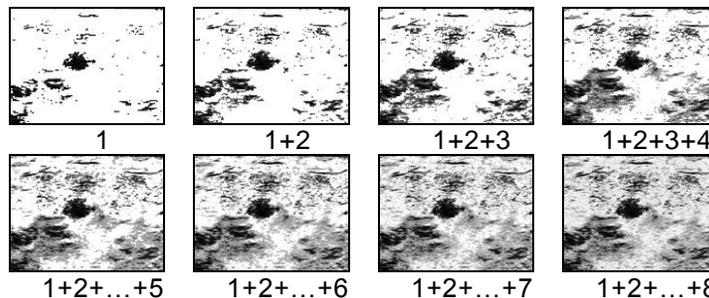


Figure 7. Cluster images obtained after sequential merging of clusters shown in Figure 5

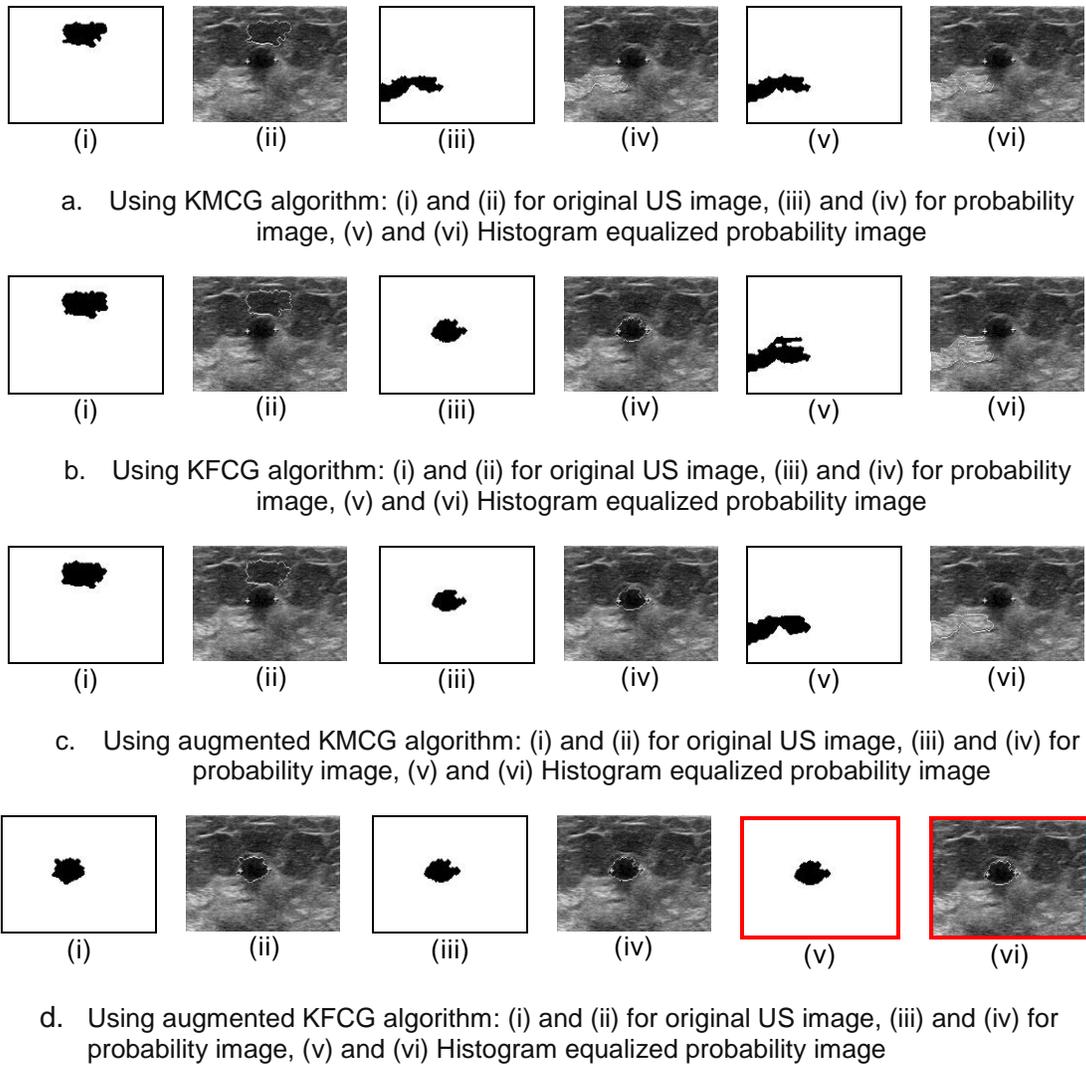


Figure 8. Segmentation results for the images shown in Figure 5 (i), (iii) and (v)

From the above results it is observed that all the codebook generation algorithms produced false positive results on original US images. Therefore these results are rejected, but KFCG and augmented KMCG / KFCG algorithms gives acceptable results on probability images. According to the expert radiologists the result obtained by augmented KFCG on histogram equalized probability image is accurate as compared to all other acceptable results. Accurate result is shown by red color box.

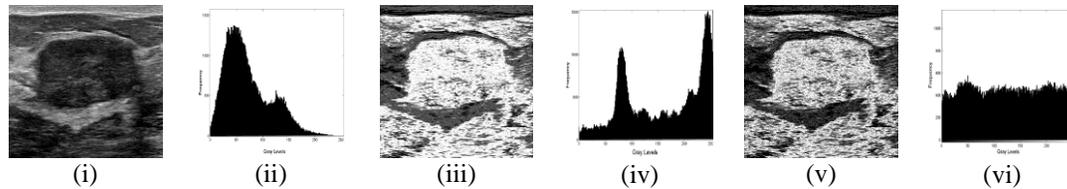


Figure 9. (i) Original US image (Image 2) (iii) Probability image (v) Histogram equalized probability image, with their respective histogram in (ii), (iv) and (vi)

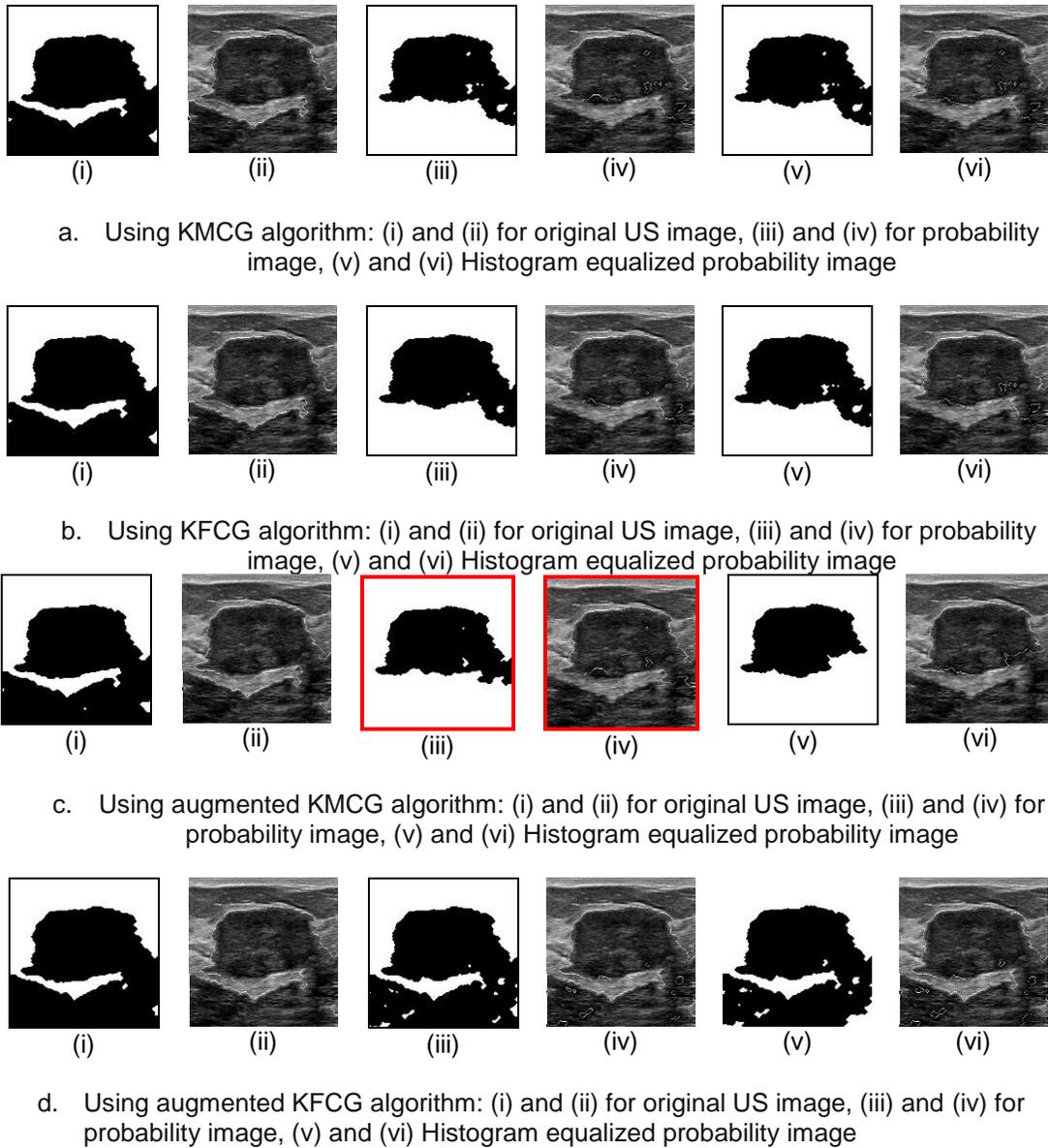


Figure 10. Segmentation results for the images shown in Figure 9 (i), (iii) and (v)

For Figure 10 observations are same as Figure 8 for the original US image, where all the algorithms produced wrong results which are rejected by the radiologists. Similarly augmented KFCG is not useful all other images including probability and its histogram equalized probability images. The accurate result is given by augmented KMCG on probability image shown by red box.

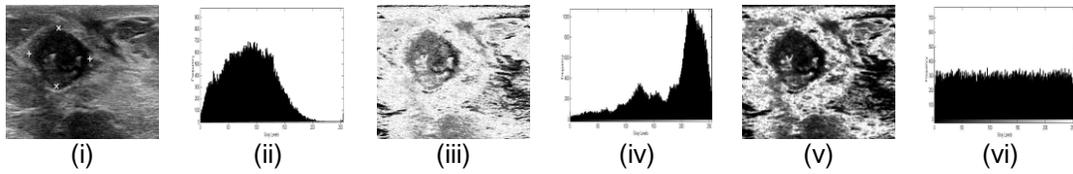
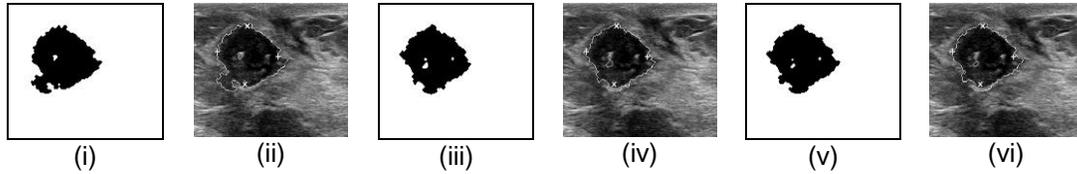
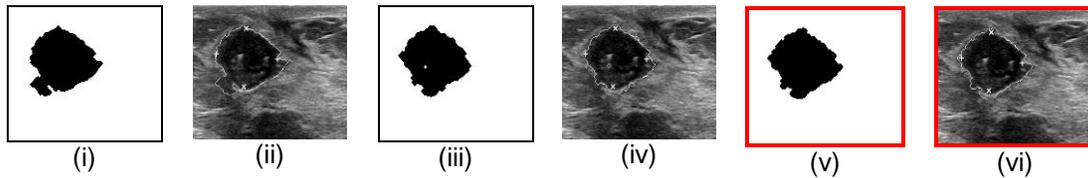


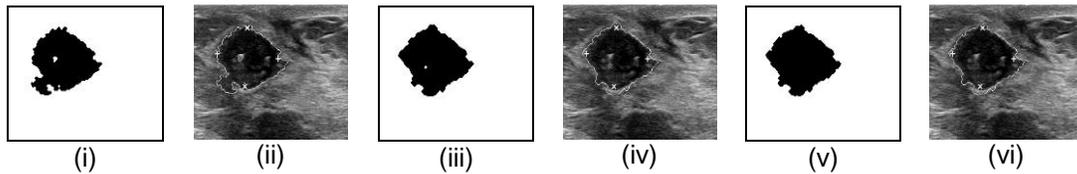
Figure 11. (i) Original US image (Image 3) (iii) Probability image (v) Histogram equalized probability image, with their respective histogram in (ii), (iv) and (vi)



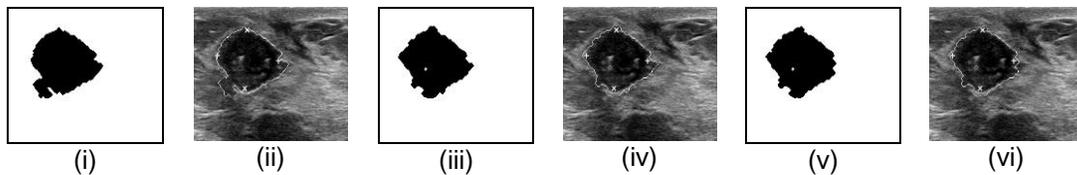
a. Using KMCG algorithm: (i) and (ii) for original US image, (iii) and (iv) for probability image, (v) and (vi) Histogram equalized probability image



b. Using KFCG algorithm: (i) and (ii) for original US image, (iii) and (iv) for probability image, (v) and (vi) Histogram equalized probability image



c. Using augmented KMCG algorithm: (i) and (ii) for original US image, (iii) and (iv) for probability image, (v) and (vi) Histogram equalized probability image



d. Using augmented KFCG algorithm: (i) and (ii) for original US image, (iii) and (iv) for probability image, (v) and (vi) Histogram equalized probability image

Figure 12. Segmentation results for the images shown in Figure 11 (i), (iii) and (v)

According to the radiologist's observations KFCG algorithm gives accurate result when it applied to the histogram equalized probability image and it is shown by red box. For other

algorithms like augmented KMCG / KFCG gives acceptable results for same type of images. The results obtained using original US images are rejected for all the algorithms, since they exhibits over segmentation.

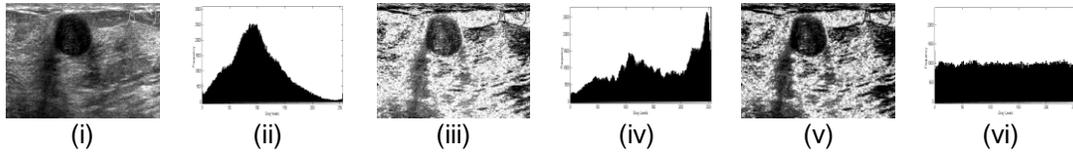
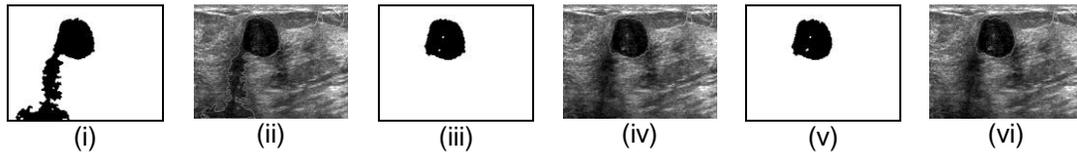
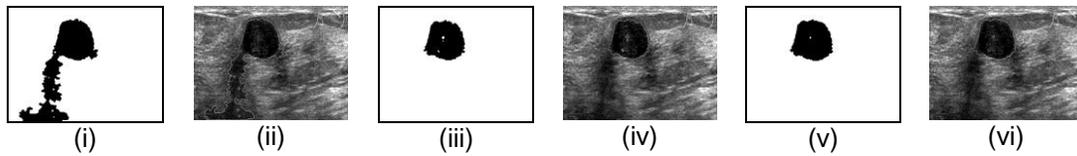


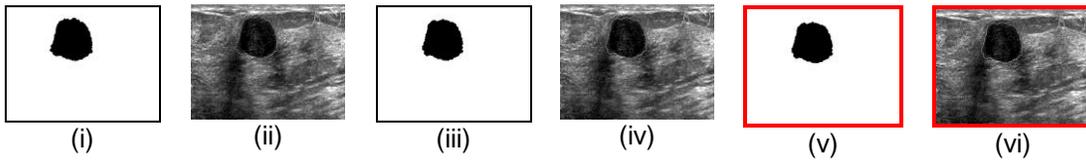
Figure 13. (i) Original US image (Image 4) (iii) Probability image (v) Histogram equalized probability image, with their respective histogram in (ii), (iv) and (vi)



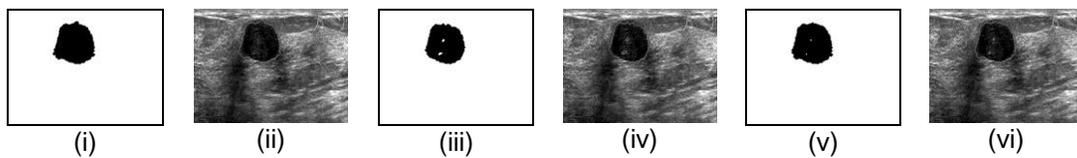
a. Using KMCG algorithm: (i) and (ii) for original US image, (iii) and (iv) for probability image, (v) and (vi) Histogram equalized probability image



b. Using KFCG algorithm: (i) and (ii) for original US image, (iii) and (iv) for probability image, (v) and (vi) Histogram equalized probability image



c. Using augmented KMCG algorithm: (i) and (ii) for original US image, (iii) and (iv) for probability image, (v) and (vi) Histogram equalized probability image



d. Using augmented KFCG algorithm: (i) and (ii) for original US image, (iii) and (iv) for probability image, (v) and (vi) Histogram equalized probability image

Figure 14. Segmentation results for the images shown in Figure 13(i),(iii) and(v)

According to the expert radiologists results obtained by using augmented KMCG on histogram equalized image is observed to be a good result as compared to all other algorithms. Unlike the other images here results of augmented KMCG / KFCG on original images is accepted. Results where tumor region contains noise are rejected such as a (iii), a (v), b (iii) and d (iii).

Table 3. Result comparison and opinion of expert radiologists

Real US Images	Types of Input Images	Tumor demarcation using codebook generation algorithms			
		KMCG	KFCG	Augmented KMCG	Augmented KFCG
Image 1 with Malignant tumor	Original	Rejected	Rejected	Rejected	Rejected
	Probability	Rejected	Good	Acceptable	Acceptable
	Histogram equalized	Rejected	Rejected	Rejected	Good
Image 2 with Benign Tumor	Original	Rejected	Rejected	Rejected	Rejected
	Probability	Acceptable	Acceptable	Good	Rejected
	Histogram equalized	Acceptable	Acceptable	Acceptable	Rejected
Image 3 with Malignant tumor	Original	Rejected	Rejected	Rejected	Rejected
	Probability	Rejected	Acceptable	Acceptable	Acceptable
	Histogram equalized	Rejected	Good	Good	Rejected
Image 4 with Benign tumor	Original	Rejected	Rejected	Acceptable	Acceptable
	Probability	Rejected	Rejected	Good	Rejected
	Histogram equalized	Rejected	Acceptable	Acceptable	Acceptable

9. Conclusion

In this paper we have proposed the algorithm for tumor demarcation for US images, which is based on clustering and statistical properties (probability) of the image. In fact we are utilizing the fundamental concept of random distribution of the gray levels present in US images. To overcome this disadvantage (i.e. speckle, attenuation) of the US images, probability images are obtained, by considering the fact of too small values of the probability, histogram equalization is also done on probability images. In these images, we observed that the tumor region is more prominent than the background with clear boundary and can easily be traceable using our VQ based codebook generation algorithms as shown in Figure 4, Figure 8, Figure 10 and Figure 12. Time complexity analysis is done for all codebook generation algorithms used for clustering in terms of M (size of training set). As shown in Table 1 and Table 2, KFCG is the fastest amongst all and as the value of M increase the difference in the execution time is also increased between algorithms. In this paper, all the segmentation results are validated through visual inspection by the expert radiologists. To select the good result amongst all, first we considered the accuracy of the results if multiple results are acceptable as good results then we considered the time required by the algorithms.

Table 3 mentioned the good segmentation results by considering accuracy and the execution time. Clustering using these algorithms is also applied to original US images and results are compared with the probability images and its histogram equalized images. Since the original US images are highly attenuated and has significant amount of speckle, results produced by the codebook generation algorithms are not good and rejected except Image 4 using augmented KMCG / KFCG.

Acknowledgements

The proposed algorithm is implemented and tested on real US images provided by the radiologists. Authors would like to thank Dr. Wrushali More, and Dr. Mushtak for their support. Authors are also grateful for their valuable inputs to understand the nature, shape of the tumor and results validation.

References

- [1] J. Ferlay, H. R. Shin, F. Bray, D. Forman, C. Mathers and D. M. Parkin, "GLOBOCAN 2008: Cancer Incidence, Mortality and Prevalence Worldwide IARC", CancerBase No. 10, Lyon, France: International Agency for Research on Cancer, (2011).
- [2] T. L. Szabo, "Diagnostic Ultrasound Imaging: Inside out", Elsevier Inc., (2004) September, pp. 19-23.
- [3] 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, (2002) January, pp. 19-26.
- [4] A. Stavros, D. Thickman, C. Rapp, M. Dennis, S. Parker and G. Sisney, "Solid breast nodules: Use of sonography to distinguish between benign and malignant lesions", *Radiology*, vol. 196, (1995), pp. 122-134.
- [5] A. Bosch, A. Kessels, G. Beets, K. Vranken, A. Borstlap, M. von Meyenfeldt and J. van Engelshoven, "Interexamination variation of whole breast ultrasound", *British Journal of Radiology*, vol. 76, no. 905, (2003) May, pp. 328-331.
- [6] H. Madjar, "Role of Breast Ultrasound for the Detection and Differentiation of Breast Lesions", *Breast Care*, vol. 5, (2010) April, pp. pp 109-114.
- [7] G. Rahbar, A. C. Sai, G. C. Hansen and J. S. Prince, "Benign Versus Malignant Solid Breast Masses: US Differentiation", *Radiology*, vol. 213, (1999) December, pp. 889-894.
- [8] K. Horsch, M. L. Giger, L. A. Venta and C. J. Vyborny, "Computerized diagnostic of breast lesions on ultrasound", *Medical Physics*, vol. 29, no. 2, (2002) February, pp. 157-164.
- [9] J. A. Noble and D. Boukerroui, "Ultrasound Image Segmentation: A Survey", *IEEE Transactions on Medical Imaging*, vol. 25, no. 8, (2006) August, pp. 987-1010.
- [10] G. Xiao, M. Brady, J. A. Noble and Y. Zhang, "Segmentation of Ultrasound B-Mode Images With Intensity Inhomogeneity Correction", *IEEE TRANSACTIONS ON MEDICAL IMAGING*, vol. 21, no. 1, (2002) January, pp. 48-57.
- [11] D. Boukerroui, O. Basset, A. Noble, A. Baskurt and G. Gimenez, "A Multiparametric and Multiresolution Segmentation Algorithm of 3-D Ultrasonic Data", *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 48, no. 1, (2002), pp. 64-77.
- [12] 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, (2003), Elsevier, Science Direct, pp. 779-790.
- [13] Y. L. Huang and D. R. Chen, "Watershed segmentation for breast tumor in 2-D sonography", *Ultrasound Med. Biol.*, vol. 30, Issue 5, (2004) May, pp. 625-632.
- [14] Y. Su, Y. Wang, J. Jiao and Y. Guo, "Automatic Detection and Classification of Breast Tumours in Ultrasonic Images Using Texture and Morphological Features", vol. 5, *The Open Medical Informatics Journal*, (2011), pp. 26-37.
- [15] D. -R. Chen, Y. -L. Huang and S. -H. Lin "Computer-aided diagnosis with textural features for breast lesions in sonograms", *Computerized Medical Imaging and Graphic*, Elsevier, vol. 35, (2011), pp. 220-226.
- [16] 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, (2002) October, pp. 1301-1310.
- [17] H. B. Kekre and P. Shrinath, "Tumor Demarcation by using Local Thresholding on Selected Parameters obtained from Co-occurrence Matrix of Ultrasound Image of Breast", *International Journal of Computer Applications (0975-8887)*, vol. 32, no. 7, (2011) October.

- [18] B. Liu, H. D. Cheng, J. Huang, J. Tian, X. Tang and J. Liu, "Probability density difference-based active contour for ultrasound image segmentation", *Pattern Recognition, Science Direct, Elsevier*, vol. 43, (2010), pp. 2028–2042.
- [19] C. W. 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, (1998) December, pp. 1673-1683.
- [20] R. M. Gray, "Vector quantization", *IEEE ASSP Magazine*, (1984) April, pp. 4-29.
- [21] P. C. Cosman, K. L. Oehler, E. A. Riskin and R. M. Gray, "Using Vector Quantization for Image Processing", *Proceedings of the IEEE*, vol. 81, no. 9, (1993) September, pp. 1326-1341
- [22] C. M. Huang and R. W. Harris, "A comparison of several vector quantization codebook generation approaches", *IEEE Transactions on Image Processing*, vol. 2, no. 1, (1993) January, pp. 108 – 112.
- [23] H. B. Kekre, T. K. Sarode and S. D. Thepade, "Color Texture Feature based Image Retrieval using DCT applied on Kekre's Median Codebook", *International Journal on Imaging (IJ)*, www.ceser.res.in/iji.html.
- [24] C. Garcia and G. Tziritas, "Face detection using quantized skin color regions merging and wavelet packet analysis", *IEEE Trans. Multimedia*, vol. 1, no. 3, (1999) September, pp. 264–277.
- [25] H. B. Kekre, T. K. Sarode and S. Gharge, "Detection and Demarcation of Tumor using Vector Quantization in MRI images", *International Journal of Engineering Science and Technology*, vol. 1, no. 2, (2009), pp. 59-66, <http://arxiv.org/ftp/arxiv/papers/1001/1001.4189.pdf>.
- [26] H. B. Kekre, T. Sarode, S. Gharge and K. Raut, "Detection of Cancer Using Vector Quantization for Segmentation", *International Journal of Computer Applications (0975 – 8887)*, vol. 4, no. 9, (2010) August.
- [27] H. B. Kekre and T. K. Sarode, "Centroid Based Fast Search Algorithm for Vector Quantization", *International Journal of Imaging and Robotics (IJIR)*, vol. 1, no. A08, (2008), pp. 73-83, <http://www.ceser.res.in/iji.html>.
- [28] H. B. Kekre and T. K. Sarode, "Fast Codebook Generation Algorithm for, Color Images using Vector Quantization", *International Journal of Computer Science and Information Technology*, vol. 1, no. 1, (2009) January, pp. 7-12.
- [29] H. B. Kekre, T. K. Sarode and S. D. Thepade, "Image Retrieval using Color-Texture Features from DCT on VQ Codevectors obtained by Kekre's Fast Codebook Generation", *ICGST-International Journal on Graphics, Vision and Image Processing (GVIP)*, vol. 9, Issue 5, (2009) September, pp. 1-8, <http://www.icgst.com/gvip/Volume9/Issue5/P1150921752.html>.
- [30] H. B. Kekre and P. Shrinath, "Tumour Delineation using Statistical Properties of the breast US images and Vector Quantization based Clustering Algorithms", *International Journal of Image, Graphics and Signal Processing (IJIGSP)*, vol. 5, no. 11, (2013) November, pp. 1-12.
- [31] H. B. Kekre and P. Shrinath, "Tumour Demarcation by using Vector Quantization and Clustering Clusters of Ultrasound Image of Breast", *International Journal of Advances in Engineering & Technology, IJAET*, ISSN: 2231-1963, (2012) July, <http://www.ijaet.org/archive>.
- [32] American Cancer Society, "Global Cancer Facts & Figures", 2nd Edition, Atlanta: American Cancer Society; (2011).
- [33] J. -H. Yu, Y. -Y. Wang, P. Chen and H. -Y. Xu, "Two-dimensional Fuzzy Clustering for Ultrasound Image Segmentation", *Proceeding of IEEE International Conference on Bioinformatics and Biomedical Engineering*, 1-4244-1120-3, (2007) July, pp. 599-603.
- [34] Q. Chen, K. Kotani, F. Lee and T. Ohmi, "VQ-based face recognition algorithm using code pattern classification and Self-Organizing Maps", *9th International Conference on Signal Processing*, (2008) October, pp. 2059 – 2064.
- [35] H. B. Kekre, T. K. Sarode, S. Gharge and K. Raut, "Image Segmentation of MRI Images using KMCG and KFCG Algorithms", *2nd International Conference and workshop on Emerging Trends in Technology (ICWET)*, Proceedings published by *International Journal of Computer Applications (IJCA)*, (2011), pp. 1-5.
- [36] H. B. Kekre and T. K. Sarode, "An Efficient Fast Algorithm to Generate Codebook for Vector Quantization", *First International Conference on Emerging Trends in Engineering and Technology*, IEEE Computer Society, 978-0-7695-3267-7/08 2008 IEEE, (2008), pp. 62-67.
- [37] H. B. Kekre and P. Shrinath, "Tumor Demarcation by VQ based clustering and augmentation with KMCG and KFCG codebook generation algorithms", *World Congress on Information and Communication Technologies (WICT)*, IEEE Conference, Trivendram, India, (2012), pp. 993- 998.

Authors



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 Engineering 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. Author's profile



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 13 years of teaching experience and currently working as Associate Professor in Computer Engineering Department, MPSTME. His area of interest is Image processing, Distributed Computing and Parallel Processing.

