TUMOUR DEMARCATION BY USING VECTOR QUANTIZATION AND CLUBBING CLUSTERS OF ULTRASOUND IMAGE OF BREAST

H. B. Kekre[1] and Pravin Shrinath[2]
[1] Senior Professor & [2] Ph.D. Scholar,
Department of Computer Engg., MPSTME, SVKM's NMIMS University, Mumbai, India

ABSTRACT

In most of the computer aided diagnosis, segmentation is used as the preliminary stage and further can be helpful in quantitative analysis. Ultrasound imaging (US) helps medical experts to understand clinical problem efficiently with low cost as compared to its counterparts. In this paper, vector quantization based clustering technique has been proposed to detect the tumour (malignant or benign) of the breast Ultrasound Images. Presence of artefacts like speckle, shadow, attenuation and signal dropout, makes image understanding and segmentation difficult for an expert. Here, we dealt with images having these artefacts and proposed fully automatic segmentation technique using clustering. Firstly well known Vector Quantization based LBG technique is used for clustering and eight clusters are obtained, sequential clubbing of these cluster are suggested to obtain segmentation results. Improvement is suggested using two new techniques over LBG to form clusters, known as KPE (Kekre's Proportionate Error), and KEVR (Kekre's Error Vector Rotation), further same method of sequential clubbing of clusters is followed here as that of LBG and their results are compared.

KEYWORDS: *Vector Quantization, Codebook, Codevector, Cluster clubbing*

I. INTRODUCTION

Ultrasound imaging (US) is very important medical imaging modality to examine the clinical problems. It has become more popular tool than its counterpart with its non invasive and harmless nature to diagnose various abnormalities present in the human organs. Ultrasonography is relatively inexpensive and effective method of differentiating the cystic breast masses from solid breast masses (benign and malignant). It is also fully established method that gives the valuable information about the nature and extent of solid masses and other breast lesions [1][2]. Detection of tumour manually is inaccurate and time consuming process for a radiologist due to random orientation of the tumour and texture (noise) present in the ultrasound images and accuracy is major concern in the medical applications. Automated (without human intervention) segmentation of US images provides detection of desired region (e.g. defected organs, abnormal masses) accurately and time efficiently. Due to some inherent characteristic artifacts such as attenuation, shadows and speckle noise, the process of segmentation of US images is quite difficult [3][4]. To acquire the accurate segmentation of US images, removal of speckle is important [5]. Many image processing algorithms (techniques) are developed and used on ultrasound image segmentation, such as texture, region growing, thresholding [6], neural network, fuzzy clustering [7] etc. Most of these methods are influenced by speckle and this makes speckle removal an important step. In this paper we are using Vector Quantization based clustering and dealing images with speckle, without any pre-processing step. In breast ultrasound images, defected area pixel (cystic or solid masses) is slightly darker than the pixel representing normal tissues, but in some cases due to limitation of acquisition process, boundary pixels of defected area is presented like normal tissue structure and this makes boundary detection difficult. Here we are exploring this phenomenon in clustering process. The other sections of this paper are organized as follows. In section II, vector quantization is discussed with its use in segmentation. In section III, three codebook generation algorithms based on VQ are explained with its use in clustering. Proposed method is explained in section IV followed by conclusion in section V.

II. VECTOR QUANTIZATION

Vector quantization (VQ) is basically designed as image compression techniques [8][9] with development of many algorithms for vector codebook generation and quantization [10-12], but now a days it has been extensively used in many applications, like image segmentation [13], speech recognition [14], pattern recognition and face detection [15][16], tumor demarcation in MRI and Mammogram images [17][18], content based image retrieval [19][20] etc. In this paper, this method has been used and implemented for demarcation of cysts and tumor (malignant or benign) in breast ultrasound images.

A two dimensional image I is converted into K dimensional vector space of size M, $V = \{V_1, V_2, V_3, \ldots\ldots\ldots, V_M\}$ (training set). VQ is used as a mapping function to convert this K dimensional vector space to finite set $CB = \{C_1, C_2, C_3, C_4, \ldots\ldots, C_N\}$. CB is a codebook of size N and each code vector from C_1 to C_N represents the specific set of vectors of the entire training set of dimensions K and size M. The codebook size N is much smaller than size of the training set M and gives the number of clusters formed. It also influences the segmentation of US images. Here optimum size codebook is designed using clustering algorithm in spatial domain.

In VQ technique, encoder divides the image into desired size blocks and these blocks then converted into finite set of training vectors. Using codebook generation algorithms as discussed in section III, the clusters are created. To form a set of clusters $CL = \{CL_1, CL_2, CL_3, \ldots\ldots\ldots, CL_N\}$ representing different regions of image, Squared Euclidean Distance (ED) between each training vector and code vector is calculated and training vector with minimum ED is then added to the respective cluster represented by particular codevector as shown in equation (1).

$$Vi \in CLj = MIN_{1 \le i \le M, 1 \le j \le N} \{d(Vi, Cj)\} \tag{1}$$

Where,

$\{d(V_i, C_j)\}$ = Euclidean Distance (ED) between training vector V_i and codevector C_j as per equation (2).

$$ED^2 = \sum_{x=1}^{K} (V_{ix} - C_{jx})^2 \tag{2}$$

III. CODEBOOK GENERATION ALGORITHMS

3.1. Linde Buzo Gray (LBG) Algorithm [8][9][10]

This algorithm is based on the calculation of the centroid as first code vector by taking the average of all vectors of training set. As shown in the Figure 1, two code vectors C_1 and C_2 are generated from this first code vector by adding and subtracting constant error 1 respectively. Euclidean distance of entire training set with respect to C_1 and C_2 is calculate as shown in equation (2) and two cluster are formed based on the closest of C_1 or C_2. This process is repeated until desired number of clusters has been formed. As shown in Figure 1 for two dimensional cases, this technique has a disadvantage, that the clusters are elongated and has constant angle with x axis of 45^0. This elongation gives inefficient cluster formation. Results of cluster images, clubbed images and superimposed images are shown in Figure 5, 6 and 7 respectively.

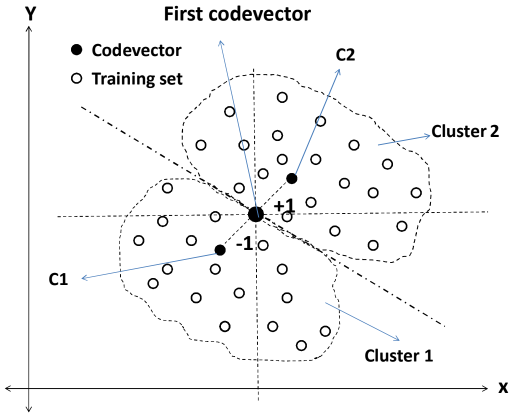


Figure 1: Clustering using LBG for two dimensional case

3.2. Kekre's Proportionate Error Vector (KPE) Algorithm [20][21][22][23]

In this technique, first code vector is generated by taking the average of entire training set, same as that of LBG, only difference is in addition and subtraction of proportionate error vector instead of constant error 1 to generate two code vectors C_1 and C_2 respectively [20]. Rest of the procedure is same as that of LBG.

Care is taken to keep code vector C_1 and C_2 within the limit of vector space while adding proportionate error. As shown in the Figure 2, unlike the LBG, clusters are not elongated and formed in different direction, so it gives efficient clustering than LBG. Results of cluster clubbing and superimposing segmentation using KPE algorithm are shown in Figure 8 and 9 respectively.

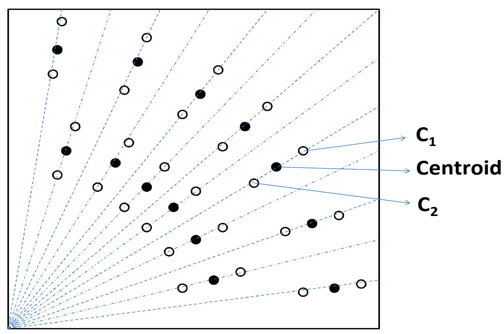


Figure 2: orientation of the line joining two code vector C1 and C2 after addition of proportionate error to the centroid.

3.3. Kekre's Error Vector Rotation (KEVR) Algorithm [24][25]

In this algorithm, two code vectors C_1 and C_2 are obtained by adding and subtracting error vector with first code vector respectively. As shown in the Figure 3, error vector matrix E is generated for dimension K and error vector e_i is the i^{th} row of the error matrix. To generate error matrix, binary sequence of number from 0 to K-1 is taken and 0 is replaced by 1, 1 is replaced by -1. With the addition and subtraction of error vector the cluster formation is rotated in different direction and elongated clusters are not formed, so cluster formation is efficient than LBG and KPE. Results of cluster clubbing and superimposing segmentation using KEVR algorithm are shown in Figure 10 and 11 respectively.

$$E = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ . \\ . \\ . \\ e_k \end{bmatrix} = \begin{pmatrix} 1\ 1\ 1\ 1 \ldots\ldots 1\ 1\ 1\ 1 \\ 1\ 1\ 1\ 1 \ldots\ldots 1\ 1\ 1\ {-1} \\ 1\ 1\ 1\ 1 \ldots\ldots 1\ 1\ {-1}\ 1 \\ 1\ 1\ 1\ 1 \ldots\ldots 1\ 1\ {-1}\ {-1} \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots\ldots\ldots \end{pmatrix}$$

Figure 3: Error Matrix generated for K dimensions [25]

IV. PROPOSED METHOD

Using codebook generation algorithms discussed in the section III, eight cluster images are obtained. Here, a method has been proposed to merge the cluster images one-by-one and forms another set of eight cluster images. Merging is done sequentially, like first cluster is added with second, resultant cluster is then added with third and so on. Eight cluster images, eight merged cluster images and eight superimposed images are shown in Figure 5, 6 and 7 respectively for LBG algorithm, implemented on original image shown in Figure 4. Same technique has been followed for KPE and KEVR algorithm, eight merged cluster images and eight superimposed images are shown in Figure 8 and 9 for KPE, Figure 10 and 11 for KEVR respectively. From Figure 6, 8 and 10, third clubbed image gives acceptable segmentation and KEVR gives better result amongst all three. This fully automatic method is implemented using MATLAB 7and tested on 30 images, from which, results of 15 images are shown in Figure 12 and only acceptable sequentially clubbed images are displayed. In Figure 12, first column shows all original images and second column gives clubbing sequence to obtain segmentation results for different algorithms, shown in column three, four, and five. This program is run on Intel Core2 Duo 2.20GHz with 1 GB RAM. Time required to get segmentation result is 2 to 3 seconds for image size 140 x 180, this is very less as compared to segmentation using manually traced method used by radiologists.

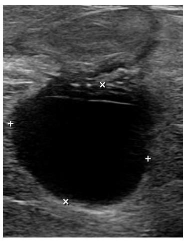
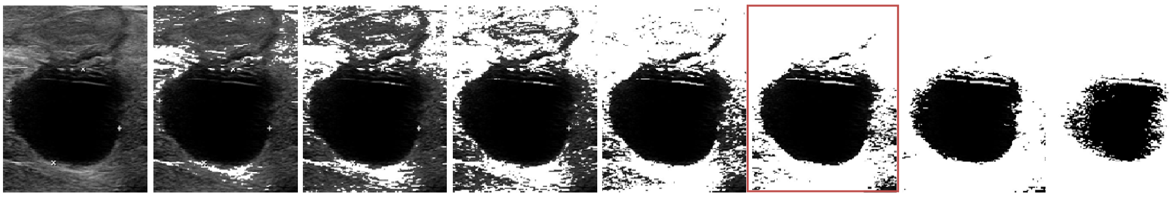


Figure 4: Brest Ultrasound image: Original

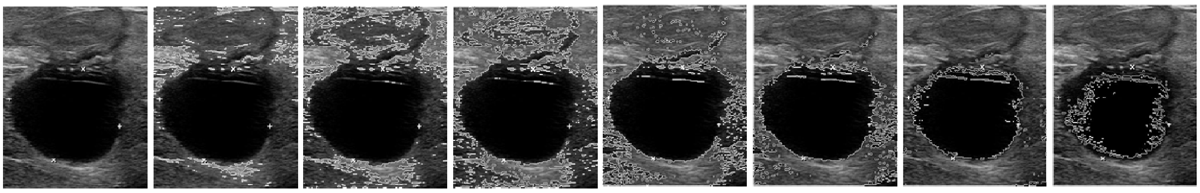
8 7 6 5 4 3 2 1

Figure 5: Eight cluster images obtained from Figure 4 using LBG: 1 to 8 from right to left



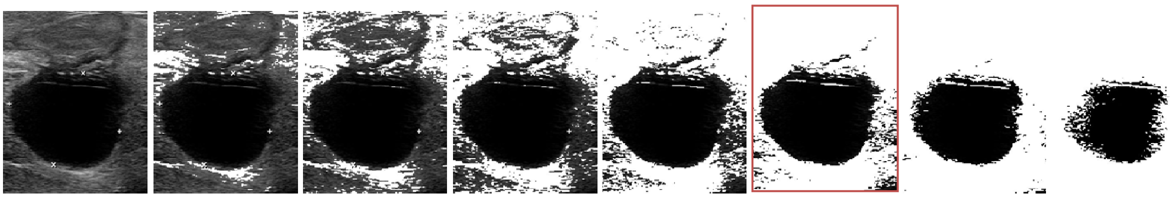
1+2.…7+8 1+2....6+7 1+2…5+6 1+2....4+5 1+2+3+4 1+2+3 1+2 1

Figure 6: Eight images obtained by clubbing clusters sequentially of Figure 5 using LBG: 1 to 8 from right to left - Best sequence 1+2+3, indicated in red box.



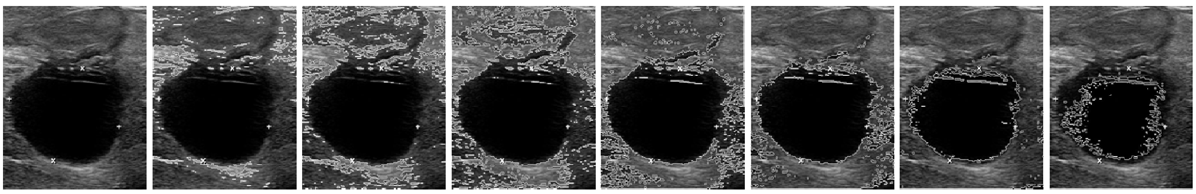
8 7 6 5 4 3 2 1

Figure 7: Eight images obtained by superimposing images of Figure 6 on original image of Figure 4: 1 to 8 from right to left



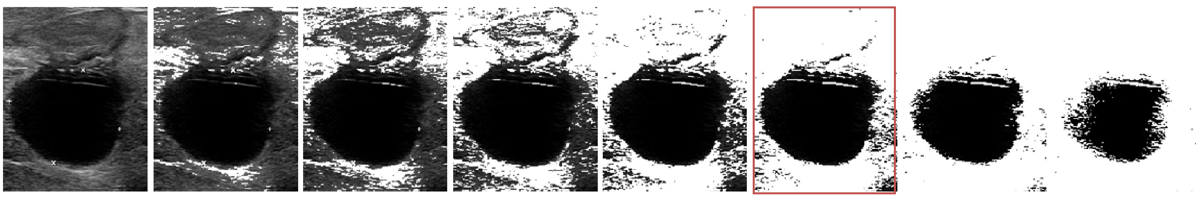
1+2.…7+8 1+2....6+7 1+2…5+6 1+2....4+5 1+2+3+4 1+2+3 1+2 1

Figure 8: Eight images obtained by clubbing clusters sequentially of KPE: 1 to 8 from right to left - Best sequence 1+2+3, indicated in red box



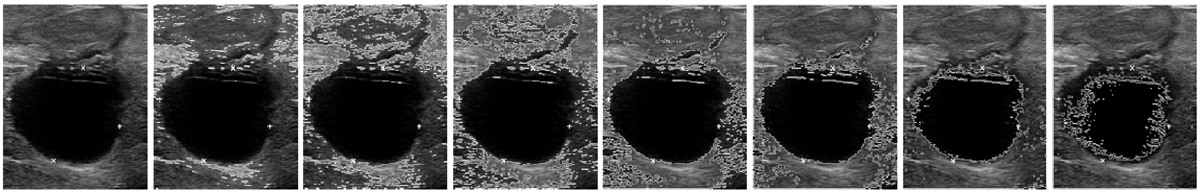
8 7 6 5 4 3 2 1

Figure 9: Eight images obtained by superimposing images of Figure 8 on original image of Figure 4: 1 to 8 from right to left



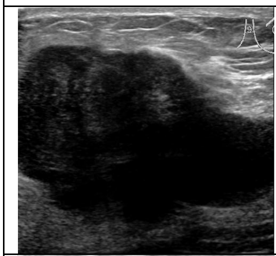
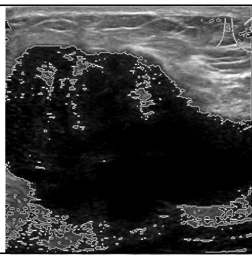
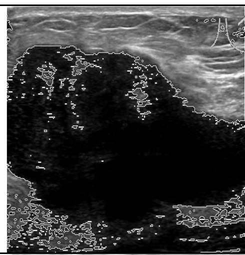
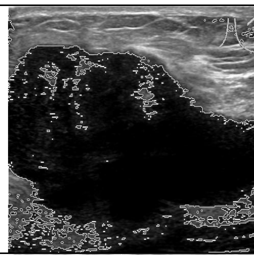
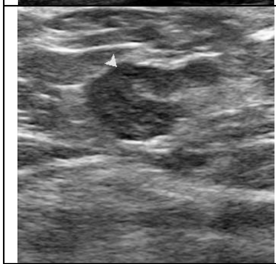
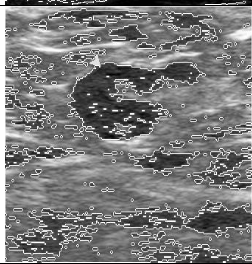
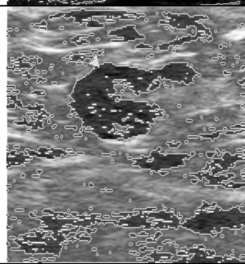
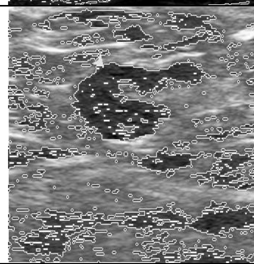
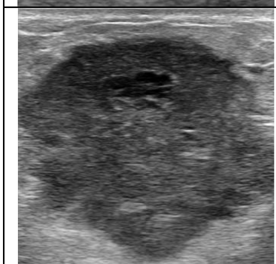
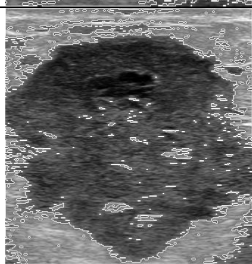
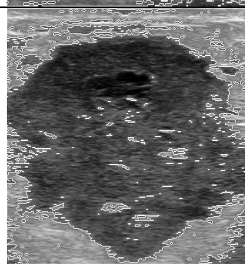
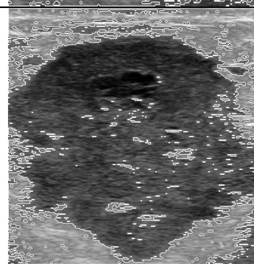
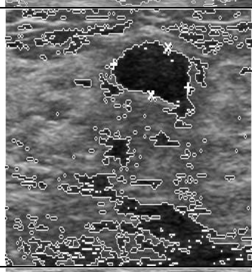
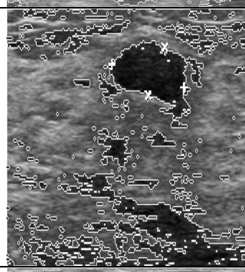
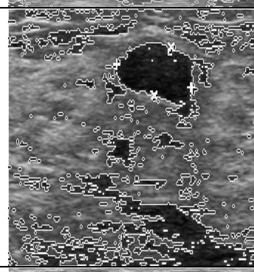
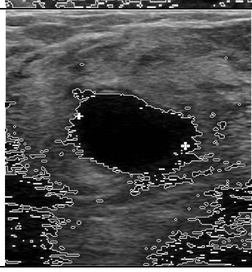
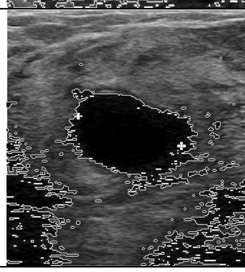
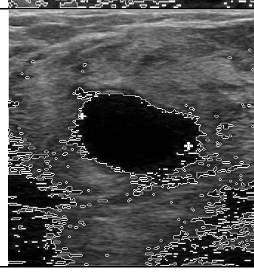
1+2.…7+8 1+2....6+7 1+2…5+6 1+2....4+5 1+2+3+4 1+2+3 1+2 1

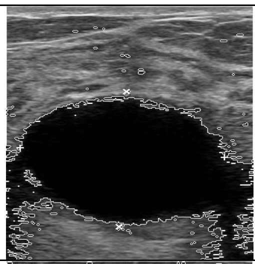
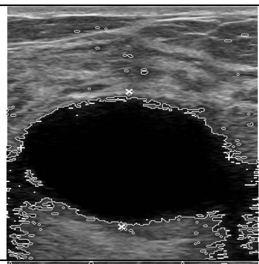
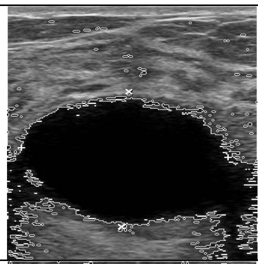
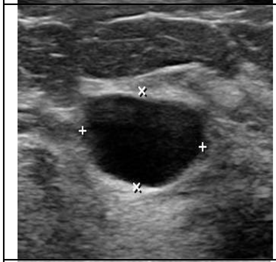
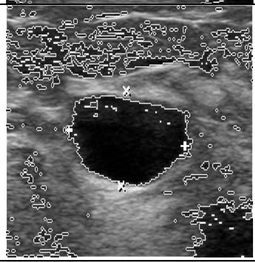
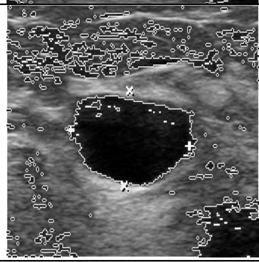
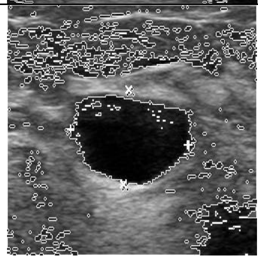
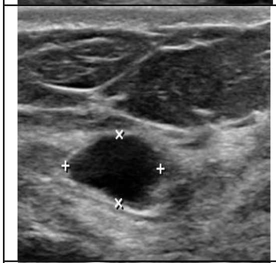
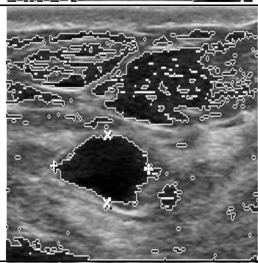
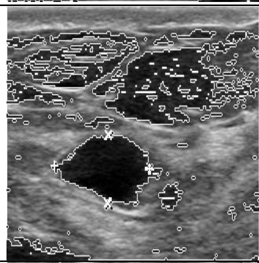
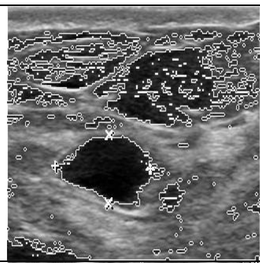
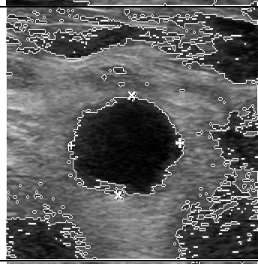
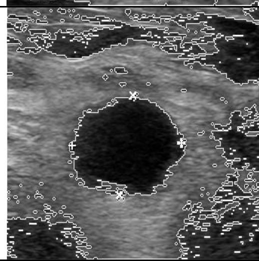
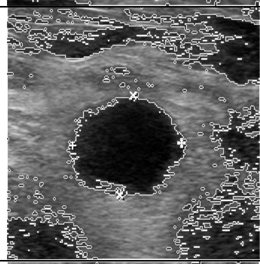
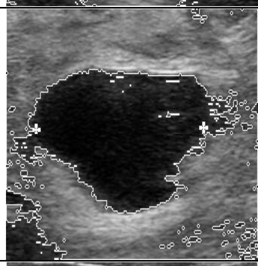
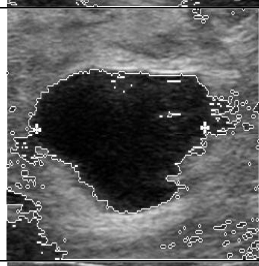
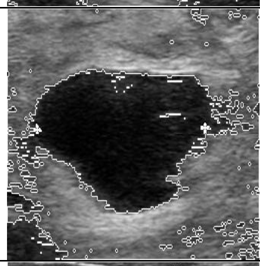
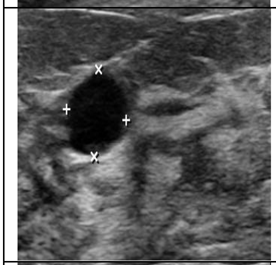
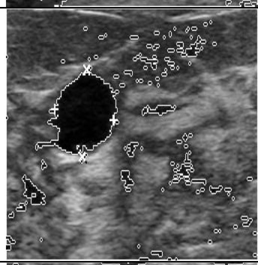
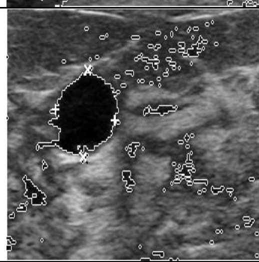
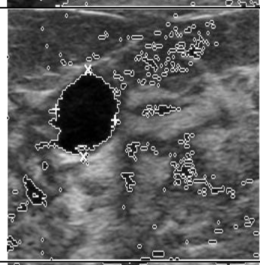
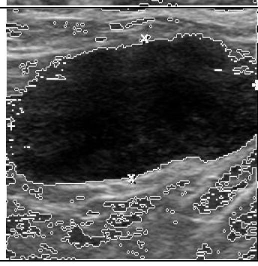
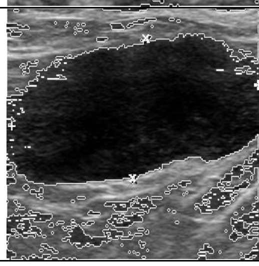
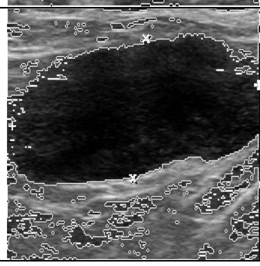
Figure 10: Eight images obtained by clubbing clusters sequentially of KEVR: 1 to 8 from right to left - Best sequence 1+2+3, indicated in red box



8 7 6 5 4 3 2 1

Figure 11: Eight images obtained by superimposing images of Figure 10 on original image of Figure 4: 1 to 8 from right to left

Original Images	Cluster Clubbing Sequence	Segmented Images: Superimposed on Original Image		
		LBG	KPE	KEVR
	1+2+..+4			
	1+2			
	1+..+5			
	1+2			
	1+2			

	1+2+3			
	1+2			
	1+2			
	1+2+3			
	1+2+3			
	1			
	1+3+4			

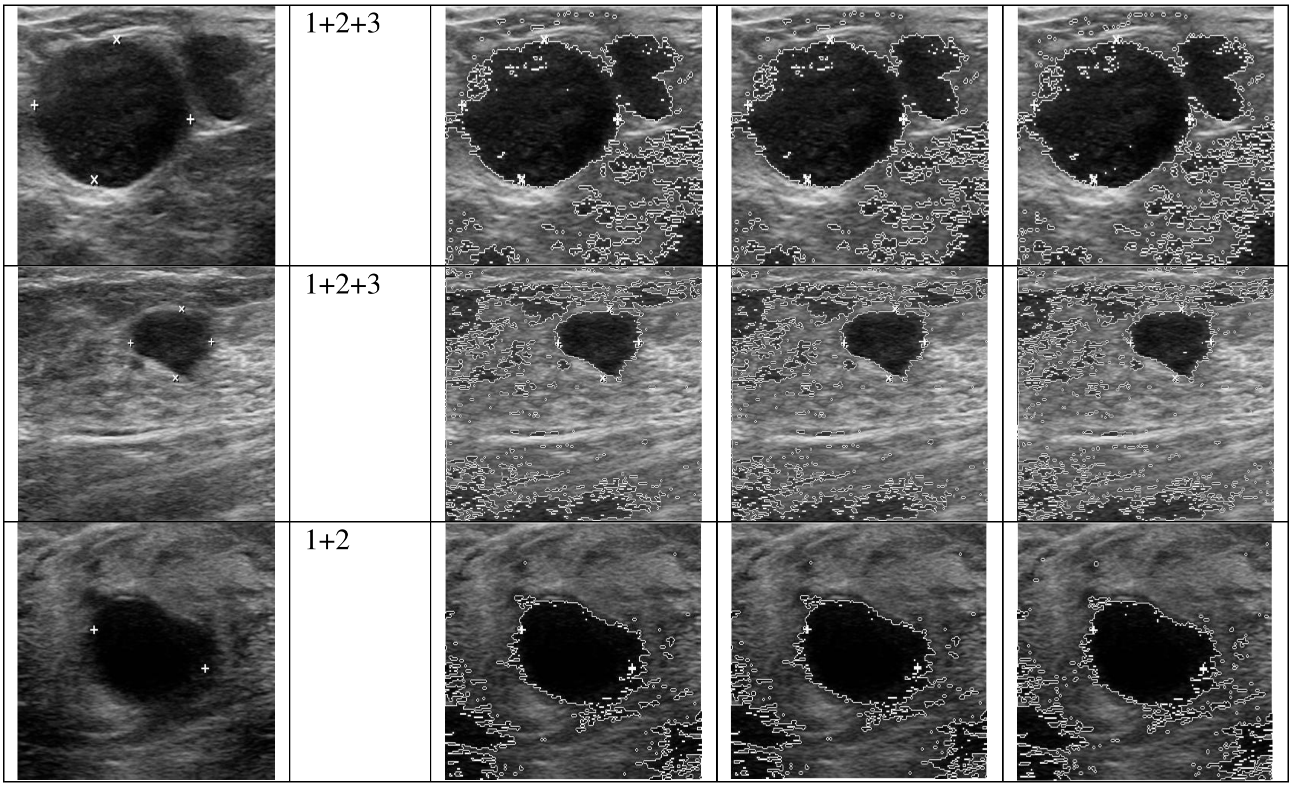


Figure 12: Segmentation result - Clubbed images superimposed on original images for LBG, KPE and KEVR algorithms

V. CONCLUSIONS

In this paper, a method has been proposed for tumour demarcation of breast ultrasound image and implemented on 30 images, out of which 16 images are shown in the paper. As shown in Figure 4, defected region (tumour) is represented by the dark pixels than the normal pixel and this phenomenon is common for all ultrasound images, so this has been explored in clustering. Clusters are formed using VQ based codebook generation algorithms and further these clusters are clubbed together sequentially to obtain the segmented image. Three methods are discussed and implemented for codebook generation, in LBG, as shown in Figure 1, the cluster elongation is unidirectional therefore cluster formation is inefficient for ultrasound images, where speckle is the dominant artefact. To overcome this drawback, in KPE, proportionate error has been used to improve the formation of clusters. As shown in Figure 2, for two dimensional vector spaces, orientation has changed but its variation is limited to the first quadrant, and proportionate error for ultrasound image would have small magnitude, so results will be similar to LBG. In KEVR, this limitation is overcome by using rotation of error vector and produced clusters with new orientation every time. Here vector is rotated in different direction and clusters are formed. Accuracy of the segmentation depends on the orientation and texture present in the image, and Clubbing sequence is varying as per representation of original image. As shown in second column of Figure 12 all images having different clubbing sequence for the best segmentation, but for all algorithms best segmented image have same clubbing sequence. As per the domain expert (Radiologist) the segmented images obtained using KEVR are better than LBG and KPE. As compared to LBG and KPE, KEVR images are having less amount of over segmentation. As shown in Figure 12, second and third clubbed images are giving the acceptable segmentation in 75 % cases and in rest of the cases, first, fourth or fifth clubbed image gives better segmentation.

ACKNOWLEDGEMENTS

The authors would like to thank Dr. Wrushali More and Dr. Anita Sable for their valuable guidance and suggestions to understand the ultrasound images and their segmentation results.

REFERENCES

[1]. Sickles EA., "Breast imaging: from 1965 to the present", Radiology, 215(1): pp 1-16, April 2000.
[2]. Sehqal CM, Weinstein SP,Arqer PH, Conant EF, " A review of breast ultrasound", J Mammary Gland Bio Neoplasia, 11 (2), pp 113-123, April 2006.
[3]. J. Alison Noble, Djamal Boukerroui, "Ultrasound Image Segmentation: A Survey", IEEE Transactions on Medical Imaging, Vol. 25, No. 8, pp 987-1010, Aug 2006
[4]. S.Kalaivani Narayanan and R.S.D.Wahidabanu, "A View on Despeckling in Ultrasound Imaging", International Journal of Signal Processing, Image Processing and Pattern Recognition, pp 85-98, Vol. 2, No.3, September 2009.
[5]. Christos P. Loizou, Constantinos S. Pattichis, Christodoulos I. Christodoulou, Robert S. H. Istepanian, Marios Pantziaris, and Andrew Nicolaides "Comparative Evaluation of Despeckle Filtering In Ultrasound Imaging of the Carotid Artery" IEEE transactions on ultrasonics, ferroelectrics, and frequency control, vol. 52, no.10, pp 46-50, October 2005.
[6]. H.B.Kekre, Pravin 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, Volume 32– No.7, October 2011, Available at: http://www.ijcaonline.org/archives
[7]. 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.
[8]. 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
[9]. R. M. Gray, "Vector quantization", IEEE ASSP Magazine., pp. 4-29, Apr.1984.
[10]. 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
[11]. 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.
[12]. 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.
[13]. 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.
[14]. Chin-Chen Chang, Wen-Chuan Wu, "Fast Planar-Oriented Ripple Search Algorithm for Hyperspace VQ Codebook", IEEE Transaction on image processing, vol 16, no. 6, June 2007.
[15]. 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.
[16]. 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.
[17]. 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.
[18]. Dr. H. B. Kekre, Dr.Tanuja Sarode, Ms.Saylee Gharge, Ms.Kavita Raut, "Detection of Cancer Using Vecto Quantization for Segmentation", Volume 4, No. 9, International Journal of Computer Applications (0975 – 8887), August 2010.
[19]. H. B. Kekre, Ms. Tanuja K. Sarode, Sudeep 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), Volume 9, Issue 5, pp.: 1-8, September 2009. Available online at http:// www.icgst.com/gvip/Volume9/Issue5/P1150921752.html.
[20]. 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)
[21]. 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
[22]. H. B. Kekre, Tanuja K. Sarode, "New Fast Improved Codebook Generation Algorithm for Color Images using Vector Quantization," International Journal of Engineering and Technology, vol.1, No.1, pp. 67-77, September 2008.

[23]. H. B. Kekre, Tanuja K. Sarode, “Speech Data Compression using Vector Quantization”, WASET International Journal of Computer and Information Science and Engineering (IJCISE), vol. 2, No. 4, pp - 251- 254, 2008. Available at: http://www.waset.org/ijcise

[24]. 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 Raisoni College of Engineering, Nagpur, India, 16-18 July 2008, Avaliable at online IEEE Xplore

[25]. Dr. 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.

AUTHORS

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.

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