Abstract

Breast cancer is the leading cause of death amongst women worldwide. However early detection can reduce the mortality rate substantially. B-mode ultrasound (US) imaging modality is one of the reliable tools used to detect and diagnose the breast cancer. In fact it can detect the cancer at an early stage where effective treatment is possible. Computer aided diagnosis system helps radiologists in the detection and diagnosis of the region of interest accurately. Image segmentation is one of the vital steps used in automated computer aided diagnosis system. Indeed accuracy of the overall diagnosis is depends on the detection and demarcation of region of interest. However US image characteristics such as varying echogenicity, heterogeneous texture patterns, irregular shape, fuzzy tumor boundary makes segmentation challenging. Moreover poor quality images due to inherent artifact such as speckle, attenuation makes segmentation more challenging. Many methods are suggested in the literature for speckle removal before implementation of region extraction algorithms. However in order to reduce the complexity of the segmentation, we suggested to omit speckle removal step deliberately from the proposed segmentation algorithm. Therefore we use original ultrasound images directly as input to the segmentation process.

During the literature survey and problem formulation we observed that US breast images have extreme random gray level distribution. This phenomenon is the major hurdle in achievement of accurate segmentation. We found that unsupervised learning (clustering) has a great potential in solving such problems. Here we proposed total six algorithms based on clustering for segmentation. Initially we proposed thresholding based clustering on texture feature images. In this
method texture has been analyzed by using selected texture parameters proposed by Haralick. Primarily texture feature images are generated using **Correlation, Variance, Sum variance** and **Sum average** parameters. Further local adaptive thresholding method has been applied to extract the region of interest. Although this method gives appropriate segmentation, due to high time complexity we anticipated new methods for segmentation using vector quantization (VQ).

Using VQ based clustering we have provided computationally efficient and accurate solution for segmentation. In this method LBG, KPE, KEVR, KMCG and KFCG codebook generation algorithms are used for clustering. Fundamentally image pixels are divided in to 8 clusters based on certain relationship. A novel cluster validation technique has been proposed based on experimentation to extract the region of interest. Eventually post processing is performed on the extracted region. Furthermore accuracy of the demarcation has been improved with reduce time complexity by proposing the modification on KMCG and KFCG algorithms. Here we tried 4 variant of this augmented algorithms. From the end results we found that augmented KFCG 3x3 is the most suitable algorithm for US breast images. It gives accurate and computationally efficient solution to the 72 (75%) images from the dataset. Eventually results obtained by this method are compared with well known Watershed and Marker Controlled Watershed algorithms. During the result validation, radiologists commented that our results are far better than these algorithms.

Furthermore to maximize the coverage of dataset we proposed statistical (i.e. probability and entropy) parameter based solution. In these techniques image gray level distribution has changed before applying VQ based clustering. These algorithms are found to be
suitable for difficult images of the dataset. Augmented KFCG 3x3 with probability and entropy images is selected as accurate solution for segmentation based on result analysis.

Afterward we suggested a novel region growing approach using VQ based clustering. In this technique a region has grown by automated initialization of the seed vector from the first cluster. Augmented KMCG 2x2 and augmented KFCG 2x2 algorithms are user for clustering and seed vector generation. This region based algorithm provides the solution for extremely difficult images where earlier proposed algorithms are failed.

To test the performance of proposed algorithms we used comprehensive dataset with the variety of images. These 92 images are categorized in to four types (i.e. Type-A to Type-D) by visually inspecting the structural characteristics in consultation with the expert radiologists. Eventually all the results are validated by the panel of three expert radiologists.