International Journal of Computer Applications (0975 – 8887)
Volume 52– No.15, August 2012

# Application of Wavelet based K-means Algorithm in Mammogram Segmentation

Shruti Dalmiya
Department of Electronics & Communication Engineering
St. Thomas' College of Engineering & technology
Kolkata, India

Avijit Dasgupta
Department of Electronics & Communication Engineering
St. Thomas' College of Engineering & technology
Kolkata, India

Soumya Kanti Datta
Research Engineer
EURECOM, France

## ABSTRACT
Research in image processing has gained lots of momentum during past two decades. Now-a-days image processing techniques have found their way into computer vision, image compression, image security, medical imaging and more. This paper presents a research on mammography images using wavelet transformation and K – means clustering for cancer tumor mass segmentation. The first step is to perform image segmentation. It allows distinguishing masses and micro calcifications from background tissue. In this paper wavelet transformation and K- means clustering algorithm have been used for intensity based segmentation. The proposed algorithm is robust against noise. In this case, discrete wavelet transform (DWT) is used to extract high level details from MRI images. The processed image is added to the original image to get the sharpened image. Then K-means algorithm is applied to the sharpened image in which the tumor region can be located using the thresholding method. This paper validates the algorithm by detecting tumor region from an MRI image of mammogram. The combination of noise-robust nature of applied processes and the simple K-means algorithm gives better results.

## General Terms
Image Processing

## Keywords
Image segmentation, Mammogram, K-means algorithm, Wavelet Transform.

## 1. INTRODUCTION
Many image analysis processes rely on image segmentation. It is a process of partitioning an image into different regions having same features [1] and is often used to extract region of interests (ROI). In medical imaging field, the technique is mostly used to detect tumor from MRI images of brain and mammogram.

The diagnosis of human being has been improved significantly with the arrival of Computed Tomography (CT), Positron Emission Tomography (PET), and Magnetic Resonance Imaging (MRI). Medical imaging provides a reliable source of information of the human body to the clinician for use in fields like reparative surgery, radiotherapy treatment planning, stereotactic neurosurgery etc. Several new techniques have been devised to improve the biomedical research. MRI is a non-invasive technique for medical imaging that uses the magnetic field and pulses of radio waves. It gives better visualization of soft tissues of human body. In this paper we have concentrated on mammogram images.

Breast cancer is one disease which threatens the lives of many women all over the world. Small clusters of micro calcifications appearing as collection of white spots on mammograms are early warnings of breast cancer. Primary prevention presently seems impossible since the causes of this disease are yet to be found out. Thus an improvement in the early diagnostic techniques of breast cancer is an extreme necessity. Mammography is the main test used for screening and early diagnosis. Early detection is performed on X-ray mammography which provides better results for breast cancer prognosis. Several computer-aided diagnosis (CAD) schemes have been developed to improve the detection of primary signatures of this disease: masses and micro calcifications. Masses are space-occupying lesions, described by their shapes, margins, and denseness properties. A benign neoplasm is smoothly marginated, whereas a malignancy is characterized by an indistinct border that becomes more speculated with time. Because of the slight differences in X-ray attenuation between masses and benign glandular tissue, they appear with low contrast and often very blurred. Micro calcifications are tiny deposits of calcium that appear as small bright spots in the mammogram.

In this paper a hybrid technique is used that makes use of DWT and K-means algorithm. By using DWT we extract the high pass image and then this image is applied as the input of K-means algorithm for segmentation. Our proposed method utilizes the advantage of noise-robust nature of wavelet and the simplicity of K-means algorithm which results in better detection of tumor in mammogram MRI images.

Our paper is organized as follows. In section 2 we present related works. Section 3 encompasses the basic aspects of wavelet algorithm, K-means algorithm and our proposed algorithm. Section 4 demonstrates the results while section 5 concludes the paper.

## 2. RELATED WORK
A comprehensive review of existing literature reveal several MRI image classification methods. In this paper we have broadly classified into three categories which are discussed below.

**Region Growing Techniques -** Reference [2] proposes a hybrid technique for breast cancer detection by combining morphological operators and fuzzy C-means algorithm. But this algorithm is not suitable for noisy MRI images. Different types of region growing techniques for tumor detection are also very common for MRI image analysis [4]. These techniques require some initial center points to grow up the regions. Li et al. proposed a watershed algorithm for brain segmentation in [5]. This is a gradient based technique and it relies on image contrast which can be degraded during image acquisition and yields to inaccurate results. So these approaches are not attractive for medical images as these could not handle in-homogeneity in MRI images. Ahmed and Mohammad [6] proposed a segmentation technique for tumor detection by employing K-means clustering and Perona Malik Anisotropic Diffusion Model. Tumors are extracted on the basis of the resultant cluster values. The main disadvantage of this algorithm is its sensitivity to false edges.

**Thresholding Based Techniques-** Many thresholding based techniques can be found in the literature. One such technique is presented by Suzuki and Toriwaki in [3] which proposes a knowledge guided thresholding technique for brain tumor segmentation. The problem with thresholding techniques is that it is normally difficult to determine any threshold value for tumor segmentation because intensities in MRI images are normally scattered; wrong threshold selection can either neglect tumor portion or label many healthy parts as tumors. So these kinds of techniques are not reliable.

**Frequency Domain based Techniques-** Now-a-days techniques in frequency domain have gained more interests as they produce better quality provision in their results. Mostafa et al. proposed image segmentation using wavelet based multi resolution Expectation Maximum (EM) algorithm [8]. But the main drawback of this algorithm is that it is based on identical and independent distribution of pixel intensities which may not be the case with noisy images.

**Table1: Comparative study of different techniques**

| **Techniques** | **Advantages** | **Disadvantages** |
|---|---|---|
| **Region Growing** | Utilizes the advantage of Morphological Operators | Not suitable for noisy image data |
| **Thresholding Based** | Very simple methods are available and time complexity is very low. | Difficult to determine any threshold value for tumor extraction as MRI images has scattered intensities. |
| **Frequency Domain based** | It can produce better quality provision. | But they are not robust for noisy images. |

From the next section, we present the details of our proposed algorithm.

# 3. BASIC CONCEPT OF PROPOSED METHOD

## 3.1 Wavelet Transform

The wavelet transform [9] is important to provide a compact description of images that are limited in time and it is very helpful in description of edge and line that are highly localized.

A 1-level wavelet transform of a discrete image f can be done by using the following two steps

- **Step1:** Perform a 1-level, 1D wavelet transform, on each row of f, thereby producing a new image.
- **Step2:** On the new image obtained from Step 1, perform the same 1D wavelet transform on each of its columns.

It can be easily shown that Steps 1 and 2 could be done in reverse order and result would be the same. A 1-level wavelet decomposition of an image f can be defined as follows

$$f \rightarrow \left(\frac{h^1 \mid d^1}{a^1 \mid v^1}\right) \qquad \ldots\ldots\ldots(1)$$

Where $h^1$, $a^1$, $d^1$, and $v^1$ are subimages each have $\frac{M}{2}$ rows and $\frac{N}{2}$ columns. Where $M$ and $N$ are the no of rows and columns respectively.

The sub image $a^1$ is created by computing trends along rows of f followed by computing trends along columns; so it is an averaged, lower resolution version of the image f .

The $h^1$ subimage is created by computing trends along rows of the image f followed by computing fluctuations along columns. We shall refer this sub image as measure of horizontal fluctuation.

The sub image $v^1$ is similar to $h^1$, except that the roles of horizontal and vertical are reversed. We shall refer this sub image as measure of vertical fluctuation.

Finally, there is the first diagonal fluctuation, $d^1$. This subimage tends to emphasize diagonal features, because it is created from fluctuations along both rows and columns.

## 3.2 K-means Algorithm

K-means [10] algorithm is a simple but elegant segmentation method. The main advantage of K-means algorithm is its simplicity. Speed of execution is very high. But the problem with K-means algorithm is that if the initial cluster centers are chosen incorrectly this algorithm may not converge. This happens in the case of noisy image mostly.

K-Means algorithm is an unsupervised clustering algorithm that classifies the input data points into multiple classes based on their inherent distance from each other. The algorithm assumes that the data features form a vector space and tries to find natural clustering in them. The points are clustered around centroids $\zeta_i \forall i = 1,2, \ldots. k$ which are obtained by minimizing the objective

$$V = \sum_{i=1}^{k} \sum_{x_j \epsilon S_i} \left(x_j - \zeta_i\right)^2 \ \ldots\ldots\ldots(2)$$

Where there are $k$ clusters $S_i$ , $i = 1,2, \ldots. k$ and $\zeta_i$ is the centroid or means point of all the points $x_j \epsilon S_i$ .

Steps of K-means algorithm is described as follows

1. Compute the intensity distribution (also called the histogram) of the intensities.
2. Initialize the centroids with k random intensities.
3. Repeat the following steps until the cluster labels of the image do not change anymore.
4. Cluster the points based on distance of their intensities from the centroid intensities.
$$C^i = arg\min_j \left\|x^{(i)} - \zeta_j\right\|^2 \quad \ldots\ldots\ldots..(3)$$
5. Compute the new centroid for each of the clusters.
$$\zeta_i = \frac{\sum_{i=1}^{m} 1\{c_i = j\} x^i}{\sum_i^m 1\{c_{i-} = j\}} \qquad \ldots\ldots..(4)$$

Where k is the parameter of the algorithm (the number of clusters to be found), $i$ iterates over the all intensities, j iterates over all the centroids and $\zeta_i$ are the centroid intensities.

## 3.3 Wavelet Based K-means Algorithm

Our proposed method is a five step process which is graphically illustrated in Figure . 1. All the steps is elucidated below

1. **Step1:** Take MRI image of an mammogram as an input.
2. **Step2:** Apply wavelet transform on the MRI image to obtain wavelet decomposed image resulting in four subbands. These are the LL (Lower resolution version of image), LH (Horizontal edge data), HL (Vertical edge data), & HH (Diagonal edge data) subbands representing approximation, horizontal, vertical and diagonal components in the form of coefficients, respectively. LL subband contains low level and the other three (LH, HL, and HH) contain high level details.
3. **Step3:** Set approximation coefficients in LL equal to zero and apply inverse wavelet transform to obtain a high pass image from the remaining (horizontal, vertical and diagonal) subbands. We call the resultant image level-1 (L1) detail image.
4. **Step4:** Add L1 to the original image to get a sharpened image.
5. **Step5:** Apply K-means algorithm for segmentation of sharpened image.
6. **Step6:** Apply thresholding method to detect tumor.

We applied Discrete Wavelet Transform (DWT) to MRI images because wavelets provide frequency information as well as time-space localization. In addition, their multi-resolution character enables us to visualize image at various scales and orientations. The multi-resolution property provides information about various high frequency components at different levels of decomposition. Over-decomposition should however be avoided, because as the decomposition levels increase, there is a great risk that lower frequencies become a part of detail components. This may restrict us to use only fewer level of decomposition because lower frequencies will become part of high pass image and reduce effective detail in an image.

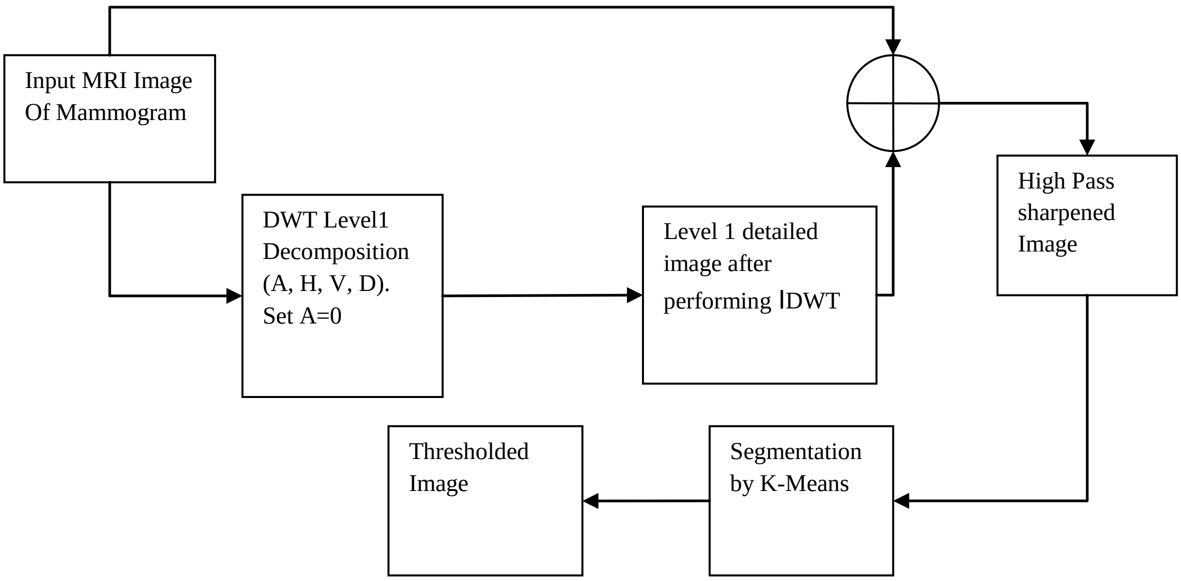


**Figure 1: Block Diagram of our Algorithm**

## 4. Results and Analysis

We have applied our proposed algorithm on a large set of MRI images of mammograms. In this paper we are discussing one example. The original input image is shown in Figure 2.

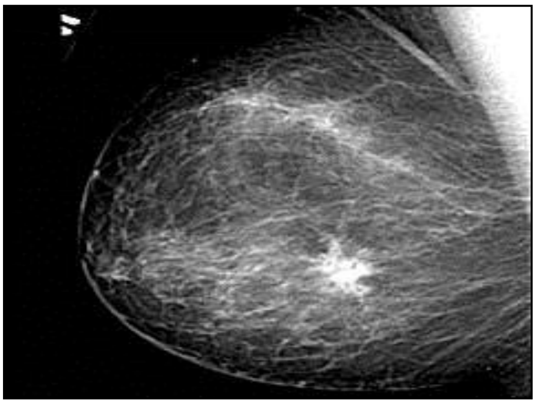

**Figure 2: Original Input Image**

This original image is decomposed by using DWT at level-1 and it gives detail image by setting LL to zero and thereafter applying IDWT on the detail image.

Then the resultant image is then added to the original image to get the sharpened image. The sharpened image is shown in Figure 3.

**Figure 3: sharpened Image**

Now the sharpened image is applied as the input of the K-means algorithm. The segmented image using K-means is shown in Figure 4.

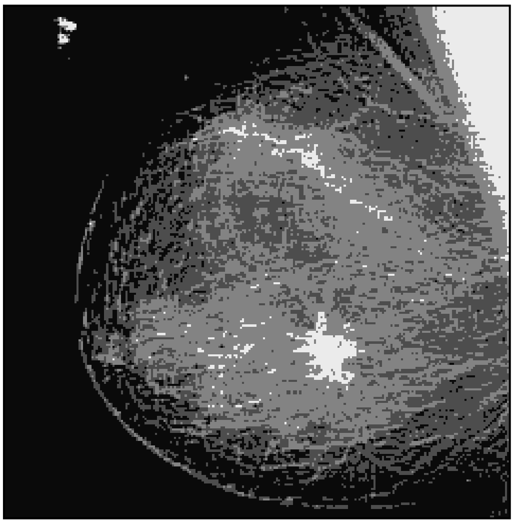

**Figure 4: Segmented image of sharpened image by proposed method**

In order to draw a comparison we presented results of segmentation of Figure .1 using only k-means and Fuzzy C-means (FCM) is shown in Figure 5 and Figure 6.

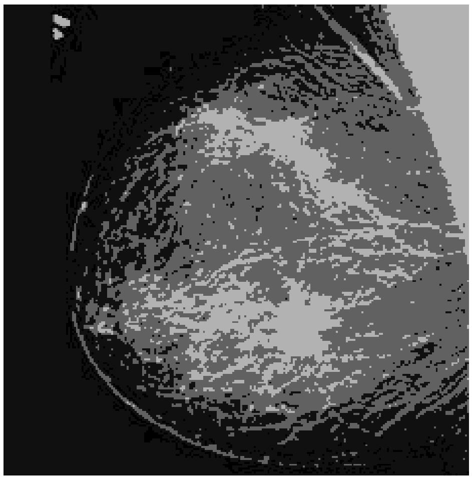

**Figure 5 Segmented by K-means algorithm**

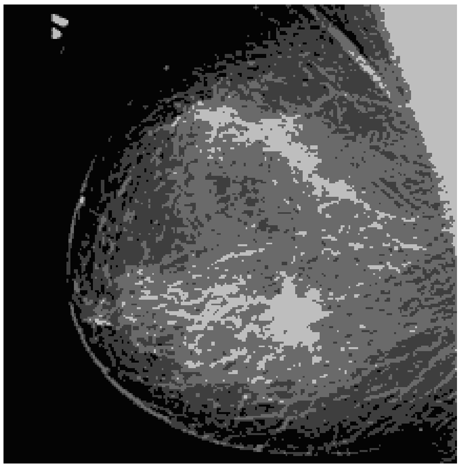

**Figure 6 Segmented by Fuzzy C-means algorithm**

It can be easily seen that segmentation of tumor region is better in our algorithm. It is slightly improved in FCM but it is also not feasible.

Now the thresholding is applied to locate the tumor. Figure 7 shows the thresholded image of Figure 5 and Figure 8 shows the thresholded image of Figure 6 and Figure 9 shows the thresholded image of Figure 4.

**Figure 7 Image obtained after thresholding Figure 5**

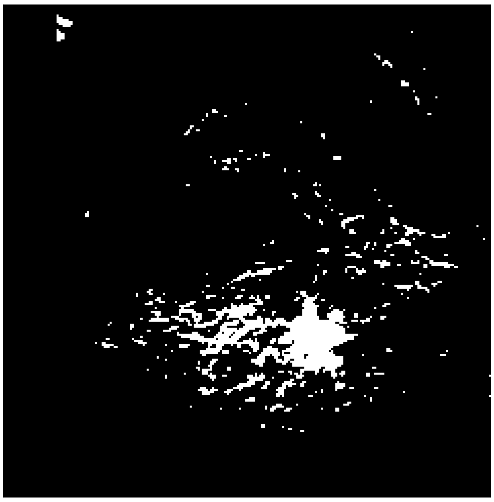

**Figure 8 Image obtained after thresholding Figure 6**

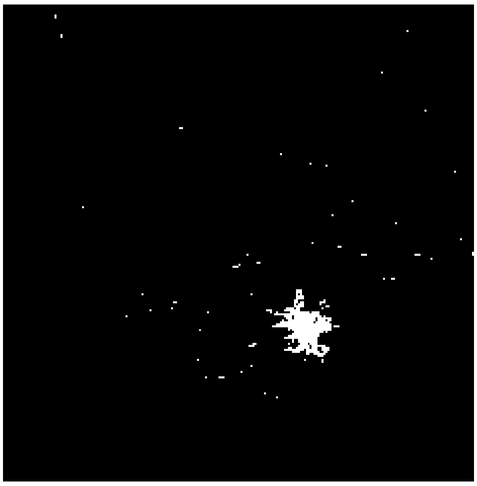

**Figure 9 Image obtained after thresholding Figure 4**

It can be easily seen that tumor region is clearly located in Figure 9. So our algorithm clearly defined the region of tumor results in better segmentation mammogram MRI images.

**Table2: Comparative analysis of results**

| Algorithm | Advantages | Disadvantages |
|---|---|---|
| K-means | Simple and Less Computation is required | Efficiency is less in noisy MRI images. |
| Fuzzy C-Means | Utilizes the advantages of Fuzzy Set over the Crisp Set | Produce better result than K-means but not Robust to noisy images |
| Proposed Algorithm | Utilizes the advantages of discrete Wavelet Transform. It gives better result if we compare it with Fig. 7,8,9. | It is robust to noisy images. |

## 5. CONCLUSIONS

This paper reported a methodology of segmentation of MRI images using wavelet and K-means algorithm. Wavelet transform made the algorithm noise free because wavelets provide frequency information as well as time-space localization. In addition, their multi-resolution character enables us to visualize image at various scales and orientations. Resolution reduction using wavelet depends on the amount of noise as well as the area of the target. Then k-means was applied to segment the mammogram. K-means provides a very simple and efficient method of segmentation. Thereafter a thresholding method has been employed to detect the tumor region. We proved our result is better by comparing with other two methods.

On the other hand, this paper has shown that advanced technique of image processing and micro calcification detection which is useful in computer aided diagnosis. The intelligent systems development combined with health specialists' knowledge improve diagnostics associated to different pathologies.

This method can be easily extended for brain tumor segmentation. In future we may seek to employ this method on SAR images and may improve this algorithm accordingly.

## 6. REFERENCES

[1] Rafael C. Gonzalez, Richard E. Woods, Digital Image Processing, Pearson Education, 2002..

[2] S. Saheb Basha, Dr. K. Satya Prasad , Automatic detection of breast cancer mass in mammograms using morphological operators and Fuzzy C-means Clustering, Journal of Theoretical and Applied Information Technology,2009.

[3] Suzuki, H. and J. Toriwaki, 1991. Automatic Segmentation of Head MRI Images by Knowledge Guided Thresholding. Computer Med. Imaging Graph: The official J. Computerized Imaging Society, 15 (4): 233-240.

[4] Robb, R.A., 2000. Biomedical Imaging, Visualization and Analysis. Wiley-Liss, USA.

[5] Li, N., M. Liu and Y. Li, Image segmentation algorithm using watershed transform and level set method. In Proc. ICASSP 2007. IEEE International Conference on Acoustics, Speech and Signal Processing (2007), 1: I-613-I-616.

[6] Ahmed, M. and D.B. Mohammad, 2008. Segmentation of Brain MR Images for Tumor Extraction by Combining K-means Clustering and Perona Malik Anisotropic Diffusion Model. International Journal of Image Processing, 2: 27-34.

[7] Mostafa, M.G., M.F. Tolba, T.F. Gharib and M.A. Mageed, Medical Image Segmentation Using Wavelet Based Mutiresolution EM Algorithm.

[8] Jung, C., R., "Combining wavelets and watersheds for robust multi scale image segmentation", Image and Vision Computing 25(2007), pp. 24–33.

[9] JA Hartigan, Clustering Algorithms, John Wiley & Sons, New York, 1975.

[10] Hartigan, J.A., Wong, M.A.: Algorithm AS136: A K-Means Clustering Algorithm. Applied Statistics 28, pp. 100–108 (1979).