ISSN 2348-375X

Unique Journal of Engineering and Advanced Sciences

Available online: www.ujconline.net

Research Article

BREAST TUMOR SEGMENTATION AND CLASSIFICATION USING SVM AND BAYESIAN FROM THERMOGRAM IMAGES

Dinsha D[1*], Manikandaprabu N[2]

[1]PG Scholar, Department of Electronics and Communication Engineering, Hindusthan College of Engineering and Technology, Anna University Chennai, India
[2]Assistant Professor, Nandha Engineering College, Erode, India

Received: 09-03-2014; Revised: 07-04-2014; Accepted: 05-05-2014

*Corresponding Author: **Dinsha D**
PG Scholar, Department of Electronics and Communication Engineering, Hindusthan College of Engineering and Technology, Anna University Chennai, India
Email: dinsha124@gmail.com

ABSTRACT

Breast cancer is one of the most important causes of death among women in the world. Mammography is so far the most common modality for the screening and diagnosis of breast tumor. However they have their limitations especially in young women with dense breasts and this necessitated the development of novel, more sensitive and specific strategies. There are no effective ways to prevent cancers and the only possible way of saving lives is early detection. Breast thermography uses thermal images of the breasts to help in the early diagnosis and detection of breast cancer. An abnormal thermogram has proven to be a reliable indicator of high risk of breast cancer in its early stages. Here initially the pre processing of the thermogram images are done wherein they are enhanced using the CLAHE method. The enhanced images after filtering are segmented using k means and fuzzy C means. The features have been extracted and these are used for classification for both the segmentation methods. Finally a comparison has been made by using the SVM and Bayesian classifiers.

Keywords: Breast tumor, Thermography, CLAHE, fuzzy C means.

INTRODUCTION

Breast cancer is the second most often cancerous cause of death among women and the second most frequent type of cancer[13]. It has been proved that breast cancer can be treated effectively if they are diagnosed at an early stage. This is the reason why the early detection is important. There are different modalities used for breast tumor diagnosis, each having their own advantages and limitations. Thermography is brilliantly simple and harmless and can detect physiological changes resulting from early cancer growth. They can be used for diagnosing abnormality in humans. Thermographs measure the infrared heat emitted by body and translates this into thermal images.

Thermography can be used as a complement for X-Ray mammography in the detection of breast cancer. A tumor can cause changes in the temperature distribution which in turn causes thermal asymmetry between breasts. It is supposed that human body has thermal symmetry, and so thermal variations in both breasts should be similar. The metabolic activity in case of both pre cancerous tissue and the area surrounding a developing breast cancer is always higher than in a normal breast tissue[14]. This is because for the cancerous tumors to grow, they require more nutrients and this is supplied by an increased blood flow. Thus the amount of blood flow into the tumor region increases. This process results in an increased regional surface temperature. These temperature variations can be captured by an ultra sensitive medical infrared camera which can then be used to analyse and produce high resolution images of these temperature variations. In comparison to breast thermography other radiology techniques are at least 8 years[14] too late because the cancer needs to be large enough to be detected. This is one major limitation of mammograms because the tumors can be detected only once they have reached a specific size however the tumor might have already been formed even years before this but remains undetected. This is where thermography plays a vital role.

The paper is structured as follows. Firstly, a small description on the existing methods is given in the related works. Then the proposed framework is explained which includes the pre-processing, segmentation, feature extraction and classification. Finally, the results and conclusion are given to evaluate the performance of the segmentation and classification techniques.

RELATED WORKS

A computer-aided approach for automating asymmetry analysis of thermograms has been developed in Detecting Breast Cancer from Thermal Infrared Images by Asymmetry

Analysis[4]. The use of thermogram images for breast cancer detection and the advantages of thermograms over traditional mammograms are studied. The Hough transform was used for segmentation. Pattern classification was done using unsupervised k means and supervised kNN.

A survey of segmentation in mass detection algorithm for mammography and thermography has been discussed in A Survey of Segmentation in Mass Detection Algorithm for Mammography and Thermography[6]. Thermal imaging followed by mammography was used to screen patients with breast cancer. Thresholding, region based methods, edge detection techniques were used in the segmentation of mammogram images. While thermal images were segmented using 3 methods based on grey threshold, RGB and K-mean technique.

An automatic diagnosis technique for the assessment of breast cancer based on thermograms using texture features and the SVM classifier has been developed in Thermography Based Breast Cancer Detection Using Texture Features and Support Vector Machine[12]. Four texture features namely moment1, moment3, run percentage, and gray level non uniformity have been used for representing the thermograms.

An objective and computational approach for detection of tumor using feature extraction and ANN has been described in Automatic analysis of breast thermograms for tumour detection based on bio statistical feature extraction and ANN[10]. An automated method for identifying ROI in thermal images is devised using canny edge detector and gradient operator.

A new approach for automatic segmentation of Region of Interest and asymmetry analysis of breast thermograms is implemented in Image Segmentation and Asymmetry Analysis of Breast Thermograms for Tumour Detection [9]. Canny edge detection operator and gradient operator are used to first segment the region of interest. Further asymmetry analysis is performed according to seven extracted features. The experimental results show that HOS parameters namely skewness, kurtosis, entropy and joint entropy combined with centre calculation and histogram analysis is promising for detection of breast abnormality.

A new approach for ROI extraction based on topological derivative is presented in Topological Derivative Applied to Automatic Segmentation of Frontal Breast Thermograms [5]. First, the method uses a low-pass filter and the topological derivative to get a rough definition of the region of interest (ROI). Then, the previous result is used to initialize the method proposed by Marques.

Proposed Framework

This section describes the flow of the proposed system. An overview of the method is shown in Fig. 1. It includes pre-processing which includes conversion of colour images to grayscale, enhancement using CLAHE, denoising, segmentation using k means and fuzzy c means, feature extraction and classification. The aim of this project is to compare the performance of segmentation using k means and FCM from thermogram images using MATLAB software.

Pre-Processing

As part of the pre processing steps the thermogram images are first converted to their gray level images which are used for the rest of the processing steps. The images obtained may be comparatively low quality images. Usually medical images are weak in terms of contrast. So in order to enhance the image pre-processing stage plays a vital role prior to segmentation.

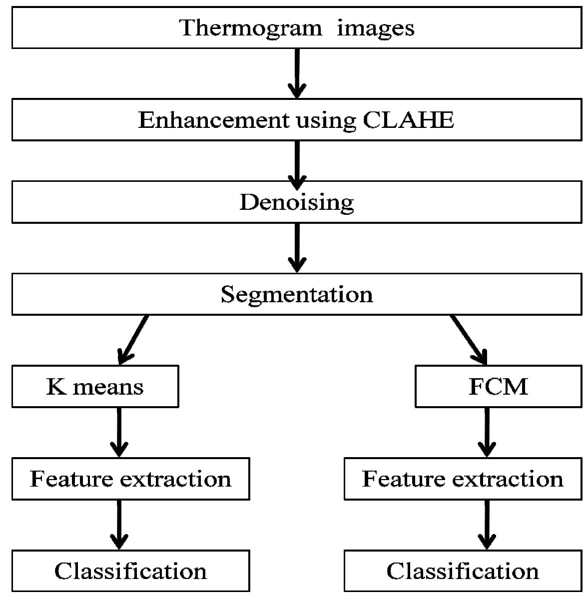

Figure 1: Schematic of proposed method

Enhancement using CLAHE

CLAHE (Contrast Limited Adaptive Histogram Equalization) is a modified version of adaptive histogram equalization and is used for the enhancement of low contrast images. It evens the distribution of used gray levels and makes the hidden features more visible. The gray channels of the histogram are enhanced by using CLAHE. It operates on small regions in images called tiles, and each tiles contrast is then enhanced. The neighbouring tiles are then combined using bilinear interpolation. This method helps to prevent over amplification of noise which is a major problem with adaptive histogram equalization (AHE).

Denoising

The contrast enhanced image will contain some noises. In order to remove the noise and at the same time enhance the edges without amplifying the noise, a bilateral filter is used. It is a non-linear filtering technique and extends the concept of Gaussian smoothing by weighting the filter coefficients with their corresponding relative pixel intensities [3]. The pixel's values are updated using a weighted average of pixels which are judged to be most similar to it. Pixels different in intensity from the center are weighted less though they are in close proximity to each other.

Segmentation

The segmentation of the enhanced filtered images is carried out using two separate methods. One using k means segmentation and the other using fuzzy C means segmentation. The main idea of the image segmentation is to group pixels in homogeneous regions and the usual approach to do this is by common feature. In image segmentation, clustering algorithms are very popular as they are intuitive and are also easy to implement[2].

The K -means clustering algorithm is one of the most widely used methods. It is an iterative technique used to partition an image into k clusters. K-means is one of the simplest unsupervised learning algorithms.
Mathematical representation for k-means:
For a given image, compute the cluster means m:

$$M=\sum_{i:c(i)=k} X_i / N_k \quad (1)$$

Now, calculate the distance between the cluster centers to each pixel:

$$D(i)=\arg\min ||x_i - M_k||^2 , i=1,...N \quad (2)$$

Repeat the above steps until mean value converge.
Algorithm for k means:

1. Give the number of cluster value as k.
2. Randomly choose the k cluster centers.
3. Calculate mean or center of the cluster.
4. Calculate the distance between each pixel to each cluster center.
5. If the distance is near to the center then move to that cluster.
6. Otherwise move to next cluster.
7. Re-estimate the center.
8. Repeat the process until the center doesn't move.

The next method is FCM. Fuzzy C-means (FCM) algorithm is one of the most popular fuzzy clustering techniques. FCM is able to determine, and in turn, iteratively update the membership values of a data point with the pre-defined number of clusters[8]. Thus, a data point can be the member of all clusters with the corresponding membership values. The FCM employs fuzzy partitioning such that a data point can belong to all groups with different membership grades between 0 and 1[11].
Mathematical representation for FCM:
It is a clustering algorithm and is based on reducing the following function:

$$Y_m=\sum_{i=1}^{N}\sum_{j=1}^{C} M_{ij}^{m} \|x_i - C_j\| \quad (3)$$

Where,
m : any real number greater than 1.
M_{ij}: degree of membership of X; in the cluster j.
x_i : data measured in d-dimensional.
R_J: d-dimension center of the cluster.
The update of membership M_{ij} and the cluster centers R, are given by:

$$M_{ij}=\frac{1}{\sum_{k=1}^{N}\left(\frac{\|x_i - x_j\|}{\|x_i - c_j\|}\right)^{2/m-1}} \quad (4)$$

$$R_j=\sum_{i=1}^{N} x_i \cdot M_{ij}^{m}/\sum_{i=1}^{N} M_{ij}^{m} \quad (5)$$

The above process ends when,

$$Max_{ij} |M_{ij}^{(k+1)} - M_{ij}^{k}| < \delta \quad (6)$$

Where,
δ : termination value or constant between 0 and 1.
k : number of iteration steps.
Algorithm for FCM:

1. Initialize $M=[M_{ij}]$ matrix, M(0)
2. At k-step: calculate the centers vectors $R^{(k)}=[R_j]$ with $M^{(k)}$
$$R_j=\sum_{i=1}^{N} x_i \cdot M_{ij}^{m}/\sum_{i=1}^{N} M_{ij}^{m} \quad (7)$$
3. Update $U^{(k)}$, $U^{(k+1)}$
$$M_{ij}=\frac{1}{\sum_{k=1}^{N}\left(\frac{\|x_i - x_j\|}{\|x_i - c_j\|}\right)^{2/m-1}} \quad (8)$$
4. If $\| M^{(k+1)} - M^{(k)} \| < \delta$ then STOP, otherwise return to step 2.

The computational time for both the segmentation methods have been calculated and tabulated.

Feature extraction

After the segmentation, the features can be extracted from each of the two segmented images for further classification. Feature extraction is done using gray Level Co-occurence matrix(GLCM). Features like contrast, correlation, energy, entropy, homogeneity can be used. Contrast measures the local variations in the gray-level co-occurrence matrix. Homogeneity measures the closeness of the distribution of elements in the GLCM. Its range is from 0 to 1. Homogeneity is 1 for a diagonal GLCM. Entropy gives the measure of randomness. The segments with hotspots should be having lesser entropy. These values provide a further insight into the type of tumor or the class of tumors which they are likely to fall into.

Classification

With the help of the features that are extracted in the previous step the classification can be done. Here two different methods of classification have been used namely SVM classifier and Bayesian classifier.
Support Vector Machine (SVM) is primarily a classier method that performs classification tasks by constructing hyperplanes in a multidimensional space that separates cases of different class labels. Bayesian classifier is a simple probabilistic classifier based on applying bayes theorem with strong independence assumptions. Depending on the precise nature of the probability model, naive Bayes classifiers can be trained very efficiently in a supervised learning setting. The results of both these classifiers can be compared by using parameters like specificity, sensitivity, accuracy.

RESULTS

The proposed algorithm is run using Matlab software and the inputs used are 9 datasets of breast thermogram images. The output images are shown as figures below.

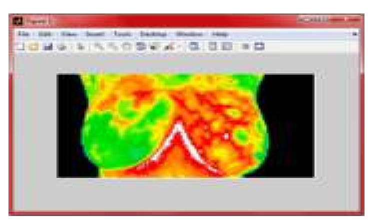

Figure 2: Input thermogram image

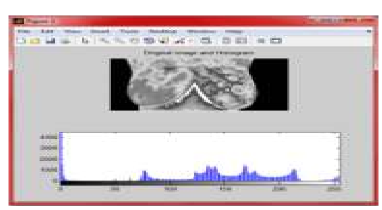

Figure 3 : Input gray level thermogram image.

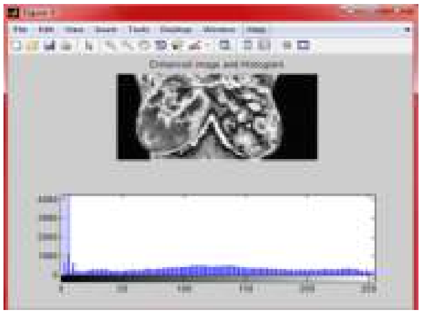

Figure 4: Enhanced image after CLAHE.

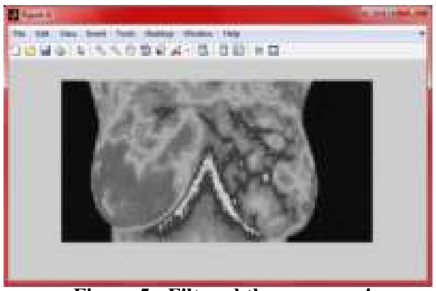

Figure 5 : Filtered thermogram image.

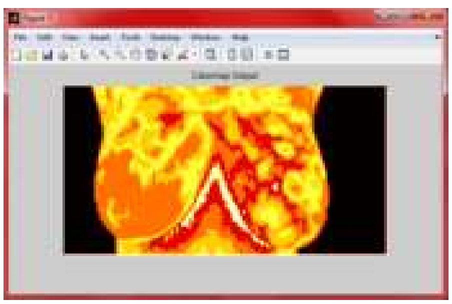

Figure 6: K means segmentation using colormap.

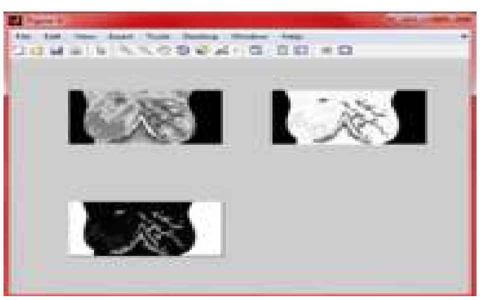

Figure 7: FCM segmented result

Figure 8: Classifier outputs

The two segmentation algorithms have been compared in terms of the time computed for different sets of thermogram images. The time taken in seconds for 5 different images have been tabulated below.

Table 1: Computational time for FCM and K Means algorithm

Algorithm	1 (s)	2 (s)	3 (s)	4 (s)	5 (s)
FCM	76.794	35.052	37.204	41.560	85.094
K Means	7.827	14.711	13.993	11.626	11.699

The classification performance can be evaluated in terms of the parameters like accuracy, sensitivity and precision and the performance of the two classifiers SVM and Bayesian have been tabulated.

Table 2: Parameter comparison for SVM and Bayesian Classifier

Parameter (%)	SVM	NB
Accuracy	85.71	92.86
Sensitivity	92.31	92.93
Precision	92.31	92.86

CONCLUSION

This paper is aimed at the segmentation of breast tumor from thermogram images using k means and FCM methods. The images are enhanced using contrast limited adaptive histogram equalization and filtered by using bilateral filters and then they are segmented. The use of thermogram images are not restricted to breast but can be used for diagnosis of other body parts to detect the problems and abnormalities. However the methods used must be varied depending upon the applications. This can be taken as a scope for future work.

REFERENCES

1. Moon WK, Shen YW, Bae MS, Computer aided tumor detection based on multi scale blob detection algorithm in automated breast ultrasound images,IEEE transactions on medical imaging, 2013; 32(7).
2. Chandhok C, Chathurvedi S, A. Khurshid AA, An approach to image segmentation using K-means clustering algorithm", International Journal of Information Technology(IJIT),2012; 1(1).
3. Goyal G, Singhal M, Bansal A, Comparison of denoising filters on colour TEM image for different noise", International Journal of Image Processing and Vision Sciences(IJIPVS) ISSN:2278-1110,2012; 1(3).
4. Qi H, Kuruganti PT, Wesley E. Snyder Detecting breast cancer from thermal infrared images by asymmetry nalysis".
5. Machado D. Giraldi AG, Novotny AA , Topological derivative applied to automatic segmentation of frontal breast thermograms,2011.
6. Manoj R and Thamarai M, A Survey of segmentation in mass detection algorithm for mammography and thermography, ISSN (Print): 2278-8948, 2012;1(2).
7. Manikandaprabu N, Lalli G, Marimuthu CN, A Development of knowledge based inferences system for detection of breast cancer on thermogram images", International conference on computer

communication and informatics (ICCCI -2014), 2014;03-05.

8. Ganeshan P, Chakravarthy P, Shweta Verma S, Segmentation of natural colour images in HSI color space based on FCM clustering", International Journal of Advanced Research In Computer Engineering and Technology, (IJARCET), 2013; 3: 3.
9. Kapoor PP, Prasad AV, Patni S, Image segmentation and asymmetry analysis of breast thermograms for tumor detection, International Journal of Computer Applications (0975 – 8887), 2012 50(9).
10. Kapoor PP, Prasad AV, Patni S, Automatic analysis of breast thermograms for tumour detection based bio statistical feature extraction and ANN, International Journal of Emerging Trends in Engineering and Development, 2012; l7(2).
11. Suganya R, Shanthi R,Fuzzy C means algorithm- A Review, International Journal of Scientific and Research Publications, 2012 ;2(11).
12. Acharya RU, Kong EY,Tan JH and Vinitha Sree S, Thermography based breast cancer detection using texture features and support vector machine, Springer ScienceBusiness Media, LLC 2010, 2012.
13. Vinitha Sree S, Kwee EY, Breast Imaging- A Survey, WJCO, 2011; 2: 4.
14. International academy of clinical thermology, Breast Thermography -2000IACT.

Source of support: Nil, Conflict of interest: None Declared