ISSN 2229-5518

Bilateral Analysis of Mammograms for False Positive Reduction Using Eigen ROI Approach

Shilpa Gopalakrishnan, Lavanya R

Abstract— Mammogram is the widely accepted imaging method used for screening breast cancer. Computer aided detection (CAD) systems intend to provide assistance to the radiologists in detecting breast cancer. However the detection of masses by CAD system is prone to a high rate of false positives. Radiologists often use bilateral analysis ie, symmetry between left and right breast to improve detection performance. In this work, a computer aided bilateral image analysis based on eigen region of interest (ROI) approach is used in order to reduce the false positive rate. It has been found that the proposed method outperforms the conventional false positive reduction techniques.

Index Terms— Bilateral analysis, CAD system, Eigen ROI, Euclidean distance, False positive, Mass, Principal component analysis.

1 Introduction

Masses are the most common form of breast cancer, and also the most difficult to evaluate due to the great variability in their appearance and shapes [1]. They are dense regions of different sizes and properties. Mammography is the best way to detect breast cancer early. However, mammograms are difficult to interpret because of their subtle and complex nature. So double reading by two different radiologists has been sometimes opted to overcome this problem. But the process is time consuming and may result in ambiguous diagnostic results. Interpretation of mammograms by radiologist usually has low sensitivity and specificity.

Computer aided detection (CAD) systems improve the sensitivity in detection of the abnormalities when compared to radiologist. This is at the cost of an increased false positive rate, leading to unnecessary biopsies [2]. However to reduce the number of false positives, a classifier is used in the second stage following detection.

Unilateral and bilateral analysis are two different methods for false positive reduction. In unilateral image analysis, information from only one of the mammogram pair is used, either left or right. But in bilateral image analysis, information from both right and left mammograms are used[3].

In this paper a new CAD system for false positive reduction based on bilateral analysis is proposed. The proposed system employs eigen region of interest (ROI) approach using one dimensional principal component analysis (1D PCA). This method is actually inspired form the eigen face approach for face recognition and face verification.

- *Shilpa Gopalakrishnan is currently pursuing masters degree program in vlsi design in Amrita VishwaVidya Peetham ,Coimbatore, India, E-mail: shilpagopalakrishnan@gmail.com*
- *Lavanya R working as assistant professor in Amrita VishwaVidya Peetham, India E-mail: laavraaj@yahoo.com*

2 Related Work

Bilateral image asymmetry analysis of mammograms is a well known procedure in breast cancer diagnosis by radiologists [4]. The asymmetries between left and right breasts may represent the signs of breast cancer. The structures that appear similar in bilateral (right and left) mammograms at corresponding locations are more likely to be false positives. To mimic this procedure CAD system compute the difference features from both left and write mammograms. The normal regions that resemble lesions will result in small difference, whereas larger value indicate asymmetry and hence abnormality. The difference features are given as input to a classifier for classification of ROIs. Many researchers have adopted bilateral analysis for automated reduction of false positives.
Yin, F.Giger et al. [5] used bilateral subtraction of mammogram to detect the masses. The breast border is tracked with a four-point connectivity tracker. The tracked border is further smoothened to define the breast region. The border matched images are subtracted. Then features are extracted from these images using various feature extraction techniques and subsequently a border test is applied to detect if it is cancerous.

Bovis and Singh [6] have also proposed bilateral subtraction technique to identify the suspicious regions. The right and left images are aligned before subtraction. Nipple is used as the reference point for breast alignment. A region splitting technique is used to segment the ROI. Once regions are identified, false positives are removed on the basis of region characteristics like shape, size, difference in homogeneity and entropy. This is followed by classification using neural networks. The use of textural features is found to be highly discriminant in differentiating between masses and non-masses yielding a very high true positive fraction (TPF) for a low false positive fraction (FPF).

Dong and Senad [7] have employed a computerized method for automated breast cancer detection using bilateral image feature subtraction. Texture features have been extracted from the mammogram pairs and the difference is determined. Finally, Naïve Bayes classifier is used to classify mammograms as normal or abnormal.

In contrast to the feature based approaches, holistic pattern recognition problem for image classification are based on the entire image rather than local features [8]. In the simplest version of the holistic approaches, the image is represented as a 2D array of intensity values. The major drawback of the holistic method is that they attempt to perform classification in a space of very high dimensionality. To counter this curse of dimensionality, several other schemes have been proposed that employ statistical dimensionality reduction methods to obtain and retain the most meaningful feature dimensions before performing recognition. Principal components analysis (PCA) is one important technique which is used to represent images economically.

In [9] eigen ROI approach for the mammographic mass detection system has been compared with the eigen face detection system. Changes in illumination and posses can be related to the changes in internal density of the breast and different compression effects suffered by the breast when the mammogram is acquired.

In this paper the proposed eigen ROI approach uses bilateral information for classification of ROIs as normal or abnormal. With bilateral information a feature space of difference ROIs are created in training phase using 1D PCA. In testing phase the difference of ROIs are projected on to this feature space. Euclidean distance based classifier is used for classification of the given ROI. Section 3 gives an overview of 1D PCA. In section 4 the implementation of the proposed method is presented. Section 5 compares the results and section 6 concludes this paper.

3 Principal Component Analysis

The Principal component analysis is one of the most successful techniques that have been used in image recognition and compression. PCA is a mathematical procedure that uses orthogonal transformation to convert a set of observations of correlated variables into a set of values of linearly uncorrelated variables [10]. These uncorrelated variables are called principal components. The number of principal components is less than or equal to the number of original variables. This transformation is defined in such a way that the first principal component has the largest possible variance and each succeeding component in turn has the highest variance possible under the constraint that it be orthogonal to (uncorrelated with) the preceding components.

3.1 Mathematics of PCA

A 2-D image can be represented as 1-D vector by concatenating each row (or column) into a long vector [11] , [12]. Suppose there are M vectors of size N representing a set of sampled images. p_j's represent the pixel values.

$$Xi = \left[P1 \cdots PN\right]^T ; \quad i = 1 \cdots M \qquad (1)$$

First the images are mean centred by subtracting the mean image from each image vector.
Let m represent the mean image and W_i be defined as mean centred image such that

$$Wi = Xi - m \qquad (2)$$

Next the covariance matrix is determined as,

$$C = W^T \times W \qquad (3)$$

where W is a matrix composed of the column vectors Wi placed side by side.

Following this eigenvectors matrix E and eigen values vector V of C are obtained. Rearrange the eigenvector columns in E in the order of descending values of V. Following this Project the centred matrix W onto E to get a matrix

$$U = E \times W^T \qquad (4)$$

Finally feature matrix is found by

$$F = U \times W \qquad (5)$$

Eigen vectors that correspond to the lower eigen value contains less information. Thus dimensionality reduction can be achieved by eliminating these vectors.

3.2 Benefits of PCA

- The basic Benefit in PCA is dimensionality reduction of the data.
- Removal of data redundancy as components is orthogonal to each other.

4 Implementation

In this paper bilateral analysis based on 1D PCA is adopted for false positive reduction. Information from both left and right (bilateral) mammograms are used for the classification of the given segmented ROI.

4.1 Training phase

A block diagram for the training phase is shown in figure 4.1. During training phase bilateral ROIs are first segmented.

To get ROIs of the same size, all ROIs are resized to the size of the largest ROI. Histogram Equalization is performed for normalization of the ROIs. The new enhanced ROI has a uniform histogram with equally distributed brightness levels over the whole brightness scale.

After resizing and normalization, difference between bilateral ROIs is calculated. The difference matrix from each case is converted into a column vector and stored. PCA is applied on to the stored data for dimensionality reduction. After applying PCA, feature matrix is determined. Average of mass feature vector and average of false positive feature vector is found out from the calculated feature vector.

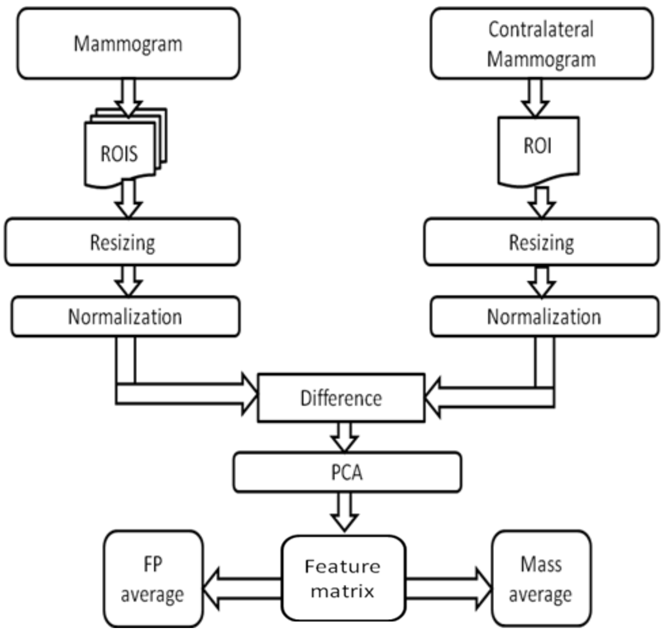

Fig. 1 Block diagram for training phase

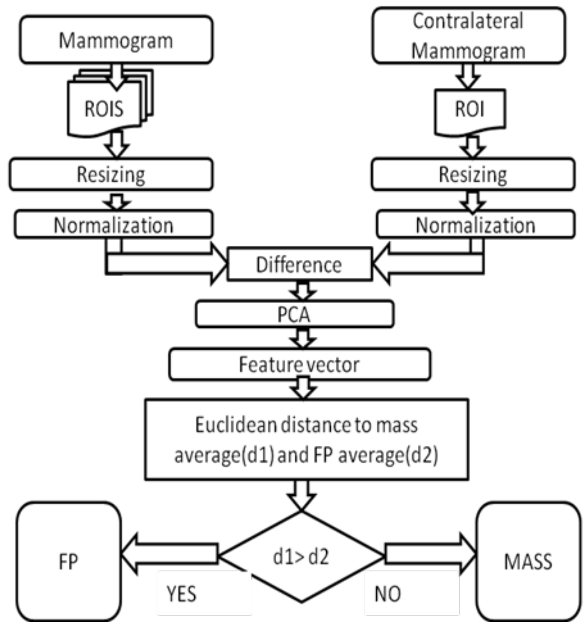

Fig. 2 Block diagram for testing phase

4.2 Testing phase

Fig 2 shows the testing block diagram. The same procedure is applied for testing images also. After finding the difference vector of corresponding test ROIs it is also projected to the training eigen space, to get test feature vector. Due to its simplicity Euclidean distance based classifier is used.

The euclidean distance of the test feature vector from both average mass feature vector as well as false positive feature vector is found out. If the distance to mass feature vector is small compared to other, the testing case indicates a mass, otherwise it is a false positive.

5 Results and Discussion

The database used to evaluate the proposed method is the Mammographic Image Analysis Society (MIAS) database. A total of 55 mammograms consisting of 85 ROIs are considered. Out of these 40 are mass cases and the rest are normal. The proposed method is implemented using Matlab R2012b software.

Three parameters are considered for system validation.

1. Accuracy: The percentage of predictions those are correct.
2. Sensitivity (TPR): It measures the proportion of actual positives which are correctly identified.

$$Senstivity = \frac{TP}{FN + TP} \quad (6)$$

where TP is true positives and FN is false negatives

3. Specificity: It measures the proportion of negatives which are correctly identified.

$$Specificity = \frac{TN}{TN + FP} \quad (7)$$

where TN is true negatives and FP is false positives

Table. 1 Comparison of classification performance of eigen ROI approach with feature based approaches

	Conventional Bilateral approach	Bilateral Eigen ROI approach
Accuracy	55.29	65.76
Sensitivity	50.00	67.5
Specificity	60.00	64.44

Table.2 Comparison of FPIs for various TPRs for the two systems

TPR		*60*	*65*	*70*	*75*	*80*	*85*	*90*	*95*
FPI	Bilateral	0.28	0.28	0.32	0.37	0.37	0.4	0.41	0.47
	Eigen ROI	0.16	0.16	0.17	0.18	0.19	0.22	0.27	0.33

Table.3 Comparison of TPRs for various FPIs for the two systems

FPI		*0.41*	*0.40*	*0.37*	*0.35*	*0.30*	*0.27*	*0.25*	*0.23*
TPR	Bilateral	86	83	83	74	65	57	52	52
	Eigen ROI	100	98	97	95	90	72	87	82

Table 1 compares the conventional bilateral approach with the proposed bilateral eigen ROI approach. This shows that accuracy of the proposed system improved to 65.75%. Also sensitivity and specificity of the proposed system increased by 17.5% and 4.44% compared to conventional approach.

A high sensitivity with low false positive per image (FPI) is of at most important for a CAD system. However the problem is that with higher sensitivity FPI also increases. An efficient system should give lower FPI for a given true positive rate (TPR).

For future validation FPIs for high sensitivities of the proposed system are determined and compared with in the table 2. Specifically the proposed eigen ROI method gives 45% less FPIs for 85% TPR. Table 3 compares TPRs for low FPIs. Specifically for 0.30 FPI the proposed eigen ROI approach gives 28% improvement in true positive rate when compared to conventional scheme.

6 Conclusion

In this work a novel approach for reduction of false positives in CAD system is proposed. The proposed eigen ROI scheme results better performance when compared to the conventional methods. The classification accuracy of the proposed method is improved by 15.92%. Also sensitivity and specificity of the proposed system increased by 17.5% and 4.44% compared to conventional feature based approach.

References

[1] da Cruz, Célia Freitas. "AUTOMATIC ANALYSIS OF MAMMOGRAPHY IMAGES." (2011).

[2] de Koning, Harry J., Jan HCL Hendriks, and Roland Holland. "Computer-aided detection versus independent double reading of masses on mammo grams."Radiology 227 (2003): 192-200.

[3] Tortajada, Meritxell, Arnau Oliver, Y. Díez, R. Marti, Joan C. Vilanova, and Jordi Freixenet. "Improving a CAD system using bilateral information." In Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE, pp. 5054-5057. IEEE, 2010.

[4] Wu, Yi-Ta, Jun Wei, Lubomir M. Hadjiiski, Berkman Sahiner, Chuan Zhou, Jun Ge, Jiazheng Shi, Yiheng Zhang, and Heang-Ping Chan. "Bilateral analysis based false positive reduction for computer-aided mass detection." *Medical physics* 34, no. 8 (2007): 3334-3344.

[5] Yin, Fang-Fang, Maryellen L. Giger, Kunio Doi, Charles E. Metz, Carl J. Vyborny, and Robert A. Schmidt. "Computerized detection of masses in digital mammograms: Analysis of bilateral subtraction images." *Medical Physics* 18, no. 5 (1991): 955-963.

[6] Bovis, Keir, and Sameer Singh. "Detection of masses in mammograms using texture features." In *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, vol. 2, pp. 267-270. IEEE, 2000.

[7] Dong, Aijuan, and Sinanovic Senad. "Comparison of Single Image Processing and Bilateral Image Feature Subtraction in Breast Cancer Detection."

[8] Jafri, Rabia, and Hamid R. Arabnia. "A Survey of Face Recognition Techniques." *JIPS* 5, no. 2 (2009): 41-68.

[9] Oliver, Arnau, Joan Marti, Robert Marti, Anna Bosch, and Jordi Freixenet. "A new approach to the classification of mammographic masses and normal breast tissue." In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, vol. 4, pp. 707-710. IEEE, 2006.

[10] Paul, Liton Chandra, and Abdulla Al Sumam. "Face Recognition Using Principal Component Analysis Method." International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) 1, no. 9 (2012): pp-135.

[11] Melišek, Ján Mazanec—Martin, and Miloš Oravec—Jarmila Pavlovicová. "Support vector machines, PCA and LDA in face recognition." *Journal of Electrical engineering* 59, no. 4 (2008): 203-209.

[12] Kim, Kwang In, Keechul Jung, and Hang Joon Kim. "Face recognition using kernel principal component analysis." Signal Processing Letters, IEEE 9, no. 2 (2002): 40-42.