Mammographic Image Enhancement, Classification and Retrieval using Color, Statistical and Spectral Analysis

Bikesh Kr. Singh
Department of Biomedical Engineering
National Institute of Technology, Raipur (C.G), India

ABSTRACT
One of the major causes of cancer death among middle aged women in developed countries is breast cancer. Mammography is one method used by radiologists for detection and interpretation of cancer in breast images. Over the past few years Content-based Image Retrieval (CBIR) approaches has received significant attention for medical images analysis. In this paper content based retrieval techniques were tested for tissue classification and analysis of breast images. The proposed method employs image enhancement, analysis and classification of mammograms using histogram, statistical, wavelet coefficients and spectral features. The implementation of proposed method was carried out using MATLAB software and hence can work on simple personal computer. Analysis was carried out on 56 images collected from open source mini-MIAS database. Euclidean distance was used to compare the features of query image with stored images in database. Results show that the suggested features can be used for both classification and retrieval of mammographic images. The retrieval efficiency was obtained to be 85.7%.

General Terms
Pattern Recognition, Medical Image Processing.

Keywords
Mammograms, image processing, spectral and statistical features, wavelet coefficients.

1. INTRODUCTION
Breast cancer ranks second to lung cancer as the leading cause of death in women diagnosed with cancer in the US. About 41,000 women in the US are expected to die from the disease in 2006[1]. The number of cases of women with breast cancer has been increasing. Mammography, making images of breast using X-rays is most widely used modality to detect and characterize breast cancer [2]. Computer aided diagnosis (CAD) has been the second opinion for radiologists for early detection of breast cancer [3]. In this paper content based image retrieval techniques are used for content retrieval and classification of mammographic images. Content based visual information retrieval (CBVIR) or Content Based Image Retrieval (CBIR) has been one on the most vivid research areas in the field of computer vision over the last years. *CBIR* is based on visual features such as colour, texture and shape [4]. Color represents one of the most widely used visual features in CBIR systems. Color spaces shown to be closer to human perception and used widely in CBIR include, RGB, LAB, LUV, HSV (HSL), YCrCb and the hue-min-max-difference (HMMD)[5,6,7,8,9,10]. The main drawback of color based CBIR for classification is that the representation is dependent on the color of the object being studied, ignoring its shape and texture. In this paper histogram is used for study of color distribution in mammographic images. To overcome the limitations of histograms use of nine statistical texture features and spectral texture features is reported. Finally breast images are classified on the basis of these extracted features. The rest of the paper is organized is follows- The next section gives brief overview features extracted followed by methodology, result and conclusions respectively.

2. OVERVIEW OF FEATURES EXTRACTED:

2.1 Histogram:
A digital image, in general, is a two dimensional mapping I: $x \rightarrow v$ from $M \times N$ *pixels* $x = [i, j]^T$ *to* values v. The histogram of an image can be found as

$$r_b = \sum_{i=1}^{M}\sum_{j=1}^{N} S_b(i, j),\ \forall b = 1, 2, .. \quad (1)$$

where $S_b(i, j) = 1$ if the value v at pixel location $[i, j]$ falls in bin b, and $S_b(i, j) = 0$ otherwise and B is number of bins in the histogram [11]. Histograms have proved themselves to be a powerful representation for image data in a region. Discarding all spatial information, they are the foundation of classic techniques such as histogram equalization and image indexing [12]. The color histogram of an image is relatively invariant with translation and rotation about the viewing axis, and varies only slowly with the angle of view [13]. Color Histograms are a commonly used as appearance-based signature to classify images for content-based image retrieval systems (CBIR) [14]. The main drawback of histograms for classification is that the representation is dependent on the color of the object being studied, ignoring its shape and texture.

2.2 Texture Analysis Using Spectral Measures:
Texture analysis is an important issue with applications ranging from remote sensing and crop classification to object-based image coding and tissue recognition in

medical images. Spectral measures of texture are based on Fourier spectrum, which is ideally suited for describing the directionality of period or almost periodic 2-D patterns in an image. These global texture patterns easily distinguishable as concentrations of high energy burst in the spectrum generally are quite difficult to detect with spatial methods because of the local nature of these techniques. Thus spectral texture is useful for discriminating between periodic and non-periodic texture patterns, and further, for quantifying differences between periodic patterns. Interpretation of spectrum features is simplified by expressing the spectrum in polar coordinates to yield a function S(r, θ), where S is the spectrum function and r and θ are the variables in the coordinate system. For each direction θ, S(r, θ) may be considered a 1-D function $S_\theta(r)$. Similarity for each frequency r, $S_r(\theta)$ is a 1-D function. Analyzing $S_\theta(r)$ for a fixed value of θ yields the behaviour of spectrum along a radial direction from the origin, whereas analyzing $S_r(\theta)$ for a fixed value of r yields a behaviour along a circle centred on the origin[15].
A global description is obtained by integrating these functions:

$$S(r) = \sum_{\theta=0}^{\pi} S_\theta(r) \tag{2}$$

and

$$S(\theta) = \sum_{r=1}^{r_0} S_r(\theta) \tag{3}$$

Where r_0 is the radius of the circle centred at origin [15]. The result of these two equations constitutes a pair of value [S(r), S(θ)] for each pair of coordinate (r, θ). By varying this coordinates we can generate two 1-D functions, S(r) and S(θ) that constitute a spectral-energy description of texture for an entire image or region under consideration. Furthermore descriptors of these functions themselves can be computed in order to characterize their behaviour quantitatively. Descriptors typically used for these purpose are the location of the highest value, mean and variance of the amplitude and axial variations, the distance between the mean and the highest value of the function.

2.3 Overview of Statistical Features Used For Texture Analysis:

The feature vector for texture analysis of a particular image is computed by calculating following nine features [15]:
1. Uniformity: If Z_i is the random variable indicating intensity, P (z) is the histogram of intensity levels in the image and L is number of possible intensity levels, and then Uniformity is calculated as:

$$U = \sum_{i=0}^{L-1} P^2(Z_i) \tag{4}$$

2. Average Entropy:

$$e = -\sum_{i=0}^{L-1} P(Z_i)\log_2 P(Z_i) \tag{5}$$

3. Relative Smoothness: If σ is the standard deviation, Relative smoothness is given by

$$R = 1 - \frac{1}{1+\sigma^2} \tag{6}$$

4. Measurement of Variance of Approximation Wavelet Coefficient: The image is subjected to discrete wavelet transform to obtain Approximation coefficient (A), Horizontal Coefficient (H), Vertical Detail Coefficient (V) and Diagonal Detail coefficient (D). The variance of these coefficients is considered to be the next four features respectively.

5. Skewness:

$$\mu_3(z) = \sum_{i=0}^{L-1} (Z_i - m)^3 P(Z_i) \tag{7}$$

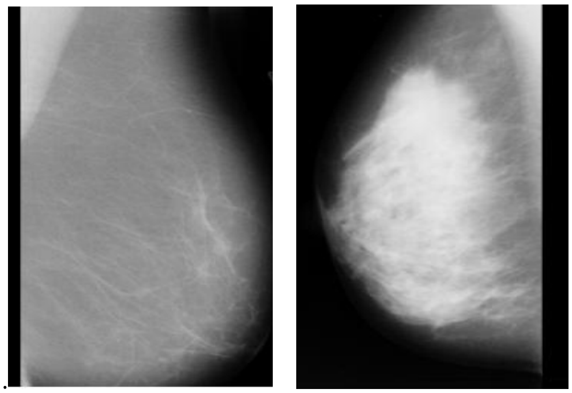
6. Kurtosis:

$$\mu_4(z) = \sum_{i=0}^{L-1} (Z_i - m)^4 P(Z_i) \tag{8}$$

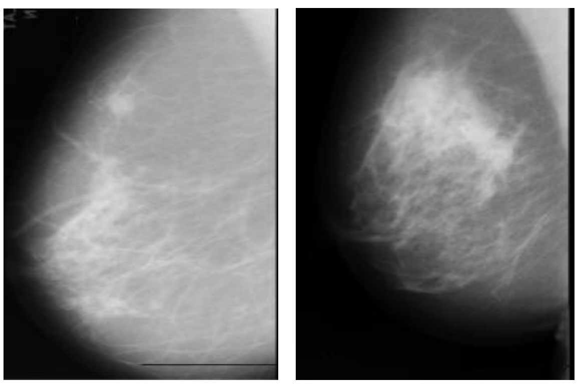
Where p(x, v) and p(v) joint PDF and marginal PDF respectively. For zeroth – order spatiogram i.e. histogram, there is no spatial dependency hence p(x, v) = p (v).

3. METHODOLOGY:

A database of 56 breast images was collected mini-MIAS open source data base [16]. Some example images are shown in figure 1. We implemented the functional code for this prototype system using MATLAB 7.1 on a Pentium dual core II, 600 MHz Windows-based PC. The first step is to select an appropriate enhancement technique for enhancing the mammogram. We implemented seven enhancement methods for enhancement of digital mammograms. The results were evaluated using CNR and PSNR as shown in table 1. The Experimental results show that the steerable filter yields maximum PSNR as well as CNR values and thus we use this method for enhancement of mammogram. Further it also consumes less time compared to counterlet transform filtering. The result of enhancement using steerable filter is shown in figure 2. After enhancement, each feature discussed in section 2 was analyzed independently for classification of various images in the database. For retrieval, a feature vector consisting of histogram bins, spectral features S(r), S(θ) and nine features discussed in previous sections is generated.



(a) (b)



(c) (d)

Figure 1: Some sample breast images: (a) Normal Fatty tissue breast (b) Normal dense glandular tissue breast (c) Fatty glandular tissue with malignant tumor (d) Fatty glandular tissue with benign tumor.

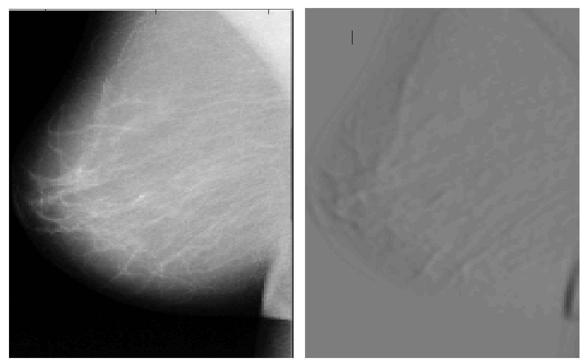
Table 1. Evaluation of Enhancement Techniques

S.NO	Enhancement Technique	CNR (dB)	PSNR (dB)
1	Contrast Stretching	0.0169	35.79
2	Histogram Equalization	0.0600	39.11
3	Mean Filter	0.0081	38.69
4	Median Filter	0.0221	19.92
5	Hybrid Technique	0.0229	20.98
6	Contourlet Transform Filter	0.0144	13.02
7	Steerable Filter	0.0716	50.81

When Query image is presented to the system, the feature vector of Query image is generated and compared with that of each image stored in the database. An important consideration in any system that retrieves images by similarity is the metric used to calculate degree of similarity of two images. Euclidean distance is used to determine the relative closeness between the extracted features of Query image and stored images.

4. STIMULATED RESULTS AND CONCLUSIONS:

As per the literature of the mini-MIAS database[16] the mammograms have been grouped in to three tissue types namely fatty, fatty glandular and dense glandular while the severity of abnormality is classified as benign or malignant tumor. In experiments conducted only two classes of abnormalities were considered namely calcification and well defined masses. Figure 3-6 below shows the histogram of four images of different categories shown in figure 1. It is observed that histogram for each of these images is different.



(a):Original Image (b): Enhanced image

Figure.2. Mammogram before and after enhancement using steerable filter.

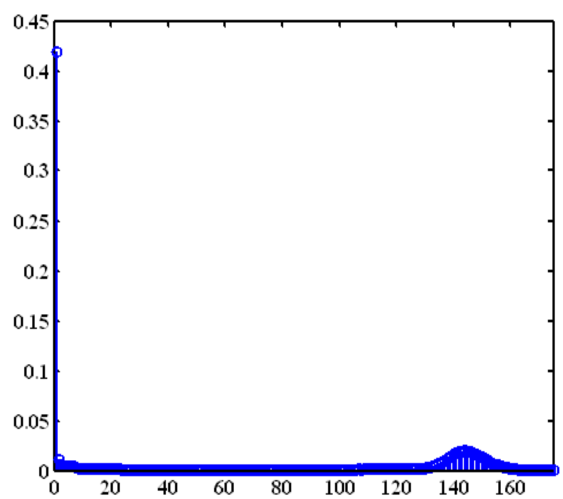


Figure 3. Histogram of normal fatty tissue breast

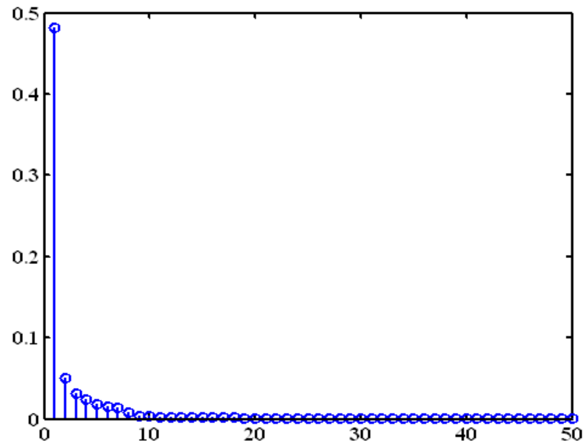


Figure 4. Histogram of normal dense grandular tissue breast

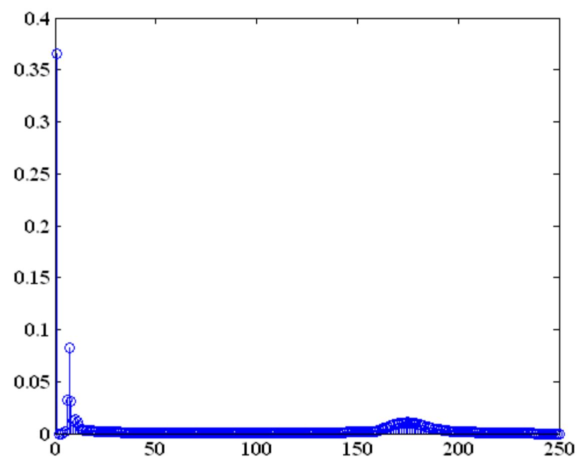


Figure 5. Histogram of grandular tissue breast tissue with malignant tumor

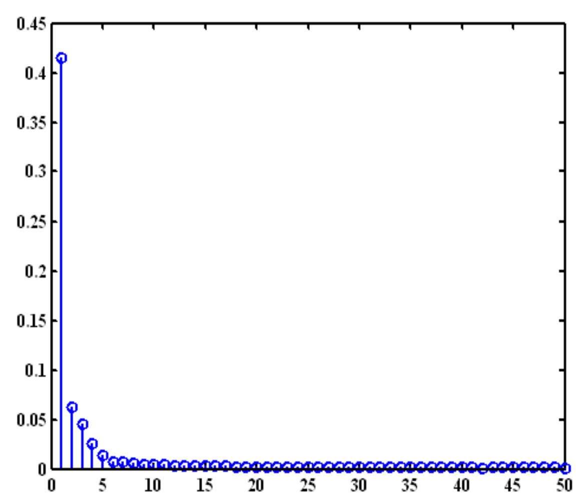


Figure 6. Histogram of fatty glandular tissue with benign tumor

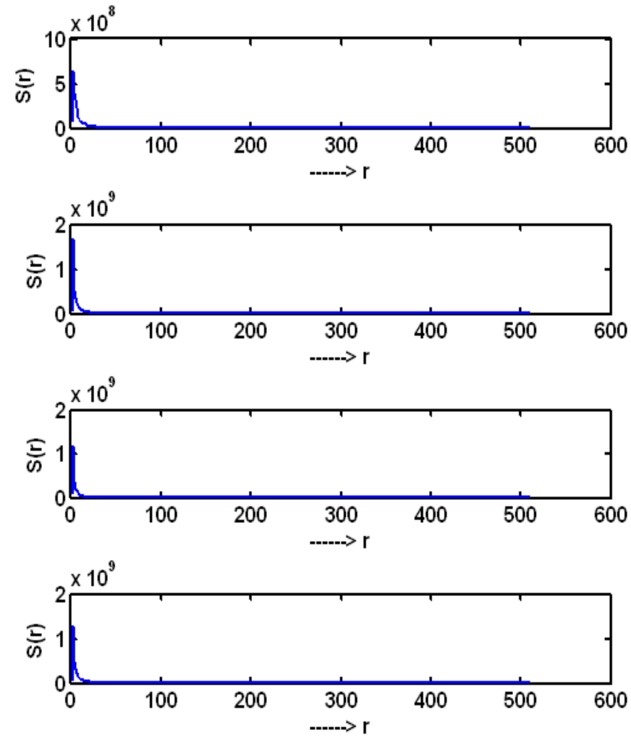


Figure 7: Radial spectral components of four breast images of different categories

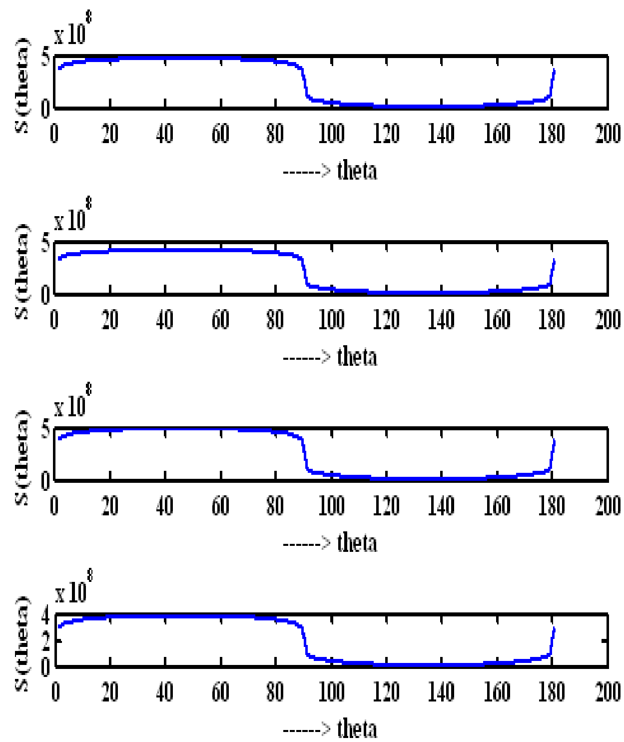


Figure 8: Angular spectral components of four breast images of different categories.

Figure 7 shows the plot of radial component and Figure 8 shows the plot of angular components of four different categories of breast images shown in figure 1. The pattern of S(r) and S(θ) obtained for each of these images is different. The plot of S(r) corresponding to Figure 7 shows no strong periodic components i.e. there are no peaks in the spectrum besides the peak approximately at the origin, which is the DC component. The peaks of these DC components are different for all images and are used to classify the images. Figure 8 shows the plot of angular component of these images respectively. The nature of energy burst is similar show strong energy components over a range of θ. However the amplitudes are slightly different which is used for classifying the images. Here from figure 7 and figure 8 we conclude that radial components may give better classification efficiency as compared to angular components. Table 2 shows the nine statistical features for different categories of breast images. A 1- dimensional feature vector composed of histogram bins, S(r), S(θ) and nine statistical features of query image is computed and compared using Euclidean distance with similar feature vector of other images in the database and image with smallest Euclidean distance is found.

5. CONCLUSIONS

In this paper CBIR features like color and texture were tested for tissue classification and analysis of breast images taken from mini-MIAS database. First an appropriate enhancement technique was selected and then histogram, spectral and statistical texture features were extracted. The implementation of proposed method was carried out using MATLAB software. Euclidean distance was used to compare the features of query image with stored images in database. Results show that the suggested features can give acceptable accuracy for both classification and retrieval of mammographic images. In our database of 56 images, 48 images were successfully retrieved which resulted in a retrieval efficiency of 85.7%. The efficiency of the proposed system for large databases can be further increased by including shape features for classification which will also help in exact determination of size of the tumor.

Table.2: Statistical Features (average) extracted for different categories of breast images

Feature Selected→	**Uniformity**	**Entropy**	**Relative Smoothness**	**Wavelet Coefficients**				Skewness	Kurtosis
Image type				**A**	**H**	**V**	**D**		
Normal dense glandular tissue	0.2379	4.5951	0.9868	202.9455	58.7549	49.9829	13.4555	488.6093	567.1971
Normal fatty tissue	0.1910	4.6683	0.9868	198.6203	65.2607	45.6971	12.0021	478.3042	557.4214
Calcified Benign tumor	0.1732	5.4479	0.9809	200.8231	69.4723	48.9771	7.6661	499.9891	581.7889
Calcified malignant tumor	0.3738	3.7550	0.9842	196.9426	52.6569	54.8716	21.5264	480.4060	552.9360
Well defined masses with Benign tumor	0.2487	4.7658	0.9877	199.3289	25.3677	41.3925	1.8182	486.6675	563.0819
Well defined masses with malignant tumor	0.1450	5.1044	0.9894	202.8468	69.1139	37.4593	11.5470	491.2150	573.7715

6. REFERENCES

[1] J. Ferlay, F. Bray, P. Pisani and D.M. Parkin. GLOBOCAN 2000: Cancer Incidence, Mortality and Prevalence Worldwide. Version 1.0. IARC Cancer Base No. 5. Lyon, IARC Press, 2001.

[2] V.D Dzung, D.T Thuan Nguyen, T.D Nguyen, T.T Nguyen and D.H Tran, "A Program For Locating Possible Breast Masses On Mammograms" Proceedings of 3rd International Conferences on the Development of BME in Vietnam, pp 109-112, 11-14 Jan 2010.

[3] Y. Ireaneus Anna Rejani, " Early Detection of Breast Cancer using SVM Classifier Technique",

[4] International Journal of Computer Science and Engineering Vol.1(3), pp. 127-130, 2009.

[5] B. K. Singh, G. R. Sinha, A. Wany and B. Mazumdar, Content based retrieval from biomedical images using color histograms, proceedings of International Conference on Nanotechnology and Biosensors, ICBN , Raghu College of Engg., Visakhapatnam 2010.

[6] P.L. Stanchev, D. Green Jr. and B. Dimitrov, " High level color similarity retrieval," Int. J. Inf. Theories Appl., 2004, pp 363–369.

[7] Y. Liu, D.S. Zhang, G. Lu and W.-Y. Ma, "Region-based image retrieval with perceptual colors,

Proceedings of the Pacific-Rim Multimedia Conference (PCM)," December 2004, pp. 931–938.

[8] 7 R. Shi, H. Feng, T.-S. Chua and C.-H. Lee, "An adaptive image content representation and segmentation approach to automatic image annotation," International Conference on Image and Video Retrieval (CIVR), 2004, pp. 545–554.

[9] V. Mezaris, I. Kompatsiaris and M.G. Strintzis, "ontology approach to object-based image retrieval," Proceedings of the ICIP, vol. II, 2003, pp. 511–514.

[10] B.S. Manjunath, "Color and texture descriptors," IEEETrans. CSVT 11 (6) (2001) pp. 703–715.

[11] G. S Linda and C. S George,"Computer Vision" Prentice Hall, 2003.

[12] M. Nilsson, J. S. Bartunek, J. Nordberg and I. Claesson, "On Histograms and Spatiograms - Introduction of the Mapogram," ICIP, 2008, page(s): 973-976.

[13] S.T. Birchfield, S, Rangarajan, "Spatiograms versus Histograms for Region-Based Tracking," 2005, In CVPR'05.

[14] G. S Linda and C. S George,"Computer Vision" Prentice Hall, 2001.

[15] M. Stricker and M. Swain, "The Capacity of Color Histogram Indexing," Computer Vision and Pattern Recognition, Proceedings CVPR IEEE Computer Society Conference, 1994, pp. 704 – 708.

[16] R. C Gonzalez, R. E Woods and S. L. Eddins, Digital Image Processing using Matlab, Pearson Education, 2006.

[17] 16. J Suckling *et al* (1994): *The Mammographic Image Analysis Society Digital Mammogram Database* Exerpta Medica. International Congress Series 1069 pp375-378.