Textural Feature Extraction and Classification of Mammogram Images using CCCM and PNN

S. Deepa[1], Dr.V.Subbiah Bharathi[2]

[1]Research Scholar, Department of ECE,Sathyabama University, Chennai, India
[2]Principal, DMI College of Engineering, Chennai, India

Abstract: *This work presents and investigates the discriminatory capability of contourlet coefficient co-occurrence matrix features in the analysis of mammogram images and its classification. It has been revealed that contourlet transform has a remarkable potential for analysis of images representing smooth contours and fine geometrical structures, thus suitable for textural details. Initially the ROI (Region of Interest) is cropped from the original image and its contrast is enhanced using histogram equalization. The ROI is decomposed using contourlet transform and the co-occurrence matrices are generated for four different directions ($\theta=0°$, 45°, 90° and 135°) and distance (d= 1 pixel). For each co-occurrence matrix a variety of second order statistical texture features are extracted and the dimensionality of the features is reduced using Sequential Floating Forward Selection (SFFS) algorithm. A PNN is used for the purpose of classification. For experimental evaluation, 200 images are taken from mini MIAS (Mammographic Image Analysis Society) database. Experimental results show that the proposed methodology is more efficient and maximum classification accuracy of 92.5% is achieved. The results prove that contourlet coefficient co-occurrence matrix texture features can be successfully applied for the classification of mammogram images.*

Keywords-*Contourlet Transform, Mammogram, SFFS, PNN, ROI, MIAS*

I. Introduction

Breast cancer is the most frequently diagnosed cancer and the leading cause of cancer death among females, accounting for 23% of the total cancer cases and 14% of the cancer deaths. Breast cancer has outnumbered other cancers and has become the leading cause of cancer deaths in women, pushing lung & bronchus to second place[1]. According to GLOBOCAN 2008[1] estimates, about 12.7 million cancer cases and 7.6 million cancer deaths are estimated to have occurred in 2008; of these, 56% of the cases and 64% of the deaths occurred in the economically developing world. Breast cancer not only endangersthe life of the patient, but also causes damage to the female sexual characteristic organ [2], [3].Early detection of breast cancer is of utmost importance: localized cancer leads to a 5-year survival rate of 97.5%, whereas cancer that has spread to distant organs has a 5-year survival rate of only 20.4% [4]. Mammography is, at present, one of the most reliable methods for early detection of breast carcinomas [5]. And it is currently the best way that can detect the clinical asymptomatic concealed breast cancer. It can reveal pronounced evidence of abnormality, such as masses and calcifications, as well as subtle signs such as bilateral asymmetry and architectural distortion [6]. Reports from the American Cancer Society (ACS) shows that early diagnosed breast cancer patient can achieve a survival rate as high as 97% before the spreading of the carcinoma cells [7]. Therefore, early detection and diagnosis of the breast cancer is very important in saving the life of the patient and it's the most effective way to increase thecure rate and lower the mortality rate. In studying mammograms specific features are sought in routine examinations as common indicators of malignancy. The specific features or abnormalities present in the mammogram images are classified into calcification, circumscribed masses, spiculated masses, ill-defined masses, architectural distortion and asymmetry. Calcifications are calcium deposits present in the breast, calcifications are further classified into Macro-calcifications and Micro-calcifications. Macro-calcifications are large calcium deposits and are not associated with cancer. Micro Calcifications (MC) appear as small white spots similar to grains of sand with a diameter of less than 0.5 mm and are grouped closely together to form clusters. Circumscribed masses mostly appear as uniform and smooth masses in the shape of irregular circles. Spiculated lesions appear as a region with segments distributed in many directions as a multi-arms star. Ill-defined masses are the masses that do not have a fixed pattern. An architectural distortion on a mammogram is basically a disruption of the normal 'random' pattern of curvilinear and fine linear radiopaque structures normally seen on the mammogram. There is no visible mass, but the distortion often appears as a 'stellate' shape or with radiating speculation. Asymmetrical distortion refers to asymmetry between the breasts (left and right breast of the same patient), it can be in shape or in certain area in the mammograms. Figure 1 shows the sample mammogram image for a normal mammogram image. Figure 2 shows the sample mammogram images for the various abnormalities discussed above.

Figure 1. Normal Mammogram Image

mdb315 – Circumscribed Mass   mdb202 –Spiculated Mass   mdb209 –Micro Calcification

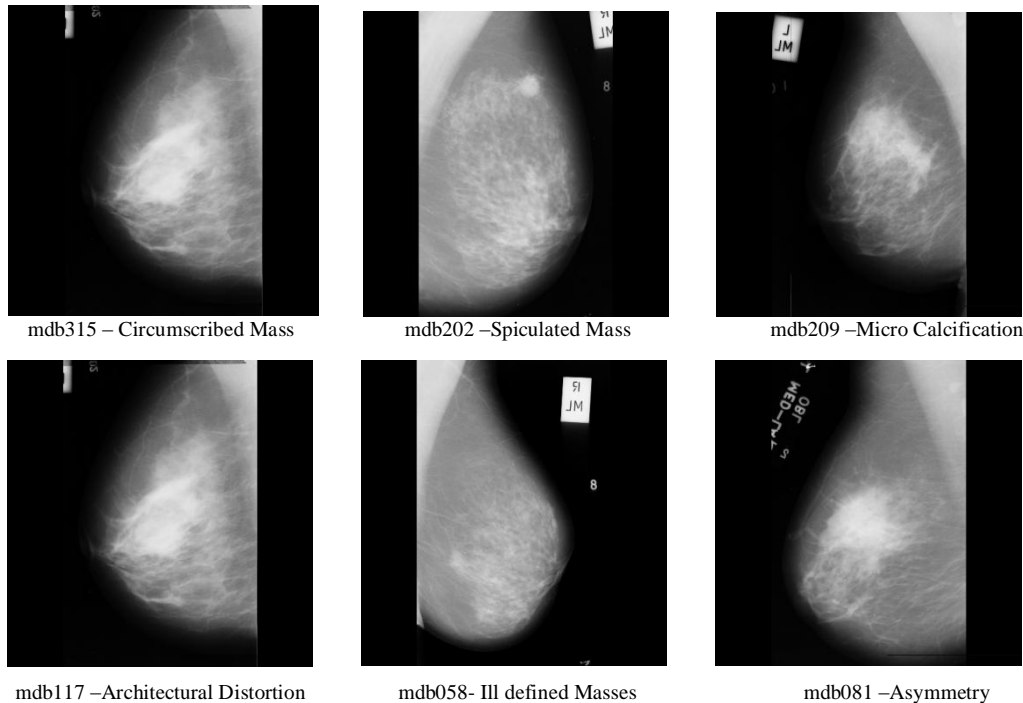mdb117 –Architectural Distortion   mdb058- Ill defined Masses   mdb081 –Asymmetry

Figure 2. Abnormal Mammogram Images

Texture is a commonly used feature in the analysis and interpretation of images. For mammographic images of mild texture the results obtained using texture analysis prove to be successful. Three main feature extraction method namely, statistical methods, model based methods and signal processing methods are discussed by Malagelda[8]. The statistical features included first order statistics, co-occurrence matrices [8] and runlengthstatistics [9]. Model based methods are based on prior models such as Markov random fields [10], autoregressive models and fractals [9]. Signal processing methods are based on pixel characteristics or image frequency spectra including Law's texture measures, Gabor filtering [8] and Wavelets[11],[12]. In the proposed method the efficiency of contourlet transform is exploited to achieve a more efficient mammogram texture feature analysis and PNN is used for classification. It has been revealed that contourlet transform has a remarkable potential for analysis of images representing smooth contours and fine geometrical structures, thus suitable for textural details.The rest of this paper is organized as follows, section 2 describes the Contourlet Transform, section 3 explains the proposed method, Section 4 discussesthe results and finally section 5 contains the conclusion.

II. CONTOURLET TRANSFORM

The major drawback for wavelets in two-dimensions is their limited ability in capturing directional information. To overcome this deficiency, researchers have recently considered multiscale and directional representations that can capture the intrinsic geometrical structures such as smooth contours in natural images. Some examples include the steerable pyramid [13],brushlets [14], complex wavelets [15], and the curvelet transform [16]. In particular, the curvelet transform, pioneered by Cand`es and Donoho, was shown to be optimal in a certain sense for functions in the continuous domain with curved singularities. Inspired by curvelets, Do and Vetterli [17], [18] developed the contourlettransform based on an efficient two-dimensional multiscale and directional filter bank that can deal effectively with images having smooth contours. Contourlets
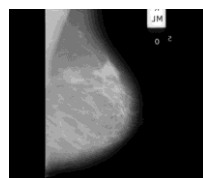
not only possess the main features of wavelets (namely, multiscale and time-frequency localization), but also offer a high degree of directionality and anisotropy. The primary goal of the contourlet construction [17], [18] was to obtain a sparse expansion for typical images that are piecewise smooth away from smooth contours. Two dimensional waveletslack directionality and are only good at catching point discontinuities, but do not capture the geometrical smoothness of the contours.Contourlets were developed as an improvement over wavelets in terms of this inefficiency. The resulting transform has the multiscale and time-frequency-localization properties of wavelets, but also offers a high degree of directionality and anisotropy.The contourlet transform is implemented via a twodimensional filter bank that decomposes an image into several directional subbands at multiple scales. This is accomplished by combining the Laplacian pyramid [19] with a directional filter bank [20] at each scale. Due to this cascade structure, multiscale and directional decomposition stages in the contourlet transform are independent of each other. One can decompose each scale into any arbitrary power of two's number of directions, and different scales can be decomposed into different numbers of directions. This feature makes contourlets a unique transform that can achieve a high level of flexibility in decomposition while being close to critically sampled(up to 33% overcomplete, which comes from the Laplacian pyramid). Other multiscale directional transforms have either a fixed number of directions, or are significantly overcomplete (depending on the number of directions).
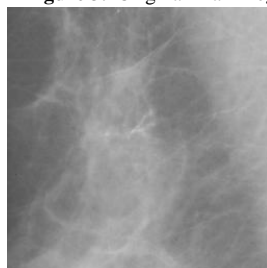
III. PROPOSED METHOD

The proposed method consists of four stages i) The Preprocessing Stage ii) Feature Extraction Stage iii) Feature Selection Stage and iv) Classification Stage. The MIAS database consists of 322 mammogram images digitized at 50 micron pixel edge, which is reduced to 200 micron edge and clipped/padded so that every image is 1024 x 1024 pixels. Of the 322 images, 207 images are normal and 115 images are abnormal. The abnormal images are further categorized into benign (non-cancerous) and malignant (cancerous) images. The case sample analyzed consists of 200 images, and the images are selected such that 120 of them are normal and 80 of them are with abnormalities. The original mammogram images taken from the mini-MIAS database is of size 1024 x 1024 pixels. In order to reduce the computation time and increase the classification accuracy of the system, instead of analyzing the entire image only the Region Of Interest(ROI) i.e., the region in the image that contains abnormalities is extracted and is subsequentlyanalyzed. The steps involved in the method are explained in detail.
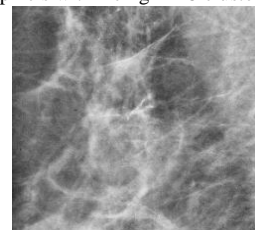
3.1 Preprocessing

With the help of an experienced radiologist the Region Of Interest (ROI) of size 256x256 pixels was manually cropped. The cropped ROI is of low contrast hence the ROI is preprocessed using a contrast enhancement technique.For enhancing the contrast of the image adaptive histogram equalization technique is used. It enhances the contrast of images by transforming the values in the intensity image. Unlike histogram equalization, it operates on small data regions (tiles), rather than the entire image. Each tile's contrast is enhanced, so that the histogram of the output region approximately matches the specified histogram. The neighboring tiles are then combined using bilinear interpolation in order to eliminate artificially induced boundaries. Figure 3 shows the original mammogram imageof size 1024x1024 pixels with benign MC cluster. Figure 4(a) shows the cropped Region Of Interest (ROI) of size 256x256 pixels and 4(b) shows the preprocessed ROI.

Figure 3. Original Mammogram Image of size1024x1024 pixels with Benign MC cluster.

(a) ROI of size 256x256 pixels

(b) Preprocessed ROI of size 256x256 pixels

Figure 4. ROI and Preprocessed ROI

3.2 Feature Extraction
3.2.1 Contourlet Decomposition
The preprocessed ROI is decomposed using ContourletTransform and the contourlet coefficients are obtianed. Contourlet Transform utilizes a double filter bank, in order to obtain a sparse expansion of typical images containing smooth contours. The Laplacian Pyramid (LP) is used first to detect the point discontinuities of the image and then a Directional Filter Bank (DFB) to link point discontinuities into linearstructures. The combined result is the contourlet filter bank, which is a double iterated filter bank that decomposes the image into directional subbands at multiple scales. The separation of directional and multiscale decomposition stages provides a fast and flexible transform with computational complexity *O(N)* for *N*-pixel images. The preprocessedROI is decomposed into four pyramidal levels, which are then further decomposed into four, eight and sixteen directional sub-bands using Contourlet transform. Figure 5 shows an example of the decomposition of a ROI using Contourlet transform.
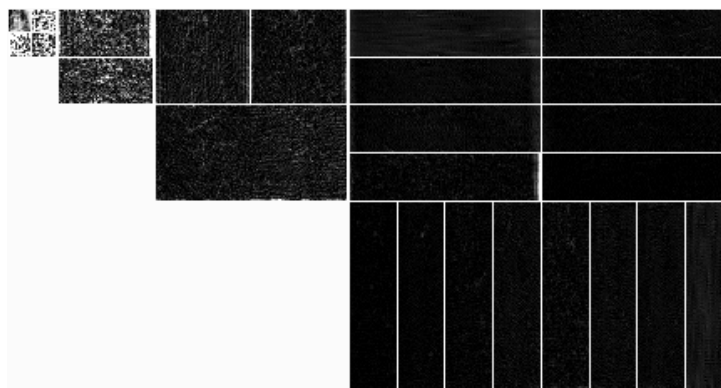

Figure 5. Decomposition of a ROI using Contourlet Transform

3.2.2 Contourlet Coefficient Co-Occurrence Matrix Features(CCCM)
First Order statistical texture measures computed using only histograms carry no information regarding the relative position of pixels with respect to each other. Co-occurrence Matrix is a well-established robust statistical tool for extracting second order texture information from images [21,22,23]. The co-occurrence matrix method of texture description is based on the repeated occurrence of pixel values with respect to relative positions of pixel values in the image. Using the contourlet coefficients generated in the previous stage, the co-occurrence matrices are computed for four different directions ($\theta=0°$, $45°$, $90°$ and $135°$) and distance (d=1 pixel). For each co-occurrence matrix the following features are determined, Autocorrelation, Contrast, correlation, cluster prominence, Cluster shade, dissimilarity, Energy, Entropy, Homogeneity, Maximum Probability, variance, sum average, sum variance, difference variance, difference entropy and Information measure of correlation.

3.3 Feature selection
Because of the high dimensionality of input features, a feature selection algorithm is deployed to reduce the redundancy and improve classification accuracy. Feature selection approaches offer more than one significant advantages,including reduction of computational complexity, improved generalization ability androbustness against outliers. A widely adopted algorithm for feature selection is theSequential Floating Forward Selection (SFFS) algorithm [24] which has been used ina broad range of applications [25]. The idea behind the SFFS algorithm consists ofconsecutive forward selection steps, followed by a number of backward steps as longas the resulting subsets are better than the previously evaluated ones at the same level.Due to its behavior, the method has the ability to correct wrong decisions made inthe previous steps in order to approximate the optimal solution as accurate as possible.This results in one of the key advantages of the SFFS algorithm: its tolerance todeviations from monotonic behavior of the feature selection criterion function [24],[25].However, in order to achieve this near optimality, the need of computational time isgreatly increased, especially in the case of data of greater complexity anddimensionality.

3.4Classification - Probabilistic Neural Network
Probabilistic Neural Network [PNN] is a network formulation of 'probability density estimation'. It is a model based on competitive learning with a 'winner takes all attitude' and the core concept based on multivariate probability estimation. The distinguishing feature of PNN is that the computational load in the training phase is transferred to the evaluation phase. The main advantage of PNN is that training is instantaneous, easy and faster compared to back propagation networks. The development of PNN relies on the

Parzen window concept of multivariate probabilities. The PNN is a classifier version, which combines the Baye's strategy for decision-making with a nonparametric estimator for obtaining the probability density function. The PNN architecture consists of an input layer, a pattern layer, a summation layer, and an output layer. Figure 6 shows the architecture of PNN. The input layer simply distributes the input to the neurons in the pattern layer and does not perform any computation.
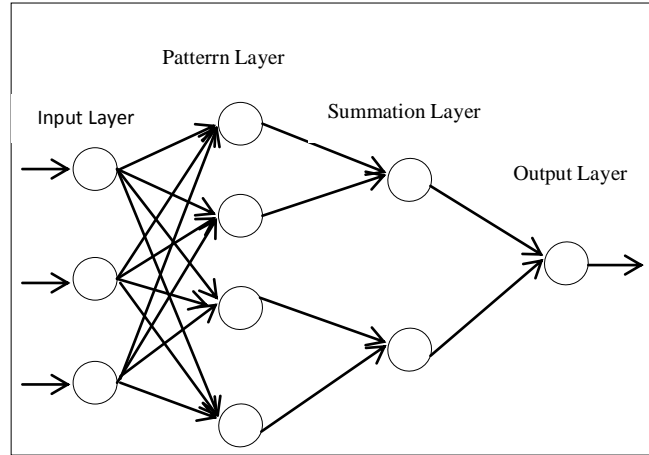

Figure 6. Probabilistic Neural Network Architecture

The neuron x_{ij} of the pattern layer receives a pattern x from the input layer and computes its output as given by the equation below,

$$\emptyset_{ij}(x) = \frac{1}{(2\pi)^{d/2}\sigma^d} \exp\left[-\frac{(x-x_{ij})^T(x-x_{ij})}{2\sigma^2}\right] \qquad (1)$$

Where
σ denotes the smoothing parameter,
x_{ij} denotes the neuron vector and
D denotes the dimension of the pattern vector x.

The summation layer neurons compute the maximum likelihood of pattern x being classified into C_i by summarizing and averaging the output of all neurons that belong to the same class using equation given below,

$$p_i(x) = \frac{1}{(2\pi)^{d/2}\sigma^d}\frac{1}{N_i}\sum_{j=1}^{N_i} \exp\left[-\frac{(x-x_{ij})^T(x-x_{ij})}{2\sigma^2}\right] \qquad (2)$$

Where N_i is the total number of samples in class C_i. The decision layer unit classifies the pattern x in accordance with the Bayes' decision rule based on the output of all the summation layer neurons by

$$C(x) = argmax\{p_i(x)\}, \qquad for\ i = 1,2\ldots,m \qquad (3)$$

Here *C(x)* denotes the estimated class of the pattern *x* and *m* is the total number of classes in the training samples.

IV. RESULTS AND DISCUSSIONS

The mini MIAS database consists of left and right breast images for a total of 161 pairs. For our experimentation purpose we have used 200 mammogram images, of which 120 images are normal and 80 images are abnormal.The PNN is trained using 60% of the samples for each category and the remaining 40% of the images are taken as the test images. TABLE 1 shows the number of training samples and testing samples used for the experimentation purpose.

TABLE 1. Distribution of Training and testing samples

Class	No. of Training Samples	No. of Testing Samples
Normal	72	48
Abnormal	48	32

The Images are decomposed using contourlet transform and the co-occurrence matrices are generated. The second order texture features are extracted and using SFFS algorithm the optimal features set is selected. The

selected features are given as input to the PNN and the image is classified as normal or abnormal. The efficiency of the system is determined using the measures such as True Positive Rate, False Positive Rate, True Negative Rate and False Negative Rate.True Positive Rate is the measure of abnormal images classified correctly as abnormal; False Positive Rate is the measure of a normal mammogram classified as abnormal. False Negative Rate is the measure of anabnormal case classified as normal. True Negative Rate is a measure of annormal mammogram classified as normal. TABLE 2 shows the confusion matrix obtained using the proposed method for a totalof 80 test samples. 30 out of 32 abnormal images are classified correctly as abnormal thus producing a true positive rate of 93.75% and 44 out of 48 normal images are correctly classified as normal to produce a true negative rate of 91.66%.The overall accuracy of the system is determined using the formula,

$$Accuracy = \frac{No.of\ images\ classified\ correctly}{Total\ no.of\ test\ images} \qquad (4)$$

Using the proposed system, 74 out of the 80 images tested are classified correctly thus producing an overall classification accuracy of 92.5%.

TABLE 2: Confusion Matrix obtained using the proposed Method for Test Images

True Class	Predicted Class	
	Normal	**Abnormal**
Normal (48)	44 (True Negative)	4 (False Positive)
Abnormal (32)	2 (False Negative)	30 (True Positive)

The results obtained using the proposed method is compared against the results obtained using other methods and the parameters obtained are shown in TABLE 3. In the first method wavelet based texture features are extracted and classified using PNN classifier. In the second method contourlet based texture features are extracted and classified using PNN classifier without using a feature selection algorithm. Thus from the results obtained it is clear that the accuracy of the proposed system with feature selection is better compared to the other methods.

TABLE 3: Comparison of results obtained

Method	Wavelet+PNN	Contourlet+CCCM+PNN	Proposed Method with SFFS
Classification Accuracy	85%	86.25%	92.5%
True Positive Rate	87.5%	87.5%	93.75%
True Negative Rate	83.33%	85.41%	91.66%
False Positive Rate	16.6%	14.5%	8.33%
False Negative Rate	12.5%	12.5%	6.25%

V. CONCLUSIONS

In this study, a texture based analysis of digital mammograms using contourletcoefficient co-occurrence matrix features is proposed. The method proves the potential of contourlet transform and second order texture features for classification problems.From the results obtained it is clear that the use of SFFS algorithm for feature selection has helped to improve the classification accuracy. The experimental dataset taken for this study consists of both normal and abnormal mammogram images taken from mini-MIAS database and a highest classification accuracy of 92.5% is achieved. The future work should include automating the segmentation procedure so that the classification process becomes fully automatic and further classifying the abnormal mammograms into benign and malignant class.

REFERENCES

[1] www.cajournal.org
[2] A. Jemal, R. Siegel, E. Ward, Y. Hao, J. Xu, T. Murray,and M. J. Thun, "Cancer Statistics, 2008," *CA: A CancerJournal for Clinicians,* vol. 58, pp. 71-96, 2008.
[3] A. Jemal, R. Siegel, E. Ward, T. Murray, J. Xu, and M. J.Thun, "Cancer Statistics, 2007," *CA: A Cancer Journal for Clinicians,* vol. 57, pp. 43-66, 2007.
[4] D. K. Espey, X. C. Wu, J. Swan, C. Wiggins, M. Jim, E.Ward, P. A. Wingo, H. L. Howe, L. A. G. Ries, and B. A. Miller, "Annual report to the nation on the status of cancer, 1975–2004, featuring cancer in American Indians and Alaska Natives," *Cancer,* vol. 110, pp. 2119-2152, 2007.
[5] C. D. Maggio, "State of the art of current modalities for the diagnosis of breast lesions," *European Journal of Nuclear Medicine and Molecular Imaging,* vol. 31, pp. 56-69, 2004.
[6] M. J. Homer, *Mammographic Interpretation: A Practical Approach*. New York: McGraw-Hill Companies, 1991.
[7] ACS, "Cancer Prevention & Early Detection Facts & Figures 2008".
[8] A. O. Malagelada, *Automatic mass segmentation in mammographic images*, PhD Thesis,Universitat de Girona, Spain, 2004.
[9] D. Raba, A. Oliver, J. Marti, M. Peracaula and J. Espunya, *Breast segmentation with pectoral muscle suppression on digital mammograms*, Springer-Verlag: Medical Imaging: Pattern Recognition and Image Analysis, 3523 (2005), 471-478.

[10]   M. Roffilli, *Advanced machine learning techniques for digital mammography*, Technical Report, Department of Computer Science University of Bologna, Italy, 2006.

[11]   L. Semler and L. Dettori, *A comparison of wavelet-based and ridgelet-based texture classification of tissues in computed tomography*, International Conference on Computer Vision Theory and Applications, 2006.

[12]   L. Semler, L. Dettori and J. Furst, *Wavelet-based texture classification of tissues in computedtomography*, IEEE International Symposium on Computer-Based Medical Systems, 2005

[13]   E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, "Shiftablemultiscale transforms", *IEEE Transactions on Information Theory, Special Issue on Wavelet Transforms and Multiresolution Signal Analysis*, vol. 38, no. 2, pp. 587–607, March 1992. IEEE TRANSACTIONS ON IMAGE PROCESSING 11

[14]   F. G. Meyer and R. R. Coifman, "Brushlets: A tool for directional image analysis and image compression," *Journal of Appl. and Comput.Harmonic Analysis*, vol. 5, pp. 147–187, 1997.

[15]   N. Kingsbury, "Complex wavelets for shift invariant analysis and filtering of signals," *Journal of Appl. and Comput.Harmonic Analysis*,vol. 10, pp. 234–253, 2001.

[16]   E. J. Cand`es and D. L. Donoho, "Curvelets – a surprisingly effective nonadaptive representation for objects with edges," in *Curve and Surface Fitting*, A. Cohen, C. Rabut, and L. L. Schumaker, Eds. Saint-Malo: Vanderbilt University Press, 2000.

[17]   M. N. Do and M. Vetterli, "Contourlets," in *Beyond Wavelets*, G. V. Welland, Ed. New York: Academic Press, 2003.

[18]   The Contourlet Transform: An Efficient Directional Multiresolution Image Representation, *IEEE Transactions on Image Processing",* VOL. 14*,* No. 12, December 2005

[19]   P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Trans. Commun.*, vol. 31, no. 4, pp. 532–540, April1983.

[20]   R. H. Bamberger and M. J. T. Smith, "A filter bank for the directional decomposition of images: Theory and design," *IEEE Trans. Signal Proc.*, vol. 40, no. 4, pp. 882–893, April 1992

[21]   Peckingpaugh SH., "An improved method for computing gray-level co-occurrence matrix based texture measures". *Graph Model Image Proc*. 1991; 53: 574-80.

22.   Haralick RM, Shanmugam K, Dinstein I. Textural features for image classification. IEEE Trans Syst Man Cybern1973;SMC-3:610–21.

23.   Walker RF, Jackway P, Longstaff ID. Improving co-occurrence matrix feature discrimination. In: Proceedings of the 3rd Conference on Digital Image Computing: Techniques and Applications (DICTA '95); 1995 December 6–8; Brisbane, Australia, 1995.

[24]   Pudil P, Novovičová J, Kittler J, " Floating search methods in feature selection" *Pattern Recognition Letters*,1994; 15(11): 1119-25.

[25]   Ververidis D, Kotropoulos C. "Fast Sequential Floating Forward Selection applied to emotional speech features estimated on DES and SUSAS data collections", *Proceedings of the 14th European Signal Processing Conference*; Florence, Italy. 2006.