

Survey Paper on Various Texture Extraction Methods Applied in Mammograms for False Positive Reduction

Varsha Baby

*Department of Computer Science and Engineering
Karunya University, Tamil Nadu, India*

Abstract

The texture extraction is one among the several steps performed in image processing. Feature extraction method is the phase which is responsible for extracting all the possible features from the region of interest in the mammograms. The texture extraction is defined as the extraction of different features from the mammogram images with the help of different methods which work with the help of different algorithms. In this paper, we studied the different texture extraction methods in the classification of mammograms for reducing the false positive. The false positive arises from the misunderstanding by the radiologists taking the normal parenchyma as lesion or the lesion containing the actual mass as normal one. All of the considered methods can be used to classify the mammogram image into mass or non mass and if it is mass, it has again classified into benign or cancer (malignant). The considered approaches include wavelet based method, ridgelet based method, co-occurrence based method, square centroid lines gray level distribution method, ranklet transform, texture and statistical feature based method, method based on intensity histogram features, contourlet based feature extraction and feature extraction using gabor filter bank method. The following texture descriptors were calculated to analyze the regions of interest (ROIs) texture patterns: entropy, energy, sum average, sum variance, and cluster tendency, mean, variance, standard deviation etc in each of the papers. It was found that each of the methods had its own advantages and disadvantages.

1. Introduction

Breast cancer is the most frequently diagnosed cancer in women all over the world. It is found in men also but only in rare cases. It has been proved that it is the second most common and leading cause of cancer death among women. There are many steps in the field of image processing. They are pre processing, segmentation, cropping, texture extraction (feature extraction), feature selection and classification. This paper gives an

overall idea about the several methods which are available for the texture extraction phase and the advantages and the disadvantages of each of the methods in comparison with the other methods. The considered methods for texture extraction phase include wavelet based method, ridgelet based method, co-occurrence based method, square centroid lines gray level distribution method, ranklet transform, texture and statistical feature based method, method based on intensity histogram features, contourlet based feature extraction and feature extraction using gabor filter bank method.

Feature extraction method is the phase which is responsible for extracting all the possible features from the region of interest in the mammograms. The main objective of all the techniques is to extract the necessary features and also to reduce the false positive in the mammogram images. The input of all the techniques is a set of mammogram images which are taken from the data sets like DDSM (Digital Database for Screening Mammography) or MIAS (Mammographic Image Analysis Society). These set of images will be passed to the pre processing and cropping phases before extracting the textures. The usually extracted features include mean, variance, standard deviation, energy, entropy, sum average, sum variance, cluster tendency, kurtosis, skewness, and maximum and minimum variance, average variance etc.

The definition of co-occurrence matrix is like it is an $M \times M$ matrix whose rows and columns are defined by the image gray levels $i = 1, 2, \dots, M$ where $M = 2^m$ for an m -bit image. The each of the normalized entry $p(i, j)$ represents the statistical frequency by which the given pair of grey level pixels i and j is found to be separated by a distance d at an angles $0, 45, 90$ and 135 degrees [10]. A wavelet is defined as a function with a zero average. The function is a set of square integrable real functions. Different families of wavelets have been developed such as Haar, Daubechies, Coiflet, cubic splines and others. The discrete wavelet transform (DWT) is obtained by the discretization of time as well as the translation and scale

parameters for discrete signals. The main concept of wavelet transform is that an original image at the j th level is wavelet transformed, yielding four sub-band images at the next level ie, at the $(j+1)$ th level [10]. Wavelets can deal successfully only with point singularities which are one dimensional in nature and which are points of bad behaviour. But on the other hand, ridgelets can deal effectively and successfully with line singularities which are two dimensional in nature. This is the main advantage of ridgelet over wavelet based method [10].

Square Centroid Lines Gray Level Distribution method is the recently developed method for extracting the features in mammograms. SCLGM is the modification of the two of the known methods Spatial Gray Level Dependence Method (SGLDM) and Run Difference Method (RDM) which can enhance their power in describing textural characteristics of the mass patterns. The smallest square that includes the segmented mass with zero background is used for this method [2]. Ranklet transform based method is one of the widely used methods for extracting textures in mammograms. The regions containing the detection module are encoded in the RFPR (Ranklet False Positive Reduction) module of both the RFPR and RFPR+RTFPR (Ranklet Texture False Positive Reduction) CAD system by means of the ranklet transform. A unique approach is adopted by the detection module. Multiresolution, orientation selective, and nonparametric analysis is performed by applying the ranklet transform to an image. Ranklet images are nonparametric because they are derived from the relative rank of the pixels, rather than their gray-scale values. Hence they are very robust to linear or nonlinear monotonic gray-scale transformations of the original image [1].

Feature is defined as a piece of information which is relevant for solving the computational task related to a certain application. Texture features, statistical features and structural features are extracted for the segmented tumor from the given input mammogram image. The features extracted are mean, standard deviation, smoothness, entropy, skewness, kurtosis, root mean square, inverse difference moment, energy, contrast, correlation homogeneity and variance [9]. Intensity Histogram analysis has been widely used in the initial stages of texture extraction. The intensity histogram features are mean, variance, entropy, skewness, kurtosis and energy etc [7]. Contourlet transform is the backbone of the contourlet based feature extraction. The contourlet transform is used as a feature extractor in order to derive the contourlet coefficients. The contourlet based feature extraction in conjunction with the state-of-art classifiers construct a powerful, efficient and

practical approach for automatic mass classification of mammograms and thus reducing the false positive. The practical approaches for image representation are mainly known as adaptive and non-adaptive representations. Curvelets and shearlets are examples for fixed non-adaptive representations which are meant to be essentially as effective as an adaptive representation. An image can be represented at different scales with the help of multiresolution analysis. Wavelet transform can successfully deal with 1-D piecewise smooth signals. The structures of many textured images such as medical images, however, are not that simple to be represented by 1-D piecewise smooth lines. But contourlets can overcome these problems [4]. The Gabor filter bank is employed in a novel way to extract the texture descriptors that characterize micropatterns (e.g. edges, lines, spots and flat areas) at different scales and orientations. Gabor filter bank allow different choices for the number of scales and orientations. Gabor filters are defined as biologically motivated convolution kernels and their response is found to be similar to receptive fields of neurons in the visual cortex [7]. They possess optimal joint localization both in frequency as well as spatial domains. Gabor filters are widely applied in the fields of computer vision and image processing e.g. face recognition, vehicle detection and texture analysis etc. A Gabor filter bank consists of multiple Gabor filters along with different parameter settings. It is found that a filter will have stronger response to an edge where the normal is parallel to the orientation [8].

2. Different texture extraction methods

Feature extraction method is the phase which is responsible for extracting all the possible features from the region of interest in the mammograms. There are several texture extraction methods available in the field of image processing to reduce the false positive in mammograms. The methods considered are wavelet, ridgelet, co-occurrence methods, SCLGM, ranklet transform method, method based on texture and statistical feature extraction ,method based on Intensity Histogram Features, contourlet based feature extraction method and Feature Extraction using Gabor Filter Bank method.

2.1 Wavelet, ridgelet and co-occurrence based methods

The wavelet, ridgelet and co-occurrence based methods are mentioned in Rodrigo et al., (2012). The data sets taken are MIAS and DDSM consisting of 120 cranio-caudal mammograms half with lesions and half with no lesions. The aim is to classify the given image into mass or non-mass and

if it is mass, then it should be again classified in to cancer or benign. The wavelet, ridgelet and co-occurrence based methods are employed for this. All these three methods were compared and wavelet is found to be the best method because of its high AUC value (AUC=0.90). The feature descriptors like energy, entropy, sum average, sum variance and cluster tendency were calculated to extract features for each of the three methods.

2.1.1 co-occurrence based method

The definition of co-occurrence matrix is like it is an $M \times M$ matrix whose rows and columns are defined by the image gray levels $i = 1, 2, \dots, M$ where $M = 2^m$ for an m -bit image. The each of the normalized entry $p(i, j)$ represents the statistical frequency by which the given pair of grey level pixels i and j is found to be separated by a distance d at an angles $0, 45, 90$ and 135 degrees. The 5 descriptors per image is calculated. The 20 descriptors are extracted from the SGLD matrices built at four pixel distances ($d = 1, 3, 6, 9$) giving four separate feature vectors. A feature matrix was developed, with each row representing a feature vector of a given ROI, and this matrix was used by the classifier. Since for the comparison of computational efforts, a full SGLD matrix, with 12 bits of pixel resolution, and a sub sampled SGLD matrix, with 8 bits of pixel resolution is used. And this matrix is obtained by reducing the image resolution to 8 bits and computing its SGLD matrix. So a total of 600 features were extracted with the help of co-occurrence based method [10].

2.1.2 Wavelet based method

A wavelet is defined as a function with a zero average. The function is a set of square integrable real functions. Different families of wavelets have been developed such as Haar, Daubechies, Coiflet, cubic splines and others. The discrete wavelet transform (DWT) is obtained by the discretization of time as well as the translation and scale parameters for discrete signals. The dyadic DWT of a signal is equivalent to its decomposition through high-pass ($g[n]$) and low-pass ($h[n]$) filter banks. The DWT can be calculated with separable wavelet functions for 2 dimensional signals or images. This means that a 2-dimensional filter can be decomposed as the product of two one-dimensional filters. A 1D-DWT is computed first in the image rows, and another 1D-DWT is applied to the columns of the two images generated.

The main concept of wavelet transform is that an original image at the j th level is wavelet transformed, yielding four sub-band images at the next level ie, at the $(j+1)$ th level. So there are three

detail images that represents horizontal, vertical and diagonal directions and one approximation image which is the original image at a coarse resolution. The approximation coefficients from the upper level are applied to the 2D-DWT bank for each desired resolution and others 4 sub-images are created. A total of 5 descriptors per band is calculated yielding 35 descriptors from one image which contains 7 bands. So a total of 4200 features were calculated with the help of wavelet method from 120 images [10].

2.1.3 Ridgelet based method

There are two types of singularities like point singularities and line singularities. Wavelets can deal successfully only with point singularities which are one dimensional in nature and which are points of bad behaviour. But on the other hand, ridgelets can deal effectively and successfully with line singularities which are two dimensional in nature. This is the main advantage of ridgelet over wavelet based method. The ridgelet based method was developed by Candes and Donoho. The discrete version is required to apply the ridgelet transform to digital images, which needs the discrete Radon transform. The discrete version of the ridgelet transform is the finite ridgelet transform (FRIT). The FRIT is defined as the summations of image pixels over a certain set of lines. It is based on the finite Radon transform. FRIT can be computed using two steps. The first is by computing the discrete Radon transform, through a 2-dimensional fast Fourier transform (FFT) followed by an one-dimensional inverse Fourier transform for each radial direction of the Radon projections and the second is by applying a one-dimensional wavelet transform over r resolutions for each radial direction of the Radon transform. So the given image is divided into 62 columns from each dimension, yielding 620 features per image. 5 descriptors are extracted from each column which comprises a total of 74400 features from 120 images [10].

2.2 Square centroid lines gray level distribution method

Square Centroid Lines Gray Level Distribution method is the recently developed method for extracting the features in mammograms. SCLGM is the modification of the two of the known methods Spatial Gray Level Dependence Method (SGLDM) and Run Difference Method (RDM) which can enhance their power in describing textural characteristics of the mass patterns. The smallest square that includes the segmented mass with zero background is used for this method. The discriminative information about the mass type is provided by the relations between

the pixels of each centroid line. There are four centroid lines (C_i) in this method. They are C_1 at $\theta = 0$, C_2 at $\theta = \pi/4$, C_3 at $\theta = \pi/2$ and C_4 at $\theta = 3\pi/4$ and these lines pass through the square's center point. The gray level points at each centroid line is represented by C_{ji} . Before extracting the textural features a set of statistics are to be computed after defining C_{gi} .

Six mathematical measures are computed by choosing four centroid lines, which are used to extract 79 features. The statistics computed include mean, variance and its difference vector, standard deviation and its difference vector, mean absolute deviation (MAD) and its differences, skewness and kurtosis. These statistics can be computed only by extracting textural features. Average variance, average difference of variance, minimum variance, minimum difference of variance, maximum variance and difference of variance, Average of Standard deviation are some of the textural features extracted.

The SCLGM algorithm's main purpose is to extract 75 features from the mammogram images. The input is enhanced segmented objects with black background and the used direction for co-occurrence $\theta = \{0^\circ, \pi/4, \pi/2, 3\pi/4\}$. The output is the SCLGM extracted features vector of the input image [2].

2.3 Ranklet transform

Ranklet transform based method is one of the widely used methods for extracting textures in mammograms. The regions containing the detection module are encoded in the RFPR (Ranklet False Positive Reduction) module of both the RFPR and RFPR+RTFPR (Ranklet Texture False Positive Reduction) CAD system by means of the ranklet transform. A unique approach is adopted by the detection module. Multiresolution, orientation selective, and nonparametric analysis is performed by applying the ranklet transform to an image. Ranklet images are nonparametric because they are derived from the relative rank of the pixels, rather than their gray-scale values. Hence they are very robust to linear or nonlinear monotonic gray-scale transformations of the original image.

Each region is decomposed into 4 resolutions for the detection and RFPR modules. This method is as arbitrary as reasonable, since it spans over a large range of resolutions, from fine to coarse. The result will be obtained as 12 ranklet images and these are then linearized into a 1×1428 ranklet feature vector invariant to linear/nonlinear monotonic gray-scale transformations of the original image which is used to discriminate through SVM classification between abnormal/normal regions.

The ranklet images corresponding to the regions containing the detection and RFPR modules are used as a starting point for the calculation of a number of ranklet texture features in the RTFPR. 11 texture features are extracted in the feature extraction step for each ranklet image derived from the ranklet decomposition at different resolutions and orientations of an image. Each ranklet image is encoded by means of a 1×11 texture feature vector with the help of texture feature extraction step [1].

2.4 Texture and statistical feature based method

Feature is defined as a piece of information which is relevant for solving the computational task related to a certain application. Features can also be defined as the result of a general neighborhood operation applied to the image and specific structures in the image itself. The structures range from simple structures such as points or edges to more complex structures such as objects. Features are mainly extracted for finding the abnormalities of mammograms. The texture feature extraction methods play vital role in the detection of abnormalities in mammograms. The texture features are proven to be useful in differentiating masses and non masses (normal breast tissues) [9].

Texture features, statistical features and structural features are extracted for the segmented tumor from the given input mammogram image. The features extracted are mean, standard deviation, smoothness, entropy, skewness, kurtosis, root mean square, inverse difference moment, energy, contrast, correlation homogeneity and variance. The definitions for each of the terms are given in a detailed manner in predeep et al.

2.5 Method based on intensity histogram features

Intensity Histogram analysis has been widely used in the initial stages of texture extraction. The intensity histogram features are calculated. The intensity histogram features include mean, variance, entropy, skewness, kurtosis and energy. Mean values gives an overall idea about the individual calcifications and standard deviations gives an overall idea about the cluster. The mean, variance, entropy, skewness, kurtosis and energy values for normal, benign and cancer images are calculated [7].

2.6 Contourlet based feature extraction

Contourlet transform is the backbone of the contourlet based feature extraction. The

contourlet transform is used as a feature extractor in order to derive the contourlet coefficients. The contourlet based feature extraction in conjunction with the state-of-art classifiers construct a powerful, efficient and practical approach for automatic mass classification of mammograms and thus reducing the false positive.

The practical approaches for image representation are mainly known as adaptive and non-adaptive representations. Curvelets and shearlets are examples for fixed non-adaptive representations which are meant to be essentially as effective as an adaptive representation. An image can be represented at different scales with the help of multiresolution analysis. Wavelet transform can successfully deal with 1-D piecewise smooth signals. The structures of many textured images such as medical images, however, are not that simple to be represented by 1-D piecewise smooth lines. Candes and Donoho [3] showed that wavelets can do well for objects with point singularities in 1-D and 2-D. The orthogonal wavelets capture only discontinuities along the directions of horizontally, vertically, and diagonally. These orientations may not preserve enough directional information in medical images. Ridgelet analysis, on the other hand, is an appropriate transform to catch radial directional details in frequency domain. Ridgelets are very effective in detecting linear radial structures. But those structures are not dominant in medical images. Ridgelets are very effective in the detection of linear radial structures. But those structures are not dominant in medical images. Curvelet transform is an extension of the Ridgelet transform introduced by Candes and Donoho. It is found that curvelets are very successful in detecting image activities along curves, while analyzing images at multiple scales, locations, and orientations.

There are mainly two types of decompositions namely mutiscale decomposition and directional decomposition. The multiscale decomposition is capable of capturing point discontinuities while the directional decomposition is capable of linking point discontinuities into linear structures. The curvelet transform can overcome the directionality lack of 2-D wavelets by geometrically representing smoothness of contours. It can also be defined like curvelets can represent a smooth contour with fewer coefficients than wavelets do. And moreover wavelet basis functions are limited to using square shaped (i.e., isotropic) along the contour, using different sizes corresponding to the multiresolution structure of wavelets. Because of that they cannot adapt to geometrical structures. The curvelet transform not only exploit its multiscale and time-frequency localization properties of wavelets, but also offers a high degree of directionality as well as an isotropic structured basis functions with effective support

shaped. So the contourlet transform is superior compared with the wavelet transform in approximating this type of 2-D piecewise smooth functions. This forms the underlying reason for the success of the curvelet transform.

A curvelet is defined by three parameters namely a scale parameter a which varies between 0 and 1, an orientation θ which varies between $-\pi/2$ and $\pi/2$ and a location parameter b . There are no significant differences in the time performance in comparison with the wavelets transform though the curvelets are more complex. To implement the idea of curvelet transform, Do and Vetterli [6] constructed a discrete domain multiresolution and multidirection expansion using non-separable filter banks for generating sparse expansions for typical images having smooth contours. The Laplacian pyramid (LP) is first used to capture the point discontinuities, followed by a directional filter bank (DFB) to link point discontinuities into linear structures in this double filter bank. An image expansion using basic elements like contour segments, named contourlets was the overall result. The contourlets have elongated supports at various scales, directions and aspect ratios. Contourlets are used for representing features of breast tissues because of its efficient properties [4].

2.7 Feature extraction using gabor filter bank

The two early signs of breast cancer are masses and microcalcifications. The segmentation of mammograms results in ROIs (regions of interest) which includes both masses and suspicious normal tissues leading to false positives. So the problem is to reduce the false positives by classifying ROIs as masses and normal tissues. The detected masses should be classified again as malignant and benign for the interpretation of mammogram. These two problems are addressed using the textural properties of masses. Gabor filter bank can be used in a novel way to extract the most representative and discriminative textural properties of masses which are present at the different orientations and scales. It makes the textural properties for a robust and discriminative representation of masses. The textural properties are useful for the representation of masses.

The Gabor filter bank is employed in a novel way to extract the texture descriptors that characterize micropatterns (e.g. edges, lines, spots and flat areas) at different scales and orientations. Gabor filter bank allow different choices for the number of scales and orientations. The major question faced here is how many scales and orientations are necessary to represent accurately the texture patterns of mass ROIs. The extracted ROIs have different sizes and it is very difficult to

deal ROIs of different sizes with Gabor filter bank. But there is the need to resize ROIs. Now another question faced is which size of ROIs will yield optimal results. Gabor filter bank are specifically applied for false positive reduction and also for the classification of benign-malignant problems. And the approach for feature extraction is taken to be a combination of local and global approaches. Each suspicious ROI are divided into overlapping windows, which form the global representation. After that, Gabor filter bank is applied on each window. Then the moments (mean, standard deviation, skewness) are extracted from the magnitudes of Gabor responses, which form the local representation of each window at different scales and orientations. The Gabor magnitude responses moments are concatenated for all windows corresponding to an ROI which is a combination of local and global feature representations.

Gabor filters are defined as biologically motivated convolution kernels and their response is found to be similar to receptive fields of neurons in the visual cortex [5]. They possess optimal joint localization both in frequency as well as spatial domains. Gabor filters are widely applied in the fields of computer vision and image processing e.g. face recognition, vehicle detection and texture analysis etc. It is found that a filter will have stronger response to an edge where the normal is parallel to the orientation. A Gabor filter bank consists of multiple Gabor filters along with different parameter settings. The different parameters of Gabor filter bank include scaling, orientation and central frequency.

The effect of Gabor filter banks with different scale and orientation settings such as GS2O3: 2 scales S x 3 orientations O), 15 filters (GS3O5), 24 filters (GS4O6) and 40 filters (GS5O8) are investigated in Muhammad et al. The initial maximum frequency is equal to 0.2 and the initial orientation is set to 0 in each of the cases [8].

3. Conclusion

The different types of texture extraction methods are studied in this paper. The methods considered include wavelet based method, ridgelet based method, co-occurrence based method, square centroid lines gray level distribution method, ranklet transform, texture and statistical feature based method, method based on intensity histogram features, contourlet based feature extraction and feature extraction using gabor filter bank method. Like the coin has two sides each of the methods have its own advantages and disadvantages in comparison with the other methods. The ridgelet

can successfully deal with two dimensional singularities which are line singularities but not effective in dealing with one dimensional singularities. On the other hand the wavelets can successfully deal with one dimensional singularities which are point singularities but not effective in dealing with two or more than two dimensional singularities. The feature extraction time taken by ridgelet and contourlet based method is somewhat same. The accuracy rate or the area under curve (AUC) value is greater for wavelet than the ridgelet method which is found to be 0.90. The best method cannot be predicted in general because of the dependence of the nature of the classifiers used along with these texture extraction methods. According to the classifiers which are used with the texture extraction methods, the false positive rate also changes.

References

- [1] Arianna Mencattini, Giulia Rabottino, Marcello Salmeri, Roberto Lojaco, and Eleonora Tamilia, "Features Extraction and Fuzzy Logic Based Classification for False Positives Reduction in Mammographic Images".
- [2] Belal K. Elfarra and Ibrahim S. I. Abuhaiba, "New Feature Extraction Method for Mammogram Computer Aided Diagnosis", International Journal of Signal Processing, Image Processing and Pattern Recognition (2013).
- [3] E.J. Candes, D.L. Donoho, "Curvelets: a surprisingly effective non adaptive representation for objects with edges in Saint-Malo Proceedings", Vanderbilt University, Nashville, TN, 2000, pp. 1–10.
- [4] Fatemeh Moayedi, Zohreh Azimifar, Reza Boostani, Serajodin Katebi, "Contourlet-based mammography mass classification using the SVM family", Computers in Biology and Medicine 40 , 373–383, (2010).
- [5] J. G. Daugman, "Two-dimensional spectral analysis of cortical receptive field profiles", Vis. Res., Vol. 20, pp. 847-856, 1980.
- [6] M.N. Do, M. Vetterli, "The contourlet transform: an efficient directional multi-resolution image representation", IEEE Transactions on Image Processing 14 (12) (2005) 2091–2106.
- [7] M.Vasantha, Dr.V.Subbiah Bharathi, R.Dhamodharan, "Medical Image Feature, Extraction, Selection And Classification",

International Journal of Engineering Science and Technology Vol. 2(6), 2071-2076, (2010).

[8] Muhammad Hussain , Salabat Khan , Ghulam Muhammad , Iftikhar Ahmad , George Bebis, “Effective Extraction of Gabor Features for False Positive Reduction and Mass Classification in Mammography”, Appl. Math. Inf. Sci. 6, No. 1, 29-33 (2012)

[9] Pradeep N, Girisha H, Sreepathi B and Karibasappa K, “Feature extraction of mammograms”, International Journal of Bioinformatics Research ISSN: 0975–3087 & E-ISSN: 0975–9115 , Volume 4, Issue 1, 2012.

[10] Rodrigo Pereira Ramos, Marcelo Zanchetta do Nascimento, Danilo Cesar Pereira, “Texture extraction: An evaluation of ridgelet, wavelet and co-occurrence based methods applied to mammograms”, Expert Systems with Applications 39 (2012) 11036–11047.

IJERT