Classification Mammogram Images Using ID3 decision tree algorithm Based on Contourlet Transform

Dr. Matheel Emaduldeen Abdulmunim
Computer Sciences Department, University of Technology/Baghdad.
Email: Matheel_74@yahoo.com
Zainab Fawzy Abed
Computer Sciences Department, University of Technology/Baghdad.

Received on: 6/4/2014 & Accepted on: 8/1/2015

ABSTRACT
Breast cancer is the most common malignancy of women and is the second most common and leading cause of cancer deaths among them. At present, there are no effective ways to prevent breast cancer, because its cause is not yet fully known. Early detection is an effective way to diagnose and manage breast cancer can give a better chance of full recovery. Therefore, early detection of breast cancer can play an important role in reducing the associated morbidity and mortality rates. In this paper, using contourlet transform that can capture the intrinsic geometrical structure that is key in visual information. The contourlet expansion is composed of basis images oriented in various directions in multiple scales, with flexible aspect ratios. The basic idea of this paper is to design and implement a proposed system that can aid the physician in reading a mammogram image by study the usage of wavelet and contourlet transform based on various operations on mammogram images and classifies them as normal, benign or malignant based on the decision tree ID3 algorithm. The experimental results show that the ID3 classifier achieves accuracy of 81% in the case of wavelet transform and 95% with contourlet transform for the same number of the test set.

Keywords: mammogram, transforming, segmentation, feature classifier, training, testing.
INTRODUCTION

Classification is a form of data analysis, which can be used to extract models describing important data classes or to predict future data trends. Data classification is a two-step process. In the first step, a model is built describing a predetermined set of data classes. The model is constructed by analyzing database tuples described by attributes. Each tuple is assumed to belong to a predefined class, as determined by one of the attributes, called the class label attribute. In the second step, the model is used for classification [1]. One of the tools to solve the problem of classification is based on decision trees. These tools have been used successfully in diverse areas such as radar signal classification, character recognition, remote sensing, medical diagnosis, expert systems, speech recognition and others. Perhaps, the most important characteristic of decision trees is their capability to break down a complex decision-making process into a collection of simpler decisions, thus providing a solution which is often easier to interpret [2]. The most important step in classification is transforming. The main restriction in transforming the two-dimensional wavelet is not receiving information from different directions. As you know the wavelet transform just can extract information in vertical, horizontal and diametrical direction. To conquer this problem, researcher use the multi-direction scales method which can conclude the main structure of picture’s geometry like contourlet transform. Contourlet transform in addition to have main features of wavelet transform which has multi scale and time frequency it also have high degree of direction and anisotropy [3].

Transformation

The transformation is a process that translates an object from a given domain to another in order to have some important implicit information, which can be used for its recognition. One of the conventional transformation is the Fourier transform which usually transforms the signal from its time domain to the frequency domain. The next form of the Fourier transform developed to an efficient transform is called the Wavelet Transform (WT) [27]. The wavelet can be regarded as the most efficient transform that deals with image, sound, or any other pattern since it provides a powerful time-space (time-frequency) representation [4]. The Contourlet Transform (CT) is a new version of wavelet which described later.

Wavelet Transform

Wavelet transform is a signal processing method that has been implemented in image processing and pattern recognition for the last decades. It is currently an important feature to be used in texture classification and has been very popularly used...
Wavelets are functions that satisfy certain requirements. The name wavelet comes from the requirement that they should integrate to zero, "waving" above and below the x-axis [6]. The purpose of wavelet transform is to change the data from time-space domain to time-frequency domain, which makes better compression results [7]. The information on the frequency domain is usually more stable than the spatial domain. Therefore, they often produce better features that lead to a higher accuracy despite being more complex and slower [5]. Wavelets are basis functions which are able to represent a signal in the time and frequency domain at the same time. They can be used to approximate an underlying trace or signal, similar to Fourier transforms[8]. Wavelets are functions defined over a finite interval. The basic idea of the wavelet transform is to represent an arbitrary function ƒ(x) as a linear combination of a set of such wavelets or basis functions [7]. These basis functions are obtained from a single function called the mother wavelet, by dilation (scaling) factor (s) and the translation (shifts) factor [9], as shown in equations (1).

\[
T = \frac{1}{\sqrt{S}} \phi \left( \frac{T-L}{S} \right)
\]  

Where
ψ_{S,L}(T): Translated and scaled version of the mother function. S: The scale factor, T: The shift factor. \sqrt{S}: The energy normalization that keeps the energy of daughter wavelet equal to that of the mother wavelet. The advantages of Wavelets are that they are localized in frequency and time and so can handle a wider range of signals than Fourier analysis [28]. A disadvantage of Wavelets is that the transform obtained only has representations of the data at a discrete number of resolution levels, each resolution level having a representation at approximately twice the frequency of the previous level [8]. Wavelet transform has been used widely in a number of useful applications, such as data compression, detection features in images, and image denoising [10].

**Discrete wavelet transform (DWT)**

The discrete wavelet transform maps an image into a set of coefficients that constitute a multiscale representation of the image [11]. An image can be analyzed by passing it through an analysis filter bank followed by a decimation operation. This analysis filter bank, which consists of a low pass and a high pass filter at each decomposition stage, is commonly used in image compression [12]. When an image passes through these filters, it is split into two bands. The low pass filter, which corresponds to an averaging operation, extracts the coarse information of the signal. The high pass filter, which corresponds to a differencing operation, extracts the detail information of the signal. The output of the filtering operation is then decimated by two. A two dimensional transform can be accomplished by performing two separate one-dimensional transforms. First, the image is filtering along the x-dimension and decimated by two. Then, it is followed by filtering the sub image along the y-dimension and decimated by two[13], as shown in equations (2) and (3).

\[
Y_{low} = \sum X[n] \ast G[2k - n]
\]  

\[
Y_{low} = \sum X[n] \ast G[2k - n]
\]  

\[
Y_{low} = \sum X[n] \ast G[2k - n]
\]
\[ Y_{\text{high}} = \sum X[n] \ast H[2k - n] \]  

...(3)

This decomposition has halved the time resolution since only half of each filter output characterizes the signal. However, each output has half the frequency band of the input so the frequency resolution has been doubled [14]. Figure (1) represent a block diagram of filter analysis.

![Figure (1): Block diagram of filter analysis [14].](image1)

This decomposition is repeated to further increase the frequency resolution and the approximation coefficients decomposed with high and low pass filters and then down-sampled (decimated by two) [14], As shown in Figure (2).

![Figure (2): A three level filter bank [14].](image2)

**Wavelet Families**

There are a number of basis functions that can be used as the mother wavelet for wavelet transformation. Since the mother wavelet produces all wavelet functions used in the transformation through translation and scaling, it determines the characteristics of the resulting wavelet transform. The main types of mother wavelet are presented, according to their signal shape and to the person who invented it, as shown in Figure (3). The Haar, Daubechies, Symlets and Coiflets are compactly supported orthogonal wavelets. These wavelets along with Meyer wavelets are capable of perfect reconstruction. The Meyer, Morlet and Mexican Hat wavelets are symmetric in shape [15].
In mathematics, the Haar wavelet is a sequence of rescaled "square shaped" functions which together form a wavelet family or basis. Wavelet analysis is similar to Fourier analysis in that it allows a target function over an interval to be represented in terms of an orthonormal function basis. The Haar sequence is now recognized as the first known wavelet basis and extensively used as a teaching example. The Haar wavelet is also the simplest possible wavelet. The technical disadvantage of the Haar wavelet is that it is not continuous, and therefore not differentiable. This property can, however, be an advantage for the analysis of signals with sudden transitions, such as monitoring of tool failure in machines. It can be used to implement the wavelet transform by convolving them with the row of one dimensional signals [15]. Thus, the four images produced from each decomposition level are LL, LH, HL, and HH. The LL image is considered a reduced version of the original as it retains most details. The LH image contains horizontal edge features, and the HL contains vertical edge features, while the HH sub-band corresponds to the diagonal edges. Only the LL image is used to produce the next level of decomposition, as shown in Figure (4) [16].
Contourlet Transform

The contourlet transform is one of the new geometrical image transforms, which represents images containing contours and textures [17]. The contourlet transform is a directional multiresolution image representation scheme proposed by Do and Vetterli, which is effective in representing smooth contours in different directions of an image [18]. The contourlet transform is a new two-dimensional extension of the wavelet transform [19]. Wavelets is offered multiscale and time-frequency localization of an image. However, wavelets are not effective in representing the images with smooth contours in different directions. Contourlet transform addresses this problem by providing high degree of direction and anisotropy.

1. Directionality: The representation should contain basis elements oriented at a variety of directions, much more than the few directions that are offered by Wavelets.
2. Anisotropy: To capture smooth contours in images, the representation should contain basis elements using a variety of elongated shapes with different aspect ratios [3,20].

To see how one can improve the 2-D separable wavelet transform for representing images with smooth contours, consider the following scenario. Imagine that there are two painters, one with a "wavelet style" and the other with a "contourlet style", both wishing to paint a natural scene. Both painters apply a refinement technique to increase resolution from coarse to fine. Here, efficiency is measured by how quickly, that is with how few brush strokes, one can faithfully reproduce the scene. Consider the situation when a smooth contour is being painted. Because 2-D wavelets are constructed from tensor products of 1-D wavelets, the "wavelet style" painter is limited to using square-shaped brush strokes along the contour, using different sizes corresponding to the multiresolution structure of wavelets. As the resolution becomes finer, we can clearly see the limitation of the wavelet-style painter who needs to use many fine “dots” to capture the contour. The "contourlet style" painter, on the other hand, exploits effectively the smoothness of the contour by making brush strokes with different elongated shapes and in a variety of directions following the contour. Figure (5) showing how wavelets having square supports that can only capture point discontinuities, whereas contourlets having elongated supports that can capture linear segments of contours, and thus can effectively represent a smooth contour with fewer coefficients [21].
The contourlet transform consists of two major stages, Laplacian Pyramid (LP) and Directional Filter Bank (DFB). At the first stage, LP is used to decompose the image into subbands, and then the second one is a DFB, which is used to analyze each detail image. This scheme results in an image expansion that uses basic elements like contour segments and thus it is named contourlet transform. The separation of directional and multiscale decomposition stages provides a fast and flexible transform for N-pixel images. A flow graph of the contourlet transform is shown in Figure (6).

Examples of the spectral split scheme achieved by such LP and the directionally-decomposed frequency split scheme achieved by DFB is shown in Figure (7) [7].
Figure (7): LP filter contents a) the spectral split scheme of LP filter. b) An example of the directional filter bank frequency partitioning[7].

Laplacian Pyramid Directional Filter Bank [18]

The LP provides multiscale decomposition. In each decomposition level, it creates a downsampled lowpass version of the original image and a bandpass image. A coarse image with the lower frequencies and a more detailed image with the supplementary high frequencies containing the point discontinuities are obtained. This scheme can be iterated continuously in the lowpass image and is restricted only from the size of the original image due to the downsampling.

The original DFB is efficiently implemented via an L-level binary tree leading to $2^L$ subbands with wedge-shaped frequency partitioning. By combining the LP and the DFB, a double filter bank named Pyramidal Directional Filter Bank (PDFB). Bandpass images from the LP decomposition are fed into a DFB in order to capture the directional information. The combined result is the Contourlet filter bank, which is a double iterated filter bank that decomposes images into directional subbands at multiple scales as shown in figure(8). In each level, the LP provides a downsampled lowpass and a bandpass version of image. The bandpass image is then fed into DFB.

Figure (8): contourlet filter bank[18].

Horizontal DFB and Vertical DFB

Directional representation that is potentially useful for image processing often with additional properties like exact reconstruction, reduced redundancy and high
computational efficiency. Directional filter banks [23]. Decompose the frequency space into wedge shaped partitions as illustrated in Figure (9). In this figure, eight directions are used, where directional subbands of 1, 2, 3, and 4 represent horizontal directions (directions between -45° and +45°) and the rest stand for the vertical directions (directions between 45° and 135°). It is realized using iterated quincunx filter banks [24].

Figure (9): Directional filter bank. a) Full frequency response of a (full tree) directional filter bank b) An Example of the Vertical Directional Filter Bank (c) An Example of the Horizontal Directional Filter Bank [24].

Multiscale

Multiscale data representation is a powerful idea. It captures data in a hierarchical manner where each level corresponds to a reduced-resolution approximation. The basic idea of the LP is the following. Based on this coarse version, predict the original (by up sampling and downsampling) and then calculate the difference as the prediction error. The sampling and downsampling operation for the LP shown in equations (4 and 5) [25].

\[
[n] = \sum_{d \in k} [m_n-k]
\]
\[\text{(4)}\]

\[
[n] = \sum_{d \in C} [k]G[n-m_k]
\]
\[\text{(5)}\]

Expressing these operations as left matrix multiplications \(P=Gc\), \(C=Hx\) where \(H\) and \(G\) correspond to \((\downarrow M)H\) and \(G(\uparrow M)\), respectively. When \(M=2\) then \(H,G\) as following:[25]
Suppose that the LP in the PDFB uses orthogonal filters and downsampling by two is taken in each dimension. Under certain conditions, the lowpass filter $G$ in the LP uniquely define an orthogonal scaling function $\varphi(t)$ via equation (6) [11].

$$\varphi(t) = 2\sum g[n]\varphi(2t-n) \quad \cdots (6)$$

$V_j$ is a subspace defined on a uniform grid with intervals $2^j \times 2^j$, which characterize the image approximation at the resolution $2^{-j}$. The difference images in the LP carry the details necessary to increase the resolution of an image approximation. Let $W_j$ be the orthogonal complement of $V_j$ in $V_{j-1}$ via equation (7). See Figure (10)[11].

$$V_{j-1} = V_j \oplus W_j \quad \cdots (7)$$

Figure (10): Multiscale subspaces generated by the Laplacian pyramid[11].

Multidirectional [11]
Suppose that the DFB’s in the PDFB use orthogonal filters. The DFB is applied to the difference image or the Wj subspace and subspaces Vj. Figure (11) illustrates the “two-direction” subspace splitting by the DFB in the frequency domain.

Figure (11): Multidirectional subspaces generated by the DFB[11].

**Multiscale and Multidirectional**

DFB is designed to capture high frequency components. The LP part of the PDFB permits sub-band decomposition to avoid “leaking” of low frequencies into several directional sub-bands, thus directional information can be captured efficiently. The number of directions is doubled at every other finer scale of the pyramid. Figure (12) graphically depicts the supports of the basis functions generated by such a PDFB. As can be seen from the two shown pyramidal levels, the support size of the LP is reduced by four times while the number of directions of the DFB is doubled. Combine these two steps, the support size of the PDFB basis functions are changed from one level to next in accordance with the curve scaling relation. In this contourlet scheme, each generation doubles the spatial resolution as well as the angular resolution. The PDFB provides a frame expansion for images with frame elements like contour segments, and thus is also called the contourlet transform [26].
Proposal System
The structure of the proposed system consists of two stages: training and testing. Each stage has specific functions, all the functions are explained in detail in the following sub sections. Figure (13) describe the block diagrams of the training and testing stages of the proposed mammograms classification system respectively.

4.1 Mammogram Images
The data set (mammogram image) used in this paper were taken from Mammography Image Analysis Society (MIAS), which is a UK research group organization related to the Breast cancer investigation. 190 images selected from the database, 100 images used for training set, 40 images are normal, 30 images are benign (non-cancerous) and 30 images are malignant (cancerous). 70 images used for testing, 30 images are normal, 20 images are benign and 20 images are malignant as shown in Table (1).

Table (1): No. Of cases tested and trained.

<table>
<thead>
<tr>
<th>Training</th>
<th>No. Of cases</th>
<th>Normal</th>
<th>Benign</th>
<th>Malignant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>100</td>
<td>40</td>
<td>30</td>
<td>30</td>
</tr>
</tbody>
</table>
Transformation
Mammogram images are difficult to explain, it is necessary to improve the quality of the image and make the feature extraction phase as an easier and more reliable. The main objective of this phase is to enhance and suppress the undesired distortion of mammogram images. This is done by using wavelet transform and contourlet transform. These transforms smooth the image by blocking detailed information. The calculation of all input mammogram images once using wavelet transform with Haar filter and again using contourlet transform with (LP and DFB) filter to extract vector of five features for each selected mammogram image. In conclusion, the enhanced image is obtained with clarity and free from noise.

Segmentation
Segmentation is the most important step in the conception of a proposed system since the efficiency of the organization depends on the accuracy of the partitioning. Segmentation of the foreground breast object from the background is a fundamental step in mammogram analysis. In MIAS data base, most of mammogram images consist of a black background with significant noise. The mammogram images are segmented using Otsu’s thresholding segmentation, which is characterized as easy to implement and widely practiced. Significant noise removed when applied Otsu’s thresholding segmentation.

Feature Extraction
Feature extraction is an essential step for classification. The segmentation results are used as a guide to extract features. For each mammogram image, Gray Level Co-occurrence Matrix (GLCM) is constructed. Various features are extracted by using GLCM. These features are Contrast, Dissimilarity, Entropy, Homogeneity, Variance, Mean and Standard deviation. These features are selected to represent the mammogram images and will be used later as an input to the classification algorithm.

Classification
In a proposed system, important steps before classification must be done. First, after getting the numerical values of five features it should be converted to categorical values by dividing the range of values of five features into k equal sized bins (equal width interval), where k is a parameter selected by a user based on length of data. Second, removing irrelevant and redundant features. Irrelevant features occur when there is more than one row in the database has the same features but it has more than class. Redundant features occur when there is more than one row in the database that has the same features and the same class. Removing irrelevant and redundant features by selecting the subset of features that can achieve the best performance in terms of accuracy and computation time.
Building classifier model is the most common task classification. This model used for prediction of the class of a mammogram image, the class can be seen as a mammogram image type.

4.5.1 Classifier model
The mammogram image data base (Training Set) consist of attribute value representation with five categorical attributes (contrast, homogeneity, entropy, energy and standard variation) and mammogram class attribute for a large number of patients. These attributes are the input of the classifier model for learning. Prediction of the new patient depending on classifier model.

The classifier model in this paper built by using decision tree based on training mammogram images. The classifier model used in prediction phase to test a new image which is not exist in the training phase. The classifier model given decision, which mean its classifies tested image as normal, benign or malignant as shown in Figure (14).

Figure (14): classifier model.

Decision Tree and Decision rules
The decision tree classification algorithm can be implemented by using Iterative Dichotomiser 3 (ID3) algorithm. ID3 algorithm has the ability to generate decision rules through a decision tree based on the attributes, which play a vital role in classification based on entropy and information gain.

As a result of computing ID3 algorithm on the training set, creation the decision rules by using if-then format which is described later. To simplify a decision tree converting it into decision rules, which are easier to understand and to implement in a computer. In the testing phase, the classifier model classifies the tested image as normal, benign or malignant based on decision rules, which are built by ID3
algorithm. All the steps in the testing phase is as in the training phase except the mammogram images tested depending on Decision rules.

**Performance Measures of the Proposed System**

The performance of the proposed system is estimated by using confusion matrix, running time and classification accuracy. The confusion matrix of system performance has been obtained from the testing part by using wavelet transforms and the contourlet transform can be shown in a table (2) and 3). The confusion matrices should be read as follows: rows indicate the object to recognize and columns indicate the label the classifiers associates at this object. The running time and classification accuracy of system performance can be shown in a table (5) and Figure (15). The running time in seconds, that is calculated by the difference between the time of the beginning of the implementation and the time of the end of the implementation of the proposed system. The running time is calculated five times to make a sure that results are clear because the CPU might be busy with another process.

**Table (2): Confusion matrix using wavelet transforms.**

<table>
<thead>
<tr>
<th>Actual Class</th>
<th>Predicate Class</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Normal</td>
</tr>
<tr>
<td>Normal</td>
<td>24</td>
</tr>
<tr>
<td>Benign</td>
<td>1</td>
</tr>
<tr>
<td>Malignant</td>
<td>2</td>
</tr>
</tbody>
</table>

**Table (3): Confusion matrix using contourlet transform**

<table>
<thead>
<tr>
<th>Actual Class</th>
<th>Predicate Class</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Normal</td>
</tr>
<tr>
<td>Normal</td>
<td>29</td>
</tr>
<tr>
<td>Benign</td>
<td>0</td>
</tr>
<tr>
<td>Malignant</td>
<td>0</td>
</tr>
</tbody>
</table>

**Table (4): Comparison between wavelet and contourlet transform**

<table>
<thead>
<tr>
<th>Class Type</th>
<th>Wavelet transforms</th>
<th>Contourlet transform</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Running time</td>
<td>Accuracy</td>
</tr>
<tr>
<td>Normal</td>
<td>90</td>
<td>80%</td>
</tr>
<tr>
<td>Benign</td>
<td>60</td>
<td>90%</td>
</tr>
<tr>
<td>Malignant</td>
<td>60</td>
<td>75%</td>
</tr>
<tr>
<td>Average</td>
<td>70</td>
<td>81%</td>
</tr>
</tbody>
</table>
The Comparison between wavelet and contourlet transform can be determined by PSNR, RMSE, value which is calculated for mammogram image clearly shows that contourlet transform when applied to mammogram image leads to best image quality can be shown in Table (5).

Table (5): Performance measures of the proposed system.

<table>
<thead>
<tr>
<th>Class Type</th>
<th>Wavelet transforms</th>
<th>Contourlet transform</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PSNR</td>
<td>RMSE</td>
</tr>
<tr>
<td>Normal</td>
<td>29.052</td>
<td>8.993</td>
</tr>
<tr>
<td>Benign</td>
<td>31.061</td>
<td>7.136</td>
</tr>
<tr>
<td>Malignant</td>
<td>30.304</td>
<td>7.785</td>
</tr>
</tbody>
</table>

Conclusions

Several conclusions are reached through the system work steps. The following items represent the important conclusions which can be drawn from the proposed system as follows:
1. The mammogram image can be used as a true indicator of breast cancer diseases and the ID3 classifier is shown as the diagnostic tool to aid the radiologists.
2. The experimental results indicate that the ID3 classifier achieves accuracy of 81% in the case of wavelet transform and 95% with contourlet transform for the same number of the test set.
3. The average of classification time (seconds) is estimated 70 s for wavelet transform 116 s for contourlet transform.
4. Automated segmentation using OTSU’s, which aims at automated extraction of object (breast), plays a central function in understanding image content for feature and classification in medical image.
5. According to the results obtained it is tough to suppose that any individual feature is superior to others. The five features used in this paper which can be extracted from GLCM matrix, and which contain information about image textural characteristics.
REFERENCES


