Improvement of Mass Detection In Breast X-Ray Images Using Texture Analysis Methods

Mohamed Abdel-Nasser†, ††, Domenec Puig† and Antonio Moreno††

†Intelligent Robotics and Computer Vision Group (IRCV)
††Intelligent Technologies for Advanced Knowledge Acquisition Group (ITAKA)
Departament d’Enginyeria Informàtica i Matemàtiques
Escola Tècnica Superior d’Enginyeria
Universitat Rovira i Virgili

Abstract. In this paper we analyse the performance of various texture analysis methods for the purpose of breast mass detection. We considered well-known methods such as local binary patterns, histogram of oriented gradients, co-occurrence matrix features and Gabor filters. Moreover, we propose the use of local directional number patterns as a new feature extraction method for breast mass detection. For each method, a Support Vector Machine is trained on the extracted features to predict the class (mass/normal) of unknown instances. In order to improve the mass detection capability of each individual method we used classifier majority voting and feature combination techniques. Some experiments were performed on the images obtained from a public breast cancer database, achieving promising levels of sensitivity and specificity.

Keywords. Texture analysis, feature combination, support vector machines.

1. Introduction

Breast cancer is one of the most dangerous diseases suffered by women around the world [1]. A breast mainly consists of lobules, ducts and stroma. Lobules are milk producing glands, ducts are small pipes that carry the milk from the lobules to the nipple, and stroma is fatty and connective tissue surrounding the ducts and lobules, blood vessels, and lymphatic vessels. Doctors usually order mammograms (X-ray images of the breast) in the diagnosis of breast cancer, especially to detect tumours that cannot be felt. A breast mammogram may show suspicious regions, but it can’t prove by itself that an abnormal area is cancer. If a mammogram presents a suspicion of cancer, a biopsy should be performed. As biopsies are expensive, it is important to reduce the number of false positives (mammograms interpreted as abnormal when they are actually normal). In the last ten years, various methods have been proposed for automatic breast mass detection using Computer Vision and Machine Learning techniques. These methods can be classified as supervised or unsupervised. In the
supervised methods, that are more accurate, feature vectors are extracted from the regions of interest (ROIs). The normal and mass regions represent the most interesting ROIs in breast mammograms. The extracted features are used to train a model, which is used to predict the class of unknown instances. Unfortunately, supervised methods usually present a high number of false positive detections.

Various Computer Vision methods have been proposed for breast mass detection. Oliver et al. [2] used Local Binary Patterns (LBP) in order to reduce the false positive rate in mass detection. They used Support Vector Machines (SVM) to predict the class of unknown instances. Charistoyianni et al. [3] proposed to use the gray level, the texture and features related to independent component analysis to train a neural network classifier. Yufeng [4] proposed to use Gabor filters to extract features for mass detection in digital mammograms. First, the Gabor filters were applied on the mammograms, and then the alarm segments (suspicious to be masses/calcifications) were extracted using a threshold that is adaptively decided upon the histogram analysis of the filtered mammogram. Martins et al. [5] used the K-means algorithm for image segmentation and co-occurrence matrix features to describe the texture of segmented structures. The classification of these structures was made by a SVM, achieving an accuracy of 85%. Oliver et al. [6] adopted an eigen faces approach for breast mass detection. Pomponiu et al. [7] used the Histogram of Oriented Gradients (HOG) for breast mass detection. HOG descriptors were used to filter the mass and normal tissue regions, and a SVM was applied to classify the identified masses. In turn, an unsupervised mass detection method, proposed by Chang et al. [8], compared the ROIs of a new image with all labelled (mass/normal) ROIs in a database. They used the gray levels and the shape of ROIs to compute a likelihood measure.

There are many factors that influence the percentage of false positive mass detection in Computer Vision systems, which should be carefully analyzed to achieve a low percentage of errors. In the previous works we identified the following factors:

- The ability of given feature extraction methods to distinguish between mass and normal ROIs.
- The robustness of the classifier used for training and prediction.
- The breast density: dense breast tissue usually produces light gray levels in mammograms, making them harder to interpret specially in younger women. In practice, it is difficult to detect masses in mammograms with high densities.
- Quality of the mammograms: both noise and physical artefacts may degrade the overall performance of the mass detection algorithms.

In this paper we analyse the performance of various texture analysis methods for breast mass detection. We utilized local binary patterns, histogram of oriented gradients, co-occurrence matrix features and Gabor filters. As a novelty on previous works, we propose Local Directional Number patterns (LDN) as a new texture feature extraction method for breast mass detection. Once the features are extracted by each method, a SVM-based binary classifier (mass/normal tissue) is trained. As commented above, these models have shown robust capabilities in classification problems and they have been widely used in medical applications. As a second contribution, we have analysed possible combinations of different texture analysis methods (both considering the majority output of the individual classifiers and building new SVMs on the concatenation of features provided by different methods).
The rest of this paper is organized as follows. Section 2 briefly presents the texture analysis methods. Section 3 summarizes SVM-based classifiers. Section 4 includes the experimental results and their discussion. Finally, conclusions and lines of further work are presented in Section 5.

2. Feature Extraction Methods

This section explains the texture analysis methods which have been usually applied for breast mass detection: Local Binary Patterns (LBP), Gray Level Co-Occurrence Matrix (GLCM), Histogram of Oriented Gradients (HOG) and Gabor filters. We also introduce Local Directional Number patterns (LDN) as a new analysis method in this field.

2.1. Local Binary Patterns

LBP is a gray scale invariant texture feature and it is considered as a good method for texture image analysis in many Computer Vision areas. The original LBP operator labels the pixels of an image by comparing the 3x3 window surrounding each pixel with the value of the central pixel. Pixels in this window with a value greater than the central pixel are labelled as 1 and the rest as 0; thus, each pixel is represented by 8 bits, as shown in the example in Fig. 1. The size of the window may vary on different applications (e.g. 5x5 or 7x7).

![Figure 1. Calculation of LBP for a given pixel in an abnormal ROI.](image)

Uniform LBP is an extension of the original LBP in which only patterns that contain at most two transitions from 0 to 1 (or vice versa) are considered. For example 00111100 is a uniform LBP whereas 00101010 is a non-uniform LBP. In uniform LBP mapping there is a separate output label for each uniform pattern and all the non-uniform patterns are assigned to a single label (the uniform mapping produces 59 output labels for neighbourhoods of 8 points). We computed the histogram with the frequency of uniform LBPs for each ROI. In order to reduce the false positive rate we utilized the histogram configuration proposed by Oliver et al. [2], in which each ROI is divided into four quadrants as shown in Fig. 2. The final feature vector $f$ is constructed by concatenating the LBP histograms ($H_1$, $H_2$, $H_3$ and $H_4$) of the four regions:

$$f = (H_1, H_2, H_3, H_4)$$  \hspace{1cm} (1)
This configuration improves the description of each ROI by adding spatial information to the original LBP method.

Figure 2. The construction of the final LBP histogram from each ROI.

2.2. Local Directional Number Patterns

Ramírez et al. [10] proposed LDN for face analysis. They showed that LDN is better than LBP because it can detect changes in regions producing different 6-bit codes, whereas LBP may produce the same pattern for pixels in different regions. Taking into account the advantages of LDN over LBP and the previous use of LBP in mass detection [2], we had intuition that LDN could be a good method for this problem. In LDN, the edge responses (M0-M7) are computed in eight different directions by convolving the Kirsch compass masks with the original image (see Fig. 3). The authors of LDN chose the location of the top positive and negative edge responses to generate a meaningful descriptor for each pixel (6-bit code). An example of LDN is presented in Fig. 3, in which the location codes that return the maximum and minimum responses are concatenated to form the LDN code of the marked pixel. This method encodes the features in a binary code shorter than the one of LBP, and its good performance has been shown in different fields, such as face expression analysis. As in the case of LBP, the feature vector of LDN is also calculated by dividing each ROI into four sub-regions and concatenating the four associated histograms.

2.3. Histogram of Oriented Gradients

Dalal and Triggs [11] proposed HOG for human detection. HOG has been considered as a robust feature extraction method because it produces distinctive features in the case of illumination change and cluttered background. In the HOG method, the occurrences of edge orientations in a local image window are counted. The image is divided into blocks (small groups of cells) and then a weighted histogram is computed for each of them. The frequencies in the histograms are normalized in the interval [0,1] to compensate changes in illumination. The combination of the histograms of all those blocks represents the HOG descriptors.
2.4. The Gray Level Co-Occurrence Matrix

In the gray level co-occurrence matrix the distribution of co-occurring gray level values at a given offset (direction and distance) is computed [12]. Different texture descriptors can be computed from this matrix. In order to increase the description given by GLCM spatial information should be added, so we divided each ROI into four sub-regions. For each sub-region we calculated 12 GLCMs by combining four orientations (0°, 45°, 90° and 135°) and three distances (3, 5 and 10 pixels). From each GLCM we calculated 22 texture features, and we concatenated all the features calculated for each GLCM for all the sub-regions into one feature vector. The extracted features were: uniformity, entropy, dissimilarity, contrast, inverse difference, correlation, homogeneity, autocorrelation, cluster shade, cluster prominence, maximum probability, sum of square, sum average, sum variance, sum entropy, difference variance, difference entropy, information measure of correlation1, information measure of correlation2, maximal correlation coefficient, inverse difference normalized and inverse difference moment normalized. Because of the limited space of this paper we do not include the mathematical expression of each feature, which can be found in [12-14]. The co-occurrence texture vector is normalized in the interval [0,1] to present attributes with higher numeric ranges from dominating those with lower numeric ranges.

2.5. Gabor Filters

Gabor filters have been widely used to extract texture features in many Computer Vision problems such as texture classification, texture segmentation, face recognition, and cancer detection [4]. A two dimensional Gabor filter \( g(x,y) \) can be viewed as a sinusoidal with a particular frequency and orientation, modulated by a Gaussian envelope

\[
g(x, y) = \exp \left( -\frac{1}{2} \left( \frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} \right) \right) \exp^{-j2\pi(\mu_0 x + v_0 y)}
\]

(2)
where \((u_0, v_0)\) is the centre of a sinusoidal function and \(\sigma_x, \sigma_y\) are the standard deviations along two orthogonal directions (which determine the width of the Gaussian envelope along the x- and y-axes in the spatial domain). Given an input image \(I(x, y)\), the filtered image \(f(x, y)\) is the result of convolving \(I(x, y)\) and \(g(x, y)\).

\[
f(x, y) = I(x, y) * g(x, y)
\]  

Eq. 2 defines a band-pass filter in the frequency domain, where the bandwidth and centre frequency of the filter are controlled by the standard deviation of the Gaussian function and the frequency of the complex sinusoid respectively. A Gabor filter bank has a number of band-pass filters (with varying centre frequencies, bandwidths and orientations) and it is controlled by the parameters of Gabor wavelets. Tuning Gabor filters to specific frequencies and directions can lead them to detect both local orientation and frequency information from an image. We can assume that local image regions are spatially homogeneous and use the mean and standard deviation of the magnitude of the filter responses to represent the region for classification. The feature vector \(\mathbf{f}\) is constructed using the mean \(\mu_{mn}\) and the standard deviation \(\sigma_{mn}\) of the filtered images.

\[
f = [\mu_1 \sigma_1 \mu_2 \sigma_2 \mu_3 \sigma_3 \ldots \mu_{mn} \sigma_{mn}]
\]  

In this expression \(m\) and \(n\) are the number of scales and orientations of the filter, respectively. In this work 6 orientations and 4 scales were considered, leading to a vector of 48 features.

### 3. Classification Stage

As commented in the introduction, Support Vector Machines have been widely used in breast cancer detection systems. Usually, the extracted features are divided into training and testing sets. The goal, as in all supervised approaches to Machine Learning, is to train a model using a training set and use it to predict the class of unknown instances [15]. Given a labelled training set of the form \((x_i, y_i), i = 1, 2, \ldots, k\), where \(x_i \in \mathbb{R}^n\), \(y_i \in \{1, -1\}\) and \(k\) is the number of samples, the SVM solves the following optimization problem:

\[
\min_{w, \xi} ||w||^2 + C \sum_{i=1}^{k} \xi_i
\]

Subject to \(y_i (w^T \phi(x_i) + b) \geq 1 - \xi_i, \quad \xi_i \geq 0.\)  

During the optimization process, the training data \(x_i\) is mapped to a higher dimension space using the kernel function \(k(x_i, x_j) = \phi(x_i)^T \phi(x_j)\). The SVM uses the kernel trick, by which the data become linearly separable in the new space. The SVM finds the hyperplane with maximum margin separation between classes in the new higher dimensional space. The regularization parameter \(C\) is used during the optimization process to control the error. The weight vector \(w\) is normal to the separating
The parameter $\xi$ is used to give a degree of flexibility for the algorithm when fitting the data and $b$ represents the bias.

In this work we used the MATLAB library libSVM [16] with the radial basis function (RBF) as a mapping kernel. This kernel is better than the linear and polynomial ones, especially with a highly dimensional feature space. The RBF kernel can be defined as:

$$k(x_i, x_j) = e^{-\gamma\|x_i - x_j\|^2}, \quad \gamma > 0$$

where $\gamma = \frac{1}{2\sigma^2}$. A grid search algorithm is required to find the optimum parameters of the RBF kernel (i.e., $\gamma$ and $C$). In this work, we preset the ranges of the grid search algorithm with steps 0.5 for the exponent. It searches for $\gamma$ in the range $[2^{-5} < \gamma < 2^3]$ and $C$ is allowed to vary in the range $[2^{-5} < C < 2^{10}]$.

4. Experiments

The Digital Database for Screening Mammography$^2$ (DDSM) is considered as one of the most comprehensive public breast cancer databases. It contains about 10,000 mammograms. Unfortunately, many of them contain artefacts and noise, hampering their use and interpretation. In our study only clear mammograms were selected. We extracted 2700 ROIs from the DDSM database (690 of them are mass ROIs and 2010 are normal ROIs). The size of each ROI is $101\times101$ pixels. The ROIs have been extracted from 205 mammograms including both Cranio-Caudal (CC) and Mediolateral-oblique (MLO) views for 80 women affected by cancer and 45 healthy women. Fig. 4 shows examples for mass and normal ROIs [9].

![Figure 4. Examples for the extracted ROIs, (left) mass and (right) normal.](http://marathon.csee.usf.edu/Mammography/Database.html)

We used the $k$-fold technique to generate the training and testing datasets. In this $k$-fold scenario, the data are randomly sorted and divided into $k$ folds (a common value of $k$ is 10). One of the folds is used for testing and the remaining folds for training a SVM classifier, and this scenario is repeated $k$ times. In order to evaluate the performance of each texture extraction method, we calculated the mean sensitivity and mean specificity of the $k$-fold. These measures are defined as follows:

$$\text{sensitivity} = \frac{TP}{(TP + FN)}$$

$$\text{specificity} = \frac{TN}{(TN + FP)}$$
TP, TN, FP and FN are true positive, true negative, false positive and false negatives of the classification for a given set of test ROIs, respectively. High values of sensitivity and specificity are required in order to obtain acceptable mass detection results. In order to perform a more robust evaluation of the feature extraction methods we used two ROI ratios (1/1 and 1/3). ROI ratio is the ratio between the numbers of mass ROIs and normal ROIs.

4.1. Parameter selection

Parameter selection plays an important role in the success of any feature extraction method. During the training stage we optimized the values of each parameter. In this section we summarize the final values of the parameters of each method and the size of the feature vectors.

- **LBP**: we used 3x3 local windows in order to generate the uniform LBP for each ROI. Each ROI was divided into four regions, and the final LBP histogram was created by concatenating the histograms of the four regions. The final dimension of the LBP feature vector was 236 (4x59).

- **LDN**: as in LBP, each ROI was divided into four sub-regions, and the final LDN feature vector (4x64=256 dimensions) was formed by concatenating the histograms of the four sub regions.

- **HOG**: we used a 3x3 cell size, 8x8 cells for the block size, and a 9-bit histogram. The dimension of the HOG feature vector was 576 (9x8x8).

- **Co-occurrence matrix**: each ROI was divided into four sub-regions, and we used four orientations (0°, 45°, 90° and 135°) and three distances (3, 5 and 10 pixels) to calculate the gray level co-occurrence matrices (GLCM) for each sub-region. Thus, there were 48 GLCMs (4 regions x 4 orientations x 3 distances), and for each GLCM we computed 22 features. We concatenated all the features calculated for each GLCM for all the sub-regions into one single feature vector of 1056 dimensions (4x4x3x22).

- **Gabor filter**: we used 4 scales and 6 orientations to calculate the responses of Gabor filters (4x6=24 responses). We calculated the mean and the variance for each Gabor filter. The feature vector is constructed by concatenating all the resulting means and variances (4x6x2=48 dimensions).

4.2. Comparative Results and Discussion

Our analysis is twofold. First we analyse each texture method individually, considering two ROI ratios. After that, we present the result of combining those methods. We utilized two combination techniques: in the first one we concatenated the features obtained by different methods, and trained a new classifier on that extended set of features, whereas in the second one we used the individual classifiers obtained by three methods and considered their majority opinion.

Tables 1 and 2 show the mean sensitivity and mean specificity of each texture extraction method over all the iterations of the 10-fold procedure. Table 1 presents the results when the ROI ratio is 1/1 (512 normal ROI and 512 mass ROI). LDN achieves the highest specificity (99.35%) with a sensitivity of 94.73%, almost equal to the maximum one (94.74%) achieved by HOG. The LBP method, applied for the first time
to breast cancer detection, improves the results of Gabor filters, which have been widely used in the past in this problem. GLCM presents results that are much worse than those of the other methods. Table 2 presents the results when there is a 1/3 ROI ratio. LDN achieves again very good results, with a specificity of 97.07% and the maximum sensitivity of 95.31%. The best specificity (98.57) is achieved by HOG, although the sensitivity drops to 89.66%. LBP improves again the results of Gabor filters. GLCM is again the worst method, presenting a very poor sensitivity, below 20%.

Table 1. Mean sensitivity and mean specificity of each feature extraction method (ROI ratio 1/1).

<table>
<thead>
<tr>
<th>Method</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDN</td>
<td>94.73</td>
<td>99.35</td>
</tr>
<tr>
<td>LBP</td>
<td>89.27</td>
<td>83.23</td>
</tr>
<tr>
<td>HOG</td>
<td>94.74</td>
<td>95.31</td>
</tr>
<tr>
<td>GLCM</td>
<td>69.80</td>
<td>33.14</td>
</tr>
<tr>
<td>Gabor</td>
<td>83.02</td>
<td>82.81</td>
</tr>
</tbody>
</table>

Table 2. Mean sensitivity and mean specificity of each feature extraction method (ROI ratio 1/3).

<table>
<thead>
<tr>
<th>Method</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDN</td>
<td>95.31</td>
<td>97.07</td>
</tr>
<tr>
<td>LBP</td>
<td>66.82</td>
<td>94.86</td>
</tr>
<tr>
<td>HOG</td>
<td>89.66</td>
<td>98.57</td>
</tr>
<tr>
<td>GLCM</td>
<td>19.42</td>
<td>84.36</td>
</tr>
<tr>
<td>Gabor</td>
<td>53.47</td>
<td>93.35</td>
</tr>
</tbody>
</table>

Based on the analysis of each texture analysis method we selected the ones with best specificity and sensitivity and tried to combine them to improve their individual accuracy. On the one hand, we concatenated the feature vectors of LDN+LBP+HOG, LDN+LBP, LDN+HOG and LBP+HOG, and trained a new SVM in each case. We used Principle Component Analysis (PCA) [17] to reduce the dimensionality of the concatenated feature vectors. On the other hand, we used the classifier majority voting technique to combine the outputs of LDN, LBP and HOG. Table 3 presents the result of the combination schemes. The best combination is LDN+LBP, with a 97.66% sensitivity and a 99.42% specificity (that improve their individual results, despite the PCA dimensionality reduction). The classifier voting technique obtains the highest sensitivity (98.05%), maintaining specificity close to the one of LDN+LBP (98.04%). LDN+LBP yields the best results because it integrates the LDN descriptor (which depends on the minimum and maximum edge responses in the neighbourhood of each pixel) with the LBP descriptor which provides a description of all the elements in the neighbourhood of each pixel. We do not discuss the effect of the rotation on the accuracy of the proposed methods because the mammograms are acquired from fixed positions (i.e., CC and MLO mammographic views).

Table 3. Mean sensitivity and mean specificity of each feature combination strategy (1/1 ROI ratio)

<table>
<thead>
<tr>
<th>Method</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LBP+HOG</td>
<td>94.92</td>
<td>94.91</td>
</tr>
<tr>
<td>LDN+HOG</td>
<td>95.50</td>
<td>94.54</td>
</tr>
<tr>
<td>LDN+LBP</td>
<td>97.66</td>
<td>99.42</td>
</tr>
<tr>
<td>LDN+LBP+HOG</td>
<td>94.53</td>
<td>95.12</td>
</tr>
<tr>
<td>Classifier voting</td>
<td>98.05</td>
<td>98.04</td>
</tr>
</tbody>
</table>
5. Conclusion and Future Work

In this paper we compared various texture analysis methods for breast mass detection, using images from a public breast cancer database. In particular, we proposed LDN as a new feature extraction method in this field. LDN improved the results of well-known texture analysis methods like Gabor filters or GLCM. In order to improve mass detection rates, we proposed to use two combination schemes. Firstly, we concatenated the features of the best texture analysis methods. Among all the concatenations, LDN+LBP obtained the best overall results. Secondly, we used the classifier voting technique in order to combine the predictions given by LDN, LBP and HOG, achieving excellent results. The future work will focus on using the proposed combination schemes to improve the performance of a computer-aided diagnostic system based on multiple mammographic views. In addition, a boosting method will be used to improve the classification results.

References