

Background Texture Extraction for the Classification of Mammographic Parenchymal Patterns

Lilian Blot*, Reyer Zwiggelaar†

Division of Computer Science, University of Portsmouth

Abstract. We have developed an approach to the separation of background texture and structures in images. The developed approach is based on the statistical difference between local and median co-occurrence matrices. It is our assertion that the classification of mammographic parenchymal patterns can be improved if anatomical structures can be removed from the image and the classification is based only on the background texture information. We compare the results of the classification between original images and images composed of their reconstructed background texture. 265 mammograms from the MIAS database [1] have been used for our experiment and the classification of the parenchymal patterns is based on Karssemeijer's model [2].

1 Introduction

Wolfe [3] has shown a relation between mammographic parenchymal patterns and the risk of developing cancer. Since the discovery of this relationship automated parenchymal pattern classification has been investigated. Karssemeijer [2] proposed a method based on grey-level histograms from selected regions and Byng *et al* [4] used fractal measures.

It is our hypothesis that classification can be improved if it is based on the background texture information only and not on a combination of the background texture and other anatomical structures which might be present. We propose a method to extract the background texture of the mammographic images (see section 2) and apply Karssemeijer's model (see section 3) to classify the image. In section 4 these results are compared with the results based on the original mammographic images.

2 Background Texture Extraction

We assume that there is a statistical difference between the Grey-Level Co-occurrence Matrices (*GLCMs*) for image regions that include image structures and regions that only contain background texture (more detail can be found in [5]).

2.1 Pre-processing

To approximate the textural properties of the texture for the *RoI* we need to compare the *GLCM* of the *RoI* with the *GLCM* of the local neighbourhood. An overlapping square regions kernel is defined by nine regions with the same area, eight of them describe the local neighbourhood and are centered around the *RoI* [5].

For each pixel in the image the following steps are taken. A kernel is centered on the pixel and for each region of the kernel a *GLCM* is computed. The *GLCM* from the neighbourhood are combined (by using a simple average) to generate the approximation of the textural property of the background (*GLCM_{Mean}*). This *GLCM_{Mean}* is compared to the *GLCM* for the *RoI* (*GLCM_{RoI}*) and for each entry of the matrices we compute the difference. The resulting *Difference Co-occurrence Matrix* (*DCM*) is given by

$$DCM(i, j) = \max\left(0, \frac{GLCM_{RoI}(i, j) - GLCM_{Mean}(i, j)}{GLCM_{RoI}(i, j) + GLCM_{Mean}(i, j)}\right) \quad (1)$$

where i and j are a particular combination of grey-levels.

If there are no structures in the *RoI* the *DCM* given by Eq. 1 is going to be sparsely populated indicating a low level of perturbation in the *RoI*. As we have normalised the *DCM* these values can be used as probability values. Lets consider the position (x, y) of the current pixel by a translation $\mathbf{t} = (a, b)$ (translation used for the *GLCMs*)

* email: lilian.blot@port.ac.uk

† email: reyer.zwiggelaar@port.ac.uk

and the corresponding grey-level value at that position is m . The resulting perturbation (or probability) for the current pixel is given by $DCM(k, m)$ where k is the intensity of the current pixel. The above approach can be used to produce a structure probability at a pixel level.

Applying the algorithm to the entire image result in a probability image. To obtain robust results a number of translations t are considered and the resulting probability images are combined (in our experiments, the mean value for each pixel) to provide an overall classification.

2.2 Background Texture Reconstruction

For the reconstruction of background texture the objective is to reduce the difference between $GLCM_{RoI}$ and $GLCM_{Mean}$. A similar technique is described in [6], here $GLCM_{Mean}$ is the textural feature to be obtained and $GLCM_{RoI}$ is the approximation to be improved. For the current pixel we have to decide if its intensity has to be changed or not (by applying a single threshold to the probability image), in the positive case the new intensity k is determined by minimising

$$\sum_{t \in T} \|GLCM_{Mean}(t) - GLCM_{RoI}^k(t)\|^2 \quad (2)$$

where $GLCM_{Mean}(t)$ is the $GLCM$ of the local neighbourhood using translation t , $GLCM_{RoI}^k(t)$ is the $GLCM$ of the region of interest using translation t and where the intensity of the current pixel has been replaced by k (the value of k is taken from all possible grey-levels) and T is the set of translations.

As structures in mammographic images tend to be of higher than average intensity, only pixels with grey-level values higher than the mean grey-level value of the RoI are considered by the model. This is the only prior knowledge incorporated in the developed approach. A more detailed discussion of the method can be found in [7], and two examples are shown in Fig. 1.

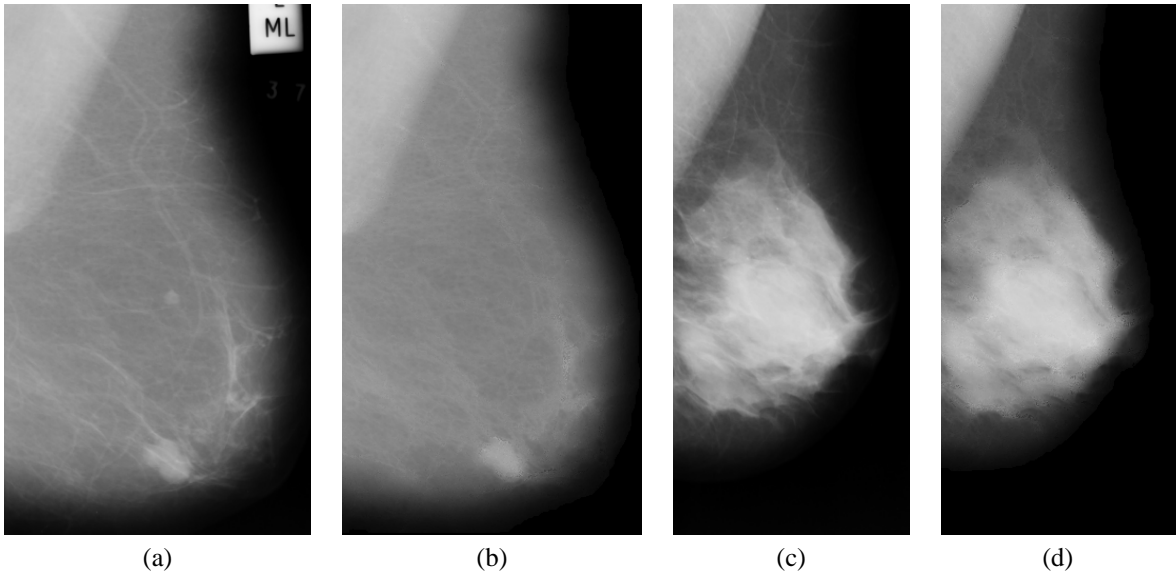


Figure 1. Original mammograms and their background texture. (a) Fatty and (c) fatty-glandular mammograms and their corresponding extracted background texture, respectively (b) and (d).

3 Classification of Parenchymal Patterns

This section presents a short introduction of Karssemeijer's model [2]. The method is decomposed in three stages, firstly the segmentation of the mammogram into three regions (background, pectoral muscle and finally breast tissue). Secondly a set of eight features is extracted from the image and thirdly the parenchymal pattern is classified using a k -nearest-neighbour approach.

3.1 Image Segmentation

To extract features related to parenchymal patterns, breast tissue must be identified. The first step of the segmentation is to remove the background. To do so a global thresholding is applied to the image, where the threshold value is determined automatically from the grey-level histogram by applying a peak detection method.

The second step is to separate the pectoral muscle from the breast tissue. A 3×3 Sobel operator is applied to the image and gradient magnitude and direction for each pixel are extracted. To detect the pectoral boundary a Hough transform is applied to the gradient image [8]. All the pixels in the tissue area and above the pectoral boundary are marked as pectoral. An example of the segmented breast tissue is shown in Fig. 2b.

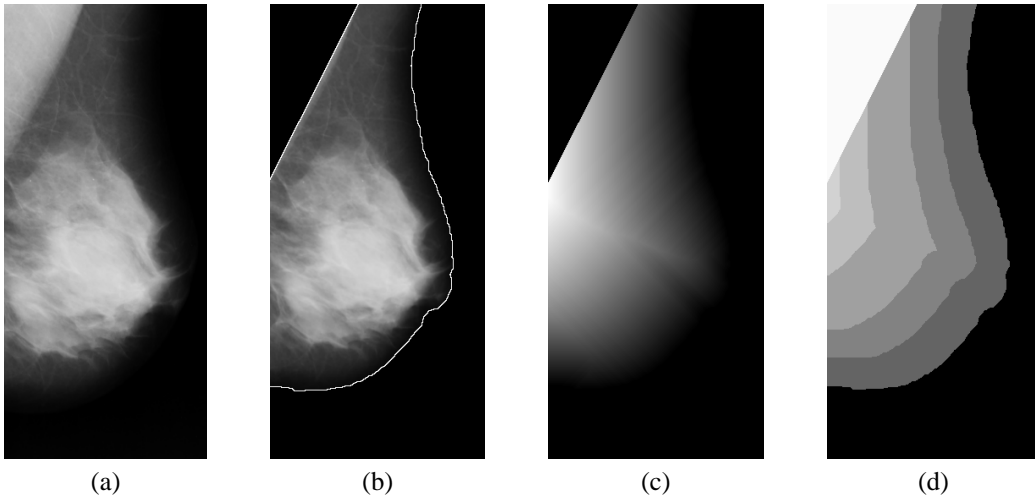


Figure 2. (a) Original mammogram, (b) the segmented breast using an automatic thresholding and the Hough transform, (c) image representing the distance to the skin and (d) the division of the breast in seven regions. The black region is the background, the white region is the pectoral muscle and the grey regions are the breast. Only the three central regions of the breast are taking into account for the feature extraction.

3.2 Feature Extraction and Classification

Features are based on the grey-level histograms from selected regions of the breast. Regions are constructed using the distance to the skin normalised from 0 to 100 (providing invariance to the size of the breast). From this *distance image* three regions are defined by:

$$s_{d_n} = \{(i, j) | d_n < d(i, j) \leq d_{n+1}\} \quad (3)$$

where $d_1 = 20$, $d_2 = 40$, $d_3 = 70$, $d_4 = 90$ and $d(i, j)$ is the distance to the skin for the pixel at position (i, j) . These regions are shown in Fig. 2d (the three central regions of the breast). Grey-level histograms are computed for $n = 1..3$ and from each of these histograms the standard deviation and the skewness are calculated. Two other features are used to compare the density of the breast tissue and the pectoral muscle (all features are normalised to unit variance and zero mean).

A k -nearest-neighbour classifier [9] with $k = 32$ is used for parenchymal patterns classification based on the eight features described previously. A leave-one-out approach is used for all classification results. This means that each sample is classified by a classifier trained on all the other samples.

4 Results

The dataset used for our experiment is composed of 265 mammograms from the MIAS database [1], which includes 75 fatty, 95 fatty-glandular and 95 dense-glandular mammograms. It must be mentioned that this classification approach differs from that of Karssemeijer [2] and Wolfe [3] who both used four classes.

Karssemeijer's model has been applied to two sets of images. The first set is composed of original mammograms and the second of their reconstructed background texture. Tab. 1 shows the performance of the k -nearest-neighbours classifiers on the two sets in the form of confusion matrices.

The performance is similar on both sets of data (the maximum difference is 2%), no degradation of information has occurred during the reconstruction step of our algorithm, neither improvement. These results are expected as Karssemeijer's model does not rely on local spatial patterns but on local grey-level histograms. Our algorithm preserves the global grey-level information stored in the histograms and changes the grey-level distribution of the image.

Most of the misclassifications are due to dense-glandular and fatty-glandular mammograms. In the future the developed approach will be tested on a large mammographic database with a classification based on 5 classes (Wolfe grades and one extra class to separate fatty mammograms).

Classified →	original mammograms			Reconstructed texture		
	Fatty	Fatty-Glandular	Dense-Glandular	Fatty	Fatty-Glandular	Dense-Glandular
True ↓						
Fatty	24%	4%	0%	23%	6%	0%
Fatty-glandular	9%	21%	6%	9%	20%	6%
Dense-glandular	3%	13%	20%	2%	14%	20%

Table 1. Classification results obtained using originals mammograms and reconstructed textures. A success rate of 65% has been achieved for the original data against a rate of 63% for the reconstructed texture.

5 Conclusions

We have developed a method, based on the full grey-level co-occurrence matrix information, to reconstruct background texture in mammograms. We have evaluated our approach using Karssemeijer's model for parenchymal pattern classification. As expected, results have shown that no loss of information has occurred during the reconstruction of the texture. However the method does not improve results based on Karssemeijer's technique. Future work will include evaluation based on fractal modelling and a mammographic database described by five parenchymal pattern classes.

References

1. J. Suckling, J. Parker, D. Dance et al. "The mammographic images analysis society digital mammogram database." In D. Gale, Astley & Cairns (editors), *Digital Mammography*, pp. 375–378. Elsevier, 1994.
2. N. Karssemeijer. "Automated classification of parenchymal patterns in mammograms." *Phys. Med. Biol.* **43**, pp. 365–378, 1998.
3. J. Wolfe. "Risk for breast cancer development determined by mammographic parenchymal pattern." *Cancer* **37**, pp. 2486–92, 1976.
4. J. Byng, N. Boyd, E. Fishell et al. "The quantitative analysis of mammographic densities." *Phys. Med. Biol.* **37**, pp. 1629–38, 1994.
5. L. Blot & R. Zwigelaar. "Enhancement of abnormal structures in mammographic images." *4th Conference on Medical Image Understanding and Analysis* **London, UK**, pp. 125–128, 2000.
6. G. Lohmann. "Analysis and synthesis of textures: a co-occurrence-based approach." *Computer & Graphics* **19(1)**, pp. 29–36, 1995.
7. L. Blot & R. Zwigelaar. "Extracting background texture in mammographic images: a co-occurrence matrices based approach." *5th International Workshop on Digital Mammography* **Toronto, Canada**, 2000.
8. D. Ballard & C. Brown. *Computer Vision*. Englewood Cliffs, NJ: Prentice-Hall, 1982.
9. J. McLachan. *Discriminant analysis and statistical pattern recognition*. New York: Wiley, 1992.