Classification of Linear Structures in Mammographic Images

Reyer Zwiggelaar[a*] and Caroline R.M. Boggis[b]

[a]Division of Computer Science, University of Portsmouth
[b]Greater Manchester Breast Screening Service, Withington Hospital, Manchester

Abstract. We have developed a novel technique for both the detection and anatomical classification of linear structures in mammographic images. Both stages are based on the statistical modelling of scale-orientation signatures. All experiments are based on a leave-one-out approach. The detection stage has been presented before [1]. For the classification stage real mammographic data was used. As the signatures incorporate scale aspects it is no longer essential to extract scale information which was previously shown to be a bottleneck in obtaining acceptable classification [2]. Spicule classification results show major improvements at low sensitivity and only moderate degradation at high sensitivity. We present spicule images based on both the detection and classification stages.

1 Introduction

Scale-orientation signatures have been used before for the detection of blob-like and linear structures in images [1, 3]. The same principles are applied here, including data selection, statistical modelling and classification. However, the main trust of this presented work is the simultaneous detection and classification of the linear structures using a single descriptor. Scale-orientation signatures provide a rich local grey-level description at a pixel level which is robust and locally stationary. Given this description, standard statistical classification methods can be used. For an overview of scale-orientation signatures we would like to refer to previous work [1, 3].

The classification of linear structures has been explored before [2]. But it has been shown that the extracted scale information, which is an indication for the width of the linear structure, is not robust enough and shows a large deviation from human observers which resulted in a degradation of the classification results [2]. The use of scale-orientation signatures for the classification of linear structures could in principle provide an improvement as the scale is incorporated in the signatures and hence is incorporated directly in the statistical modelling.

The UK Breast Screening Programme alone generates 1.5 million mammograms per annum. Normal mammograms contain a variety of linear structures: vessels, ducts, fibrous tissue, skinfolds, edges and others. In abnormal mammograms linear structures called 'spicules' may also be present. Abnormalities are non-accidentally associated with these linear structures. For example, microcalcifications are more likely to imply malignancy if they are located in ducts and spicules are always associated with lesions (called spiculated lesions).

2 Methods

A Directional Recursive Median Filter performs a smoothing operation that removes (sieves) image peaks or troughs of less than a chosen size [3]. By applying sieves of increasing size to an image and taking the difference between the output image from adjacent size sieves, it is possible to isolate image features of a specific size. Signatures at different positions on the same structure are similar (local stationarity) and the interaction between adjacent structures is minimised. The signature is a 2-D array in which the columns represent measurements for the same orientation and the rows represent measurements for the same scale.

Principal component analysis (PCA) can be used to obtain data generalisation and efficiency for classification purposes, by reducing the dimensionality of the data, instead of using the full signature information [4]. We intensity normalise signatures (each column is normalised independently) since there is no reason to believe that high-contrast features are more important than those of low contrast. Indeed, it is particularly important to detect low contrast structures of characteristic appearance. With the above approaches a signature is entered once whilst building the PCA model and the classifier. However, to improve the robustness with respect to the orientation of the detected structures the signatures can be entered several times. This is done once for each orientation contained in the signatures by applying a simple horizontal shift, which means that successive orientations are entered first (this approach will be referred to as orientation generalisation). This is based on the assumption that the appearance of the structures is independent of orientation. This might not be entirely correct as in general linear structures are expected to be orientated towards the nipple, but this is not necessarily the case for the spicules.

* email: reyer.zwiggelaar@port.ac.uk

3 Results

The developed approach was applied to mammographic data. A non-linear line operator [5] was applied to 29 mammograms from the MIAS database [6]. Once the linear structures were detected, a selection was labelled by an expert radiologist into anatomically distinct classes: *ducts*, *edges*, *fibrous tissue*, *skin folds*, *spicules*, *vessels* and *others*. Linear structures were randomly selected, taking the detected line strength into account, so that for each class the total number of samples obtained approximately represented their prevalence in mammograms. In total the data-set comprised the position of almost 50.000 pixels from linear structures. An example of a section from a mammographic images containing a spiculated lesion is shown in Fig. 1a. For the training data the positions of the annotated linear structures were used to extract signatures from the mammographic images. The signatures were based on twelve orientations and covering the 1 to 128 pixels scale. Experimental results are based on two sets of data: the original signatures and the normalised signatures. For the detection stage an equal number of signatures were randomly selected from the background. The classification stage was based on signatures from the annotated linear structures.

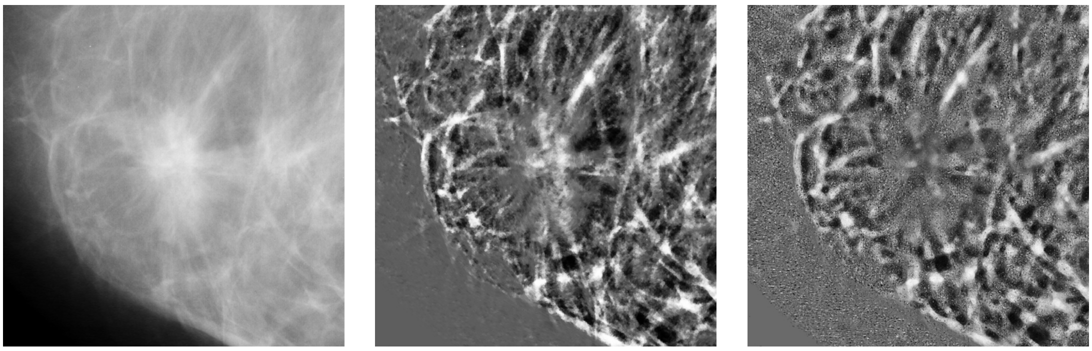
3.1 Statistical Models

For the original signatures the principal component model indicated that the mean exhibited a band across the signature. The first principal component indicated an intensity aspect, whilst the second and fifth principal component showed scale aspects. Only with the third and fourth principal component directional (as could be expected for linear structures) aspects became clear. The apparent blob-like attributes of the mean and the first two principal components are explained by the use of orientation generalisation. The resulting principal component model based on the normalised signatures showed similar, but slightly more pronounced, aspects as that for the original signatures.

To reduce noise a limited number of principal components can be used for reconstruction. Results indicated that for the original signatures based model noise is removed and structure information is preserved. However, reconstruction based on the normalised signature model produces different results, especially for the reconstruction of the background signatures, where a low number of principal components results in a signature that seems to contain a distribution not dissimilar to that expected for a (low-contrast) linear structure signature.

3.2 Detection Probability Images

For the detection stage it is our aim to distinguish between linear structures and the background. Based on the described approach signatures can be extracted at a pixel level. Each signature can be given a linear structure probability and results based on the image shown in Fig. 1a of such an approach can be found in Fig. 1b,c. This shows a clear enhancement of the linear structures, but also that the original signatures provide clearer enhancement of the linear structures than detection based on the intensity normalised signatures. No attempt at classification is made and the enhanced linear structures are likely to be ducts, vessels and spicules in this case.



(a) (b) (c)

Figure 1. (a) Original mammographic image. (b) Line strength image based on the original scale-orientation signatures. (c) Line strength image based on the normalised scale-orientation signatures.

3.3 Classification

It was shown [2] that the width is an important factor for the classification of linear structures and that the width as extracted by the non-linear line operator did not correlate with the width as indicated by the radiologist. The proposed use of the signatures has the potential of circumventing this as the scale of image structures (and hence the width of linear structures) is incorporated in the signatures and there should be no need to explicitly extract it.

In Fig. 2 ROC classification results based on the signatures from the annotated linear structures are shown. The anatomical probabilities were thresholded and correct/in-correct classification levels were determined. This are classification results for the detection of spicules (i.e. the linear structures associated with spiculated lesions) using a two-class classifier. For comparison classification results based on the cross-sectional profiles of the linear structures have been included. This indicates that at a high sensitivity the signatures based methods have a slightly inferior performance, but the opposite is true at lower sensitivity where the presented approach shows a significant improvement over previous results. The results indicate an improved classification when the scale-orientation signatures are intensity normalised. Additional results (graphs not included) indicated that the use of principal component analysis for noise/dimensionality reduction did not show any significant improvement in the classification results.

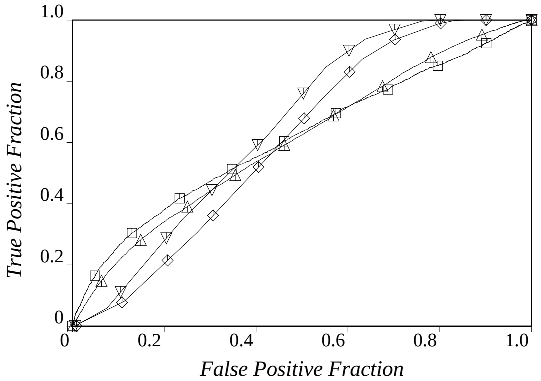
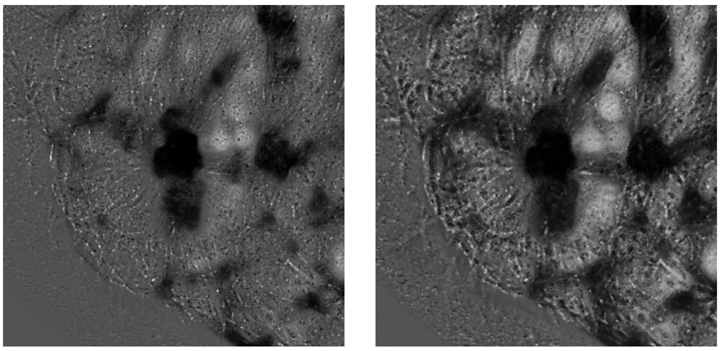


Figure 2. ROC classification curves, where the results are based on: the original signatures (△), the normalised signatures (□), median width cross-sectional profile (◇) and annotated width cross-sectional profile (▽).

3.4 Classification Probability Images

Here we are interested in determining the spicule probability at a pixel level. Using the same approach as described in the previous section a set of resulting probability images can be found in Fig. 3. There are two aspects which should be noted. The first aspect is the enhancement of particular line-like structures and the second aspect is the clear segmentation of the image in areas with a low (close to zero) probability, where clearly no spicules are expected. When comparing this with the original image (see Fig. 1a) these low probability areas constitute the centres of blob-like structures.

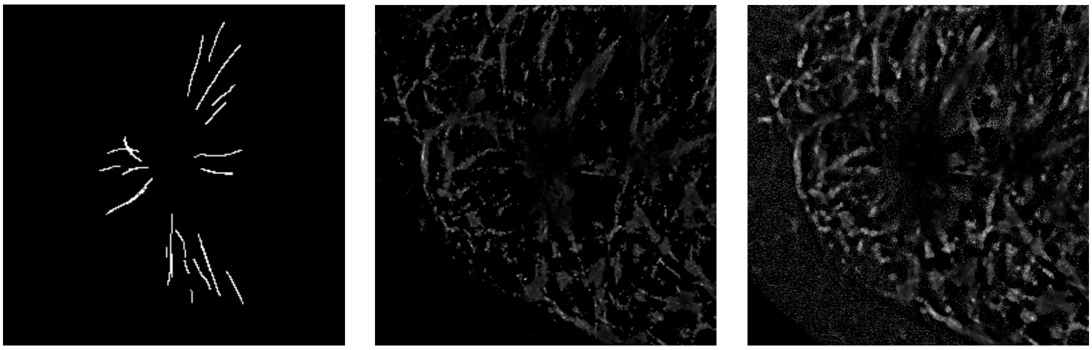


(a) (b)

Figure 3. (a) Spicule strength image based on the original scale-orientation signatures. (b) Spicule strength image based on the normalised scale-orientation signatures.

3.5 Spicule Images

It is possible to combine the information (the probability images) from the detection and classification stages. The results of such an approach (here we have used multiplication of the probabilities with an individual threshold equal to 0.5) can be found in Fig. 4. For comparison we have included an image indicating the spicules as annotated by the radiologist (see Fig. 4a). It should be mentioned that not necessarily all the spicules have been annotated. There seems to be a correlation between the annotated linear structures and some of the structures in the spicule images. However, other linear structures also have a probability of being a spicule and some of the spicules have been broken up and do not form clearly defined linear structures any more.



(a) (b) (c)

Figure 4. (a) Spicule annotation image. (b) Spicule image based on the original scale-orientation signatures. (c) Spicule image based on the normalised scale-orientation signatures.

4 Conclusions

It is expected that both the detection and classification of the linear structures can be improved by the introduction of non-linear classification methods (such as neural networks). In addition the described approach might benefit from additional pre-processing of the scale-orientation signatures which could be based on transportation aspects [7].

For mammographic data there is currently only visual assessment of the detection of linear structures and for the classification of spicules improvements only appeared at selective sensitivity levels. However, the resulting spicule images (based on both the detection and classification stages) look promising, although some additional processing is needed.

We have presented a novel approach to the detection of linear structures in images. The described approach uses statistical modelling and linear classification based on scale-orientation signatures to obtain a line-probability at a pixel level. The classification of mammographic linear structures based on the same scale-orientation signatures did not show an overall significant improvement. However, the presented initial results look promising and at selected sensitivity levels perform better than previous approaches. In addition the resulting spicule images show correlation with the linear structures annotated as spicules by an expert radiologist.

References

1. R. Zwiggelaar, R. Marti & C. Boggis. "Detection of linear structures in mammographic images." 4^{th} *Conference on Medical Image Understanding and Analysis* **London, UK**, pp. 97–100, 2000.
2. R. Zwiggelaar & C. Boggis. "Automatic classification of linear structures in mammographic images - is it possible?" 3^{rd} *Conference on Medical Image Understanding and Analysis* **Oxford, UK**, pp. 185–188, 1999.
3. R. Zwiggelaar & C. Taylor. "The detection of abnormal masses in mammography: Part ii." 2^{nd} *Conference on Medical Image Understanding and Analysis* **Leeds, UK**, pp. 109–112, 1998.
4. I. Jolliffe. *Principal Component Analysis*. Springer Verlag, 1986.
5. R. Zwiggelaar, T. Parr & C. Taylor. "Finding orientated line patterns in digital mammographic images." In *Proceedings of the* 7^{th} *British Machine Vision Conference*, pp. 715–724. Edinburgh, UK, 1996.
6. 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.
7. A. Holmes & C. Taylor. "Developing a measure of similarity between pixel signatures." 10^{th} *British Machine Vison Conference* **Nottingham, England**, pp. 614–622, 1999.