Improving te Brake's mammographic mass-detection algorithm using phase congruency

Keith Yates, Carolyn Evans* and Michael Brady
Medical Vision Laboratory
Engineering Department
University of Oxford
Oxford United Kingdom
OX1 3PJ

Abstract

We propose a method for improving te Brake's mass-detection algorithm [4, 5]. The method involves a pre-processing step. In this step the locally linear fine detail structure is removed from the image, whilst retaining the larger underlying mass structure. The removal technique for the fine structure is based on wavelets and a feature detection measure known as phase congruency [3]. The resulting ROC curve shows the pre-processing step improves the mass detection rate.

Introduction and Motivation

There are numerous ways to motivate our work. From a medical perspective, breast cancer is one of the largest killers of women in the Western world, approximately 1 in 10 will develop breast cancer at some time in their life. From a technological viewpoint, mammography is moving from being film-based to digital, with charge coupled devices, CCD's, replacing the traditional film cassette and allowing the direct creation of digital images. Hence there is considerable interest in computer-aided detection and the development of algorithms which will aid radiologists in their search for mass structures.

We explain more fully what we mean by mass. A mass is defined to be a region of non-normal, not necessarily malignant, breast tissue. Mass regions have an X-ray attenuation factor marginally higher than normal tissue, consequently they appear slightly brighter than normal tissue on a mammogram. Masses have a range of sizes and shapes. Further masses are often associated with spicules : long thin strands of tissue acting as anchors holding the mass in position.

The difficult task we face is the detection of a structure of variable shape and size, with an image intensity only marginally brighter than the surrounding tissue.

*Corresponding author - cje@robots.ox.ac.uk

Method

Our method combines two ideas, te Brake's gradient orientated mass detection algorithm, [4, 5] , and Kovesi's feature detection method,[3], which is developed around the concept of phase congruency. We summarise both.

Mass detection by Gradient Orientation Analysis

Te Brake's algorithm assigns to each pixel a value, which is a measure of the level of suspicion of that pixel. The method is best explained by taking an explicit structure which has a strong response to the algorithm. As such, consider a mass with an intensity distribution given by a two-dimensional Gaussian profile,

$$I(x, y) = I_o \exp(-A(x^2 + y^2)). \quad (1)$$

In fig 1 we construct the contours for such a shape and display their gradient vectors for a representative set of points. If we extrapolate each point along its gradient vector we find they all cross at the origin.

te Brake's algorithm visits each pixel in turn counting the number of surrounding pixels that satisfy the following two criteria,

1. They are within a distance R_{size} of the central pixel. This bounds the size of the mass structure for which we are looking to be less than R_{size}.

2. The extrapolated gradient vector at the pixel passes with a distance $R_{teBrack}$ of the central pixel, fig 1.

If the resultant value is higher than that obtained from vectors orientated randomly, then this is indicative of an underlying structure. A fuller description of the method can be found in [5].

No one is suggesting that masses are Gaussian structures with axial symmetry, however modelling a mass as an approximately circular structure whose image intensity decays away from a relatively bright region into a background value is a feasible starting point.

One problem with the above method is that masses are often held in place by spicules, which give the mass a star like structure. Such spicules 'confuse' the gradient operator and reduce the te Brake measure at that point. Spicules are part of a larger set of locally linear breast structure: blood vessels, Cooper's ligaments, milk ducts and fibrous tissue are all locally linear. Collectively they are known as curvilinear structure or CLS. CLS can obscure masses and alters the local gradient orientation in a mammogram.

To show quantitatively, whether the presence of curvilinear structures affects the performance of te Brake's algorithm, we aim to detect and remove these structures from mammograms prior to running the detection algorithm.

Detection and Removal of CLS using Kovesi's feature detection

CLS is locally linear with a well defined orientation, however on a larger scale it curves. CLS appears on mammograms as locally bright ridges, of various widths and lengths. CLS adds a 'whispy' effect to mammograms and obscures potential masses.

The detection of CLS has been considered previously. For completeness we mention Cerneaz [1], Cerneaz and Brady [2] and Zwiggelar [6]. Our approach to multiscale CLS detection is based on the ideas of phase congruency, first applied to feature detection by Kovesi [3].

Phase Congruency

The local energy model of feature detection postulates that features are perceived at points where the local Fourier components of a signal are maximally in phase, *i.e.* where phase congruency is a maximum [3]. Consider the local Fourier decomposition of a one-dimensional signal,

$$f(x) = \sum_{j=1}^{N} E_j \cos(jx) + iO_j \sin(jx) \qquad (2)$$

with $i = \sqrt{-1}$. Phase congruency is defined as,

$$PC = \frac{\sqrt{(\sum_{j=1}^{N} E_j)^2 + (\sum_{j=1}^{N} O_j)^2}}{\sqrt{\sum_{j=1}^{N}(E_j^2 + O_j^2)}}. \qquad (3)$$

The numerator of Eq. (3) is the square root of the local energy of the one-dimensional signal, and the denominator is the square root of the maximum possible local energy of the signal if all the Fourier components were in phase. The greater the agreement between phases of the individual Fourier components, the closer PC is to one. Equation 3 is invariant to the overall magnitude of the signal and so PC can be used to detect low-contrast features in a signal or image. However, it is also very sensitive to image noise and is ill-conditioned if all the local Fourier components of the image are very small. Kovesi alters the formulation of PC in [3] to avoid both problems, and uses oriented log-Gabor wavelet filters in symmetric/antisymmetric quadrature pairs to obtain local frequency information at an image pixel. Further details can be found in [3].)

We make several assumptions about the CLS that we wish to detect. The intensity profile perpendicular to the orientation of a strand of CLS is a one-dimensional peak that can be approximated to a scaled version of $\cos(\bar{\phi})$, where $-\pi/2 \leq \bar{\phi} \leq \pi/2$. We also assume that, because

the CLS is longer than it is wide, the PC perpendicular to the CLS is greater than the PC along the CLS at scales shorter than the CLS. We calculate the weighted mean local phase at each pixel in a mammogram, $\bar{\phi}_\theta(x)$, over a range of orientations, θ, according to,

$$\bar{\phi}_\theta(x) = \arctan\left(\sum_{j=1}^{N} O_{\theta,j}(x), \sum_{j=1}^{N} E_{\theta,j}(x)\right). \quad (4)$$

Here $O_{\theta,j}(x)$ and $E_{\theta,j}(x)$ are the local results of convolving the mammogram with the odd and even parts of the log-Gabor filter at scale j and orientation θ [3]. Since we are looking for ridge structures we only calculate local phase congruency at a pixel x and orientation θ if the following inequality holds,

$$|\bar{\phi}_\theta(x)| \leq \pi/2. \quad (5)$$

Otherwise $PC_\theta(x)$ is set to zero.

After determining values of $PC_\theta(\vec{x})$ at intervals over the full range of θ, we consider the pixel $\vec{x}$ to be part of a CLS if the absolute difference $|PC_\theta(\vec{x}) - PC_{\theta+\pi/2}(\vec{x})| > 0$ for any θ. The final output of the CLS detection procedure is a binary image with CLS features marked, as shown in Fig. 2b.

Removing CLS from Mammograms

The binary image of the CLS obtained in the previous section is now used as a template for removing CLS from the original mammogram. We first dilate the binary image using a 3×3 structuring element so that all the boundary pixels in the dilated image are positioned just outside CLS in the original mammogram. Finally, the CLS are removed from the original mammogram by replaces the original values of pixels located in the interior of detected CLS, with values smoothly interpolated from the boundaries of detected CLS.

Results

The results are displayed in fig 3 using a ROC format. We explain how the results are generated. Te Brake's method assigns a value to each pixel - a measure of suspicion. If this value is less than (greater than) some threshold value, α, the pixel is deemed to be normal (mass) tissue. Every image in our database [1] has an accompanying truth file - an image file with the mass structures marked by a radiologist. Given this ground truth we calculate for every image in the database the following two numbers,

- TP - The true positive fraction, the ratio of the number of pixels which the algorithm and the radiologist have both marked as masses divided by the total number of mass pixels.
- FP - The false positive fraction - the ratio of the number of pixels marked as masses by the algorithm and normal by the radiologists divided by the total number of normal mass pixels.

Each point on the graph is, for a fixed threshold α, the mean $(\bar{TP}, \bar{FP})$ evaluated over our set of seventy cases,

$$(\bar{TP}, \bar{FP}) = \left(\frac{\sum_{i=1}^{N} TP_i}{N}, \frac{\sum_{i=1}^{N} FP_i}{N}\right). \quad (6)$$

Where TP_i and FP_i denote the true positive fraction and false positive fraction respectively for image i. As the threshold is lowered, more and more of each image is deemed to be mass and the TP value rises. Of course the increase in TP has to be counterset against the rise in FP. Radiologists would quickly lose confidence if the number of false positives were unacceptably high.

The results are shown in the ROC curve. fig 3. Three lines are displayed,

1. - The results of the algorithm on the raw image data.
2. - The results of the algorithm with the pre-processing CLS removal step
3. - The results of the algorithm with a pre-processing averaging step. Each pixel is replaced by the mean of its own value and its eight nearest neighbours.

We see that the pre-processing CLS removal step improves te Brake's algorithm, for any value of FP its TP value is always greater than both the averaged and the original images.

[1] We acknowledge the database at the University of South Florida http://marathon.csee.usf.edu

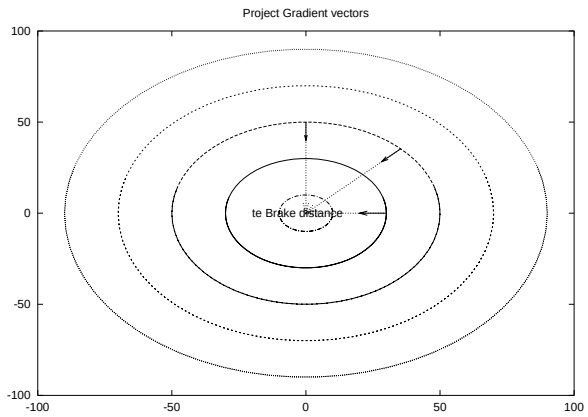


Figure 1: Gradient vectors and their projection back into the te Brake distance.

Conclusions

The main conclusion to be drawn from this work; CLS removal improves the performance of te Brake's algorithm.

A natural question arises, would te Brake's algorithm be improved by other more conventional ways of removing fine detail structure? The answer to this question is yes, as can be seen in the ROC curve for the averaged data. However averaging offers a smaller improvement in performance than our CLS removal technique.

References

[1] N.J. Cerneaz. *Model-based Analysis of Mammograms*. PhD thesis, Department of Engineering Science, University of Oxford, 1994.

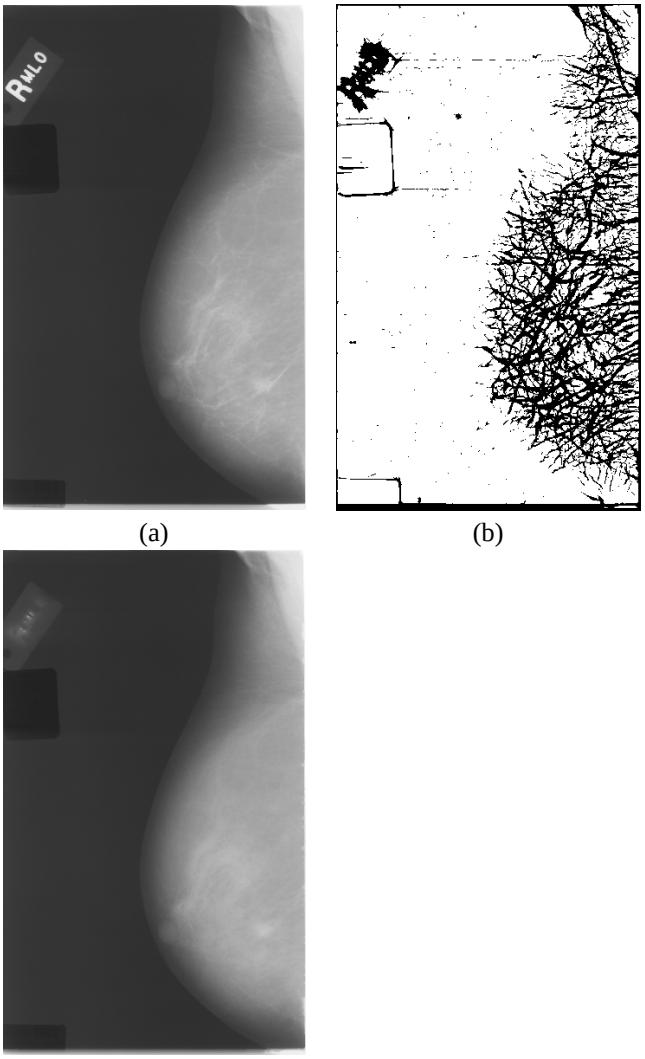
[2] N.J. Cerneaz and M. Brady. Finding curvilinear structures in mammograms. In *First International Conference on Computer Vision, Virtual Reality and Robotics in Medicine, CVRMed'95*.

[3] P. Kovesi. Image features from phase congruency. *Journal of Computer Vision Research*, 1999.

[4] G.M. te Brake, N. Karssemeijer, and J.H.C.L. Hendriks. Automated detection of breast carcinomas not detected in a screening program. *Radiology*, 1998.

[5] G.M. te Brake, N. Karssemeijer, and J.H.C.L. Hendriks. An automatic method to discriminate meligant masses from normal tissue in digital mammograms. *Phys. Med. Biol.*, 2000.

[6] R. Zwiggelaar, R. Marti, and C.R.M. Boggis. Detection of linear structures in mammographic images. In *Proceedings of Medical Image Understanding and Analysis MIUA 2000*.



(a) (b)

(c)

Figure 2: (a) Original Mammogram. (b) CLS Detected in (a) using phase congruency. (c) Mammogram with CLS removed.

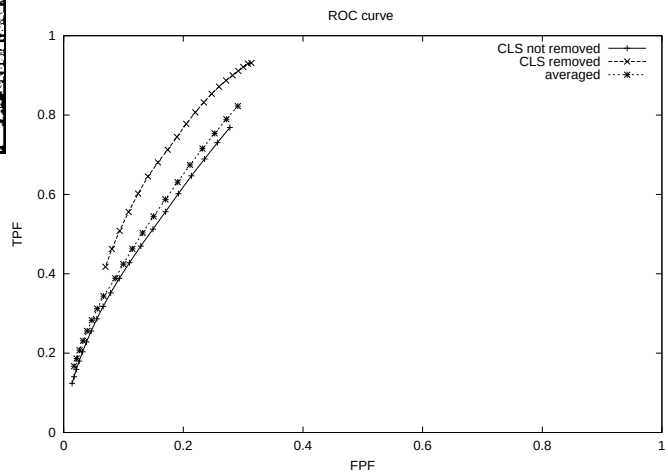


Figure 3: Results