BMVC99

Generating Realistic Mass Lesions In Digital Mammograms Using Statistical Models

Steven Caulkin and Sue Astley
Imaging Science and Biomedical Engineering
Stopford Building
University of Manchester
Oxford Road
Manchester, M13 9PT, UK
sjc@server1.smb.man.ac.uk

Abstract

We describe a model-based method of generating realistic mass lesions and placing them in appropriate locations on normal mammogram backgrounds. These lesions are highly irregular in shape and have no consistent internal landmarks. Thin plate splines are used both in modelling the background to each lesion, and to warp examples when training an appearance model.

1 Introduction

The aim of this work is to generate realistic lesions for the purposes of training and testing radiologists in a computer-aided learning system. In order to achieve this, it was neccessary to develop a model of the shape and grey levels of a set of example lesions. The model had to be capable of coping with examples which have no consistent landmark points. In addition, a method was required for isolating each lesion from the background structures upon which it is projected during imaging. The model may then be used to generate new example lesions consistent with the training data.

A data set of consecutive abnormal examples was used to provide a representative sample of mass lesions. A statistical model of where lesions occur within the breast was built in order to select realistic locations in normal mammograms at which to place synthetic lesions. The background was subtracted from the region of each lesion in order to model the attenuation due to the lesion alone. The edges of the annotated lesion boundaries were refined using the subtracted images to ensure that they were in the correct positions. Landmark points were placed around the lesion boundaries and a combined model of shape and grey level was built using thin plate splines. The model was then used to generate new example lesions, which were superimposed on normal mammogram backgrounds.

BMVC 1999 doi:10.5244/C.13.29

2 Background

One in twelve women develop breast cancer within their lifetimes. In the United Kingdom women between the ages of 50 and 65 are screened for breast cancer every three years. The screening procedure involves taking an X-ray of each breast, and these are examined by expert radiologists in order to detect any abnormalities. Each film is viewed independently by two radiologists, which significantly improves the detection rate compared to that of a single radiologist. The radiologists are specially trained in viewing mammograms but their judgements are still prone to error. Computer vision has the potential to aid or even replace the radiologist if sufficient detection performance levels can be achieved. Aid for the radiologist may be in the form of prompts placed on the screen at locations regarded by the computer system as suspicious [1] or as part of the training and performance evaluation process.

Mammography is highly bound by quality assurance procedures and evaluation of radiologists' performance is a key part of the process. It has been found that radiologists need to maintain a high throughput in order to perform well. The aim of any training system is to improve the performance of radiologists so fewer cancers are missed and fewer healthy women are subjected to unnecessary surgery. Traditional training methods involve a time consuming process of pulling films from an archive and radiologists may only have access to a limited number of films. Computer-aided learning has the potential to provide a computer-adaptive approach, which is more flexible.

A model-based method of describing lesions allows the generation of a very large number of new examples limited only by the variation within the training set. There will be no repetition of images, as may be the case during conventional training, which would lead to some examples being remembered. If the training set is representative of the screening population, a full range of abnormalities may be generated. The size of a model is small compared with storing very large numbers of images, although a range of normal images on which to place the generated lesions is required. Quantitative analysis of the performance of the radiologist is possible in a computer-based system. The radiologist's ability to detect certain types of lesions or lesions in certain locations may be analysed and training sets loaded to focus on specific areas of weakness. This approach would be difficult without the use of synthesized abnormalities.

Previous attempts have been made to model mass lesions in mammograms. Claridge et al [2] used a simple spherical model of a lesion and blurred the edges according to measured values of edge blur. Highnam et al [3] used the h_{int} representation to produce more complicated synthetic lesions. A spherical model of attenuation was used along with measured attenuation coefficients to calculate the X-ray attenuation due to a lesion. Morphological operations were then used to transform the distribution onto an existing shape. Neither of these approaches attempts to generate new, realistic lesion shapes and consequently neither would be suitable for our application.

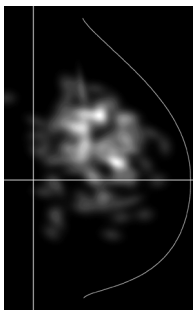

Figure 1: Sites of occurrence of lesions

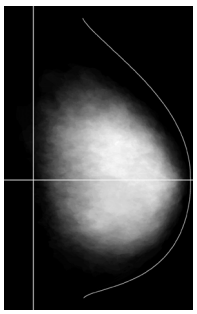

Figure 2: Glandular tissue distribution within the training set

3 Data

102 images were taken from a consecutive sequence of screening mammograms showing cancerous masses from the Greater Manchester Breast Screening Centre. The data includes roughly equal numbers of incident, prevalent and interval cancers. Incident, prevalent and interval cancers are those detected at the first screening visit, at later screening visits and between screening visits respectively. The abnormalities are all biopsy-proven malignant mass lesions. The lesions may be spiculated, with radiating linear structures, well-defined or ill-defined, referring to the sharpness of the edge. For each film the lesion, breast border, pectoral muscle, glandular region and nipple position have been annotated. In the case of spiculated lesions, the spicules have not been annotated at this stage because the process is time consuming and error prone. They will, however be incorporated into the model later. Images were excluded from the sequence if the lesions overlapped the edge of the film. The films were digitised on a Lumisys 100 scanner at a spatial resolution of 100μ and an 8-bit grey level resolution.

4 Method and Results

4.1 Placing Synthetic Lesions

We have previously modelled the spatial distribution of abnormalities within the breast [4]. The annotated images were used to warp each image onto the mean breast shape. The warped lesion outlines were used to create a smooth distribution of lesions within the breast, shown in figure 1. The axes represent a straight line fitted to the pectoral muscle and its perpendicular passing through the nipple. The results were consistent with radiologists' observations that cancers occur most frequently in the upper outer quadrant of the breast. A model of underlying glandular tissue distribution was also developed, since most lesions lie within the gland disc and this is shown in figure 2. The distributions may be used to select realistic locations for generated lesions which are representative of the screening population.

4.2 Subtracting Background

4.2.1 Introduction

In order to model mass lesions and superimpose them on normal mammogram backgrounds, the X-ray attenuation due to the lesion alone must be calculated. The mammogram is created from the exposure of each part the film to the X-ray beam, which is proportional to the energy imparted by the X-ray beam at that point. The beam is attenuated by the mammographic structures, which absorb a fraction of the beam as it passes through. In the lesion region, the beam is attenuated by a fraction due to the background and a further fraction due to the lesion. The optical density of the film is proportional the log of the exposure. Thus multiplying the exposure by a fraction (i.e. attenuating the exposure) is equivalent to subtracting a value from the optical density. The relationship between the film optical density and the grey level values obtained by scanning the film with a Lumisys scanner is linear so adding a value onto the optical density is equivalent to adding a value onto the grey level. Therefore attenuating the beam by a further fraction at a given point is equivalent to adding a value onto the grey level of the digitised image. Alternatively, subtracting a grey level from the image is equivalent to removing a certain attenuation factor. In order to calculate the attenuation pattern due to each lesion, a smooth surface was fitted to the background grey levels around the lesion and subtracted from the original image. If these subtracted images are modelled, lesions generated from the model may be superimposed on another mammogram to simulate the resulting increase in attenuation.

4.2.2 Method

A region of interest was taken to encompass each lesion and its surrounding background. The background may consist of a range of tissue types from fatty (dark) to glandular (bright). Smaller scale linear structures such as vessels and ducts may also be present. The grey level values within the lesion must be estimated using only information from the grey levels outside it. A three dimensional thin plate spline [5] was used to fit a surface to the background around the lesion and to interpolate inside the lesion region. The thin plate spline is defined by two sets of corresponding landmarks in three dimensional space. A transformation is then defined from one co-ordinate system to the other for any point.

In this case, the dimensions are the x and y co-ordinates and the grey level values, and the image is represented as a three-dimensional grey level surface. Landmark points were placed outside the boundary of the lesion and the spline was used to define the transformation from the set of points $(x, y, 0)$ to (x, y, z), where z is the grey level and x and y represent the two-dimensional co-ordinates of the landmark points. This transformation could then be used to estimate a grey level value at any point in the region bounded by the landmarks.

4.2.3 Parameter Selection

There are several parameters which may be chosen when defining the spline. These relate to both placing landmark points and smoothing. Quantitative tests were carried out on patches of normal background in order to optimise the technique and

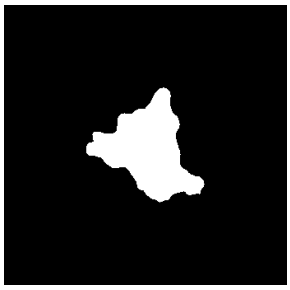

Figure 3: Mask

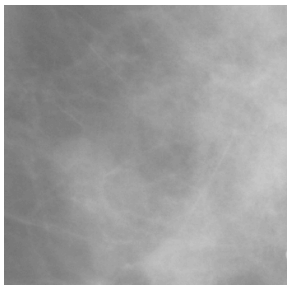

Figure 4: Normal mammogram

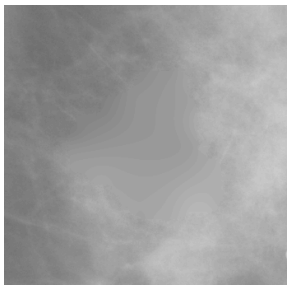

Figure 5: Background fitted within the mask

determine which parameters to use. Twenty normal background regions of interest (ROI) were selected from the images in the training set. They were chosen to avoid the abnormality in each image and to cover a wide range of background tissues. To define the ROI, the first 20 lesion boundaries were taken and superimposed on the normal background regions. The grey levels within this region were calculated by fitting a spline to pixels outside the boundary. The difference between the real and fitted grey levels was calculated and the mean absolute error in grey level was taken as a measure of error in the fitting process.

Tests were carried out by selecting a set of landmarks and smoothing the image before fitting the spline. Landmarks were placed at equal intervals throughout the background region. A permitted region was then defined and any landmarks outside this region were ignored. The permitted region was defined by repeated dilation of the lesion boundary, to produce a band the same shape as the boundary. The standard deviation of the Gaussian filter used for smoothing, the spacing of the landmark points and distance of the band from the lesion boundary were varied in a series of experiments. The same method was used for all 20 of the test images and the mean and standard deviation of the error were calculated. The error increased as the band was moved further away from the lesion boundary, as expected. The spacing of the landmark points had little effect on the fitted surface but the error decreased as the standard deviation of the Gaussian filter increased. The best result was a mean grey level error of 4.74 ± 1.41 on a scale of 0 to 255. This was achieved with the band of pixels placed as near to the lesion boundary as possible without the filter encroaching into the lesion region. An example normal background, the mask used and the fitted background are shown in figures 3,4 and 5 respectively. The error in this case was 5.02 grey levels.

When fitting a background to images containing real lesions, it is essential that the lesion does not contribute to the fitted background. It was observed that the annotated lesion boundaries sometimes encroached upon the lesion region, although they otherwise seemed correct. The landmarks defining the spline were placed a fixed distance outside the lesion boundary to allow for any inaccuracies. The band of landmarks was between 20 and 40 pixels from the boundary. The image was smoothed using a 21 by 21 pixel Gaussian filter with a standard deviation of 5 pixels. Thus a band of 10 pixels thickness around the lesion is left untouched.

Figure 6: Lesion

Figure 7: Fitted Background

Figure 8: Grey levels due only to lesion

Using these settings on the normal background images resulted in a mean error of 5.28 ± 1.66 grey levels. An example lesion, the fitted background and the lesion image after background subtraction are shown in figures 6,7 and 8 respectively. The subtracted image has been superimposed on a flat grey background to improve visibility.

4.3 Edge refinement

As mentioned in section 4.2.3, it became apparent that the parts of some lesions had been annotated incorrectly. The subtracted images were calculated using a dilated boundary to ensure that the incorrect annotations did not have an effect. However, if an accurate model of the subtracted lesion grey level and boundary is to be built, the boundary must be in the right place. The majority of malignant lesions have blurred edges, and in any case a small amount of blurring is inherent in the imaging process, so one would expect the attenuation contribution at the edges of lesions to be very low. In some cases the annotated border overlaps part of the lesion; this would result in the model generating unrealistically high grey level values at the edges of lesions. Since we are currently unable to obtain expert refinement of the annotations, we have used a simple, automatic method, which alters the boundary at pixels which deviate from the mean grey level around the boundary. In this way, bright boundary pixels are identified as being within the lesion and the boundary moved outward accordingly. This process is repeated for a number of iterations and eventually converges.

4.4 Appearance Modelling Using Thin Plate Splines

4.4.1 Introduction

An appearance model [6] is a combined statistical model representing the shape and grey level in an image. A training set of images is required in which all of the images have a consistent set of corresponding landmarks defined. The first step is to build a statistical model of shape variation. The landmark points are all aligned into a common co-ordinate frame and each shape may be represented by a vector x. Principal component analysis (PCA) is then performed on the normalised vectors. Any example x can then be represented as a weighted sum of

orthogonal modes

$$x = \bar{x} + P_s b_s \tag{1}$$

where $\bar{x}$ is the mean shape, P_s is a set of orthogonal modes of shape variation and b_s is a set of shape parameters.

A similar model of grey level is then built. Each example image is warped to the mean shape and a grey level sample vector g is taken from the image. However, the vectors are not normalised before PCA is performed. This is because the sample vectors contain attenuation information, which would be distorted by scaling and shifting them. A grey level sample vector g may then be expressed as a weighted sum of orthogonal modes

$$g = \bar{g} + P_g b_g \tag{2}$$

where $\bar{g}$ is the mean shape, P_s is a set of orthogonal modes of grey level variation and b_s is a set of grey level parameters. Shape and grey level variations may be correlated so the two models are combined to form a model of appearance. A concatenated vector is generated for each example and PCA is again applied to the data to give orthogonal modes of combined shape and grey level variation. The shape and grey level vectors must be weighted in order to allow for the difference in units between the shape and grey level models. The concatenated vector b is calculated for each example as follows

$$b = \begin{pmatrix} W_s b_s \\ b_g \end{pmatrix} = \begin{pmatrix} W_s B_s^T (x - \bar{x}) \\ P_g^T (g - \bar{g}) \end{pmatrix} \tag{3}$$

After applying PCA to these concatenated vectors, each example vector b may be represented as follows

$$b = Qc \tag{4}$$

where Q are the eigenvectors of b and c is a vector of appearance model parameters. An example image is regenerated from the model by warping the shape-free grey level sample vector onto the original landmark points. Note that the shape vectors have been normalised; to reconstruct the original image the inverse transform must be applied to the shape vectors by scaling and shifting them accordingly.

4.4.2 Adapting the Models to Lesions

Applying appearance models to lesions poses several problems. Ideally, landmarks should be placed on consistent features present in all images, for example the eyes, nose and mouth for face images. Lesions have no consistent internal features, so points were placed at equal intervals around the boundary. An orientation was defined for each lesion by drawing a straight line from the nipple through the centre of the lesion. This was chosen as the most reasonable orientation due to the branching nature of the ducts which open at the nipple; many lesions develop in these structures. Two landmarks were defined by taking the intersection of the line and the boundary and a further 75 were placed between the two landmarks at equal spacings along the boundary on either side to give a total of 152. The number of

landmarks was chosen to give a spacing of about 5 pixels on a boundary of average length, which should be sufficient to describe irregularities in the boundary.

Previous applications of appearance models have used Delaunay triangulation to warp each shape onto the mean before sampling grey levels and reconstructing examples from the sample vector [6]. Triangulation involves drawing a mesh of triangles between the landmark points in each image. Triangles are then matched in different images and warping one image onto another involves bilinear interpolation of the grey levels within matching triangles. Appearance models have been applied to faces [6], which have regular outlines and internal landmarks. Defining a triangulation for lesion outlines is more difficult, as there are no internal landmarks and a zigzag pattern of long, thin triangles would be the result. The triangulation could be improved by defining the centre of the lesion as a landmark and drawing triangles radially outward from it. This results in triangles which are still long and thin but which are all of a fairly similar size. A more serious problem is the fact that the lesions sometimes have bulbous emanations. This would result in triangles disappearing or being flipped over. It was decided to use thin plate splines [5] to perform warping. This means that the landmarks on the boundary can be used and no internal points need to be defined. The method does not suffer from any of the drawbacks which occur with the conventional method, Delaunay triangulation.

4.5 Generating New Examples

Each example in the training set may be represented as a weighted sum of combined model parameters (4) so for each parameter there is a distribution of weights. A new parameter vector was generated by sampling each element of the vector from the corresponding cumulative distribution function of the training set parameter vectors. The advantage of this method is that no values outside the range of the data are selected and the distribution is not assumed to be normal. However, interdependencies between modes are not yet taken into account. In order to reconstruct a new example from the parameter vector, the shape vectors must be rescaled. Currently, the scale value is sampled from the distributions of scales in the training set. In the final system, it will be necessary to model the relationship between the scale and modes. An example normal background is shown in figure 9 and the same background with a synthetic lesion superimposed upon it is shown in figure 10.

5 Discussion

The main advantage of this approach is that everything is based on real data. The shape and grey level of lesions is modelled in such a way as to allow generation of a very large number of examples. Using the thin plate spline as part of an appearance model leads to much better results with this data than if a triangulation had been used. There is potential to use this in other applications to improve the warping, although it is more computationally expensive, which may be important when using the active appearance model for searching. The background subtraction method proposed here allows the extra attenuation due to the lesion to be

Figure 9: Normal Background

Figure 10: Synthesized lesion

modelled directly. There are, however, errors involved in the background subtraction process, as linear structures or parenchymal patterns which only appear within the lesion region will not be modelled. The method may be improved by trying to detect such structures in the lesion region and correcting the subtraction accordingly.

The method currently used to generate new examples is very simple and doesn't take into account any interdependency between modes or any relationship between the size and the feature vector. The brightness of a lesion is related to the thickness of the tissue, so bigger lesions should normally appear brighter. This could be improved by modelling scale in the same way as shape and grey level, which would result in a set of modes containing scale parameters. Although the generated parameters are all within observed ranges, combinations may be selected which do not actually occur. Further work will be carried out to investigate whether generated sets of parameters are realistic.

Another problem with this approach is the loss of resolution due to sampling grey levels from lesions of varying size. This can be minimised by using a large sample vector; since the generated lesions are superimposed on noisy mammogram backgrounds, the loss of resolution may not be easily visible.

We have not yet considered spicules. The spicules will be annotated by an expert radiologist and modelled separately from the central mass. This is because a boundary with spicules has a spiky appearance, which would be difficult to model as the spikes can occur at any position around the boundary. A central mass could, however, be generated using the technique described and the number, position and appearance of spicules could then be modelled and joined to the generated mass at appropriate points on the boundary by distorting the generated boundary at those points.

6 Conclusions

We have addressed the problem of producing training and test images for radiologists. These are necessary for computer-adaptive CAL systems and for quality

assurance in breast screening. The model could be developed further to aid both detection and diagnosis of mass lesions.

The appearance modelling approach has been adapted to cope with cases with irregular, highly variable boundaries and no consistent internal landmarks. The technique appears to work well and is capable of producing a wide range of realistic synthetic lesions. Problems with the approach, such as the modelling of scale parameters, will be addressed and the model will also be expanded to include spiculated lesions.

We are currently carrying out a study to determine whether radiologists can distinguish between real and synthetic lesions; preliminary results suggest that the synthetic lesions will be sufficiently realistic for training and test purposes.

7 Acknowledgements

We are grateful to the Greater Manchester Breast Screening Centre and the National Health Service Breast Screening Programme for their collaboration.

References

[1] Astley S, Zwiggelaar R, Wolstenholme CB, Davies K, Parr TC, Taylor CJ (1998) Prompting in mammography. How good must prompt generators be? Digital Mammography, Ed. Karssemeijer, Thijssen, Hendriks and van Erning, Kluwer Academic Publishers, Dordrecht, pp 347-354.

[2] Claridge E, Richter JH (1994) Characterisation of Mammographic Lesions. Proc. 2nd International Workshop on Digital Mammography pp 241-250, Excerpta Medica International Congress Series 1069.

[3] Highnam R, Brady M, Shepstone B (1996) A representation for mammographic image processing. Medical image analysis 1(1), pp 1-18.

[4] Caulkin SJ, Astley S (1998) Sites of occurrence of malignancies in mammograms. Digital Mammography, Ed. Karssemeijer, Thijssen, Hendriks and van Erning, Kluwer Academic Publishers, Dordrecht, pp 279-282.

[5] Bookstein FL (1989) Principal warps and the the decomposition of deformations. IEEE transactions on pattern analysis and machine intelligence 11(6), pp 567-585.

[6] Edwards GJ, Taylor CJ, Cootes T (1997) Learning to identify and track faces in images sequences. Proc. British Machine Vision Conference 1997, Colchester, UK.