Elimination of Linear Structures as an Attempt to Improve the Specificity of Cancerous Mass Detection in Mammograms

Marcin Bator and Leszek J. Chmielewski

Institute of Fundamental Technological Research, Polish Academy of Sciences,
Świętokrzyska 21, PL 00-049 Warsaw
{mbator, lchmiel}@ippt.gov.pl

Summary. In the screening mammographic examination of large parts of populations thousands of mammograms are analysed. The Computer Aided Diagnosis methods available still tend to produce too many false positive (FP) detections with respect to the number of true positive (TP) detections, which makes it impractical to use such methods to support the human observer in the analysis of mammograms. In this paper an attempt has been made to decrease the number of FP errors in the hierarchical correlation-based cancerous mass detection method by eliminating the images of linear structures (LSs) from the mammograms. The LSs were detected with an accumulation-based line detector and the image intensity function in the regions of the LSs was interpolated with an anisotropic membrane. Examples of images representing typical detection problems caused by the LSs selected from the MIAS database suggest the feasibility of the proposed approach.

1 Introduction

Breast cancer is an important social problem. It can be estimated that in the US in one out of eight women the breast cancer will develop at some point during her life [11]; the corresponding number for Poland is one in 16 women [6]. Early detection of breast cancer makes it possible to apply a sparing treatment and elongates the survival. Therefore, screening the groups of elevated risk (women above the age of 50) is justified, which involves the analysis of thousands or millions of mammograms. The development of Computer Aided Diagnosis (CAD) methods might help the radiologists in carrying out this task. In less than 10% of women the result of the screening mammography is positive, that is, it suggests further investigation. If a CAD system could classify at least a part of 90% of mammograms as negative prior to the investigation by a human, it would greatly reduce the workload of radiologists and make it possible for them to pay more attention to the remaining mammograms, potentially carrying the features of an abnormality. A CAD system, to be useful in improving the performance

of a radiologist, should have a sufficiently large specificity at a given sensitivity level, comparable to that of a human observer. In [17] a discussion has been presented which leads to the conclusion that there should be not more than 1.5 false positive detections (FP) for each true positive detection (TP). In the present paper we report our attempt to increase the TP/FP ratio in a cancerous mass detection algorithm described in [1] by eliminating the linear structures from the mammograms.

Comprehensive studies of all aspects of breast cancer imaging can be found in [8]. Reviews of cancerous masses detection and classification can be found in [11] and to some extent also in [15]. Linear structures (LS) [7, 16], also called curvi-linear structures [10, 12], are the blood vessels, milk ducts and connective tissue fibers (Cooper ligaments) in normal mammograms, and the spicules in pathologic mammograms. Some authors report that their classification should improve the subsequent detection and classification of the features of cancer, like in [16]. Others, like in [12], state that the attenuation of the LSs improves the results of analysis. The LSs related to cancerous masses are spicules; other LSs interfere in the detection of masses. In this study a mass detection algorithm using no information on the spicules is used, so we go along the line of elimination of all the LSs without their classification.

2 Detection of cancerous masses

The cancerous masses were detected with a multiscale template matching algorithm reported previously in [1]. A template is an auxiliary image representing the model brightness distribution in a mass. The template is shifted across the image and the similarity measure between the template and the corresponding window of the image is calculated. As the similarity measure the correlation coefficient was used, as expressed by the equation

$$c(T, I) = \frac{\sum_{j=1}^{N}(t_j - \bar{t})(i_j - \bar{i})}{\sqrt{\sum_{j=1}^{N}(t_j - \bar{t})^2} \sqrt{\sum_{j=1}^{N}(i_j - \bar{i})^2}} \; , \tag{1}$$

where T is a template and I is a window in the image, both containing N pixels numbered consecutively, t_j is a grey level in a template pixel, i_j is a grey level in a window pixel, and $\bar{t}, \bar{i}$ are the mean grey levels of the template and the window, respectively. The correlation coefficient belongs to an interval $\langle -1, 1 \rangle$ and depends only on the relation of shapes of the template and the window, not on their grey scale. A template with a hemispherical brightness was used:

$$t_R(x, y) = \sqrt{R^2 - x^2 - y^2} \; \text{ for } \; x^2 + y^2 \leq R^2 \; , \tag{2}$$

where x, y are pixel coordinates in a local coordinate system of the template having origin at the template centre. As in [1], the the Fast Fourier Transform was used and the correlation was calculated in the Fourier domain, while unlike in [1], the template was circular.

As advised in [14] and according to the previous experience [1], the template matching combined with a multiscale approach was used. The templates with 8 radii from 36 to 108 pixels, differing by a factor of $\sqrt{2}$, were used. To reduce the amount of calculations, images in four scales were analysed: 1:1, 1:2, 1:4 and 1:8, with two radii in each: 36 and 51 pixels. Templates with such radii detect objects having radii of 30-43 pixels and 43-60 pixels, respectively. The multiscale approach was designed according to the observation made in [9] that instead of using a template with two times larger radius one could use an image with two times smaller linear dimensions, obtaining approximately the same results. Combining such a resolution pyramid with FFT calculations mentioned above reduced the calculation time for a single mammogram to less than 10 min on a 2 GHz computer. At the image resolution of 50 μm/pix, the masses of diameters from 3 to 48 mm could be detected.

The final detection result was calculated by thresholding the resulting correlation maps at the subsequent pyramid levels with a single threshold and aggregating the partial results received onto the full scale level 1:1.

3 Linear structures

3.1 Detection

The method used for detection of the linear structures was the evidence accumulation-based line detection algorithm already described in sufficient detail in [4, 5], and previously in [3]. In the present application, linear structures of width between 2 and 16 pixels were sought. Structures having the line intensity maxima smaller than 0.1 of the maximum intensity of the strongest LS in the given image were neglected and the ridges of the remaining, stronger structures were followed (some less important details were omitted for the sake of compactness; for details see [4], p. 360 or [5], Chapt. 6.3.6 and 6.8).

The line detector used finds the line width w_0 as the distance between the loci of maximum image intensity gradients, as is typically done by detectors. However, the actual width w of the LS in the image is larger. Therefore, the width found was multiplied by a factor slightly larger than one: $w = 1.1\,w_0$.

The primary results of the detection process for each mammogram were the binary masks of the LSs. In the LS elimination process described in the following Section, also the the map of the directions of the LSs and the relative distance transform, defined further, were used. These were easy to generate as the byproduct of the detection process, due to that each detected pixel is related to its respective pixel in the skeleton of the line, in which the local line width and direction has been stored by the line detection algorithm. The direction is simply copied from that stored in the relative skeleton pixel. The relative distance transform in a given pixel of the LS is the ratio of its distance from the line edge to the distance of its relative skeleton pixel from the edge, equal to a half of the width.
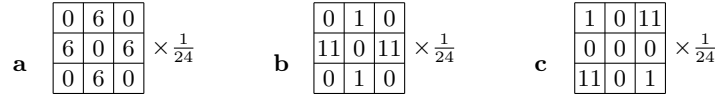
a $\begin{array}{|c|c|c|}\hline 0 & 6 & 0 \\\hline 6 & 0 & 6 \\\hline 0 & 6 & 0 \\\hline\end{array} \times \frac{1}{24}$ **b** $\begin{array}{|c|c|c|}\hline 0 & 1 & 0 \\\hline 11 & 0 & 11 \\\hline 0 & 1 & 0 \\\hline\end{array} \times \frac{1}{24}$ **c** $\begin{array}{|c|c|c|}\hline 1 & 0 & 11 \\\hline 0 & 0 & 0 \\\hline 11 & 0 & 1 \\\hline\end{array} \times \frac{1}{24}$

Fig. 1. Examples of masks used for interpolation in the membrane spanning algorithm. The sum of elements is divisible by four to make it easier to construct the isotropic mask. (**a**) Isotropic membrane; (**b**) membrane stronger in the W-E direction; (**c**) membrane stronger in the SW-NE direction.

3.2 Elimination

A linear structure in the breast projects itself on the mammogram together with other tissues and manifests itself as a bright object in the image. The image intensity function of a mammogram can be split as follows

$$I(x, y) = I_0(x, y) + I_{LS}(x, y) , \tag{3}$$

where I is the image intensity of the mammogram, x, y are coordinates, I_0 is the intensity without the LS, and I_{LS} is the intensity of the LS alone.

The possible approaches to the elimination of the LSs are: 1° subtraction of the I_{LS} component of the intensity from the image, 2° replacement of the image intensity in the regions occupied by the LSs with the intensity I_0, and 3° suppression of the component I_{LS} so that the resulting intensity is closer to I_0. The first approach necessitates for the model of the LS. The images of the LSs are very different so this approach seems unrealistic. The second approach makes it necessary to know I_0 in the regions of the LSs. In the third approach it is assumed that the method of suppressing the LSs is known. For example, in [12] the LSs were attenuated by averaging the image intensity in the regions occupied by the LSs in a 11×11 pixels window.

In this paper the second approach was used. The image intensity I_0 without the LSs in the regions of the LSs was estimated as the interpolation of the image intensity from outside the regions occupied by the LSs. This was done by spanning a *membrane* across the regions of the LSs. An iterative algorithm was used, as described in [2], Chapter 6.2.1. In the present application that algorithm becomes radically simplified to finding the averages of the neighbouring pixels, with a 4-neighbourhood. This corresponds to the iterative convolution of the image with the mask shown in Fig. 1a, until the result stabilizes. The intermediate calculations were performed in the floating-point arithmetic. The method was similar to that used in [12] in that the convolution equivalent to averaging was used, while the main difference was that the convolution was repeated until the surface received stabilized.

Spanning an isotropic membrane across the regions occupied by the LSs yields undesirable results where the linear structure crosses the regions having different signs of the curvature, that is, the "hills" and "valleys", as shown in Fig. 2a and b. To reduce the saddles near the ridges and bottoms an anisotropic membrane was used. The direction of larger stiffness was across the LS, as found by the line detector. These directions were rounded to four characteristic
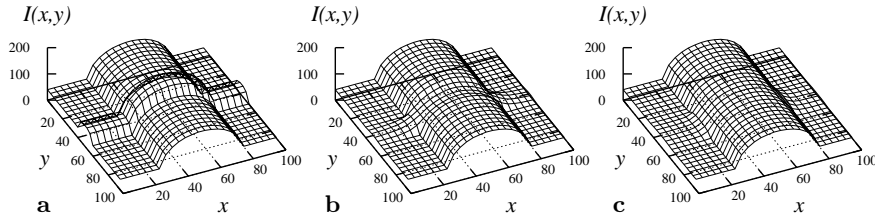
Fig. 2. Using the anisotropic membrane to interpolate an LS crossing a "hill" in the image intensity function. (**a**) Original image intensity: detected LS along Ox, 'hill' along Oy; (**b**) LS eliminated with the isotropic membrane: saddles present in the ridges and bottoms of the intensity graph; (**c**) LS eliminated with the anisotropic membrane, stronger in the direction across the LS: saddles greatly reduced.

directions differing by 45°, and for each pixel in the LS the appropriate mask was used. The examples of the masks are shown in Fig. 1b and c and the result of using this approach can be seen in Fig. 2c. The ratio of the stiffnesses in the perpendicular directions was close to 10, which was enough to receive a clear improvement in relation to the described effect.

In the final algorithm each interpolated surface was inserted into the image intensity function with a narrow transient stripe in which the two functions were gradually interchanged by linear weighting. The weighting functions used could be easily calculated from the relative distance transform generated by the line detection algorithm. The stripe width corresponded to the difference between w and w_0, mentioned above.

According to the multiscale organisation of the correlation-based mass detector, the line elimination process was also performed in four scales. However, the line detection was done only in the scale 1:4 and its results were appropriately recalculated to the remaining three scales.

4 Examples of results

The feasibility of the method has been tested on the images from the MIAS database [13]. The results for two images have been selected for presentation due to that the correlation method yields unsatisfactory results in these images because of linear structures passing over cancerous masses. Detection results before and after the elimination have been shown in Fig. 3 and Table 1. The detected masses have been marked with circles according to the shape of the templates used. The radii of the circles correspond to the radii of the masks, recalculated to the full scale 1:1, for which the maximum correlation reached the threshold. Thresholds used were equal to correlations shown in Table 1.

It can be observed that in the examples shown: 1° the correlation after the elimination of the LSs grew, 2° the number of false positive detections decreased (effect visible in the window shown only for `mdb124rm`), and 3° the location of the detected mass became closer to that determined by a human observer. A cautious conclusion can be drawn from these observations that it could be
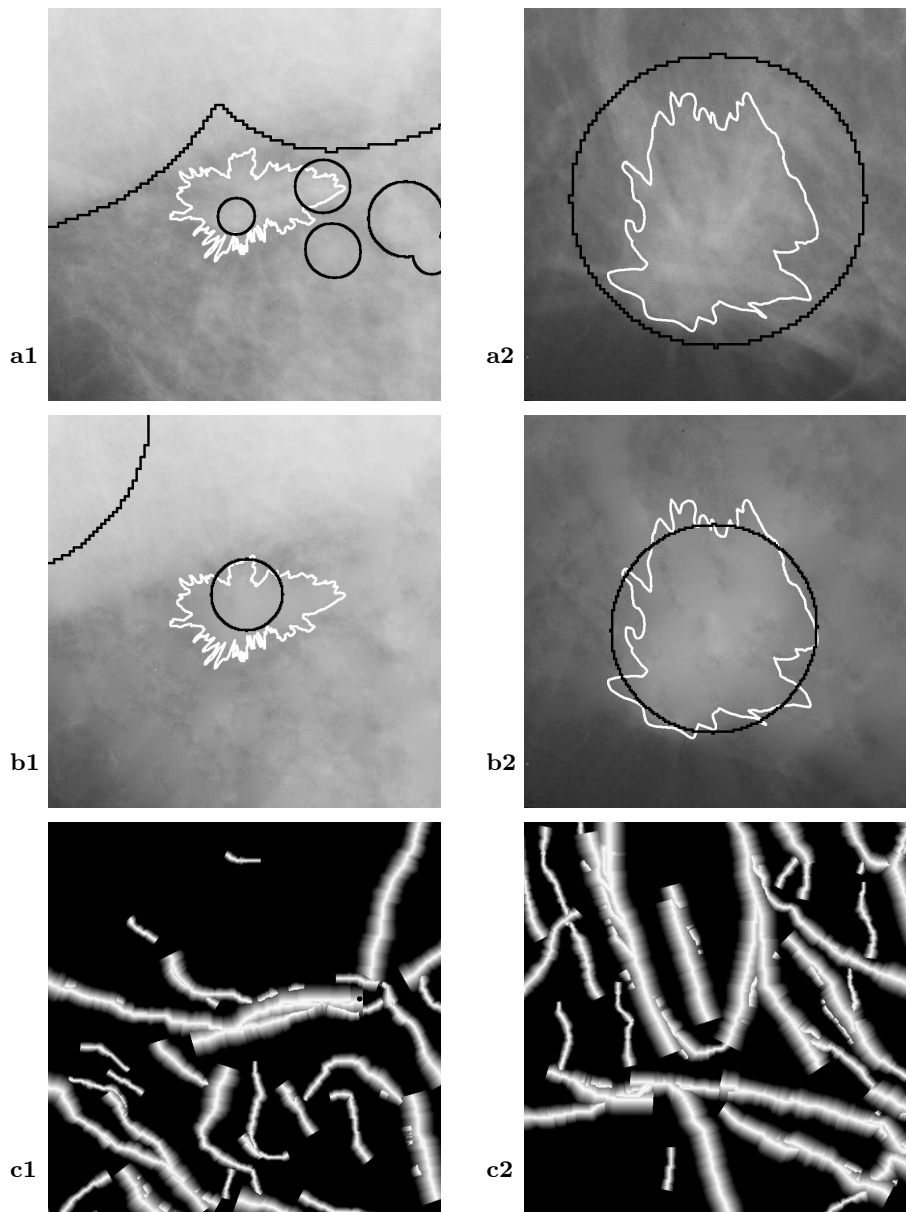
Fig. 3. Examples of results – windows 800×800 of mammograms from the MIAS database: (**1**) mdb124rm, upper left corner at $(1340, 24)$; (**2**) mdb264rm, upper left corner at $(2156, 836)$. (**a**) Original images; (**b**) images with the LS eliminated; (**c**) relative distance transforms of the LSs. The *black circles* indicate the detected cancerous masses, in the original and processed images, respectively, and the *white contours* indicate the shape of the masses marked by a human observer.

Table 1. Change of parameters of the masses in the mammograms shown in Fig. 3.

mammogram	before elimination		after elimination	
	correlation	radius [pix]	correlation	radius [pix]
mdb124rm	0.63	36	0.76	72
mdb264rm	0.76	288	0.80	204

possible to increase the threshold used in the detection of cancerous masses, which in effect should reduce the number of false positive detections at a given level of true positive detections, thus improving the relation of specificity to sensitivity of the method.

5 Conclusion and further work

To reduce the number of false positive detections while maintaining the number of true positive detections in the cancerous mass detection method applied to mammograms, the elimination of linear structures was applied. The mass detection method used was the the hierarchical correlation-based detection described previously in [1]. The linear structures were detected with the accumulation-based line detector [3, 4, 5] and the elimination was performed by interpolation of the image intensity in the regions of the LSs with an anisotropic membrane being a simple derivation of the model described in [2].

The results for images representing typical detection problems caused by the LSs passing over cancerous masses, selected from the MIAS mammographic database [13], suggest the feasibility of the proposed approach. The trials with a large set of benchmark mammographic images are necessary to validate the proposed method. Preliminary results of such trials being carried out at present with the entire MIAS database are promising.

References

1. Bator M, Nieniewski M (2006) The usage of template matching and multiresolution for detecting cancerous masses in mammograms. In E. Piętka et al (eds), *Proc. 11th Int. Conf. Medical Informatics & Technology*, Wisła-Malinka, Poland, 25-27 Sept 2006, pp 324–329
2. Blake A, Zisserman A (1987) *Visual Reconstruction.* MIT Press, Cambridge, MA, London

3. Chmielewski L (2006) Detection of non-parametric lines by evidence accumulation: Finding blood vessels in mammograms. In K. Wojciechowski et al (eds) *Computer Vision and Graphics: Proc. Int. Conf. Computer Vision and Graphics ICCVG 2004*, volume 32 of *Computational Imaging and Vision*, Warsaw, Poland, 22-24 Sept, 2004. Springer, Berlin Heidelberg New York, pp 373–380

4. Chmielewski L (2005) Specification of the evidence accumulation-based line detection algorithm. In Kurzyński M., Woźniak M, Puchała E et al. (eds). *Proceedings of international conference on computer recognition systems CORES 2005*, Rydzyna, Poland, 22-25 May 2005. Volume of Advances in Soft Computing. Springer, Berlin Heidelberg New York, pp 355-362

5. Chmielewski LJ (2006) *Metody akumulacji danych w analizie obrazów cyfrowych.* Akademicka Oficyna Wydawnicza EXIT, Warszawa

6. Dziukowa E, Wesołowska E (eds) (2006) *Mammografia w diagnostyce raka sutka.* Medipage, Warszawa, second edition

7. Hong B-W, Brady M (2003) Segmentation of mammograms in topographic approach. In *Proc. IEE Int. Conf. on Visual Information Engineering VIE'03*, Guildford, U.K., July 2003

8. Kopans DB (1998) *Breast Imaging.* Lippincott-Raven, Philadelphia

9. Liu S, Babbs CF Delp EJ (2001) Multiresolution detection of spiculated lesions in digital mammograms. *IEEE Trans Image Proc* 10(6):874–884

10. Rangayyan RM, Ayres FJ (2006) Gabor filters and phase portraits for the detection of architectural distortion in mammograms. *Med Bio Eng Comput* 44(10):883–894

11. Sampat MP, Markey MK, Bovik AC (2005) Computer-aided detection and diagnosis in mammography. In Bovik AC (ed) *Handbook of Image and Video Processing.* Academic Press, pp 1195-1217

12. Sheshadri HS, Kandaswamy A (2005) Detection of breast cancer tumor based on morphological watershed algorithm. *ICGST Int J Graph Vision and Image Proc* 05(V5):17–21

13. Suckling J, Parker J et al (1994) The Mammographic Images Analysis Society digital mammogram database. In: Gale AG, Astley SM et al (eds) *Digital Mammography. Exerpta medica international congress series*, vol 1069, pp 375-378

14. te Brake GM, Karssemeijer N (1999) Single and multiscale detection of masses in digital mammograms. *IEEE Trans Med Imaging* 18(7):628–639

15. Thangavel K, Karnan M et al (2005) Automatic detection of microcalcifications in mammograms – A review. *ICGST Int J Graph, Vision and Image Proc* 05(V5):31–61

16. Zwiggelaar R, Astley SM et al (2004) Linear structures in mammographic images: detection and classification. *IEEE Trans Med Imaging* 23(9):1077–1086

17. Zwiggelaar R, Parr TC et al (1999) Model-based detection of spiculated lesions in mammograms. *Med Image Anal* 3(1):39–62