

Finding regions of interest for cancerous masses enhanced by elimination of linear structures and considerations on detection correctness measures in mammography

Marcin Bator · Leszek J. Chmielewski

Received: 29 February 2008 / Accepted: 26 May 2008 / Published online: 1 August 2008
© Springer-Verlag London Limited 2008

Abstract Cancerous mass detection methods for mammographic images still miss some malignant cases on the one hand, and produce too many false-positive (FP) detections with respect to the number of true-positive (TP) detections on the other. An attempt has been described to improve the TP ratio per image and to decrease the number of FP errors in the hierarchical template matching detector of regions of interest (ROIs) for cancerous masses by eliminating the images of linear structures (LSs) from the mammograms. The LSs were detected with an accumulation-based line detector. The measure of correctness of the ROIs detection was discussed and the quality of the detector, represented by free receiver operating characteristics curves, was compared with the human-eye observations. The result is that the widely used measure of detection correctness seems to underestimate the detection quality made by a human. Tests were performed on the mammograms from the MIAS database.

Keywords Mammograms · Cancer · Tumour · ROI · Detection · Correctness measure · Linear structures · Elimination · Accumulation line detector

1 Originality and contribution

- The commonly used measures and criteria of correctness of a single result of cancerous mass detection

versus the reference, ground truth data have been discussed. A minimum requirement has been formulated that a measure should not yield a positive result for a detector which always returns the whole image as a single detection. The cases in which the criteria found in the literature could violate this basic requirement have been pointed out.

- A trial to improve the quality of results of a template-matching hierarchical cancerous mass detection algorithm by eliminating the images of linear structures from the mammograms has been presented.
- The results without and with the elimination of the linear structures have been compared. The comparison has been carried out with two methods. First, a formally defined correctness criterion and the free receiver operating characteristics (FROC) curves were used. Then, two results chosen on the grounds of that comparison were further compared by an assessment with a human eye. The methods gave different results: the human eye assessment appeared to be more in favour of the results with LSs elimination than the assessment with the algorithmic criterion, although the eye comparison was strict. This led to a conclusion that more effort should be put to the design of the correctness criteria.
- The assessments made it possible to formulate the ranges of application of the presented algorithms.

2 Introduction

Breast cancer is an important medical and social problem. It can be estimated that in the US one out of eight women will develop the breast cancer at some point during her life [1]; the corresponding number for Poland is 1 out of 16 [2].

M. Bator · L. J. Chmielewski (✉)
Institute of Fundamental Technological Research,
Polish Academy of Sciences, Świątokrzyska 21,
00-049 Warsaw, Poland
e-mail: lchmiel@ippt.gov.pl

M. Bator
e-mail: mbator@ippt.gov.pl

Early detection of breast cancer makes it possible to apply a sparing treatment and makes the survival time longer. 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 as negative at least a part of the 90% of mammograms having no signs of pathology 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 [3], 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). However, in the automatic mammogram analysis programs existing at present, the false positive ratio per image (FPI) is usually larger than this.

In the present paper we report our attempt to improve the detection in an algorithm for detecting the regions of interest (ROIs) for cancerous masses, described in [4], by eliminating the linear structures from the mammograms. The quality of detection is considered as a high value of the true-positive detection ratio (TPR), commonly named *sensitivity*, together with a possibly low value of the false-positive detections ratio per image (FPI), which can be related to the *specificity* in that the lower the FPI, the higher the specificity of the detector. To measure the TPR and FPI ratios it is necessary to know whether a particular indication returned by the detector is true or false. We discuss the virtues and shortcomings of the commonly used measures of correctness of a single result of mass detection and indicate how the choice of this measure can improve or impair the measures of the detection quality.

Comprehensive studies of all the aspects of breast cancer imaging can be found in [5]. Reviews of cancerous masses detection and classification can be found in [1] and to some extent also in [6]. Linear structures (LSs) [7, 8], also called curvi-linear structures [9, 10], are blood vessels, milk ducts and connective tissue fibers (Cooper ligaments) in normal mammograms, and the spicules in pathologic mammograms. The LSs directly related to cancerous masses are spicules; other LSs interfere in the detection of masses. Some authors report that the classification of LSs should improve the subsequent detection and classification of the features of cancer, like in [8]. Recent publications [11, 12] suggest that the strength of the LSs in the

mammogram is strictly related to the risk of the development of cancer and can be used directly in the risk assessment analysis. Other authors, like [10], state that the attenuation of the LSs improves the results of analysis.

In this study, a mass detection algorithm using no information on the spicules is used. Moreover, in some mammograms the view of a cancerous mass is intercepted by a view of a LS which passes over or under the mass and interferes in the detection process. Therefore, we go along the line of elimination of all the LSs without their classification as an attempt to improve the detection outcome. The results are evaluated on all the mammograms from the Mammographic Image Analysis Society (MIAS) database [13].

The method of detecting the ROIs used in this paper is a multiscale template matching algorithm described earlier in [4]. Its brief explanation is repeated here for better clarity. The linear structures are detected with an accumulation-based line detector already presented in [14–17]. The method of line elimination by image interpolation is presented here.

The paper is a revised and extended version of our previous paper [18].

This paper is organised as follows. In Sect. 3, the questions related to the choice of the measure of correctness of the detection results are discussed. In Sect. 4 the ROI detector used in this study is described. In Sect. 5 the methods of detecting and eliminating the linear structures are presented. Next, the results of the detection of cancerous masses with and without the elimination of the LSs from the mammograms are described and their comparison with the results received in the way of an assessment by a human is given in Sect. 6. The final conclusions come in Sect. 7.

3 Measures of correctness of the automatic detections

In the literature on cancer detection in mammograms, as the ground truth information necessary for verifying the quality of the results of automatic detection, usually the indications made by qualified human observers are used. Such indications made by radiologists are treated as veritable evidence in at least two senses. The first is that inside the indicated region there is only the tumour, and the second is that outside this region there is only the normal tissue.

The expert's indication is relatively easy to use if the suspected region is outlined, but this is not always the case.

It is enough to show the tumour with an arrow if the observer is an experienced person. Cancerous masses were marked in such a way for example in [2]. An arrow is usually slightly outside the indicated region, not to obscure

that what is interesting, so its end does not point directly to any characteristic point of the tumour. There can be more arrows for one mass, each pointing to a different direction or parallel. Such indications are not at all appropriate for algorithmic verification.

However, even with an explicit, closed outline the situation is made complex by that the human indications are not objective. For example, in the paper [19] the relative accuracy of the indications of two experienced radiologists was tested. In the set of images containing 100 cancerous masses, the average ratio of the common areas to the sum of areas of the indications made by the radiologists was 0.76. The subjectivity of the tumour borders has at least two reasons: the image of the tumour is always superimposed on the image of the normal tissue, and the infiltration of normal tissues by the neoplasm is always irregular.

It should be always made clear which rules were applied in marking the interesting regions, for example, whether only the central mass was outlined, or together with the spicules, or even together with the surrounding area of increased lung translucence. Such rules are not always explicitly reported. In some of the available mammographic databases, free shaped regions are allowed, like in DDSM [20], or only circles, like in MIAS [13]. Any constraint imposed on the shape of an indication excludes its “pixel to pixel” accuracy.

Despite the above mentioned drawbacks of the indications made by experts there is no other source of information available to be used in checking or training an automatic detector or classifier. The regions found automatically should then be compared to the reference regions. If such comparison is done by human experts then it would have the same drawbacks as the process of forming the reference regions. Therefore, some objective measure of correctness of an automatic detection result should be introduced.

Many authors neglect to state what measure they have used, for example [21–26]. In the literature, definitions using the area of the common part of the detected region and the reference region can be found. This common part is related to the detected or reference areas themselves. This can be done in several ways, and the resulting measures, which we shall call here the *coverage factors*, can be expressed by one of the following formulae:

$$f_D = \frac{s(R \cap D)}{s(D)}, \tag{1}$$

$$f_R = \frac{s(R \cap D)}{s(R)}, \tag{2}$$

$$f_{RD} = \frac{s(R \cap D)}{s(R \cup D)}, \tag{3}$$

where R is the reference region as indicated by the radiologist, D is the region found by the detector, and $s(\cdot)$

denotes the area (in pixels). The consequences of using any of the above formulae will be discussed in the sequel.

The detection is treated as correct if the factor is larger than a threshold:

$$f > \alpha. \tag{4}$$

Several values of α were used in the literature, for example in [27–32] it was assumed $\alpha = 0.5$. Other arbitrarily chosen values were, for example, $\alpha = 0.4$ in [33], $\alpha = \frac{1}{3}$ in [34] or $\frac{1}{4}$ in [35]. In an extreme case, the detection was considered as correct if there was any nonempty common part of the detector finding and the reference region [3], so $\alpha = 0$.

The choice of the threshold α is vital and can lead to undesirable effects in the extreme cases. For example, if $\alpha = 0$ then a detector with numerical complexity $O(0)$, sensitivity 100%, and an average false positive rate per image (FPI) below one, can be easily constructed. How this can be done? It is enough for the detector to indicate the whole breast, or even the whole image as the tumour region. Let us call the detector which always returns the whole image as a single indication the *trivial detector*. It works properly as just described irrespective of which of the formulae (1)–(3) is used. There is another way to make the trivial detector work. If the formula (2) is used to define the coverage factor, then for any $\alpha \in [0, 1]$ the output from the elementary detector is always classified as a single true positive. This seems to disqualify this definition. Although in the papers [27, 29, 32] the formula has not been given explicitly, the description suggests this unfortunate definition (2).

It can be suggested that any measure of detection correctness should be checked against the trivial detector. If it is possible for the trivial detector to yield a positive result, then such a measure should be dismissed.

Let us now consider a following example illustrating the drawbacks of using the coverage factor to differentiate between the correct and incorrect detections. Assume that the radiologist has indicated a tumour with a circle and the detector returned an indication also as a circle having the same centre, but with the radius larger by a factor of at least $\sqrt{2}$. If $\alpha = 0.5$ then such an indication is classified as erroneous according to (1) as well as to (3). Although with the definition (2) it would be classified as correct, but this definition has just been dismissed. The case becomes more complex if the centres do not coincide. In the Fig. 1a an example which illustrates this problem is shown. The image is a fragment of the mammogram MDB264RM from the MIAS database in which the detection was attempted. The reference regions are white: the white circle is the reference region as found in the database and the white irregular contour has been drawn by an experienced radiologist who knew the information from the database. The

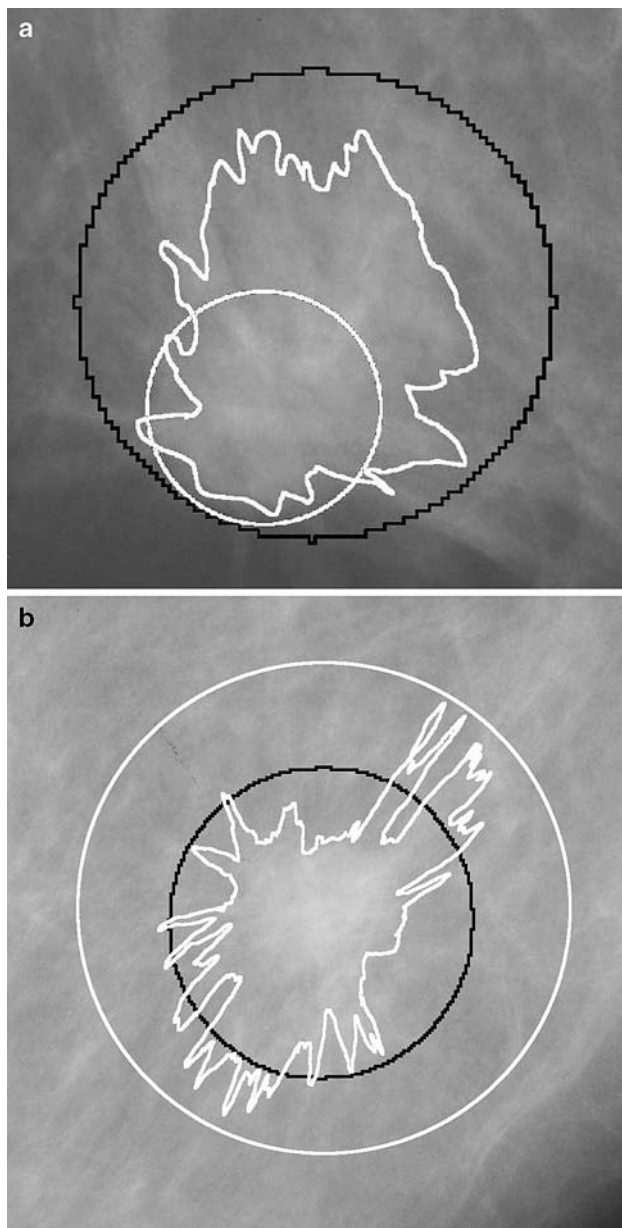


Fig. 1 Fragments of the images from the MIAS database which illustrate the problems with the coverage factors according to (1)–(3) at $\alpha = 0.5$. *White circle* reference region from the database; *white irregular contour* reference region indicated by the radiologist; *black circle* region found by the detector. **a** MDB264RM : an example of a cancerous mass tightly circumscribed by the reference contour from the database; **b** MDB155LL : an example of a mass loosely circumscribed by the reference contour from the database (see also the text.)

black circle is the result from the detector. Despite that the detector result is clearly in good conformity both with the radiologist's indication and the database reference region, it will be classified as false positive due to that less than half of its area lies inside the database reference region. It should be noted that according to the considered criterion also the radiologist's indication will be classified as false.

The solution to this problem is to reduce α , but as stated before, the vicinity of the limiting case of $\alpha = 0$ should be avoided.

It should be noted that when the definition (3) is used, then at $\alpha = 0.5$ the radiologist's indication with respect to the database reference, and vice versa, are incorrect.

Using the coverage factor f_D according to (1) has an important advantage when the reference indications are generated inconsequently as far as their tightness around the cancerous mass is considered, like in the MIAS database. In the just considered contour in Fig. 1a the mass was circumscribed very tightly, while in the Fig. 1b there is a loosely circumscribed mass. It is better then, that in the denominator of the coverage coefficient there is the detected area, not the reference one.

A further advantage of using the coverage factor as defined with (1) is its suitability to the assessment of the detectors which yield the results of thresholding a map of similarity of the region to an object sought. Examples of such an approach can be found in [27, 32, 36]. The detected region decreases together with the increase of the threshold and can become a one-pixel indication.

Another approach is to use as an indication only the centre of the detected region and to neglect its dimensions. However, even with this approach the detection result shown in Fig. 1a is considered as erroneous.

In the two above approaches, the indication quality is the same irrespective whether it is located near the middle or near the border of the reference region. Also, inside one reference region more than one detection can appear. In the literature the ways of treating such multiple indications are different, sometimes even in one paper, like in [33].

Bearing in mind all the above critical remarks we have used in this paper the quality measure of the detection results using the coverage factor f_D , according to (1), (4), with the most commonly used value of $\alpha = 0.5$.

4 Detection of regions of interest for cancerous masses

The detector used was a multiscale template matching algorithm [4]. This was a single-feature detector which yielded the results characterised by a very high sensitivity (over 0.90), but with a considerable number of false-positive detections (over seven per image), as it will be seen in further sections. Therefore, this detector could be successfully used as a ROI detector for further stages of any cancerous mass detection and classification algorithm, due to that it is clear in structure, efficient in calculation, and makes it possible to reduce to a large extent the area of the image which should be further processed.

In the sequel this detector of the ROIs will be simply referred to as the detector.

A template is an auxiliary image which represents a model brightness distribution in a mass. A similarity measure between the template and the corresponding window of the image is calculated while the template is shifted across the image. As the similarity measure the correlation coefficient was used, as expressed by the equation

$$c(T, I) = \frac{\sum_{j=1}^N (t_j - \bar{t})(i_j - \bar{i})}{\sqrt{\sum_{j=1}^N (t_j - \bar{t})^2} \sqrt{\sum_{j=1}^N (i_j - \bar{i})^2}}, \quad (5)$$

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 $(-1,1)$ and depends only on the relation of shapes of the template and the window, not on their grey scale. Negative values would be obtained for the template being equal or close to a negative of the image. A template with a hemispherical brightness was used:

$$t_R(x, y) = \sqrt{R^2 - x^2 - y^2} \quad \text{for} \quad x^2 + y^2 \leq R^2, \quad (6)$$

where x , y are pixel coordinates in a local coordinate system of the template having origin at the template centre. The basic rationale for using a sphere as the template was that in an unconstrained environment a tumour would develop into a sphere, due to that it can be safely assumed that in the absence of obstacles it would develop isotropically. A wider discussion on the spherical shape of the template can be found in [37, Section III.A].

As in [4], the Fast Fourier Transform was used and the correlation was calculated in the Fourier domain, while unlike in [4], where the domain of the template was complemented into a square circumscribed on the great circle of the hemisphere, in the present paper the template was circular.

As advised in [37] and according to the previous experience [4], the analysis was carried out in a multiscale setting. 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 and 43–60 pixels, respectively. The multiscale approach was designed according to the observation made in [27] that instead of performing a detection of an object having two times larger dimensions 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 under MATLAB.

At the image resolution of 50 $\mu\text{m}/\text{pix}$, 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 the same threshold and aggregating the partial results received onto the full scale level 1:1. In this way the regions having arbitrary shapes could be received. As a single detection an eight-connected set of pixels was considered. For the presentation in the images however, each detection for a given radius of the template (recalculated to the full resolution in the pyramid) was dilated with a disk of the same radius. This was done solely for the presentation purposes, to show the possible extent of the indicated ROI for a tumour (as it will be seen in the Sect. 3), while in the formulae for checking the quality of detection (1)–(4), the detections without dilation were taken.

5 Linear structures

5.1 Detection

The method used for detection of the linear structures was the evidence accumulation-based line detection algorithm already described in sufficient detail in [15, 17], and previously in [14]. A brief description can be found also in [16], Sect. 6.2. In the present application, linear structures of width between 2 and 16 pixels, related to the scale 1:4, were sought (the question of choosing this scale will be further explained in Sect. 5.2.4). This was enough to capture all the LSs visible in the mammograms. Structures having the line intensity maxima smaller than 0.3 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 [15], p. 360 or [17], 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, due to that the line width extends beyond the loci of maximum gradient. Therefore, the width found was multiplied by a factor f_w slightly larger than one: $w = f_w w_0$, $f_w = 1.25$. This is explained in Fig. 2a. Note that if the profile of the line were sinusoidal, the factor f_w should be equal to two. The profile of a typical blood vessel is much different from sinusoidal, so a value much nearer to one was chosen.

The primary results of the detection 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.

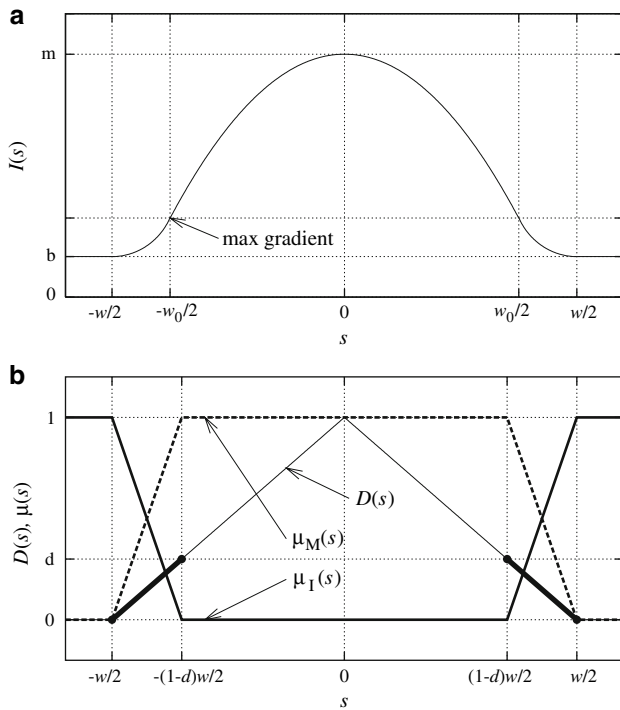


Fig. 2 In both subimages, s is the variable running across the line and w is the actual width of the LS. **a** Extending the width of a linear structure in the mammogram. w_0 line width found by the line detection algorithm; b , m local background and maximum image intensity, respectively. **b** Weighting functions for insertion of the membrane into the image. $D(s)$ relative distance transform (*thin solid line*), $\mu_I(s)$, $\mu_M(s)$ weighting functions of the image and membrane (*medium solid line* and *dashed line*), respectively, d dimensionless width of the transient stripe between the image and the membrane, measured in the terms of the distance transform. *Thick lines* parts of the $D(s)$ graph corresponding to the transient stripe

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.

5.2 Elimination

5.2.1 The basic approach

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

$$I(x, y) = I_0(x, y) + I_{LS}(x, y), \tag{7}$$

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 [10] 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 [38], 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. 3a, until the result stabilises. The intermediate calculations were performed in the floating-point arithmetic. The method was similar to that used in [10] in that the convolution equivalent to averaging was used, while the main difference was that the convolution was repeated until the surface received stabilised.

5.2.2 Directional analysis

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. 4a–c. 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

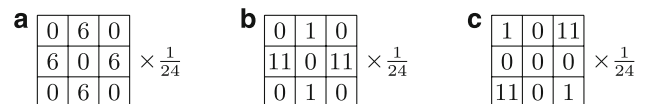
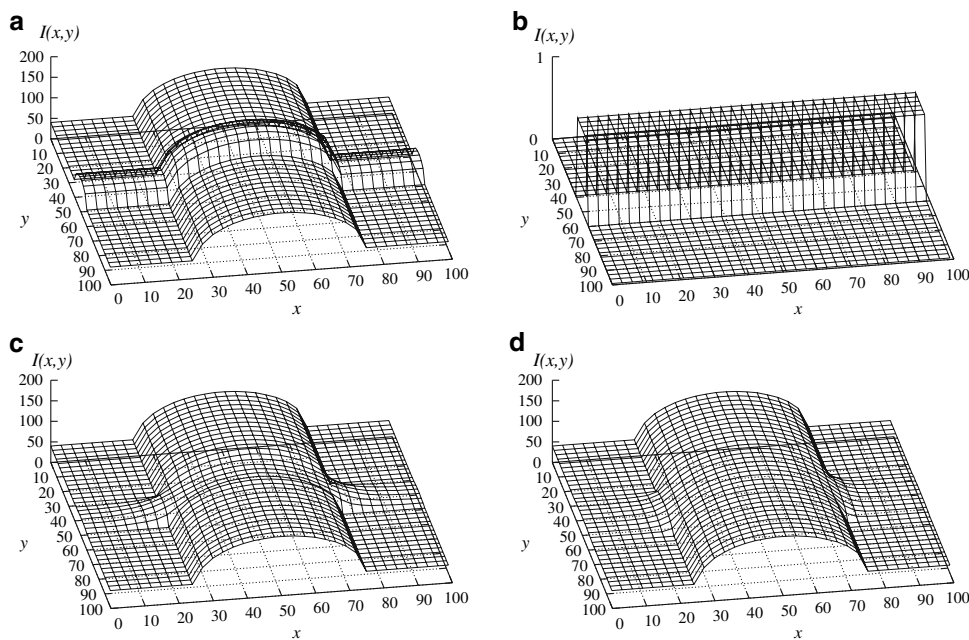


Fig. 3 Examples of masks used for interpolation in the membrane spanning algorithm. The sum of elements is divisible by four to make it easier to construct the isotropic mask. **a** Isotropic membrane; **b** membrane stronger in the W–E direction; **c** membrane stronger in the SW–NE direction

Fig. 4 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** binary mask of the detected LS; **c** LS eliminated with the isotropic membrane: saddles present in the ridges and bottoms of the intensity graph; **(d)** 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 a convolution mask which expressed the anisotropic membrane stiffness was used. The examples of the masks are shown in Fig. 3b and c and the result of using this approach can be seen in Fig. 4d. As it can be seen in the masks, 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.

5.2.3 Improving the continuity

The membrane maintains C_0 continuity with the image intensity function, but due to the multiple averaging process it is much more “smooth” (in the commonsense meaning) than that function. To avoid abrupt changes of smoothness of the resulting intensity function, a superposition in a transient stripe was used. Let us start with the formulation without the superposition:

$$I(x, y) = I_M(x, y)b_{LS}(x, y) + I_0(x, y)b_I(x, y), \tag{8}$$

where I_M is the intensity found with the membrane, b_{LS} is the binary mask of the LSs, $b_I(x, y) = 1 - b_{LS}(x, y)$ is the binary mask of the image out of the LS regions, and the remaining notations are as described before. Now, let us superimpose the intensity functions I_M and I_0 so that they pass from one to another gradually:

$$I(x, y) = I_M(x, y)\mu_{LS}(x, y) + I_0(x, y)\mu_I(x, y), \tag{9}$$

where μ_{LS} , μ_I are real weighting functions such that in a narrow transient stripe belonging to the LS and neighbouring its boundary, they pass from one to another gradually (in the simplest case, linearly), as shown in Fig. 2b. These weighting functions can be found in a

straightforward way from the relative distance transform of the LSs found in the line detection phase. The width of one transient stripe can be calculated as a fraction of the line width: $w_t = f_t w$. On the other hand, given the relative distance transform $D(s)$ of the line, the extent of the transient stripe can be found from this transform by specifying the limit d for the transform: $w_t = d w/2$. There are two stripes, one at each side of the line. In the presented calculations the transient stripe extended through the part of the LS width added to compensate for the too small width of the lines originally found by the detector, as shown in Fig. 2a, that is, between w_0 and w . Assuming that $f_w = 1.25$ it can be calculated that $f_t = 0.1$.

5.2.4 Multiscale analysis

According to the multiscale organisation of the mass detector used, the line elimination process was also performed in four scales. However, the line detection was done only in the scale 1:4 (the second least detailed scale) due to that this scale was fine enough to make the LSs detection accurate, but still coarse enough to keep this detection time-efficient. Its results were recalculated to the remaining three scales. This was done for each line by appropriately deleting pixels of the skeleton belonging to each second image line and column, or inserting interpolated pixels into the skeleton of the detected line, and plotting the line by creating its width in each such obtained skeleton pixel. The line elimination was performed with the primary image as the initial value at the lowest scale level and with the value obtained at the previous scale as the initial value at the remaining three scales.

6 Results and discussion

6.1 The FROC graphs

As described in our previous paper [18], the first observation indicated that the elimination of the linear structures from the mammograms improved the results of detection in a number of cases, especially when the tumour was occluded by a strong view of a blood vessel or milk duct. However now the feasibility of the method has been tested on the whole MIAS mammographic database. First, the cancerous masses were detected in the original 321 images from this base with the detector described in Sect. 4. Then, the linear structures were detected and eliminated, as described in Sect. 5 and the masses were detected one more time. Thus, two series of detection results were obtained. Both series of the results were evaluated against the reference data from the MIAS database as described above in Sect. 3 for a series of thresholds of the cancerous mass detector, and overall characteristics of the results were aggregated in the FROC curves [1, 34, 39]. The quantity on the abscissa is the ratio of false positive detections per image, FPI. On the ordinate, there is the true positive ratio, TPR. The graphs are shown in Fig. 5. The following observations can be made. First, the elimination of LSs did not yield a general improvement in the performance of the detector. Second, the only range where the results with line elimination are better, is the range of large FPI rates for the malignant tumours. The TPI value approaches 0.95 in this range (namely, it is 0.9487 for $FPI \geq 12.77$).

This second observation can suggest that the elimination of LSs could improve the sensitivity of the correlation-based detector, so this detector could be used more successfully as a preprocessor which finds the regions of interest for other, more computationally demanding classifiers, without producing a significant number of omissions of malignant tumours. This is possible in spite of that as a stand-alone classifier the considered simple and efficient detector yields too many false positive results.

6.2 Assessment by a human eye

The first one of the above described observations evoked our doubts towards the encouraging preliminary results reported in [18]. However, bearing in mind the discussion from Sect. 3, we decided to verify this by carrying out a review and classification of the results for the database considered by a human eye. The results shown in the FROC graphs in Fig. 5 were obtained from a number of series of the images taken at different thresholds. In the detection we are interested primarily in the malignant tumours, so for the review we have chosen two series which should have very similar results as far as the

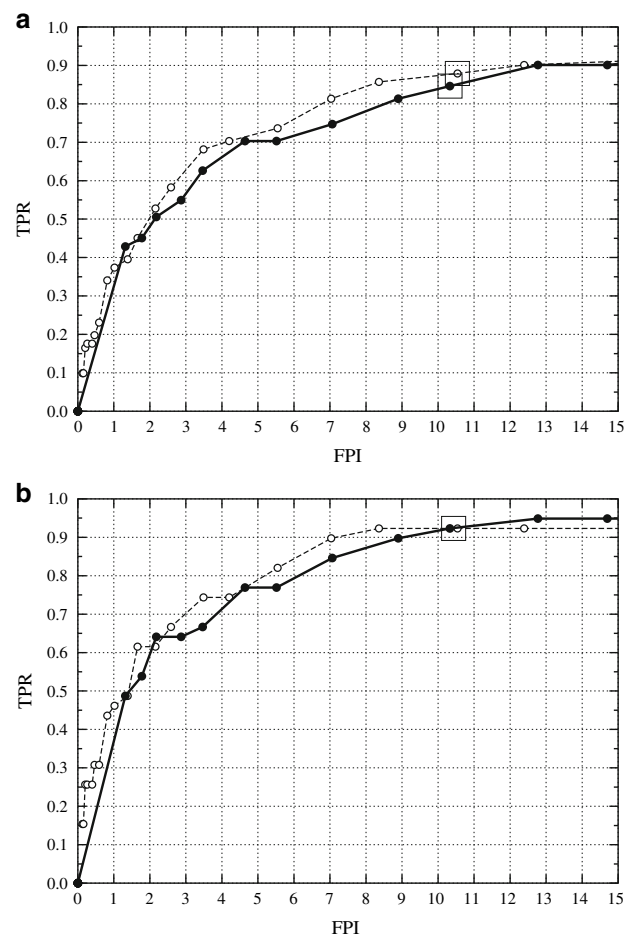


Fig. 5 Free receiver operating characteristics for the correlation-based cancerous mass detection for the whole MIAS database: **a** all tumours; **b** malignant tumours. *Solid lines, full circles* detection results for the original images; *dashed lines, empty circles* detection results for the images with linear structures eliminated. *Large squares* Series of images chosen for the assessment by a human eye described in Sect. 2

malignant tumours are considered. These results have been marked in the FROC graph for the malignant cases in Fig. 5b with a large square, while in the graph for all the tumours in Fig. 5a the results are less similar and have been marked with two squares. For malignant tumours the series have nearly equal FPI measures and should have equal overall numbers of true-positive detections, while for all the tumours the number of true-positive detections should be lower in the case of eliminated LSs than in that with no elimination. The pairs of images compared were the images after detection, without and with the LSs elimination, with marked detection results and reference regions from the database and from the radiologist. The degree to which each result has been changed by the LSs elimination was assessed. The following grades were given to each pair: *strong improvement, medium improvement, weak improvement; strong deterioration, medium*

deterioration, weak deterioration, and three types of no change in the quality of detection: *good to good*, *weak to weak* and finally *bad to bad*. The task was relatively easy so we considered the review made by ourselves as sufficient. The detection of cancerous masses was assessed, so 25 images containing calcifications were not taken into account. In this way, the analysis was performed on $321 - 25 = 296$ images, of which 91 contained the symptoms of abnormalities of various types and 205 were normal (see [13] or <http://peipa.essex.ac.uk/info/mias.html>). The normal images were not taken into account in the assessment by a human eye, as by definition they contained only the false-positive detections, which were approximately equally numerous in both series, as it can be seen in the graphs in Fig. 5. The results of the assessment are gathered in the Table 1.

Examples of pairs of the images which received some of the grades used will be presented in the following subsection to give an insight into how the grades were assigned. We tended to be rather critical than permissive in qualifying the improvements.

The results in Table 1 show that in the results with the elimination of the linear structures, the improvements tend to prevail over the deteriorations. The numbers characterising the improvements are small, but also the numbers of the images in which the differences could be seen are not large. The low numbers make it unreasonable to calculate any statistics on such data, but it can be noticed that the differences in the FROC graphs are also grounded upon the numbers of the same order of magnitude. Nevertheless, if an attempt of finding out whether the improvement was larger in the malignant cases alone or in all the cases together, then a following calculation could be done: the

prevalence of the improvements over deteriorations versus all the image pairs with differences was $(6 - 2)/(6 + 2) = 50\%$ in malignant cases and $(17 - 8)/(17 + 8) = 36\%$ in all the cases. If these prevalences were related to the totals of image pairs, the numbers would be $(6 - 2)/39 = 10.3\%$ and $(17 - 8)/91 = 9.9\%$, respectively. These results seem to be above rather than below the negligibility level.

The meaning of the presented numbers is contradictory with respect to what can be seen in the FROC graphs for the image series considered. In the human eye assessment, the results with the LSs eliminated are better than those without elimination, and this difference is similar in both the malignant cases and in the malignant and benign cases merged together.

The conclusions will be given in Sect. 7.

6.3 Examples of images used in the assessment

The pairs of images presented in this Section have been selected from all the images used in the FROC analysis and the human-eye assessment. Typical images showing

Table 1 Comparison of the results of cancerous mass detection in pairs of images: without and with the elimination of LSs

Malignant Σ39				Benign Σ52				All Σ91			
↑↑↑	↑↑	↑	Σ	↑↑↑	↑↑	↑	Σ	↑↑↑	↑↑	↑	Σ
2	1	3	6	5	4	2	11	7	5	5	17
↓↓↓	↓↓	↓	Σ	↓↓↓	↓↓	↓	Σ	↓↓↓	↓↓	↓	Σ
2	0	0	2	2	0	4	6	4	0	4	8
Good → Good				Good → Good				Good → Good			
29				32				61			
Weak → Weak				Weak → Weak				Weak → Weak			
0				1				1			
Bad → Bad				Bad → Bad				Bad → Bad			
2				2				4			

Each pair was graded as belonging to one of the nine classes: ↑↑↑: strong improvement, ↑↑: medium improvement, ↑: weak improvement; ↓↓↓: strong deterioration, ↓↓: medium deterioration, ↓: weak deterioration; and three classes with no change in the quality of detection: *good to good*, *weak to weak*, *bad to bad*

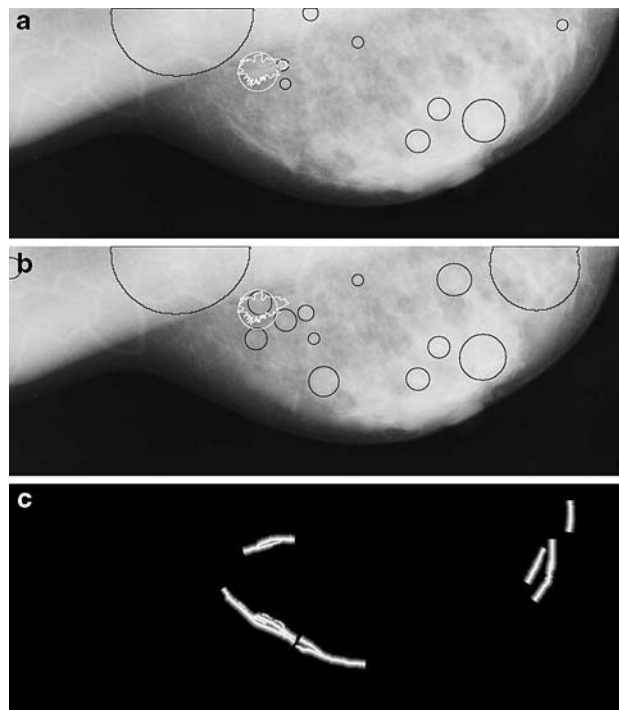
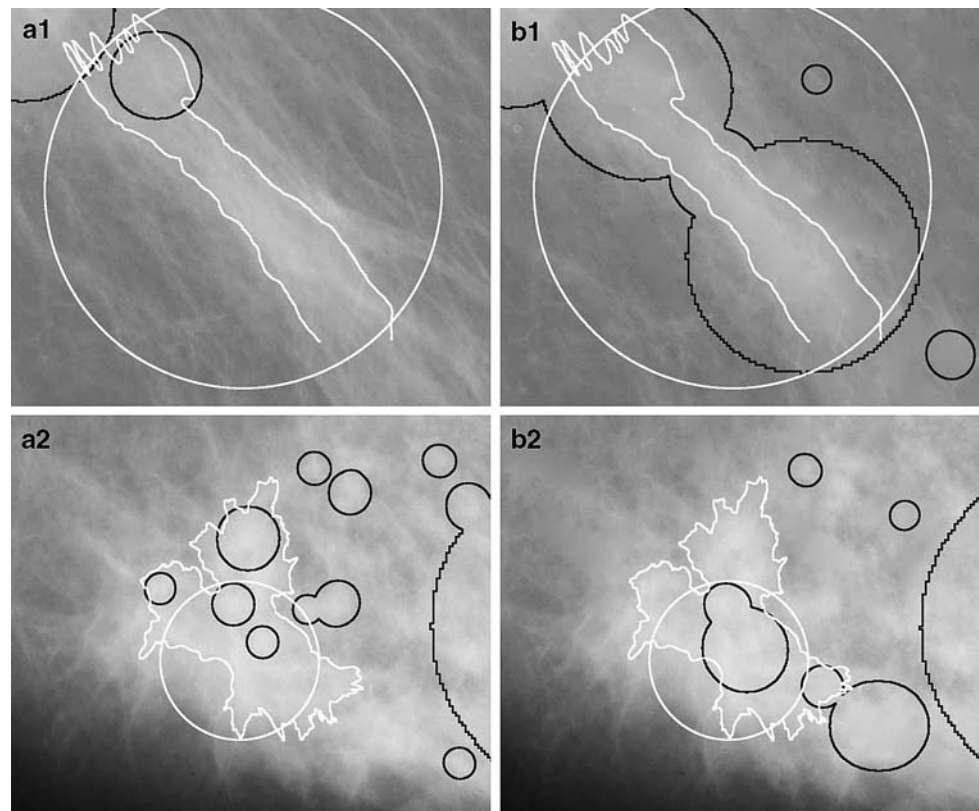


Fig. 6 Examples of results for mammogram MDB124RM from the MIAS database, malignant case, strong improvement of the detection result: **a** in the original image; **b** in the image with LSs eliminated; **c** relative distance transforms of the LSs. *White circle* reference region from the database; *white irregular contour* reference region indicated by the radiologist; *black circles* regions found by the detector. The lower, black part of the image has not been shown

Fig. 7 Examples of results for fragments of mammograms from the MIAS database, malignant cases: **a** original images; **b** images with LSs eliminated; **1** MDB274RX: strong improvement; **2** MDB090RM: medium improvement. *White circle*: reference region from the database; *white irregular contour*: reference region indicated by the radiologist; *black circles*: regions found by the detector



differences between the detection results with and without the LSs elimination were chosen.

In Fig. 6 an example of the results for a *strong improvement* for a malignant case is shown. A map of the detected linear structures is added to this pair of images to make it easier to see where are the differences between the images. In Fig. 7 there is an example of one *strong* and one *medium improvement* for a malignant case. Examples of a *strong deterioration* in malignant cases is shown in Fig. 8. Further, the benign cases are displayed: a *strong improvements* in Fig. 9, and a *medium improvement* and a *strong deterioration* in Fig. 10. The next Fig. 11 needs an explanation. The image is problematic due to that the radiologist's indication and the ground truth data from the database do not conform. The pair of images has been graded as *strong deterioration* due to that a small indication by the detector located in the centre of the database reference region disappeared after LSs elimination. In fact, according to the README file of the MIAS database, this mammogram shows an *architectural distortion* of a breast, not a cancerous mass, so it could be excluded from the analysis as well.

The examples of no change of the quality of detection do not have to be shown.

The locations and dimensions of windows presented in Figs. 6–10 have been collected in Table 2.

7 Conclusion

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 hierarchical correlation-based detection described previously in [4]. The linear structures (LSs) were detected with the accumulation-based line detector [14, 15, 17] 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 [38].

The results were assessed with a commonly used measure of correctness of a single result of detection with respect to a reference indication using the notion of the coverage factor (1), (4). The FROC graphs were used to this end. The results with and without the elimination of LSs which should appear similar according to the FROC analysis were assessed with a human eye. The assessment were made for all the relevant images from the MIAS database (i.e., only the images containing the calcifications were not considered in the human-eye assessment).

The conclusions which can be drawn from the presented considerations, mainly those given in Sects. 3 and 6, are as follows.

Fig. 8 Examples of results for fragments of mammograms from the MIAS database, malignant cases: **a** original images; **b** images with LSs eliminated; **1** MDB125LL: strong deterioration; **2** MDB141LX: strong deterioration. *White circle* reference region from the database; *white irregular contour* reference region indicated by the radiologist; *black circles* regions found by the detector

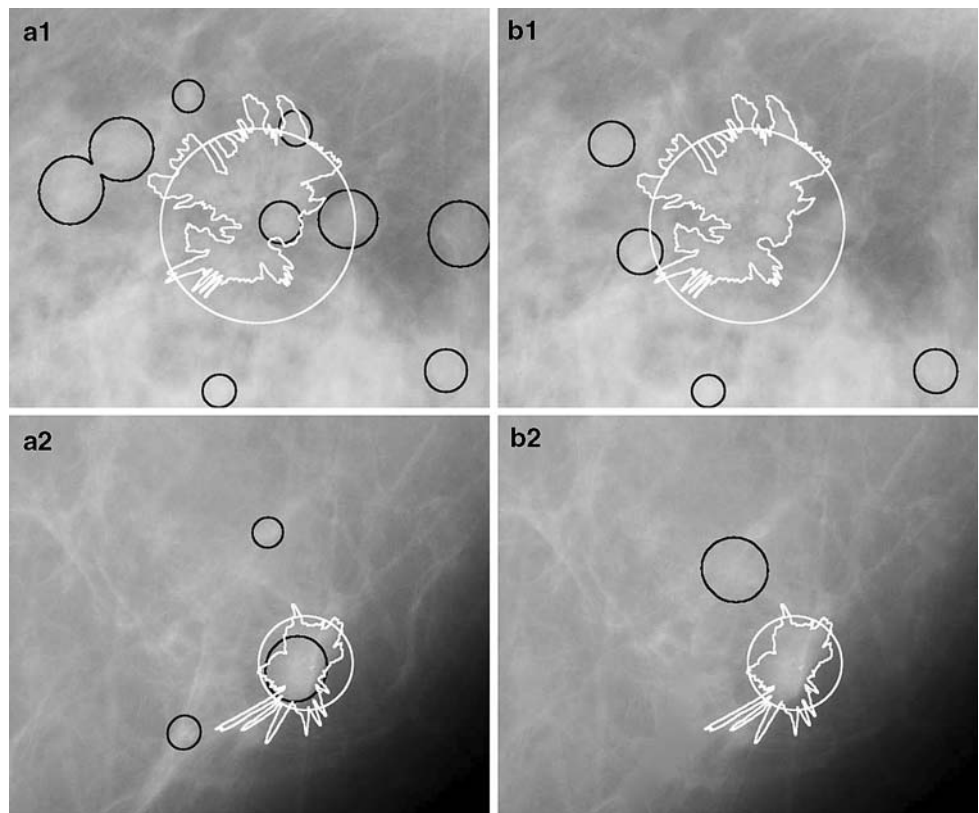


Fig. 9 Examples of results for fragments of mammograms from the MIAS database, benign cases: **a** original images; **b** images with LSs eliminated; **1** MDB148RX: strong improvement; **2** MDB175LM: strong improvement. *White circle* reference region from the database; *white irregular contour* reference region indicated by the radiologist; *black circles* regions found by the detector

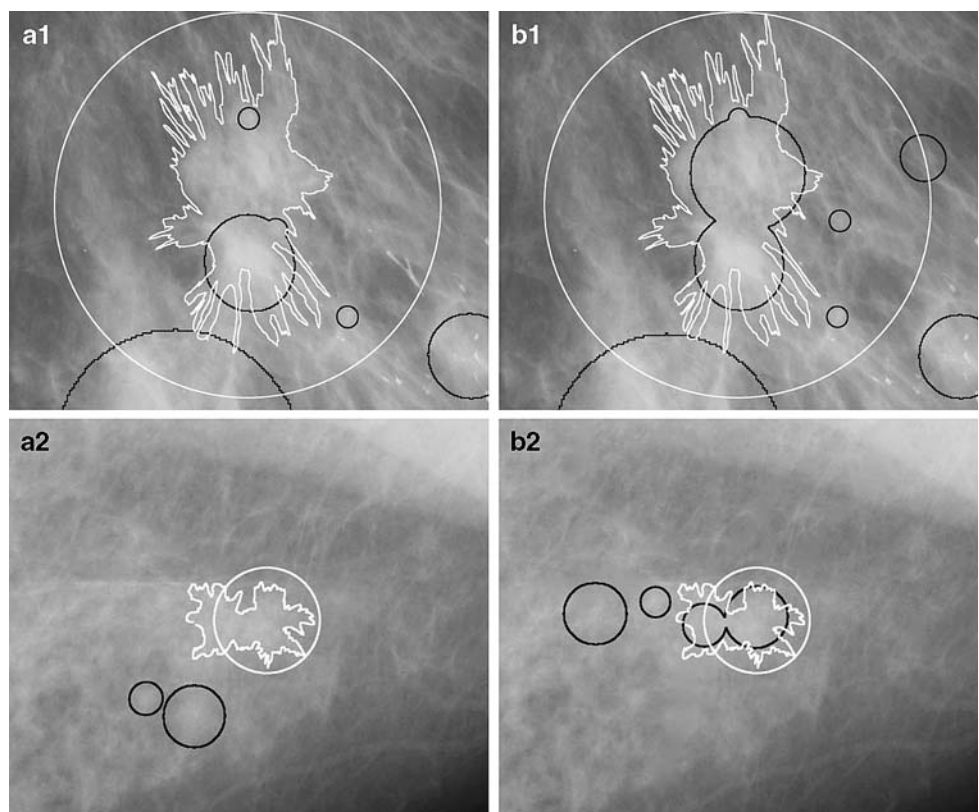
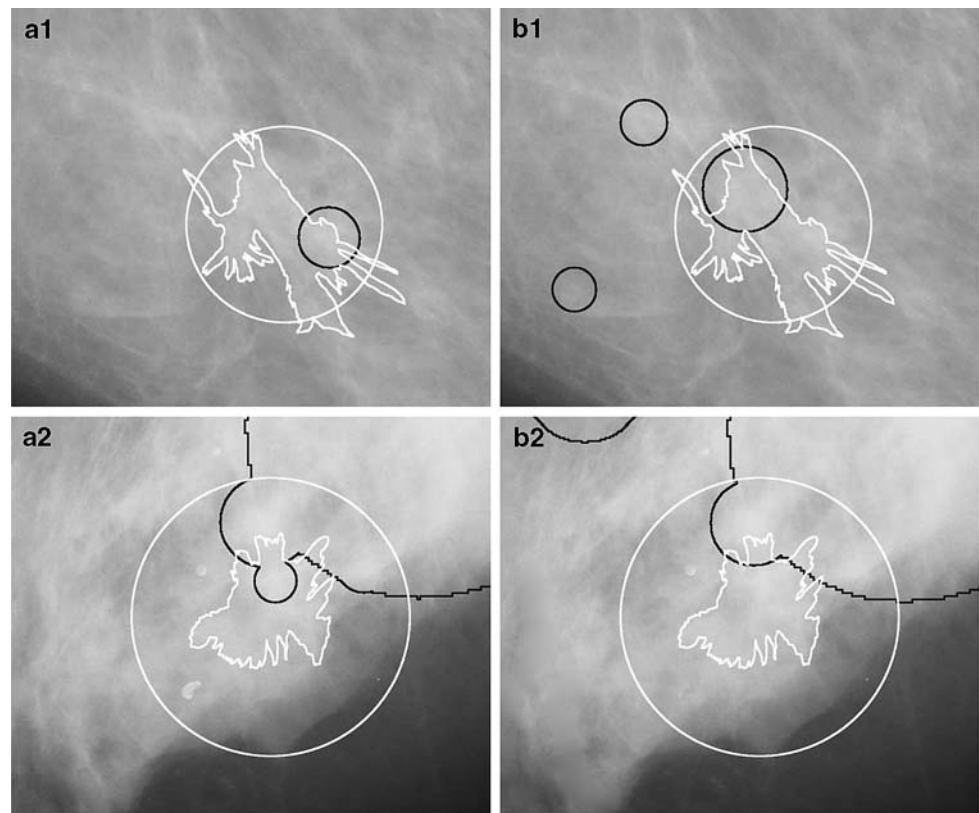


Fig. 10 Examples of results for fragments of mammograms from the MIAS database, benign cases: **a** original images; **b** images with LSs eliminated; **1** MDB16ORL: medium improvement; **2** MDB121LL: strong deterioration. *White circle* reference region from the database; *white irregular contour* reference region indicated by the radiologist; *black circles* regions found by the detector



The results of the human eye assessment are not in good conformity with those obtained with the measure using the coverage factor. The human eye assessment indicated that the impact of the LSs elimination on the result is positive, while this did not result from the coverage factor measure. A possible reason for this fact is that the human eye captures the similarities in the location and size of irregular shapes better than simple formulae. The design of the measures using the coverage factor strongly depend on the definition of this factor and the threshold used, which can make the results impossible to compare or even unreasonably optimistic. The strict definition of the measure of results quality should always be given, which is not the case in some publications. This measure should at least be checked for its reasonability against what has been called the *trivial detector* which is the detector that always returns the whole image as a single indication. The measure used should not yield a positive result for this unreasonably simple detector. Finally, a suggestion can be made that more effort is necessary to develop a better measure of correctness of the indications of detectors versus the reference, ground truth data.

The elimination of the linear structures from the mammograms improved the results of cancerous mass detection

to some extent. This improvement is visible in the human eye assessment at least at the false positive per image (FPI) index around 10.5. In the assessment with the measure (1) and (4), it is visible only for the malignant tumours, in the range of large values of the false positive per image ratio: $FPI > 13$. In this range the value of the true positive ratio is $FPI \approx 0.95$.

The presented template-matching cancerous mass detector is clear in structure, simple in implementation and efficient in calculation. The hitherto experience with this single-feature detector is that it is capable of significantly reducing the area of the mammogram which should be processed further while missing very few suspected regions and keeping a very high sensitivity of 0.90–0.95. Therefore, it can be considered as a good candidate as a front-end detector of the regions of interest (ROI) for any complex cancerous mass detection and classification system.

Acknowledgments The research was partly financed by the Ministry of Education and Science, Poland, as the research project No 3 T11C 050 29 in 2005–2008. We thank Dr Ewa Wesolowska from the Center of Oncology, Warsaw, for providing the contours indicating the shapes of cancerous masses shown in Figs. 1, 6–11.

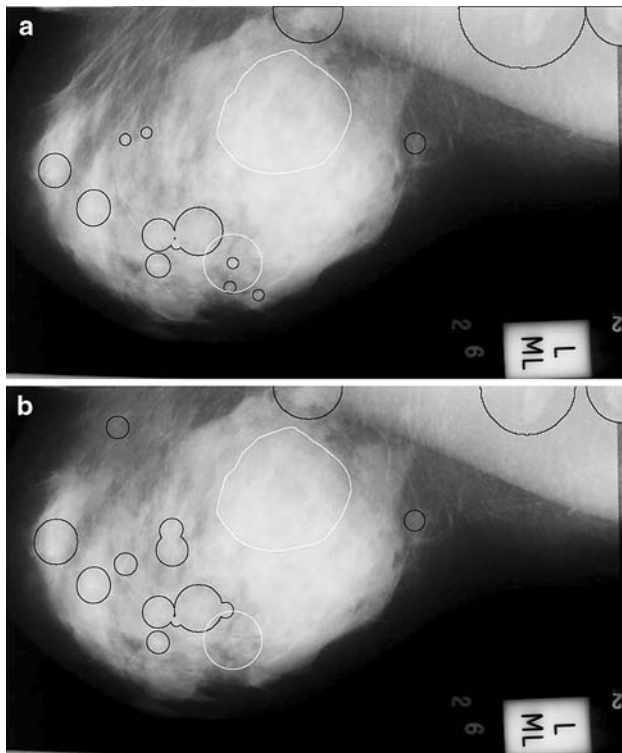


Fig. 11 Examples of results for mammogram MDBMDB163LL from the MIAS database, benign case; the image is problematic due to that the radiologist and the database ground truth data do not agree. The case has been graded as a strong deterioration. **a** Original image; **b** image with LSs eliminated. *White circle* reference region from the database; *white irregular contour* reference region indicated by the radiologist; *black circles* regions found by the detector

Table 2 Locations and dimensions of windows of the images from the MIAS database shown in Figs. 6–10

Fig.	MIAS index	Upper left corner	Dimensions
6	MDB124RM	0, 0	4,320 × 1,600
7(1)	MDB274RX	2,050, 0	1,200 × 1,000
7(2)	MDB090RM	1450, 400	1,200 × 1,000
8(1)	MDB125LL	1,700, 0	1,200 × 1,000
8(2)	MDB141LX	2,850, 1,550	1,200 × 1,000
9(1)	MDB148RX	1,355, 550	1,740 × 1450
9(2)	MDB175LM	2,150, 200	1,200 × 1,000
10(1)	MDB160RL	1,450, 850	1,200 × 1,000
10(2)	MDB121LL	1,200, 880	1,200 × 1,000

Numbers in pairs are *x* (column), *y* (row)

References

1. 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, New York, pp 1195–1217
2. Dziukowa J, Wesolowska E (eds) (2006) Mammography in breast cancer diagnosis (in Polish), 2nd edn. Medipage, Warszawa

3. Zwiggelaar R, Parr TC, Schumm JE et al (1999) Model-based detection of spiculated lesions in mammograms. *Med Image Anal* 3(1):39–62
4. Bator M, Nieniewski M (2006) The usage of template matching and multiresolution for detecting cancerous masses in mammograms. In: Pięka E et al (eds) Proceedings of 11th international conference on medical informatics and technologies, Wisła-Malinka, Poland, pp 324–329
5. Kopans DB (1998) Breast imaging. Lippincott-Raven, Philadelphia
6. Thangavel K, Karnan M, Sivakumar R, Mohideen AK (2005) Automatic detection of microcalcifications in mammograms—a review. *ICGST Int J Graph Vis Image Proc* 05(V5):31–61
7. Hong B-W, Brady M (2003) Segmentation of mammograms in topographic approach. In: Proceedings of IEEE international conference on visual information engineering VIE’03, Guildford
8. Zwiggelaar R, Astley SM, Boggis CRM, Taylor CJ (2004) Linear structures in mammographic images: detection and classification. *IEEE Trans Med Imaging* 23(9):1077–1086. doi:10.1109/TMI.2004.828675
9. 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. doi:10.1007/s11517-006-0088-3
10. Sheshadri HS, Kandaswamy A (2005) Detection of breast cancer tumor based on morphological watershed algorithm. *ICGST Int J Graph Vis Image Proc* 05(V5):17–21
11. Hadley EM, Denton ERE, Zwiggelaar R (2006) Mammographic risk assessment based on anatomical linear structures. In: Astley SM et al (eds) Proceedings of 8th international workshop digital mammography IWDM 2006, vol 4046 of LNCS, Manchester, Springer, Berlin, p 626–633. doi:10.1007/11783237_84
12. Hadley EM, Denton ERE, Zwiggelaar R et al (2007) Risk classification of mammograms using anatomical linear structure and density information. In: Proceedings of 3rd Iberian conference on pattern recognition and image analysis IbPRIA 2007, vol 4478 of LNCS, Girona, Springer, Berlin, pp 186–193. doi:10.1007/978-3-540-72849-8_24
13. Suckling J, Parker J, Dance D et al (1994) The mammographic images analysis society digital mammogram database. In: Gale AG, Astley SM, Dance DR et al (eds) Digital mammography, vol 1069 of Exerpta Medica International Congress Series, pp 375–378. <http://www.wiau.man.ac.uk/services/MIAS/MIASweb.html>; later moved to peipa.essex.ac.uk/info/mias.html
14. Chmielewski L (2006) Detection of non-parametric lines by evidence accumulation: finding blood vessels in mammograms. In: Proceedings of international conference on computer vision and graphics ICCVG 2004, Warsaw, Poland, Sep 22–24, 2004, vol 32 of computational imaging and vision, Springer, Berlin, pp 373–380. doi:10.1007/1-4020-4179-9_54
15. Chmielewski L (2005) Specification of the evidence accumulation-based line detection algorithm. In: Kurzyński M et al (eds) Computer recognition systems. Proceedings of international conference on computer recognition systems CORES 2005, vol 30 of Advances in Soft Computing, Rydzyna, Poland, 22–25 May 2005. Springer, Berlin, pp 355–362
16. Chmielewski LJ (2006) Fuzzy histograms, weak fuzzification and accumulation of periodic quantities. Application in two accumulation-based image processing methods. *Patt Anal Appl* 9(2–3):189–210. doi:10.1007/s10044-006-0037-7
17. Chmielewski LJ (2006) Evidence accumulation methods in digital image processing (in Polish). Akademia Oficyna Wydawnicza EXIT, Warsaw. <http://www.ippt.gov.pl/~lchmiel/akum06/>
18. Bator M, Chmielewski LJ (2007) Elimination of linear structures as an attempt to improve the specificity of cancerous mass

- detection in mammograms. In: Kurzyński M et al (eds) *Computer Recognition Systems 2. Proceedings of international conference on computer recognition systems CORES 2007*, vol 45 of *Advances in Soft Computing*, Wrocław, 22–25 Oct 2007, Springer, Berlin, pp 596–603. doi:10.1007/978-3-540-75175-5_75
19. Sahiner B, Petrick N et al (2001) Computer-aided characterization of mammographic masses: accuracy of mass segmentation and its effects on characterization. *IEEE Trans Med Imaging* 20(12):1275–1284
 20. Heath MK, Bowyer D, Kopans R et al (2000) The digital data base for screening mammography. In: *Proceedings of 5th international workshop digital mammography*, Toronto, pp 212–218. <http://figment.csee.usf.edu/Mammography/DDSMM>
 21. Bovis K, Singh S (2000) Detection of masses in mammograms using texture features. In: *Proceedings of 15th international conference on pattern recognition (ICPR'00)*, vol 2, pp 2267–2270. doi:10.1109/ICPR.2000.906064
 22. Cahoon TC, Suttah MA, Bezdak JC (2000) Breast cancer detection using image processing techniques. In: *Proceedings of 9th IEEE international conference on Fuzzy Systems 2000 FUZZ IEEE 2000*, vol 2, pp 973–976
 23. Li HD, Kallergi M, Clarke LP, Jain VK, Clark RA (1995) Markov random fields for tumor detection in digital mammography. *IEEE Trans Med Imaging* 14(3):565–576
 24. Kobatake H, Murakami M, Takeo H, Nawano S (1999) Computerized detection of malignant tumors on digital mammograms. *IEEE Trans Med Imaging* 18(5):369–378
 25. Zheng L, Chan AK (2001) An artificial intelligent algorithm for tumor detection in screening mammogram. *IEEE Trans Med Imaging* 20(7):559–567
 26. Zheng L, Chan AK, McCord G et al (1999) Detection of cancerous masses for screening mammography using DWT based multiresolution Markov random fields. Technical report, Texas A&M University
 27. Liu S, Babbs CF, Delp EJ (2001) Multiresolution detection of spiculated lesions in digital mammograms. *IEEE Trans Image Proc* 10(6):874–884
 28. Christoyianni I, Dermatas E, Kokkinakis G (2000) Fast detection of masses in computer-aided mammography. *IEEE Signal Proc Mag* 17:54–64
 29. Kegelmeyer WP, Pruneda JM, Bourland PD et al (1994) Computer-aided mammographic screening for spiculated lesions. *Radiology* 191:331–337
 30. Lai S, Li X, Bishof WF (1989) On techniques for detecting circumscribed masses in mammograms. *IEEE Trans Med Imaging* 8(4):377–386
 31. Li H, Wang Y, Ray Liu KJ et al (2001) Computerized radiographic mass detection—part II: decision support by featured database visualization and modular neural networks. *IEEE Trans Med Imaging* 20(4):302–313
 32. Liu S (1999) The analysis of digital mammograms: spiculated tumor detection and normal mammogram characterization. Thesis, Purdue University
 33. Heath M, Bowyer K (2000) Mass detection by relative image intensity. In: *Proceedings of 5th international workshop on digital mammography*, Toronto, Medical Physics Publishing, Madison, pp 219–225
 34. Bornefalk H, Bornefalk Hermansson A (2005) On the comparison of FROC curves in mammography CAD systems. *Med Phys* 32(2):412–417. doi:10.1118/1.1844433
 35. Wei J, Sahiner B, Hadjiiski LM et al (2005) Computer-aided detection of breast masses of full field digital mammograms. *Med Phys* 32(9):2827–2838
 36. Sallam MY, Bowyer KW (1999) Registration and difference analysis of corresponding mammogram images. *Med Image Anal* 3(2):103–118
 37. te Brake GM, Karssemeijer N (1999) Single and multiscale detection of masses in digital mammograms. *IEEE Trans Med Imaging* 18(7):628–639
 38. Blake A, Zisserman A (1987) *Visual reconstruction*. MIT Press, Cambridge
 39. Edwards DC, Kupinski MA, Metz CE, Nishikawa RM (2002) Maximum likelihood fitting of FROC curves under an initial-detection-and-candidate-analysis model. *Med Phys* 29(12):2861–2870

Author Biographies

Marcin Bator received his MSc degree from the Gdansk University of Technology, Poland, in 2000. He has 7 years of experience in software engineering for a number of enterprises. Since 2002, he has been with the Institute of Fundamental Technological Research, Polish Academy of Sciences, as a PhD student. His research interests include image analysis and evolutionary computing.

Leszek J. Chmielewski received his MSc and PhD degrees from the Warsaw University of Technology, Poland, in 1981 and 1985, and the DSc degree from the Institute of Fundamental Technological Research, in 2008. Since 1986, he has been with the Institute of Fundamental Technological Research as an assistant professor. His research interests include image analysis in measurements and defectoscopy, medical image analysis and accumulation-based methods. In 1992–2000, he was the Executive Editor of *Machine Graphics & Vision*, and in 1994–2003, the President and Vice-President of the Association for Image Processing, Polish member society of the IAPR. In 1996–1998, he was the member of the Committee on Biocybernetics and Biomedical Engineering, Polish Academy of Sciences.