Detection of Cancerous Masses in Mammograms by Template Matching: Optimization of Template Brightness Distribution by Means of Evolutionary Algorithm

Marcin Bator · Mariusz Nieniewski

Published online: 12 July 2011
© Society for Imaging Informatics in Medicine 2011

Abstract Optimization of brightness distribution in the template used for detection of cancerous masses in mammograms by means of correlation coefficient is presented. This optimization is performed by the evolutionary algorithm using an auxiliary mass *classifier*. Brightness along the radius of the circularly symmetric template is coded indirectly by its second derivative. The fitness function is defined as the area under curve (AUC) of the receiver operating characteristic (ROC) for the mass *classifier*. The ROC and AUC are obtained for a teaching set of regions of interest (ROIs), for which it is known whether a ROI is true-positive (TP) or false-positive (F). The teaching set is obtained by running the mass *detector* using a template with a predetermined brightness. Subsequently, the evolutionary algorithm optimizes the template by classifying masses in the teaching set. The optimal template (OT) can be used for detection of masses in mammograms with unknown ROIs. The approach was tested on the training and testing sets of the Digital Database for Screening Mammography (DDSM). The free-response receiver operating characteristic (FROC) obtained with the new mass detector seems superior to the FROC for the hemispherical template (HT). Exemplary results are the following: in the

case of the training set in the DDSM, the true-positive fraction (TPF)=0.82 for the OT and 0.79 for the HT; in the case of the testing set, TPF=0.79 for the OT and 0.72 for the HT. These values were obtained for disease cases, and the false-positive per image (FPI)=2.

Keywords Cancer detection · Image analysis · Mammography · Breast

Introduction

Computer-aided detection of cancerous masses in mammograms has been a research problem in recent years all over the world. There is a wealth of ideas, concepts, and research results, as exemplified by four review papers recently published [1–4]. Various algorithms have been proposed and even their classification is a matter of personal experience and preferences. Below, we follow the terminology developed by Oliver et al. [1]. Detection of cancerous masses is defined as the identification of potential lesions within all the parenchymal background. Usually, a mass detector generates a marker/prompt at a suspicious region in a mammogram. The next processing step is segmentation, which is defined as a method able to detect the precise outline of the potential lesion. Further steps include classification of masses as benign or malignant, as well as rejection of FP indications.

In general, mass detection and classification algorithms can be divided into two approaches: supervised segmentation and unsupervised segmentation. Supervised segmentation includes mainly model-based methods. The template matching described in this paper belongs to this group of methods.

The model-based methods rely on the prior knowledge about the object and background regions to be segmented.

M. Bator
Faculty of Applied Informatics and Mathematics,
Warsaw University of Life Sciences,
ul. Nowoursynowska 159,
02776 Warsaw, Poland
e-mail: marcin_bator@wp.pl

M. Nieniewski (✉)
Department of Mathematics and Informatics,
University of Lodz,
ul. Banacha 22,
90238 Lodz, Poland
e-mail: mnieniew@math.uni.lodz.pl

The prior information is used to determine if specific regions are present within an image or not. Several papers on model-based methods are mentioned below. The early paper was written by Lai et al. [5]. Constantinidis et al. [6, 7] used normalized cross-correlation for detection of cancerous masses. Tourassi et al. [8] used mutual information for determining whether a given ROI contains a cancerous mass. Freixenet et al. [9] proposed a probabilistic template matching to detect masses. In that paper, the shape and deformations of a deformable template are learnt from real mass examples. Subsequently, a Bayesian scheme is used to adapt the learnt deformable template to the unknown masses in the mammogram. In the paper by Székely et al. [10], a global segmentation method is used to find ROIs as a first step. This method utilizes texture features, decision trees, and a multiresolution Markov random field model. Subsequently, the output from the global segmentation is used by a combination of three different local segmentation methods and relevant features are extracted. These features describe the shape of the object or texture parameters. The final mass detection is accomplished by a linear combination of the calculated features.

Unsupervised segmentation might be represented by a great many papers. It works by partitioning an image into several regions which are distinct and uniform with respect to specific properties, such as gray level, texture, or color. There are three groups of methods of unsupervised segmentation:

1. Region-based methods, which divide the image into homogeneous and spatially connected regions. An example of these methods is shown in the paper by Wei et al. [11]. In that paper, a prescreening method identifies the mass candidates. The suspicious structure in each identified region is extracted by clustering-based region growing. Morphological and spatial gray-level dependence texture features are extracted for each suspicious object. Finally, rule-based and linear discriminant analysis classifiers are used to differentiate masses from normal tissues.
2. Contour-based methods, which find the boundaries of regions. An example is the paper by Domínguez and Nandi [12], where a dynamic programming method is developed for tracing the contour of a lesion. In particular, the authors propose a new local cost function to be used in the dynamic programming algorithm. This cost function is a combination of three components: edge strength, gray level deviation, and shape component.
3. Clustering methods, which group together those pixels having the same properties and might result in non-connected regions. An example of these methods is given in the paper by Suliga et al. [13], where Markov random fields are used for clustering pixels belonging to lesions. In the paper by Heath and Bowyer [14], a

mass detection algorithm is developed based on the area of fraction under minimum (AFUM) filter. The AFUM filter finds the degree to which the surrounding region of a given point decreases radially in intensity. The final step in the algorithm consists in thresholding the image to identify suspicious regions.

The citations included here may be based on personal preferences of the authors, and the review papers [1–4] should be consulted for more comprehensive information.

In pattern or template matching, the training usually includes a presentation of a collection of images containing the object to detect. Various modifications of templates were proposed, beginning with the paper by Lai et al. [5]. In that paper, tests are made on 17 mammograms containing circumscribed masses. The paper by te Brake and Karssemeijer [15] considers a larger dataset of 132 mammograms containing malignant masses. The paper is mainly concerned with comparison of single and multiscale detection of masses and concludes that multiscale gives a slight improvement. Various results are given for various sizes of masses, which are modeled by the upper hemisphere. The overall result is TPF=0.65 for FPI=2, which is approximately the same as obtained with the HT described below. The paper by Hatanaka et al. [16] analyzes 335 mammograms from a proprietary database. The paper actually describes a system utilizing several methods of mass detection, one of which is template matching by means of the correlation coefficient. The authors' primary concern is the detection of masses partially visible within the mammogram. Hence, the mask of the template has a shape of the sector cut out of a full disk. The template brightness is described by a Gaussian function. The global results presented in the paper are highly satisfactory; however, they were obtained within specific detection areas rather than full breasts. Due to lack of details, these results seem to us rather hard to reproduce.

Detection of cancerous masses is also one of the tasks performed by commercial computer-aided diagnosis (CAD) systems. The current state of the art of the CAD systems is described by Uchiyama [17] in the paper, which nominally describes the CAD utilization in Japan but in fact is concerned with CAD for mammography in the USA and Europe as well. The author shows the FROC curve for the CAD system known as the Image Checker 9.3 manufactured by Hologic. The included FROC may be described by the three representative points TPF=0.83, 0.88, and 0.90 for which FPI=0.7, 1, and 1.4, respectively. No information on the database used for obtaining this curve is provided. The results seem impressive, nevertheless, the author writes that "regarding the present performance of CAD, the detection rate is over 90% in microcalcifications, whereas

the detection rate for distortions and masses is not satisfactory compared with experienced readers.” That means that further investigations in this area are necessary. Uchiyama also states in his introduction that “at present there are still few institutions where computer-aided detection (CAD) is used for digital mammography in clinical settings.” Perhaps more important is Uchiyama’s detailed presentation of the conclusions drawn by various researchers evaluating the performance of the CAD systems for screening mammography. In fact, five of the evaluations are positive whereas two are quite negative. Obviously, the current paper cannot resolve the controversy regarding the evaluation of the CAD systems. However, the review of CAD systems in Ref. [17] confirms that the problem of automatic assistance in mammographic diagnosis needs further investigation.

The purpose of the current paper is an improvement of the specific mass detector based on template matching, which is also called pattern matching in a more general situation. The interest in this problem is confirmed by Oliver et al. who write in their recently published paper [1] that pattern matching is one of the most commonly used model-based segmentation methods. The particular novelty of the current paper is the use of the evolutionary algorithm for finding the optimal brightness distribution in the template.

Materials and Methods

Database Used

In the research described below, we use the mammograms from the DDSM as described by Heath et al. in Refs. [18, 19]. In particular, experiments on the training test BCPRP_MASS_0 and the testing set BCPRP_MASS_1 of the DDSM are carried out in order to make possible direct comparisons with publications by Heath and Bowyer [14, 20]. In particular, the paper [14] presents the FROC curves obtained for the AFUM filter and was chosen as a reference here since it contains a complete specification of mammograms and we could consider exactly the same set of mammograms in our research. The intended purpose of Ref. [14] was to serve as a reference in further publications and comparisons of mass detectors. Since that time, the DDSM has been used by a number of other researchers. However, subsequent papers either do not specify exactly which mammograms from the DDSM have been analyzed or use a mixture of mammograms from the DDSM and their own databases so the reproduction of the presented results is impossible.

As specified in Ref. [14], the training set contains 39 disease cases, that is 156 mammograms. Each case consists

of four mammograms: left and right breasts in craniocaudal and mediolateral oblique projections. The cases start with number 1118 and end with number 1908. The testing set contains 40 cases. The cases start with number 1112 and end with number 1999. The case no 1658 was unavailable at the web site. So in practice, the testing set also contains 156 mammograms. Full lists of mammograms in both sets are not provided here since they can be found in Ref. [14]. All cases contain at least one malignant mass with spiculated margins. The ground truth areas for all masses were outlined by a radiologist and are provided in the DDSM. For our purposes, the number of gray levels in mammograms was reduced to 8 bits.

Mass Detector

The described mass detector finds a mass by matching the template to a window in the mammogram. The mammogram is scanned, and the correlation coefficient between the template and the window is calculated similarly to the description in the paper by Filev et al. [21]. It is assumed that the mass should have at least some minimal radius in order to be detected. The described algorithm takes into account the possible range of sizes by calculating the correlation coefficient for various template sizes and various mammogram resolutions. Experiments have shown that a reasonable way of increasing the radius of the HT is according to the equation $R_n = \sqrt{2}R_{n-1}$, where R_n and R_{n-1} are respectively a new and previous radii of the template. This problem is discussed in more detail in the paper by Bator and Nieniewski [22], but in practice it turned out necessary to use only two sizes of the template. As regards the mammogram resolution, each coarser resolution is obtained by first smoothing the mammogram with a 3×3 Gaussian filter and then sampling the result by taking every other pixel in every second row. In practice we use three resolutions. The two template sizes together with three resolutions give a total of six various relations between the mammogram and the template. It is shown in [22] that the smallest template should have the radius equal to $\sqrt[4]{2}R_{\min}$, where $R_{\min}$ is the radius in pixels of the smallest mass to be detected.

It is well known that the correlation coefficient, denoted here by w , does not change if brightness of the template is multiplied by a constant or some constant is added to brightness of every pixel. This property of the correlation coefficient can be expressed by the equation

$$w(a\mathbf{I}_{x,y} + b, c\mathbf{T} + d) = w(\mathbf{I}_{x,y}, \mathbf{T}). \quad (1)$$

The $a, \dots, d$ in Eq. 1 are arbitrary constants such that $ac > 0$. $\mathbf{T}$ denotes the template, and $\mathbf{I}_{x,y}$ is a window positioned at pixel coordinates x and y . It follows from Eq. 1 that the

absolute value of brightness is unimportant and relative values can be used instead. The values of the correlation coefficient are calculated for each pixel in the mammogram and compared for various template sizes and mammogram resolutions so that the maximum out of six values is chosen. This maximum is subsequently compared with a chosen threshold, and pixels for which the correlation coefficient is higher than the threshold are considered as belonging to a lesion.

Optimization of Brightness Distribution

The templates described in the literature typically have a square or circular region of definition, possibly with some modifications. For example, te Brake and Karssemeijer [15] extend the circular template to $1.3R$, where R is the radius of the model of the lesion, and brightness of the ring surrounding the model is zero. A more general template using a sector-form model is considered by Hatanaka et al. [16], who analyze detection of masses with a partial loss of region in the vicinity of film edge.

Our basic assumption concerning the template is that it has a circular symmetry, so that the consideration of brightness distribution is reduced to the optimization along the radius. The brightness distribution in the template should model brightness of a typical mass. In most cases the mass has a flat bright center, as exemplified by the ROI shown in Fig. 1a and by the brightness distribution across the center of the ROI as shown in Fig. 1b. The center of the lesion is quite often surrounded by a background area which has a darker appearance with respect to the center of the mass as well as with respect to the areas farther away from the mass. Obviously, this is a tentative assumption, the usefulness of which in a statistical sense has yet to be proved.

Theoretically, the border of a mass might be found by setting the Laplacian of brightness to zero. By analogy, we assume that the second derivative of brightness with respect to the radius R from the center of the template is equal to zero at any point on the border of the lesion. It follows then that for the circularly symmetric template the brightness inside the mass should have a negative second derivative, that is

$$\frac{d^2 f(R)}{dR^2} < 0 \quad \text{for } R < R_m, \quad (2)$$

and outside the mass it should have a positive second derivative. The R_m in Eq. 2 denotes the hypothetical radius of the mass.

We find the optimal brightness distribution in the template by means of an evolutionary algorithm. Due to computational difficulties, however, this optimization is accom-

Fig. 1 Example of brightness distribution across the lesion. **a** Window taken from the original mammogram. **b** Brightness distribution along the horizontal line in the middle of **a**

plished not in the process of mass detection but rather in the auxiliary process of mass classification. Of course, mass classification requires some teaching set of ROIs containing examples of both cancerous masses and normal tissues. This teaching set is described below and at the moment, we concentrate on the evolutionary algorithm. The evolutionary algorithm evaluates individuals in the current population of candidate solutions and generates a new population. Details concerning typical evolutionary algorithms can be found in the review paper by Jin and Branke [23]. In the following, only the basic facts related to the application at hand are given. The search for the optimal brightness distribution depends on the definitions of the operations used for generating a new population and on the choice of information stored in the genotype. Moreover, it depends on the fitness function comparing the candidate brightness distributions.

It would be possible to directly code the template brightness in the genotype for successive values of the

radius. However, this approach would result in a violation of condition expressed by Eq. 2. An example of such inconsistency is shown in Fig. 2, where crossing between two parents generates a descendant for which the condition imposed on the second derivative is not observed. On the other hand, this condition can be satisfied if the genotype contains a sequence of second differences of brightness function along the radius, divided by the square of the increment of radius that is a digital approximation of the second derivative of brightness. Further simplification results from the fact that by virtue of Eq. 1, dividing the second differences by the increments of the radius may be omitted.

Table 1 shows an example of the genotype containing second differences of brightness. The last two rows in the genotype specify the minimum and maximum positions (61 and 233, respectively) along the radius that are currently used in the genotype. In addition to second differences, the evolutionary algorithm also changes the maximum and minimum positions. The values in positions preceding 61 and following 233 in Table 1 are currently not in use. The full Table 1 would be too lengthy to include in print, and only the points in the vicinity of minimum and maximum positions, as well as around the zero crossing of the second derivative are shown here.

The calculations based on Table 1 for an exemplary genotype are illustrated in Fig. 3. In particular, Fig. 3 shows the approximation of the second derivative of the brightness function together with the reconstructed first derivative and the original function, that is the brightness distribution. In order to have all three curves roughly in the same range, the first derivative and the original function are divided by their respective maximum absolute values. As a result, the zero

Table 1 Example of a genotype

Position	Value
1	-0.512
—	—
60	-0.982
61	-0.789
62	-0.609
—	—
166	-0.723
167	-0.696
168	0.000
169	0.325
—	—
233	0.726
234	0.875
—	—
241	0.250
242	61
243	233

Successive positions correspond to the increment of radius equal to one fourth of a pixel. The values specified for each position are the second differences of brightness. The values in the last two rows specify the range of positions in use that is from 61 to 233

crossing of the second derivative is clearly visible as well as the variability or noisiness of the second derivative.

The radii corresponding to individual positions in the genotype depend on R_m expressed in pixels, which varies with the size of the template. We assumed that R_m is represented by position 168 in Table 1. The radii for which brightness is calculated are given by the equation $R = iR_m/168$, where $i = 0, 1, \dots, 241$. We put $R_m = 42$ pixels, which with the resolution $43.5 \mu\text{m}$ per pixel for the mammograms under consideration gives 3.65 mm for the diameter of the mass. Dividing this number by a $\sqrt[4]{2}$ we

Fig. 2 Example of crossover on function values that violates the constraint imposed on the second derivative. The parents are shown by dotted-dashed lines and the descendant by a continuous line

Fig. 3 Result from the evolutionary algorithm—the second derivative of brightness together with the first derivative and the original brightness function

obtain approximately 3.1 mm for the size of the smallest mass to be detected.

The radius along which brightness is to be calculated in general is rotated with respect to the coordinate system of the template, hence brightness at a given pixel cannot be obtained directly from the genotype and has to be calculated by a linear interpolation between the neighboring positions specified in the genotype.

The standard operations in the evolutionary algorithm are crossing, mutation, and selection. It is noteworthy that the second differences of brightness in the genotype are not integers but rather real numbers. The assumed genotype allows parent crossing in the form of a simple combination of parts taken from each of the parents. The mutation is implemented by multiplication of a given real number by a random number from the interval (0, 2]. In the case of the minimum and maximum positions, the mutation takes form of incrementing or decrementing the position by unity. Selection of genotypes for the next iteration follows a classical roulette. The number of iterations in our experiments turned out to be less than 100.

Note that the second differences of brightness are used exclusively in the genotype and related operations of the evolutionary algorithm. Any interaction of the template with a mammogram requires the original function that is brightness itself. Using the digital approximation, one can easily reconstruct the derivative of brightness and the brightness itself from the second differences.

The assumed number of individuals in the population that is effectively of brightness distributions is 100. The comparisons between individuals are made by means of the experimental ROC obtained for an auxiliary, teaching set of ROIs in a manner described by Fawcett [24]. The ROIs are generated by running the mass detector using an HT, for which brightness $f(R)$ is described by the equation

$$f(R) = \sqrt{R_m^2 - R^2}. \quad (3)$$

In our experiments we assumed a relatively low empirically selected threshold 0.65 and obtained 70 TPs and 1139 FPs for the training set BCPRP_MASS_0 in the DDSM for all template sizes and mammogram resolutions.

The teaching set of ROIs makes possible the optimization of the template used in the classifier of the obtained ROIs. When a given template is evaluated, then the correlation coefficient between this template and each position in each ROI in the teaching set is calculated. It is known which of these ROIs are TPs and which are FPs by comparing the ROIs with the ground truth data for the training set in the DDSM. By changing the threshold in the mass classifier one controls the ratio of TPF and false-positive fraction and obtains an experimental ROC. A meaningful quantity for comparison of ROCs is the AUC

discussed in the paper by Fawcett [24]. The AUC is a single number, hence suitable for a fitness function. In the example shown, the AUC=0.53 for the OT, whereas AUC=0.51 for the HT. It is noteworthy that these numbers were obtained for the case of heavy imbalance of TPs against FPs, for which AUC=0.5 does not have the same meaning as for equal probabilities of TP and FP. Once the OT has been found by the mass classifier, we can use the same template in the mass detector for arbitrary mammograms for which no ROIs are known. The validity of this generalization is confirmed by the mass detection results presented below.

The OT generated by the evolutionary algorithm is shown in Fig. 4a, and the optimal brightness distribution is shown in Fig. 4b. The optimal brightness distribution has been adjusted so that it is equal to 0 for radius R_m , and with the brightness of the HT following a quarter of a circle, both curves meet again at the inner radius R_1 . This adjustment is done for better visualization of the results and follows the equations

$$f_1(R) = f(R) - f(R_m), \quad (4)$$

$$f_2(R) = f_1(R) \times \frac{\sqrt{R_m^2 - R_1^2}}{f(R_1)}. \quad (5)$$

The R in the above equations denotes the variable radius. The function $f(R)$ is the original brightness as obtained from the data such as in Table 1 by reconstructing the original function from the second differences. The function $f_2(R)$ is the adjusted brightness.

Approximate results of the adjustment of brightness distribution for the example of Fig. 4 are shown in Table 2.

Results of Mass Detection

Free-Response Receiver Operating Characteristics

Some details of the proposed approach to the mass detection are illustrated by Fig. 5. In particular, this figure shows the original mammogram. The white contour around the breast indicates the analyzed area of the mammogram. Calculations of the correlation coefficient are the most time-consuming part of computations, hence it is imperative to reject irrelevant areas. A smaller white contour around the local brightening in the mammogram indicates the ground truth area specified in the DDSM. Black, almost circular contours indicate areas from which information was gathered for the calculation of the correlation coefficient and this coefficient exceeded the chosen threshold. If the correlation coefficient in a single pixel exceeds the threshold, then the area from which information is gathered

Fig. 4 Optimal template. **a** Gray-level image representing the template. The white color inside and outside the ring corresponds to areas not included in the template. **b** Template brightness distribution along the radius. R_1 and R_2 are the inner and outer radii of the template

is circular and the respective pixel is at the center of the circle. Usually, the correlation coefficient exceeds the threshold in a group of neighboring pixels, and there is a

Table 2 Results of the adjustment of the optimal brightness distribution

R	f	f_1	f_2
15.25	-0.8	2,853.1	39.1336
42	-2,853.9	0	0
58.25	-4,985.2	-2,131.3	-29.2324

Fig. 5 Example of mass detection for the mammogram A_1112_1_LEFT_CC at the threshold 0.68 for the correlation coefficient

superposition of several circles so the black contour in general is not a circle. More complicated shape of the black contour may also result from calculating the correlation coefficient for various resolutions of the image and various sizes of the template. In a particular example in Fig. 5, there are one TP and one FP findings.

The overall presentation of the mass detection results is carried out by means of an FROC showing the relation between the TPF and FPI. Both TPF and FPI depend on the parameters of the mass detector as well as on the definitions of TPs and false negatives (FNs), which are discussed in more detail by Nishikawa [25]. A complete FROC curve is obtained by changing the threshold in the mass detector, similarly as was the case with the ROC.

The FROC curves obtained for the OT and HT for the training set and the testing set under consideration are shown in Figs. 6 and 7, respectively. Figures 6a and 7a refer to individual mammograms, whereas Figs. 6b and 7b refer to disease cases. Figures 6b and 7b also show the FROC curves for the AFUM filter.

We used the ROIs from the training test BCPRP_MASS_0 for optimizing the template and later the same training set was used by the mass detector. In order to avoid any

possible bias, we used the mammograms from the testing set BCPRP_MASS_1 solely in the mass detector. A numerical comparison of the TPF results for two exemplary points FPI=2 and 3 for the OT, HT, and the AFUM filter is given in Table 3.

The results for the OT indicate a reduction of the FPI for the same TPF in comparison with the AFUM filter and with the HT for a wide range of FPIs. It is noteworthy that the FROCs for the training and testing sets in Figs. 6 and 7 are quite similar so the fact that the training set was used both for the optimization of the template as well as for the subsequent mass detection does not have a bearing on the results.

Several remarks are in order as to the evaluation of the mass detectors. Heath and Bowyer [14] assume that a

Fig. 6 Comparison of the FROC curves obtained with the optimal template, hemispherical template, and the AFUM filter for the training set in the DDSM. **a** Results for individual mammograms. **b** Results for disease cases

Fig. 7 Comparison of FROC curves obtained with the optimal template, hemispherical template, and the AFUM filter for the testing set in the DDSM. **a** Results for individual mammograms. **b** Results for disease cases

ground truth area must contain at least one detection for the detection to be counted as TP. The area with no detection counts as FN. In the case of multiple, say N , detections corresponding to a single ground truth area, they are counted as one TP and $(N-1)$ FNs. Detections in Ref.

Table 3 Comparison of TPFs for disease cases

DDSM set	FPI=2		FPI=3	
	Training	Testing	Training	Testing
Optimal template	0.82	0.79	0.87	0.90
Hemispherical template	0.79	0.72	0.90	0.82
AFUM filter	0.61	0.72	0.72	0.77

[14] are in the form of single pixels. Less suspicious detections at a distance of up to 5 mm from more suspicious ones are removed. In our case, the detection may have a form of several or more connected pixels since it is obtained by thresholding the correlation coefficient, and all multiple TPs count as a single TP. The detection has to overlap at least 50% of the ground truth area in order to be counted. In both Ref. [14] (page 223 therein) and in our experiments, the FROC is calculated for the detection of disease cases. It means that a case is detected whenever at least one of the four mammograms constituting a single disease case reveals a cancer. A minor difference is that our method is based on the assumption that $R_{\min}=3.1$ mm whereas in [14] the minimal radius is not specified. It should be noted that we read the relevant data off the diagram in the paper by Heath and Bowyer [14], so there may be some small inaccuracies.

Confidence Intervals for Points on the FROC

In order to get a deeper insight into the performance of the proposed mass detection system, a statistical analysis of the obtained results is performed. This analysis was inspired by the paper by Bornefalk et al. [26], which discusses the use of parametric and nonparametric bootstrap for calculating the confidence intervals for the points on the FROC curve. We chose the nonparametric bootstrap since we were not certain about the validity of assumptions with respect to underlying probability functions.

The details of the nonparametric bootstrap are well explained in the book by Martinez et al. [27]. The bootstrap treats the original sample, in our case a limited collection of mammograms, as a pseudo-population and samples with replacement from this population are drawn in order to get estimates of the parameters of the original population, in our case a very large collection of mammograms, for which we do not have the necessary data.

For a given $TP = TP'$, the threshold θ as well as FP are random variables, however, θ depends on TP directly, whereas FP depends on TP through θ . We aim at calculating the 95% confidence interval for FP given $TP = TP'$. The following notation will be used for the case of discrete random variables rather than continuous variables since we use the discrete variables. We denote by f_{θ} the probability mass function of the variable θ . Similarly, f_{TP} denotes the probability mass function of TP. Using the Bayes' formula, we obtain the following expression for the conditional probability $f_{\theta|TP=TP'}$ of θ given $TP = TP'$

$$f_{\theta|TP=TP'} = \frac{f_{\theta} f_{TP=TP'|\theta}}{\sum_{\theta} f_{\theta} f_{TP=TP'|\theta}}, \quad (6)$$

where $f_{TP=TP'|\theta}$ is the conditional probability of $TP = TP'$

given θ . The threshold θ is a parameter used for the generation of the FROC and we can assume that f_{θ} is uniform hence it effectively drops out of the RHS of Eq. 6

$$f_{\theta|TP=TP'} = \frac{f_{TP=TP'|\theta}}{\sum_{\theta} f_{TP=TP'|\theta}}. \quad (7)$$

We denote by $f_{TP,FP|\theta}$ the joint conditional probability mass function of TP and FP given θ . Assuming that TP and FP are independent we obtain

$$f_{TP,FP|\theta} = f_{FP|TP,\theta} f_{TP|\theta} = f_{FP|\theta} f_{TP|\theta}. \quad (8)$$

The conditional probability mass function $f_{FP|TP=TP'}$ of FP given $TP = TP'$ is obtained from the equation

$$f_{FP|TP=TP'} = \sum_{\theta} f_{\theta|TP=TP'} f_{FP|TP=TP',\theta}, \quad (9)$$

where

$$f_{FP|TP=TP',\theta} = f_{FP|\theta}. \quad (10)$$

The practical calculations were carried out on the training set from the DDSM described in the previous subsection. The original sample is an array of size 156×21 , where 156 is the number of the mammograms in the set and 21 denotes the number of thresholds used for obtaining the points on FROC. The values of thresholds are: 0.60, 0.61, ..., 0.80. The number of bootstrap samples is set equal to 1,000. The output array of the bootstrap method at this stage is of size $1,000 \times 21$ and represents the distribution of the total of TPs for each bootstrap sample and threshold. A corresponding array is then generated which contains the relative numbers TPF instead of TP. The results from the latter array are collected in an array of size 25×21 which contains histograms for the columns of previously calculated array of TPFs. More specifically, the whole range of TPFs is divided into 25 bins, where each bin has a width of 0.04, and respective middle values of the bins are 0.02, ..., 0.098. Based on the above mentioned 25×21 array, the probability mass function $f_{\theta|TP=TP'}$ can be drawn. Examples of such a function are shown in Fig. 8.

The probability mass function $f_{FP|\theta}$ is calculated by means of the bootstrap sampling from the original training set data using another pseudo-population represented by an array of size 156×21 , which specifies how many FPs occur for a given image and threshold. 1,000 bootstrap samples are generated from the above mentioned array, and the result is an array of size $156 \times 21 \times 1,000$, which specifies FP for a given image, threshold, and bootstrap sample. Summing FPs over all 156 images one can calculate the FPI as a function of threshold and bootstrap sample. The results are stored in the array of size $21 \times 1,000$. Each

Fig. 8 Probability mass function $f_{\theta|TP=TP^*}$ according to Eq. 7 for $TPF=0.30, 0.38, \dots, 0.86$

column of this array contains samples from the probability mass function $f_{FPI|\theta}$ (Eq. 10). The next step is the generation of the array of size $25 \times 1,000$ with each column containing samples from the probability mass function of FPI for a given TPF, as defined by Eqs. 9 and 10. This means that a single sample from $f_{FPI|\theta}$ is now effectively replaced by the weighted sum of samples taken for several θ s.

The calculation of confidence interval for a given TPF is carried out using the bootstrap percentile interval method ([27]) and is based on the quantiles of the distribution of the bootstrap samples. For this purpose, the recently calculated $25 \times 1,000$ array of FPIs is sorted columnwise. The 95% confidence interval is obtained by choosing the values of FPI in positions 25 and 975 in a given column. These values are respectively the lower and upper ends of the confidence interval. Examples of the obtained confidence intervals are shown in Fig. 9 for the FROC repeated from Fig. 6a. Because the original data in Fig. 9 were obtained for selected values of θ , we cannot simply use these TPFs but rather have to find the values of θ for assumed sequence of $TPF=0.26, 0.30, \dots, 0.86$ by interpolating linearly between the original data points of the FROC. The resulting interpolated points are marked with crosses located inside respective confidence intervals. For example, for $TPF=0.7$ we get $FPI=2.44$, and the respective confidence interval is (2.213, 3.129). Comparing the area indicated by the FROC for the OT together with the respective end points of the confidence intervals in Fig. 9 with the FROC for the HT in Fig. 6a, one might observe that the latter FROC in general lies outside the specified area.

Fig. 9 95% confidence intervals for points on the FROC for optimal template, calculated for individual mammograms in the training set of the DDSM. The FROC curve is the same as in Fig. 6a

Conclusions

By using the template with the brightness distribution generated by an evolutionary algorithm, one can get some improvement of cancer detection results in comparison with the usual template for which brightness is described by a simple geometry-based formula, for example hemispherical distribution. Furthermore, the obtained FROC curves are better when compared with those for the AFUM filter. Our results confirm the well-known fact that most of the information about a lesion is contained in the area near the lesion boundary and not in the center of the mass. The algorithmic procedure for the optimization of the template is rather straightforward and needs to be carried out only once for the given working conditions. The statistical analysis shows that the obtained FROC points are restricted by reasonable confidence intervals, and we may expect that the obtained results are relatively independent of our choice of mammograms. The described optimization of the template does not involve any additional calculations in the process of mass detection once the optimized template has been found.

It seems at present time that no single method of mass detection will be sufficient for a full-fledged CAD system, and most probably, some mixture of methods will have to be used. In addition, particularly imperative is the further development of methods for the reduction of FPIs for a given TPF.

References

- Oliver A, Freixenet J, Martí J, et al: A review of automatic mass detection and segmentation in mammographic images. *Med Image Anal* 14(2):87–110, 2010
- Domínguez AR, Nandi AK: Towards breast cancer diagnosis based on automated segmentation of masses in mammograms. *Pattern Recogn* 42(6):1138–1148, 2009
- Tang J, Rangayyan RM, Xu J, et al: Computer-aided detection and diagnosis of breast cancer with mammography: recent advances. *IEEE T Inform Techn in Biomed* 13(2):236–251, 2009
- Elter M, Horsch A: CADx of mammographic masses and clustered microcalcifications: a review. *Med Phys* 36(6):2052–2068, 2009
- Lai SM, Li X, Bischof WF: On techniques for detecting circumscribed masses in mammograms. *IEEE T Med Imaging* 8(4):377–386, 1989
- Constantinidis AS, Fairhurst MC, Rahman AFR: Detection of circumscribed masses in digital mammograms using behaviour-knowledge space method. *Electron Lett* 36(4):302–303, 2000
- Constantinidis AS, Fairhurst MC, Rahman AFR: A new multi-expert decision combination algorithm and its application to the detection of circumscribed masses in digital mammograms. *Pattern Recogn*. 34(8):1527–1537, 2001
- Tourassi GD, Vargas-Vorecek R, et al: Computer-assisted detection of mammographic masses: a template matching scheme based on mutual information. *Med Physics* 30(8):2123–2130, 2003
- Freixenet J, Oliver A, Martí R, et al: Eigendetection of masses considering false positive reduction and breast density information. *Med Physics* 35(5):1840–1853, 2008
- Székely N, Tóth N, Pataki B: A hybrid system for detecting masses in mammographic images. *IEEE T Instrum Meas* 55(3):944–952, 2006
- Wei J, Sahiner B, Hadjiiski LM, et al: Computer-aided detection of breast masses on full field digital mammograms. *Med Physics* 32(9):2827–2838, 2005
- Domínguez AR, Nandi AK: Improved dynamic-programming-based algorithms for segmentation of masses in mammograms. *Med Physics* 34(11):4256–4269, 2007
- Suliga M, Deklerck R, Nyssen E: Markov random field-based clustering applied to the segmentation of masses in digital mammograms. *Comput Med Imag Graph*, 32(6):502–512, 2008
- Heath M, Bowyer K: Mass detection by relative image intensity. In: MJ Yaffe, Ed. *Proc. 5th international workshop on digital mammography*. Madison: Medical Physics Publishing, 2001, pp 219–225
- te Brake GM, Karssemeijer N: Single and multiscale detection of masses in digital mammograms. *IEEE T Med Imaging* 18(7):628–639, 1999
- Hatanaka Y, Hara T, Fujita H, et al: Development of an automated method for detecting mammographic masses with a partial loss of region. *IEEE T Med Imaging* 20(12):1209–1214, 2001
- Uchiyama N: Current status of CAD utilizing digital mammography in Japan. *Breast Cancer* 17:169–179, 2010
- DDSM: Available at: <http://marathon.csee.usf.edu/Mammography/Database.html>
- Heath M, Bowyer K, Kopans D, et al: The Digital Database for Screening Mammography. In: MJ Yaffe, Ed. *Proc. 5th international workshop on digital mammography*. Madison, WI: Medical Physics Publishing, 2001, pp 212–218
- FROC of Training and Testing Data. Available at: http://marathon.csee.usf.edu/Mammography/DDSM/BCRP/bcrp_mass_01.html
- Filev P, Hadjiiski L, Sahiner B, et al: Comparison of similarity measures for the task of template matching of masses on serial mammograms. *Med Physics* 32(2):515–529, 2005
- Bator M, Nieniewski M: The usage of template matching and multiresolution for detecting cancerous masses in mammograms. *Machine Graphics & Vision* 16(3/4):329–345, 2007
- Jin Y, Branke J: Evolutionary optimization in uncertain environments—a survey. *IEEE T Evolutionary Comput* 9(3):303–317, 2005
- Fawcett T: An introduction to ROC analysis. *Pattern Recogn Letters* 27(8):861–874, 2006
- Nishikawa RN: Current status and future directions of computer-aided diagnosis in mammography. *Comput Med Imag Graph* 31(04–05):224–235, 2007
- Bornefalk H, Bornefalk Hermansson A: On the comparison of FROC curves in mammography CAD systems. *Med Physics* 32(2):412–417, 2005
- Martinez WL, Martinez AR: *Computational statistics handbook with MATLAB*. Boca Raton: Chapman & Hall, 2007