Mammogram,
detection of cancerous masses,
template matching

Marcin BATOR*, Mariusz NIENIEWSKI*

THE USAGE OF TEMPLATE MATCHING AND MULTIRESOLUTION FOR DETECTING CANCEROUS MASSES IN MAMMOGRAMS

The paper describes the usage of template matching and multiresolution for detecting breast cancers containing the main mass. It was assumed that the template has a hemispherical brightness distribution and a square region of definition. The multiresolution images were obtained by a Gaussian pyramid. The correlation coefficient (CC) was thresholded to generate the mask of the center of the mass. The approach described was tested on the complete Mammographic Image Analysis Society (MIAS, UK) database giving the results which are easy to compare to the database information as well as to some other papers investigating the same database. The Free Receiver Operating Characteristics (FROCs) were obtained by varying the threshold used with the CC. The calculation of the CC was accelerated by using the Fast Fourier Transform (FFT).

1. INTRODUCTION

Breast cancer is the most common form of cancer in Polish women. It is also the major case of death. Mammography is currently the best technique for early detection of breast cancer. However, it is not perfect and more than 20 % of cancers may be missed by using single-view screening mammography ([4] pp. 722). A frequent appearance of cancer is a round main mass, possibly with spicules extending from the mass. The purpose of this paper is to present the usage of template matching together with multiresolution for detection of cancerous masses. [1] analyzed three methods for detecting masses in mammograms using a single scale as well as multiscale approach and showed that the multiscale was advantageous in comparison with a single scale. [6] made the assumption that a small spiculated mass in higher resolution is similar to a bigger one in lower resolution. We used a spherical template for template matching combined with multiresolution. This is not a complete Cancer Detector (CD) but rather a means for generating a sensitive feature to be incorporated into a CD using several features.

2. METHODS

Once the template has been defined, it is shifted across the image and the CC between the template and the appropriate window is calculated. The CC can be written in the usual form

* Institute of Fundamental Technological Research, ul. Swietokrzyska 21, 00-049 Warsaw

$$w(\mathbf{T},\mathbf{I}) = \frac{\sum_{i=1}^{N} t_i\, i_i - N\,\bar{t}\,\bar{i}}{\sigma_{\mathbf{T}}\,\sigma_{\mathbf{I}}}, \tag{1}$$

where **T** is a template; **I** is a window in the image; $\bar{t}, \bar{i}$ are the mean values of **T** and **I**, respectively; $\sigma_{\mathbf{T}}, \sigma_{\mathbf{I}}$ are standard deviations of **T** and **I**, respectively; *N* is the number of pixels in the template; t_i, i_i are pixel values. The template can be described by its brightness distribution and by its size. It can be proved that the CC is independent of scaling the gray values in the template and image

$$w(a\,\mathbf{T}+b, c\,\mathbf{I}+d) = w(\mathbf{T},\mathbf{I}), \tag{2}$$

where *a, b, c,* and *d* are constants. Equation (2) is valid when **T** and **I** contain real numbers. We investigated how discretization of the pixel values influences the CC. In particular, we assumed that the brightness distribution in the template is described by the equation of the hemisphere

$$t_R(x,y) = \sqrt{R^2 - x^2 - y^2} \quad \text{for} \quad x, y \in [-0.7R,\ 0.7R], \tag{3}$$

where x and y are the pixel coordinates measured from the center of the template. The radius *R* was varied in the range from 35 through 497 pixels. We computed the CC between the template containing real numbers and the image obtained by approximating the hemisphere by varying number of integer gray levels (Table 1). Inspection of this table reveals that little information is lost if we look for masses in 8-bit images instead of in 12-bit images. The maximum of the local gray level increase caused by a mass may be significantly less than 255. In fact 20 or so gray levels is a typical maximum. Once the brightness distribution in the template has been assumed, masses should be detected in various tissues and under changing mammogram exposure parameters.

Table 1 The CC as a function of a number of image gray levels for the template corresponding to R=35.

Number of gray levels	21	12	10	8	7	6	5	4	3	2
CC	1.00	0.99	0.98	0.97	0.95	0.94	0.91	0.88	0.81	0.58

Without the aforementioned restrictions of x and y we would have to put $t_R(x,y) = 0$ for pixels for which $x^2 + y^2 > R^2$. Because we were interested in the spherical brightness distribution we took only a part of the hemisphere (Fig. 1(c)). As shown in Fig . 1 non-square templates were considered in the literature. The main reason why we chose a square template is that the term $\sum_{i=1}^{N} t_i\, i_i$ in equation (1) can be calculated using the FFT and transform multiplication for the entire image at one time rather than performing the summation $\sum_{i=1}^{N} t_i\, i_i$ for each position of the window.

It can be shown that the summation has the complexity O(*MN*), whereas the FFT has the complexity O($M \log M$), where M and N are the pixel sizes of the image and the template, respectively.

a) b) c)

Fig. 1 Examples of templates. (a) diameter 9 used in [5], (b) diameter 53 used in [1], (c) size 51x51 used in this paper.

When looking for the mass of unknown size, we can change the size of the template ([1]) or the resolution of the image while keeping the template unchanged ([6]). The multiresolution approach reduces the computation time, whereas increasing the template would increase the computation time. The multiresolution was implemented by low-pass filtering images by means of the Gaussian filter and subsequent sampling which results in dividing the linear dimensions by 2. The template matching was then performed for each level of the pyramid (Fig. 2).

3. EXPERIMENTS AND RESULTS

We used in our experiments 321 mammograms from the MIAS database (image 295 in our copy of MIAS was defective). All classes of images were included: circumscribed masses, spiculated masses, architectural distortions, asymmetrical distortions, ill-defined masses, other masses, microcalcifications and normal cases. In principle we intended to detect circumscribed masses and spiculated masses possibly with a round core. Experiments were conducted with templates described by equation (3) with R=35 pixels and five levels of resolution in which we looked for masses of diameters: 3.5, 7, 14, 28, and 56 mm.

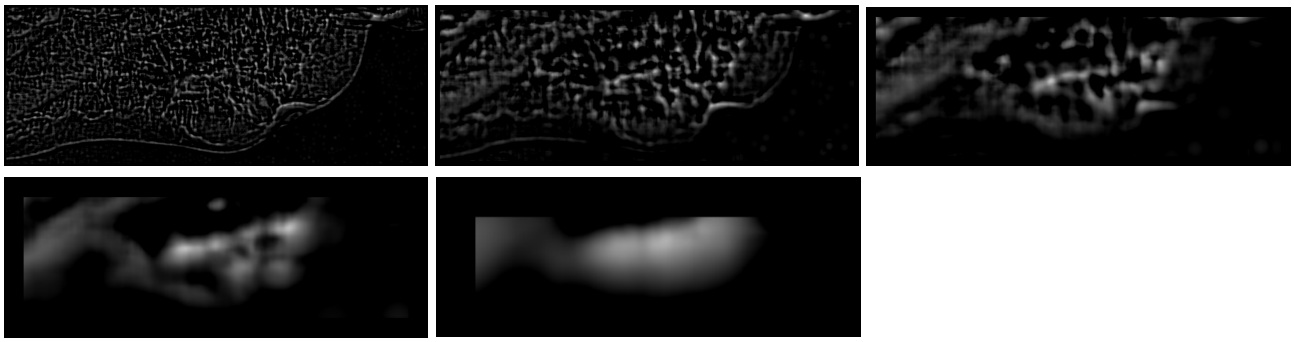

Fig. 2 The CC calculated in five resolutions for MIAS image 178. Smaller images were increased to the size of the maximal resolution by pixel duplication.

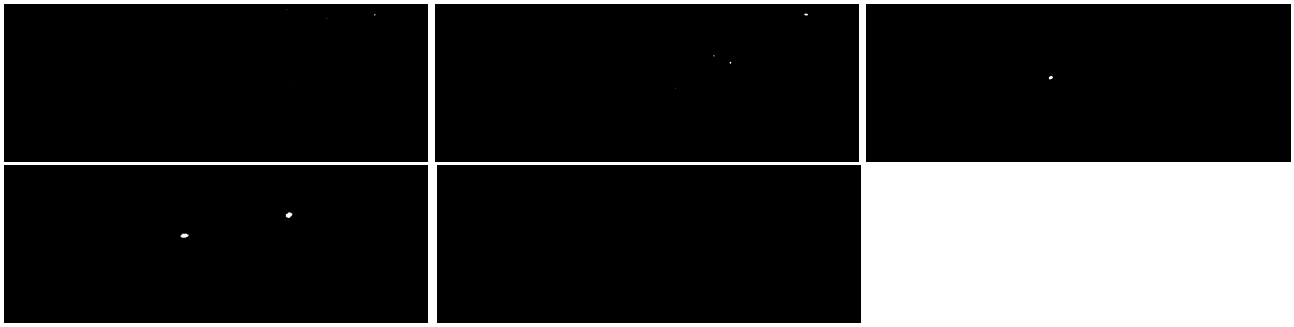

Fig. 3 Masks of the center of the mass obtained by thresholding respective images in Fig. 2 with threshold 0.7.

The obtained masks of the centers of the cancerous masses are shown in Fig. 3. Masks having at least 50 % of their area within the circle given in the MIAS database were defined as True Positive (TP) ([3], [6]). Masks not satisfying this condition were defined as False Positive (FP). Masks having at least 50 % of their area within the circle given in the database for microcalcifications were ignored. Masks outside the breast area were removed. Although detection of the breast contour is known from the literature ([2]), we obtained and cleaned the breast contours in a simplified manner carrying out the following operations: (1) thresholding images with a threshold approx. equal to 10, (2) smoothing the obtained contours by morphological closing and opening, (3) manual removal of labels and artefacts from the image.

Table 2 Number of detected masses with thresholding the CC at 0.7 (B – benign case, M – malignant case).

Abnormality	Circumscribed		Spiculated		Asymmetrical distortion		Architectural distortion		Miscellaneous	
	B	M	B	M	B	M	B	M	B	M
To detect	19	4	11	8	6	9	9	10	7	8
Detected	15	4	6	8	5	6	4	6	5	5
Sensitivity	79 %	100 %	55 %	100 %	83 %	67 %	44 %	60 %	71 %	63 %

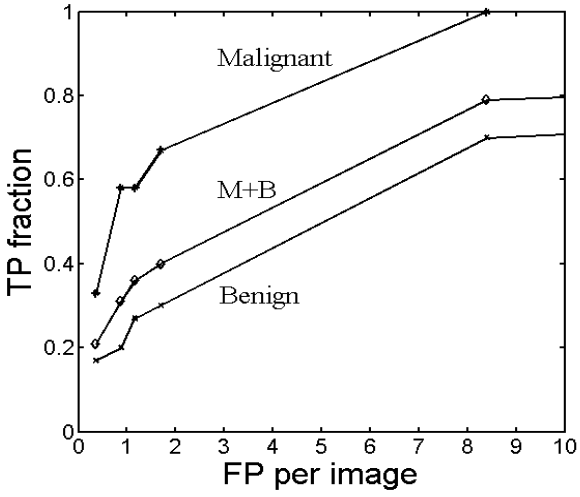

Fig. 4 FROCs for the MIAS database.

As illustrated by Table 2 the sensitivity for circumscribed and spiculated masses was 100 %. It was 74 % for all malignant masses, which means that 29 out of all 39 such masses were detected. The FROC curves obtained by varying the threshold for the CC are shown in Fig. 4. The annotation “M+B" in Fig. 4 means that all benign or malignant, circumscribed or spiculated masses are included.

For better visualization the masks obtained at each resolution were dilated with a disk of the radius 35 pixels. Subsequently images of lower resolution were increased by pixel duplication to the size of the maximal resolution image, and finally all the masks were logically added. The contours of the resulting masks are shown in black in Fig. 5. The white circle representing the reference area was taken from the MIAS database. Examples of correct and missing detections are shown in Figs. 6 and 7, respectively.

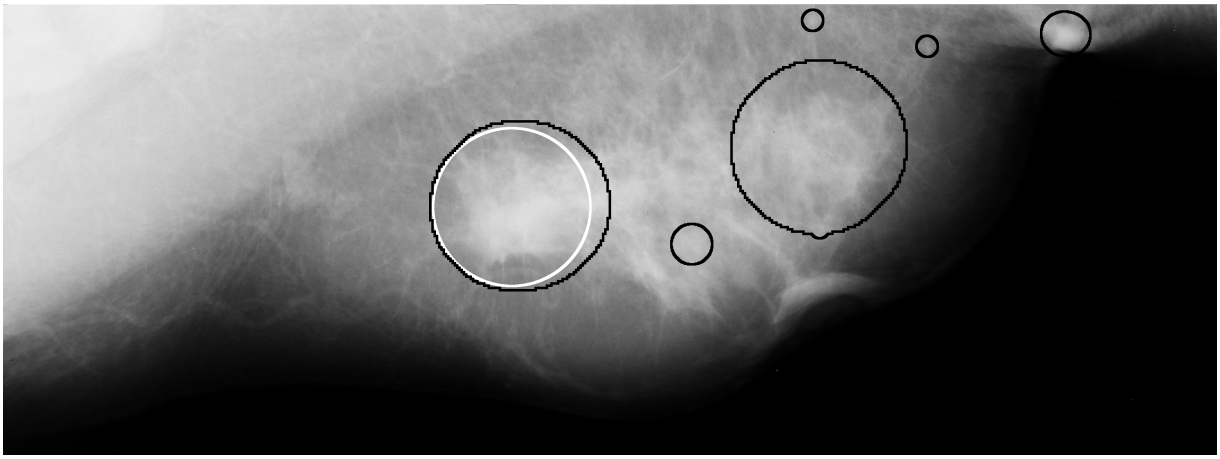

Fig. 5 Results for MIAS image 178. White circle is given in the database; black contour was obtained by the CD.

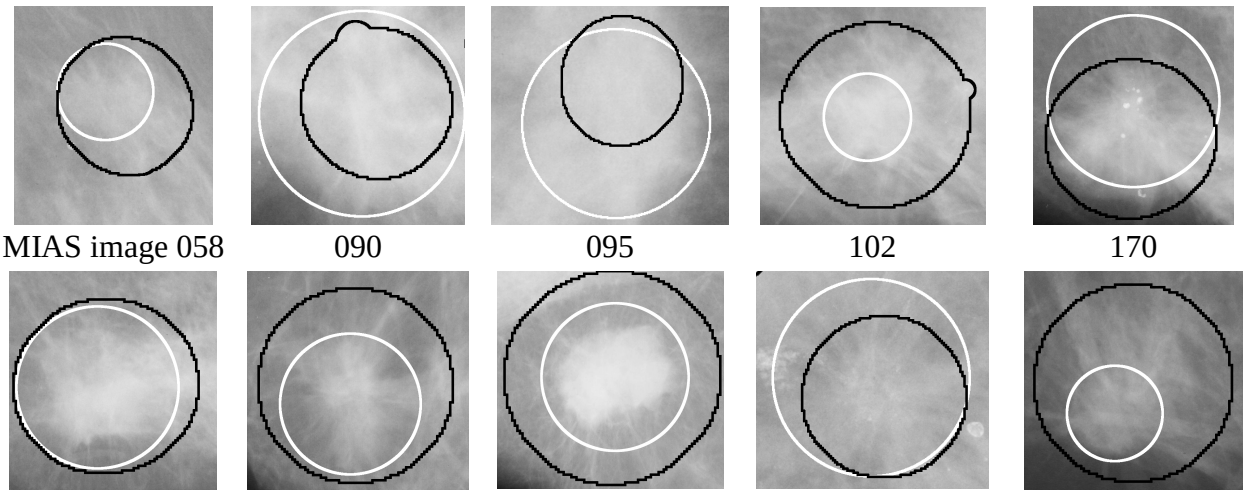

Fig. 6 Examples of properly detected masses.

Fig. 7 Examples of missed masses.

4. CONCLUSIONS

[7] achieved sensitivity 74.4 % for M+B (as defined above) and 100 % for malignant masses with 2.2 FP per image using 56 images (23 circumscribed, 19 spiculated masses and 14 normals) from the MIAS database. We had sensitivity 79 % for M+B and 100 % for malignant (circumscribed and spiculated) with 8.4 FP per image using all images from the MIAS database. At the same time we achieved sensitivity 74 % for all malignant masses and 70 % for all the masses in the database. High values of FP per image indicate that template matching cannot be used as an exclusive means of cancerous mass detection. However, large sensitivity confirms that this approach could be valuable when used with other cancer indicators. The obtained computation time was quite short: on the order of three minutes for one mammogram in five resolutions using 2 GHz Pentium computer and MATLAB environment.

ACKNOWLEDGMENT

This research was financed by the (Polish) Ministry of Education and Science as the research project. No 3 T11C 050 29 in 2005-2008.

BIBLIOGRAPHY

[1] TE BRAKE G. M., and KARSSEMEIJER N., Single and Multiscale Detection of Masses in Digital Mammograms, IEEE Transaction on Medical Imaging, Vol. 18, No. 7, pp. 628-639, 1999.
[2] FERRARI R. J., RANGAYYAN R. M., DESAUTELS J. E. L., and FRERE A. F., Segmentation of Mammograms: Identification of the Skin-Air Boundary, Pectoral Muscle, and Fibroglandular Disc, IWDM 2000, Fifth International Workshop on Digital Mammography, pp. 573-579, Medical Physics Publishing, Madison, 2001.
[3] KEGELMEYER W. P., PRUNEDA J. M., BOURLAND P. D., HILLIS A., RIGGS M. W., and NIPPER M. L., Computer-aided Mammographic Screening for Spiculated Lesions, Radiology, Vol. 191, pp. 331-337, 1994.
[4] KOPANS D. B., Breast Imaging, Lippincott-Raven Publishers, Philadelphia, 1998.
[5] LAI S.-M., LI X., and BISHOF W. F., On Techniques for Detecting Circumscribed Masses in Mammograms, IEEE Transaction on Medical Imaging, Vol. 8, No. 4, pp. 377-386, 1989.
[6] LIU S., BABBS C. F., and DELP E. J., Multiresolution Detection of Spiculated Lesions in Digital Mammograms, IEEE Transactions on Image Processing, Vol. 10, No. 6, pp. 874-884, 2001.
[7] MUDIGONDA N. R., RANGAYYAN R. M., and DESAUTELS J. E. L., Detection of Breast Masses in Mammograms by Density Slicing and Texture Flow-Field Analysis, IEEE Transaction on Medical Imaging, Vol. 20, No. 12, pp. 1215-1227, 2000.