

An Iris Detector for Tumoral Masses Identification in Mammograms

Arianna Mencattini, Giulia Rabottino, Marcello Salmeri, and Roberto Lojacono
University of Rome Tor Vergata
Rome, Italy
Email: mencattini@ing.uniroma2.it

Abstract—Radiologists that analyze screening mammographic images miss the 10-20% of the diagnosis since this kind of images are very difficult to interpret. In this paper, we present the first step of a CADx (Computer Aided Diagnosis) system that, from the original mammogram, extracts suspicious regions on which the radiologists have to focus their attention. The procedure often succeeds also in case of very low contrast, because it depends only on the orientation of the gradient vectors in the image but not on their amplitude.

I. INTRODUCTION

With the use of mammography as a screening tool and better treatments for breast cancer the death rates from breast cancer are steadily decreasing [1]. However, nowadays breast cancer is still the most common form of women cancer and every year it is diagnosed to more than one million women worldwide with the 25% of deaths. The mistakes occurring in the radiologist analyses of the image are principally due to human factors such as the difficulty to appreciate very low contrast objects or very small details. For this reason, it is possible to improve the survival rate with a second reader that evaluates the mammographic image independently. The aim of a CADx (Computer Aided Diagnosis) system [2] is to take the place of the second human reader to save time and costs but never to substitute the first radiologist in the image interpretation [3].

Starting from the original idea of the Iris filter proposed in [4], we present in this paper a new method for the automatic identification of masses in mammographic images that succeeds also when the contrast is weak because it depends only on the orientation of the gradient vectors in the image but not on their amplitude.

II. THE DATABASE

The images used in this work are screen-film mammographies belonging to the public DDSM database [5]. Attached to every image, the database contains also information about the lesion, if present, as the subtlety degree. In particular, there is a chain code identifying a contour around the masses manually extracted by an expert radiologist. In this way, we can compare our results with the position of the lesion specified by the radiologist. In the following figures, the white markers represent the position of the masses signed by the radiologists and the black ones the output of our algorithm.

III. MASSES IDENTIFICATION

A mass can be roughly considered as a circle with a luminance that grows from its border to its center. Recognize masses in a mammogram is difficult because of their low contrast with respect to the surrounding healthy tissue. In the following, we present the steps of the implemented procedure that indicates the regions in the mammographic image that could be tumoral signs.

- *Decimate the original image.* All the mammographic images considered in this study are represented by matrices of about 4000×6000 pixels. To reduce the computational time of the algorithm, we consider 1 pixel every 10. In this way, the algorithm can identify also the smallest masses that at least have a diameter of 3 mm, that, with a spatial resolution of $42 \mu\text{m}$, corresponds approximately to 70 pixels. In this case, a preprocessing of the image is not necessary because, as explained below, the algorithm considers only the luminance differences between pixels and so the information on the pixels of the decimated image has to be the same of that in the not-decimated image. Moreover, this kind of images are affected by low noise contributions [6] so that a preprocessing step could be more onerous in terms of computational time than useful. Obviously, this step is not applicable for the identification of microcalcifications clusters whose sizes can be only some pixels.
- *Segment the background.* To reduce the computational time, we do not consider the film background, isolating it simply with a threshold. Considering the characteristics of the mammographic images, we have experimentally set a threshold so that also the lighter background in the database can be segmented. We have also proven that this threshold value segments the background without altering the breast tissue.
- *Choose a grid.* Applying the algorithm on every pixel of the original image requires a lot of time. So, we apply the algorithm only on some pixel taken by gridding the original image with a step s that have to be chosen carefully. In fact, a dense grid could significantly increase the computational time, but a sparse grid could not center the mass, as shown in Fig. 1.
In this figure, one can see how, with the grid on the left, the system finds two suspicious regions but not the mass

identified by the radiologist and how, with a more dense grid, it finds only the region reported by the radiologist. In the section IV the reader can find some simulation results and the consequent choice of this parameter.

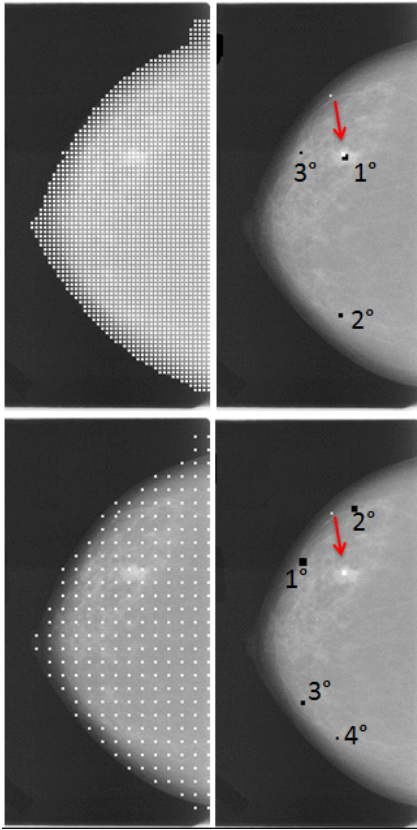


Fig. 1. Two different choices for the parameter s and the consequent results of the algorithm

- Consider a circle with radius R around every point (i, j) of the grid and compute on the points of the circle, every 15 degrees, this means that $N = 24$ equispaced points on the circle are considered:

$$x_R(i, j) = \frac{\sum_{k,l \in R} \cos \theta(k, l)}{N},$$

where $\theta(k, l)$ is the angle between a gradient vector in (k, l) and the straight line connecting the pixels at (i, j) and (k, l) (see Fig. 2 right). The term $\cos \theta(k, l)$ represents a measure of the convergence of gradient vectors in the circle on the pixel of interest (i, j) . When $x_R(i, j) = 1$ all the gradient vectors from the circle are oriented toward the same point. This occurs when the surfaces with the same values of luminance are concentric. It occurs when the pixel (i, j) is near the center of a massive lesion as one can see in Fig. 3.

- Repeat for different radii and compute, for every point (i, j) of the grid:

$$x(i, j) = \underset{R_{min} < R < R_{max}}{\text{mean}} x_R(i, j).$$

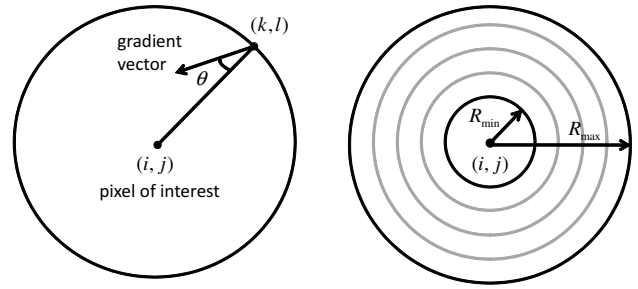
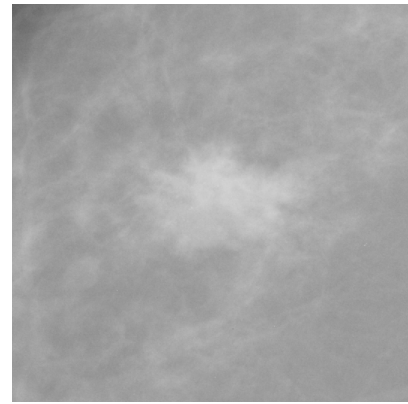
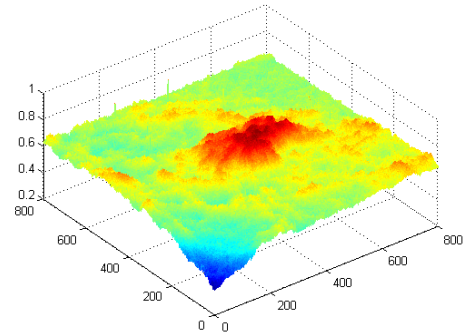


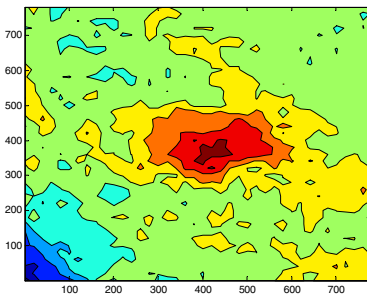
Fig. 2. Definition of the quantities (left) and iteration for different radii (right)



(a)



(b)



(c)

Fig. 3. a) A Region of Interest (ROI) containing a massive lesion, b) a 3D representation and c) the isolines of the same ROI

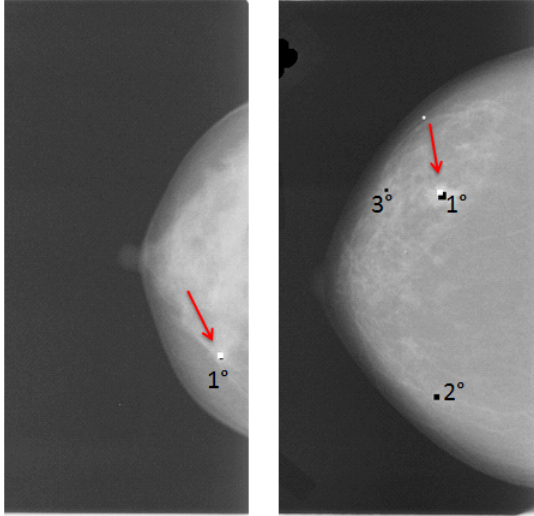


Fig. 4. Two examples of the algorithm output

Because masses have different diameters, from 3 to 40 mm, to identify all the possible masses the region of support have to be adapted to all the possible sizes of the masses. In particular, we know that the radius of typical masses is in the range [1.5 – 20] mm that for the considered not decimated image corresponds to [30 – 400] pixels.

- *Sort the results.* In a preliminary phase, we show to the radiologist more than one option. All pixels (i, j) which exceed a certain percentage p of the maximum value of $x(i, j)$ are presented to the radiologist and labeled in a ascending order: the more suspicious structures have the bigger markers. So we have to choose the parameter p so that:

$$x(i, j) > p \cdot \hat{x}$$

where \hat{x} is the maximum value of x for every i, j . Figure 4 shows two examples of the algorithm results: in the first case, the algorithm finds only one mass that corresponds to the radiologist marker and in the other case it finds 3 masses and the correct result is the mass labeled as the first. The final decision belongs to the radiologist that can examine with more attention the labeled regions. In paragraph IV there are some simulation results and the final choice of the parameter p .

- *Optimize.* All the markers that fall into adjacent centers are grouped to have a more readable result.

IV. SIMULATIONS AND RESULTS

A simulation on 48 images with one mass, taken from the DDSM database has been performed. We consider 4 different subtlety degrees (12 images for each subtlety degree), from 2 to 5. This rating has been specified by an expert radiologist in the DDSM database, according to the ACR BiRADS standard. We evaluate for each image:

- True Positives (TP): number of correctly signed areas;

- False Positives (FP): number of not founded masses;
- False Negatives (FN): number of signed areas that not corresponds to a mass.

In Fig. 5 are shown the results for all the images. It is possible to note that when the grid step and the value of p increase, the number of suspicious regions presented to the radiologist decreases but less regions are correctly classified. Using the

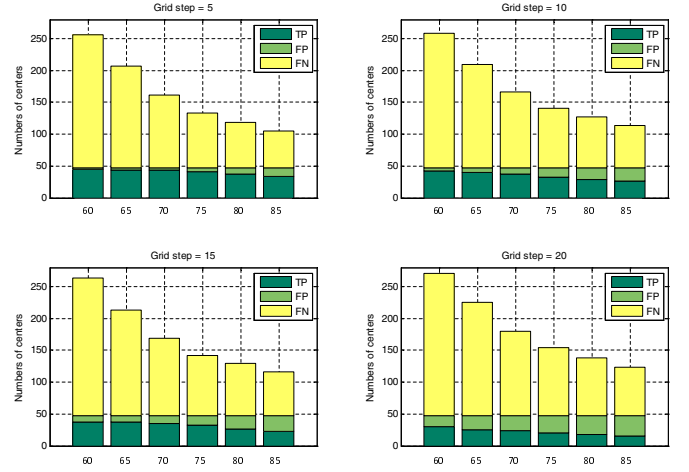


Fig. 5. TP, FP and FN for different grid steps and values of the parameter p

information about TP, FP and FN it is possible to evaluate the compactness and correctness of the algorithm, for different grid steps (form 5 to 20) and for different values of p (from 60 to 85), as shown in Fig. 6.

$$Compactness = \frac{TP}{TP + FN}$$

$$Correctness = \frac{TP}{TP + FP}$$

Basing on these results it is possible to choose the more suitable grid step. In particular, we note that compactness and correctness, for a grid step of 5 pixels, are higher with respect to the other grid steps, for all the possible values of the parameter p .

Because we want not only to present suspicious regions to the radiologist, but also to assign to them an index of suspiciousness, as explained in the previous paragraph, the found regions are labeled for different percentages p for the chosen grid step, as shown in Fig. 7. One can note how the mass is marked as the most suspicious area in the 30-35% of cases, depending on the value of p , as shown in the following table.

	p					
	60	65	70	75	80	85
1°	30%	32%	36%	36%	32%	32%
1°–3°	66%	64%	68%	70%	63%	60%
not found	4%	8%	8%	12%	19%	28%

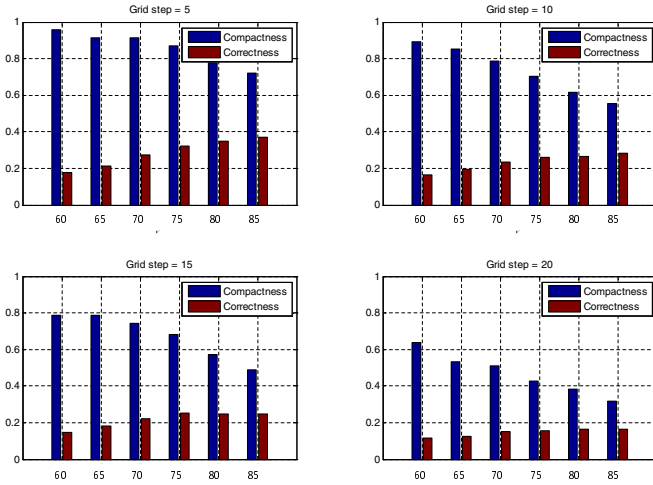


Fig. 6. Compactness and correctness for different grid steps and values of the parameter p

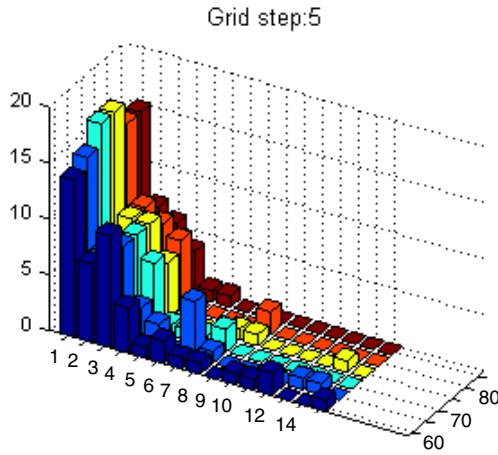


Fig. 7. Classification using the chosen grid step for different values of the parameter p

Moreover, the mass is found in the first three degrees of suspiciousness in the 60-70% of cases. It is not found in the 4-27% of cases.

These results permit some considerations about the choice of the parameter p . This is a difficult task because, varying p , compactness and correctness have an opposite behavior: when the compactness increases the correctness decreases. This means that if we want to fail less, we have to present to the radiologist more suspicious areas. But, presenting many suspicious areas can confuse the radiologist and makes the CAD no more useful. For these reasons we choose $p = 75$: it is the best trade off between compactness and correctness and the value that perform the higher number of success in the first three degrees of suspiciousness.

V. FUTURE WORKS

Although the algorithm performs good results, some aspects can be studied in deep and improved.

- A study about the evaluation of the algorithm performance with respect to the density of the breast tissue and the subtlety of the lesion has to be done.
- A study about the normal mammograms, where is proved that a mass is not present, has to be performed.
- A high gradient is naturally present on the profile of the breast and can influence the algorithm results. It has to be taken into consideration to improve the algorithm results.
- A rough classification step could be added so that structures, that surely are not masses, are not labeled.

We also have to integrate the identification step in a more complex CADx system [7]–[9] for the identification and classification of tumoral signs.

VI. CONCLUSION

We have implemented an automatic system for the identification of tumoral masses that recognize suspicious areas in the mammographic image searching for circular structures with a growing luminance towards the center. It could be a useful aid for the radiologist, specially if integrated in a whole CADx system, in the analysis of mammographic images, since this kind of images are very difficult to interpret.

ACKNOWLEDGMENT

The authors would like to thank the student Dora Sesti of the Department of Electronic Engineering of University of Rome “Tor Vergata” for her contribution in the validation of the procedure.

REFERENCES

- [1] J. Ferlay, F. Bray, P. Psiani, and D. M. Parkin, “Cancer incidence, mortality and prevalence worldwide iarc cancerbase no.5, version 2.0,” *IARCH Press*, 2004, Lyon, France.
- [2] L. J. Warren Burhenne, S. A. Wood, and C. J. D’Orsi, “Potential contribution of computer aided detection to the sensitivity of screening mammography,” *Radiology*, no. 215, pp. 554–562, 2000.
- [3] M. Bazzocchi, F. Mazzarella, C. Del Frate, F. Girometti, and C. Zuiani, “CAD systems for mammography: a real opportunity? a review of the literature,” *La Radiologia Medica*, vol. 112, no. 3, pp. 329–353, October 2007.
- [4] H. Kobatake and M. Murakami, “Adaptive filter to detect rounded convex regions: Iris filter,” *IEEE Pattern Recognition*, vol. 2, pp. 340–344, 1996.
- [5] University of South Florida, “University of south florida digital mammography home page,” 2000.
- [6] A. Mencattini, M. Salmeri, R. Lojaco, and M. Arnò, “Noise estimation in mammographic images for adaptive denoising,” in *EFOMP European Conference on Medical Physics (EFOMP ’07)*, Castelvechio Pascoli, Italy, September 2007.
- [7] M. Salmeri, A. Mencattini, G. Rabottino, and R. Lojaco, “Signal-dependent noise characterization for mammographic images denoising,” in *IMEKO TC4 Symposium (IMEKOTC4 ’08)*, Firenze, Italy, September 2008.
- [8] A. Mencattini, G. Rabottino, M. Salmeri, R. Lojaco, and E. Colini, “Breast mass segmentation in mammographic images by an effective region growing algorithm,” *Lecture Notes on Computer Science (Springer-Verlag)*, vol. 5259, pp. 948–957, October 2008.
- [9] A. Mencattini, M. Salmeri, R. Lojaco, G. Rabottino, and S. Romano, “Mammographic image analysis for tumoral mass automatic classification,” in *EFOMP European Conference on Medical Physics (EFOMP ’07)*, Castelvechio Pascoli, Italy, September 2007.