

Monte Carlo generation of benign and malignant masses inside mammograms

Georgios K. Leontidis

University of Athens, Department of Informatics and Telecommunications, Athens,
Greece Panepistimioupoli Zografou, Ilissia Email: geleonti@di.uoa.gr

Abstract

It is a fact that breast cancer is the most common cancer among women in the western world. The most reliable radiographic tool for the detection and diagnosis of breast cancer is mammography. Interpreting a mammographic image is a difficult task and prone to many errors, thus physicists-radiologists are trained very well. Many CAD (computer-aided detection) systems have been developed in order to provide assistance to radiologists by classifying any mammographic lesions as benign or malignant but their reliability is yet to be proved. The most accurate way of spotting masses on the breast, until now, is radiologist's opinion, making their training even more crucial. In this paper a method which simulates all the possible masses and place them on a mammography randomly using Monte Carlo technique, is proposed. After a comprehensive study of many different mammographic images, the final algorithm generates masses within a mammography, which in most cases a radiologist expert could not make the difference between a real mass and a generated one. The proposed method takes into account as many as possible variables for the generation of masses. There are definitely some improvements that could be developed in the future, mostly regarding the position of masses within the breast.

Keywords: Monte Carlo, Mammography, Simulation, Masses, Benign, Malignant

1. Introduction

According to the American Cancer Society, breast cancer is the most common type of cancer among women in the USA. 25% of all cancer cases diagnosed in US women account for breast cancer. Until today, mammography is regarded as the most reliable method for breast cancer detection and diagnosis. Depending on the morphology of the mass it can be either malignant or benign. The primary features that indicate malignancy are related to the tumor's density, size, shape and borders [2]. The boundary of the mass extends irregularly outward from the central mass. Most of the times, the shape of the mass can be used to discriminate between malignant and benign masses. Because a mammogram is a two-dimensional image of a three-dimensional breast, superposition of breast tissue produces sometimes patterns that appear like suspicious masses to a radiologist. It is a common situation where a lesion with typical malignant characteristics may represent benign lesions and vice versa. Figure 1[2] illustrates the morphological spectrum of breast masses frequently seen on mammograms.

According to the description of Gallager and Martin, the shapes of cancerous masses often fall into two classes: nodular or stellate [6]. They have found that the outline, the margins and the shape of the mammographic masses could be used in the classification of the masses as either benign or malignant. Biopsies used to actually

prove the malignancy or not, cause major mental and physical discomfort for the patients and many unnecessary expenses as well.

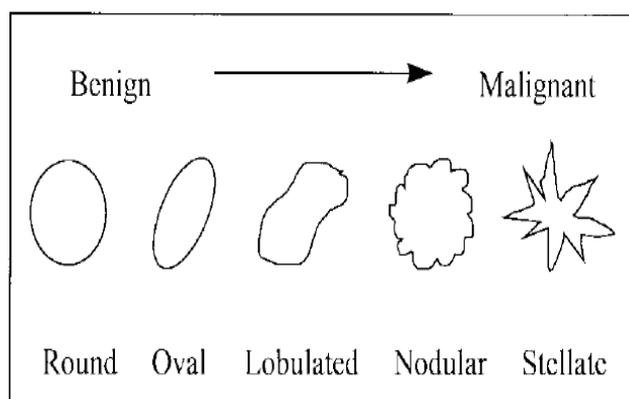


Fig. 1. Morphologic spectrum of mammographic

Recently a comprehensive study [1] and method was developed for the emulation of benign and malignant masses using four categories (benign: oval circumscribed and oval obscured; malignant: irregular ill-defined and irregular speculated). Their method was based on the measured mass characteristics taken from mammographic images from the university of Florida database. Their process consists of three stages: 1) creation of an array with elliptical rings; 2) modification of the initial array and 3) conversion of the modified elliptical array into pixel values. Their results were evaluated from an expert radiologist showing quite realistic appearances.

In this study, a new technique to simulate mammographic lesions was developed, using Monte Carlo methods to produce random variables for the different parameters that constitute a mammographic mass and also to place them on random positions within the breast tissue. Four characteristics were taken into account: a) the shape of the mass, b) the optical density of the mass, c) the size of the mass and d) the texture of the mass.

2. Monte Carlo

Monte Carlo is a method [5] that relies on repeated random sampling to obtain numerical results. This method is extremely useful for simulating systems with a high degree of freedom. It is used in Physical and mathematical problems for three main problems: a) optimization, b) numerical integration and c) generation of samples from a probability distribution. The Monte Carlo technique is mostly used when we want to produce random samples and not pseudorandom. The difference between pseudo randomness and randomness is that a pseudorandom sequence typically exhibits statistical randomness while it is entirely generated by a deterministic causal process, something that is not happening with a truly random process.

3. Method

The first step before starting developing the method was to study many different mammographic images from an existing Digital Database for Screening Mammography from the University of South Florida, where different kind of masses were evaluated and the main characteristics were extracted. The most challenging part is to make the simulated masses to look as normal as possible and also produce masses that are not clear if they are benign or malignant, just like in normal

mammographic images. Given the fact that in real mammographic images the size varies a lot, from very small to quite large masses, the range of possible mass sizes that can be generated via the Monte Carlo randomness has been optimized to generate very small difficult to distinguish masses as well as larger ones that are very obvious.

4. Shape extraction

The shape of each mass has been chosen to be circular, oval or random shape. The first two shapes are mostly used for benign masses but the latter, the random, is used mainly for malignant masses, since in real mammographic images the malignant masses have irregular shape. For that purpose for each of the three shapes, a mask was created, using as parameters the size and the shape, in order then to use the created mask to embed the texture and optical density variables. The size has been chosen to produce very small almost indistinguishable masses, as well as larger, about 50% of the breast size.

For the circular shape the process is pretty clear. But for the other two shapes, some arrangements were made, so as to enhance their randomness. As far as the oval shape is concerned, the direction of the oval shape has to change randomly, so the code does not create just an Oval shape, but it rotates it in a variety of angles in order to visualize every possible combination.

For the random shape, a random walk process was created, so as to make it totally random. As can be seen in figure 2, a square was created and then a coordinate system was used within the square, to divide it in four different areas. Then for each quadrant a random walk was executed, where the end of the first quadrant shape would be the beginning of the second quadrant shape and so on. As an example, if we suppose that the mass size is called size, then the first quadrant has the following coordinates (0,0), (0,size/2), (size/2,0), (size/2, size/2). So the shape begins to create at (X0,size/2) point, where X0 is a random value between 0 and size/2. The X0 can only move right or preserve its value, whereas the Y value can move up or preserve its value. For example if we started from (5,size/2) point, then for the next step we will have four different possibilities: a) (5, size/2), b) (5,size/2-1), c) (6, size/2), d) (6,size/2-1). The process terminates when we reach the second quadrant, where the X0 value would be size/2.

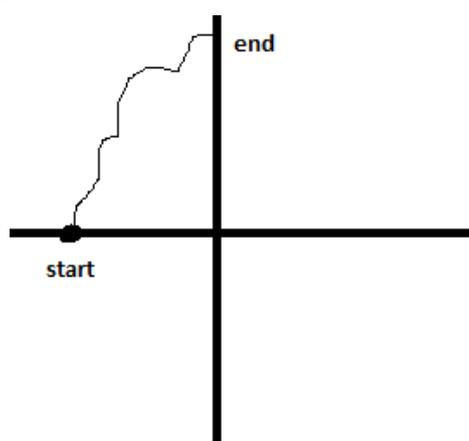


Fig. 2. Explanatory figure for the random walk



Fig. 3 A randomly shaped mass

Texture [3] is a set of metrics designed to quantify the perceived texture of an image. So the texture of an image gives us information about the spatial arrangement of color or intensities in an image or a specific region of interest.

Optical intensity[4] is the amount of light reflected by a certain point on the object in the direction of the viewer, which is multiplied, in most cases, by some constant factor that depends on the parameters of the image-forming system.

In this method both the values for texture and optical density have been experimentally adjusted to simulate the way that each generated mass suits with the image background so as not to look unreal. But even in that case, some real masses are very different from the background of the image, making them quite obvious and distinguishable. Even the texture of the mass can be optimized to fit the surrounding of the image, but in some real cases with malignant masses, the mass texture is much different from the surrounding tissue. The code that has been developed has been optimized to produce the whole range of optical density and texture, in order to cover all the possible situations.

In figure 3 we can see a generated a randomly shaped mass, which at the next stage will be filtered with a Gaussian filter making the mass shape smoother[7]. Then, before the mass is positioned on the mammographic image, it will be adjusted according to the randomly chosen optical density.

5. Position of the masses

Another challenging part is the positioning of each generated mass on random places within the breast. There are some complicated method that could be used to accurately detect the edges of each mammographic image like Canny detection or Hough transform, but it is not an essential part for my study.

The method used to constraint the places that a mass can be positioned, was developed by making a window, within each mass can be positioned depending on the size of the mass, as well as the size of the mammographic image. So the final dimensions of the window are chosen in a way that covers almost the whole area of the breast and avoiding the black background. So the algorithm selects random positions within that window and places the generated mass.

6. Comparison

In figure 4 we can see 2 simulated mammographic images from my method which are compared visually with another proposed method in literature[1] as well as with some real images (figure 5).

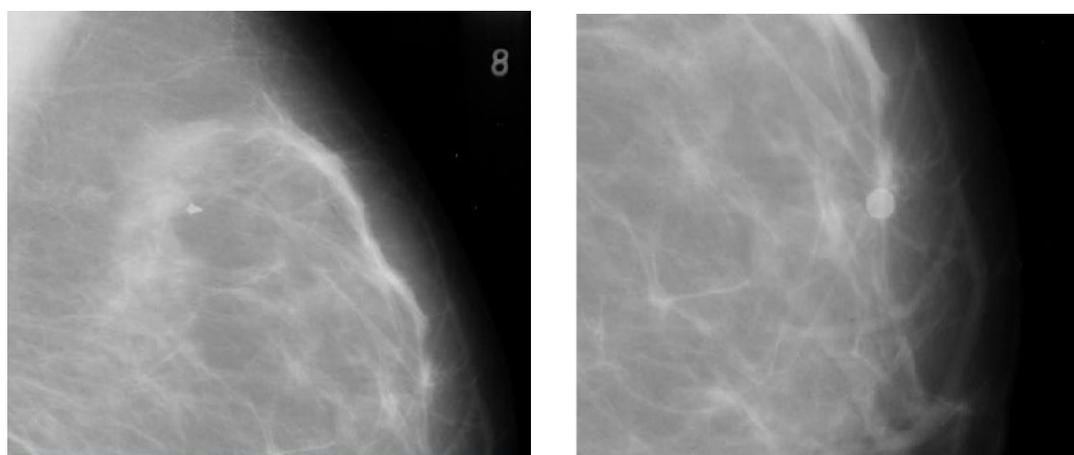


Fig. 4. Generated masses from the

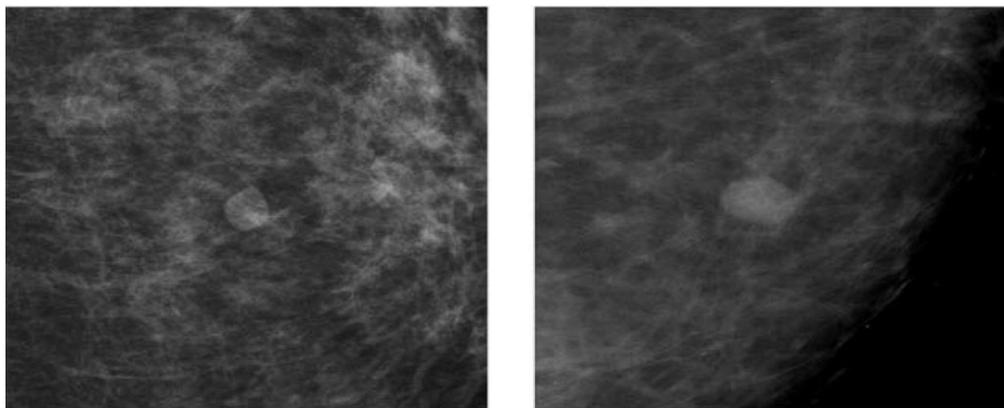


Fig. 5. On the left is a mass generated from another method in literature[1] and on the right is a real mass in a mammographic image

7. Discussion and conclusion

In this paper, a method to generate benign and malignant masses using Monte Carlo technique has been developed which aims to give radiologists a reliable method to train themselves without needing to find real mammographic images which are often inadequate and do not cover some rare and difficult situations where the masses are almost indistinguishable.

8. Future Research

A logical next step would be to improve the algorithm so as to detect the edges of each mammographic image breast and then position the mass within the margins of each breast. This requires slightly different approach to my subject since it requires mostly image processing techniques.

References

1. Robert Saunders, Ehsan Samei, Jay Baker, David DeLong, Simulation of mammographic lesions, *Academic Radiology* 08/2006; 13(7):860-70. DOI:10.1016/j.acra.2006.03.015
2. Lori Mann Bruce, Reza R. Adhami, Classifying Mammographic Mass Shapes Using the Wavelet Transform Modulus-Maxima Method, *IEEE transactions on medical imaging*, Vol. 18, No. 12, December 1999
3. Matthias Elter, Alexander Horsch, CADx of mammographic masses and clustered microcalcifications: A review, *Medical Physics*, Vol. 36, No. 6, June 2009
4. Berthold K.P. Horn, *image intensity understanding*, Massachusetts institute of technology, 1975
5. Nicholas Metropolis; S. Ulam, The Monte Carlo Method, *Journal of the American Statistical Association*, Vol. 44, No. 247. (Sep., 1949), pp. 335-341
6. H. S. Gallager and J. E. Martin, "The study of breast carcinoma by correlated mammography and subserial whole organ sectioning: Early observations," *Cancer*, vol. 23, pp. 855-873, 1969
7. Pei-Yung Hsiao, Generic 2-D gaussian smoothing filter for noisy image processing, *TENCON 2007 - 2007 IEEE Region 10 Conference*

Περίληψη

Ο Καρκίνος του μαστού αποτελεί την πιο συχνή μορφή καρκίνου στον δυτικό κόσμο. Το πιο αξιόπιστο εργαλείο για την ανίχνευση και τη διάγνωση του καρκίνου του μαστού αποτελεί η μαστογραφία. Η αξιολόγηση μιας μαστογραφίας αποτελεί απαιτητική διαδικασία και είναι επιρρεπής σε πολλά λάθη, καθιστώντας την εκπαίδευση των ακτινολόγων ιατρών ακόμα πιο ουσιαστική. Αρκετά υπολογιστικά συστήματα έχουν αναπτυχθεί που υποβοηθούν το έργο των ιατρών, όσον αφορά την ταξινόμηση οποιασδήποτε περιοχής εντός της μαστογραφίας, είτε ως καλοήθης είτε ως κακοήθης. Παρόλα αυτά, η αξιοπιστία των μεθόδων αυτών τίθεται εν αμφιβόλω. Η πιο ακριβής μέθοδος για τον εντοπισμό των μαζών στις μαστογραφίες, αποτελεί η διάγνωση του ιατρού, ακόμα και σήμερα. Στην παρούσα εργασία επιχειρείται η ανάπτυξη μιας μεθόδου που θα παράγει διαφορών ειδών μάζες, και θα τοποθετούνται εντός των ορίων του μαστού, χρησιμοποιώντας την τεχνική του Μόντε Κάρλο. Κατόπιν ενδελεχούς μελέτης της σχετικής βιβλιογραφίας αλλά και πολλών διαγνωσθέντων μαστογραφιών, ο αλγόριθμος που αναπτύχθηκε, παράγει πολλές διαφορετικές μάζες, με μεταβλητό μέγεθος, σχήμα, υφή και οπτική πυκνότητα, τις οποίες συνεργαζόμενη ακτινολόγος ιατρός έκρινε αληθοφανείς. Υπάρχουν περιθώρια βελτίωσης του αλγορίθμου στο μέλλον κυρίως στον τομέα τοποθέτησης των μαζών εντός του μαστού.