Computers in Biology and Medicine 43 (2013) 987–999

Contents lists available at SciVerse ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/cbm

Detection of masses based on asymmetric regions of digital bilateral mammograms using spatial description with variogram and cross-variogram functions

Daniel Rodrigues Ericeira [a], Aristófanes Corrêa Silva [a,*], Anselmo Cardoso de Paiva [a], Marcelo Gattass [b]

[a] Federal University of Maranhão—UFMA, Applied Computing Group—NCA, Av. dos Portugueses, SN, Campus do Bacanga, Bacanga 65085-580, São Luís, MA, Brazil
[b] Pontifical Catholic University of Rio de Janeiro—PUC-Rio, R. São Vicente, 225, Gávea, 22453-900 Rio de Janeiro, RJ, Brazil

ARTICLE INFO

Article history:
Received 12 April 2012
Accepted 29 April 2013

Keywords:
Medical image
Computer-aided detection (CAD)
Breast cancer
Mass detection
Cross-variogram function
Variogram function

ABSTRACT

A mammogram is an examination of the breast intended to prevent and diagnose breast cancer. In this work we propose a methodology for detecting masses by determining certain asymmetric regions between pairs of mammograms of the left and the right breast. The asymmetric regions are detected by means of structural variations between corresponding regions, defined by a spatial descriptor called cross-variogram function. After determining the asymmetric regions of a pair of images, the variogram function is applied to each asymmetric region separately, for classification as either mass or non-mass. The first stage of the methodology consists in preprocessing the images to make them adequate for registration. The following step performs the bilateral registration of pairs of left and right breasts. Pairs of corresponding regions are listed and their variations are measured by means of the cross-variogram spatial descriptor. Next, a model is created to train a Support Vector Machine (SVM) using the values of the cross-variogram function of each pair of windows as features. The pairs of breasts containing lesions are classified as asymmetric regions; the remaining ones are classified as symmetric regions. From the asymmetric regions, features are extracted from the variogram function to be used as tissue texture descriptors. The regions containing masses are classified as mass regions, and the other ones as non-mass regions. Stepwise linear discriminant analysis is used to select the most statistically significant features. Tests are performed with new cases for the final classification as either mass or non-mass by the trained SVM. The best results presented in the final classification were 96.38% of accuracy, 100% of sensitivity and 95.34% of specificity. The worst case presented 70.21% of accuracy, 100% of sensitivity and 67.56% of specificity. The average values for all tests were 90.26% of accuracy, 100% of sensitivity and 85.37% of specificity.

1. Introduction

Breast cancer is the irregular and uncontrolled growth of cells which originates in the breast tissue. A group of such cells may form an extra mass of tissue (tumor). According to the American Cancer Society (ACS), breast cancer is the commonest type of cancer among women and, in general, the second type of cancer which causes more deaths (behind lung cancer). According to the ACS, in western countries the cases of breast cancer have increased about 30% in the last 25 years. This increase may be explained by the improvement of the detection systems, which are able to detect cancer in its initial stages. Still according to the ACS, the rates of deaths by breast cancer have been falling steadily since 1990. This fact is also a result of better treatments and detection systems [1].

A mammogram is a breast exam used to prevent and diagnose breast cancer. This exam, which consists of a radiograph of the breasts, allows the early detection of cancer by showing lesions in their initial stage. Despite the fact that a mammogram exam is able to detect small cancer formations even years before they are tangible in physical exams, it is estimated that most lesions are not detected by the specialists who analyze them. The slow and gradual evolution of cancer can be identified more easily and earlier with the help of computer vision techniques associated to image processing, which can improve the efficiency of the preventive exams.

Mammograms of the left and the right breast of the same patient tend to present a high degree of symmetry [2]. Although there is clearly a wide variation in breast size and parenchymal

---

* Corresponding author. Tel.: +55 98 32729286; fax: +55 98 32728241.
E-mail addresses: daniel.ericeira@gmail.com (D.R. Ericeira), ari@dee.ufma.br (A.C. Silva), paiva@deinf.ufma.br (A.C. de Paiva), mgattass@tecgraf.puc-rio.br (M. Gattass).

0010-4825/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.compbiomed.2013.04.019

pattern, the breasts are generally symmetric structures with similar density and architecture. However, asymmetric breast tissue is encountered quite often. Asymmetric breast tissue is usually benign and secondary to variations in normal breast tissue, postoperative changes, or hormone replacement therapy. However, an asymmetric area may also indicate a developing malignant lesion [3].

Scutt et al. [2] observed that the group considered normal (i.e. did not develop cancer) presented volumetric asymmetry with mean of 52.99 ml, while the group which developed cancer presented mean of 63.17 ml. In the floating asymmetry (FA) analysis, which identifies small deviations from perfect symmetry in any type of organism with bilateral symmetry [4], the normal group presented mean breast FA of 2.5%, while the cancerous group presented mean of 2.7%. Hence, we can see that symmetry analysis can indicate possible anomalies. The regions where the breasts present greater disparities (asymmetries) may be pointed as suspect of having a neoplasm.

This work presents a methodology for the detection of masses by identifying asymmetric regions between mammograms of patients' left and right breasts. The asymmetric regions are detected by means of structural variations between corresponding regions, defined by a spatial dataset descriptor known as cross-variogram function. After determining the asymmetric regions in a pair of images, the variogram function is used in each individual suspect region, for classification as either mass or non-mass.

2. Related work

Many different methodologies have been proposed for the development of tools to assist in the early detection and diagnosis of cancer.

Costa et al. [5] compared the efficiency of the Support Vector Machine (SVM) to that of Linear Discriminant Analysis (LDA), classifying 200 regions of interest (ROIs) from mammogram images supplied by the MIAS database and 3600 ROIs from the DDSM. The results using MIAS were 85% and 97% for LDA and SVM, respectively. Using the DDSM, the authors achieved 89.2% and 99.6% for LDA and SVM, respectively.

For the segmentation of mass candidates, Oliveira et al. [6] proposed the use of Growing Neural Gas (GNG) and SVM combined with Ripley's K function to detect masses in mammograms. Using 997 images from the DDSM, they obtained a sensitivity of 89.3%, 0.93 false positives per image and 0.02 false negatives per image. Also, Nunes et al. [7] proposed a methodology for the detection of masses that uses the K-means clustering method and the template matching technique. They used 650 mammogram images from the DDSM and achieved an average accuracy of 83.94%, sensitivity of 83.24%, and 84.14% of specificity, with a rate of 0.55 false positives per image and 0.17 false negatives per image. Pereira et al. [8] analyzed the performance of the random forest method for the detection of masses using information extracted through the ridgelet transform from craniocaudal and oblique mediolateral views. They used the DDSM, from which 270 regions of interest containing masses and normal tissues were selected. This methodology achieved a performance of 94.4% of sensitivity, 96.9% of specificity and 91.8% of accuracy.

Sahba et al. [9,10] proposed schemes for detecting masses based on the idea of clustering the pixels of an image by using a mean shift algorithm. Both works used the MIAS database. The results obtained in one of the studies [9] were a true positive detection rate of 90% with a false positive fraction of 1.9 per image, and an estimated Az value of the Receiver Operating Characteristic (ROC) curve of 0.88. In the other study [10], the authors obtained a true positive detection rate of 88% with a false positive fraction of 2.1 per image, and an Az value of the ROC curve of 0.86. With a similar objective, morphological component analysis was introduced by Gao et al. [11], who decomposed a mammogram into a piecewise-smooth and a texture component. The proposal was evaluated using the DDSM database, achieving a sensitivity of 99% for malignant masses, 88% for benign masses, and 95.3% in all types of cases. Finally, Terada et al. [12] proposed a method which consists of applying mean shift segmentation to detect masses in mammograms. After the segmentation, the concentration of gradient vectors is computed using Iris Filter and then mass regions are detected. In the results, a sensitivity of 81% was obtained, with 5.0 false positives per image, and 75% of the masses were detected with an Area Overlap Measure (AOM) of more than 60%.

Zheng et al. [13] used Gabor features. After a preprocessing stage, they applied a Circular Gaussian Filter (CGF) that makes the masses appear as a bright region, extracted by means of adaptive thresholding. Thus, a set of Gabor-filtered images with edge histogram descriptors (EHD) was extracted. These descriptors were used with the fuzzy C-means clustering technique and k-nearest neighbor (KNN) to classify the suspicious regions. Using the DDSM database, they achieved a true positive rate of 90% and 1.21 false positives per image in mass detection.

The relation between the symmetry of the breasts and the occurrence of cancer has been the object of analysis in several studies. Scutt et al. [2] presented an initial observation of this connection. After comparing 250 patients with cancer and 250 healthy patients with the same age, they concluded that the group with cancer presented higher asymmetry (mean of 87.39 ml) than the healthy group (mean of 59.27 ml). More recently, the same authors [14] verified this relation between volumetric asymmetries and breast cancer with 252 healthy women who did not develop cancer and 252 women who developed the disease. It was observed that the group of women who did not develop cancer presented mean volumetric asymmetry of 52.99 ml, while the group that developed cancer presented mean of 63.17 ml.

Methods to analyze the differences between pairs of corresponding mammograms and identify suspect regions were proposed by Sallam et al. [15], Georgsson et al. [16] and Wu et al. [17], who achieved an improvement in accuracy by 15% to 20% while reducing the number of false positives. Also, works seeking to detect tumors by means of the bilateral registration of breasts and asymmetry analysis have been developed. Lau et al. [18] proposed one of such methods by searching for intense structural asymmetries between left and right breast mammograms. First the images were aligned. Next, each asymmetry was evaluated considering brightness, directionality and roughness. The method achieved accuracy of 92.3%. Wang et al. [19] developed an automated scheme to detect breast tissue asymmetry depicted in bilateral mammograms and predict the likelihood (or the risk) of women having or developing breast abnormalities or cancer. The authors used a proprietary dataset of full-field digital mammography images, with 200 cases. The asymmetry in breast tissue was identified by means of the differences between related features computed from bilateral images, and using a neural network classifier. The results obtained were of 0.754 for AUC, and at 90% specificity, the classifier yielded 42% sensitivity.

Tzikopoulosa et al. [20] presented a segmentation and classification scheme for mammograms based on breast density estimation and detection of asymmetry, using the miniMIAS database. The asymmetry is characterized by the difference between statistical features computed from the pair of mammograms and classified by a one-class SVM, achieving a success rate of 84.47%.

Stamatakis et al. [21] proposed two methods for comparing left and right breasts. The first method defines an intensity differentiation threshold which determines corresponding areas of the