Generation of clustered microcalcifications' atlases for benign and malignant cases

Ioannis I. Andreadis, George M. Spyrou, Panos A. Ligomenides and Konstantina S. Nikita

Abstract— **Breast microcalcifications are one of the most important mammographic findings related to the existence of the breast cancer. Radiologists usually characterize microcalcifications based on their morphologies, the distribution within the cluster they form, the shape of the cluster and its relative location inside the breast. In this study, we focus on the latter factor and we study its effect on the probability of malignancy. The main purpose of our study is to generate probabilistic breast cancer atlases for clusters of microcalcifications in order to visualize the influence of cluster location on cancer probability. We propose a framework for the generation of such atlases, including segmentation of important breast landmarks and projection of different clusters of microcalcifications on a reference breast shape. The generation of the atlases is implemented using mammograms from the Digital Database of Screening Mammography. The obtained probabilistic atlases reveal specific areas in the breast of higher occurrence of clusters and higher risk of malignancy.**

I. Introduction

COMPUTER AIDED DIAGNOSIS (CAD_x) systems for the diagnosis of clusters of microcalcifications (MCs) have been recently introduced to help the radiologists in their diagnostic process by supporting them with a reliable second opinion [1]. The basic pipeline implemented in a CAD_x system includes the following steps [2]: segmentation of the MCs from the surrounding tissue, feature extraction to compute objective features to describe the findings and classification of the findings based on the features extracted in the previous level.

The most critical phase of a CAD_x system is probably the feature extraction phase as the features extracted in this step have to mimic radiologists' strategy and will form the basis for the final categorization of the finding. Many studies have been focused on the development of various feature extraction methods for the analysis of MCs, including features describing the shape of MCs and the morphology of the whole cluster [3], features concerning the distribution of the MCs inside the cluster [4] and finally features related to the texture of the background tissue of the Region of Interest (ROI) in which the MCs are embedded [5]. In a previous work [6], we have performed a large scale comparative study of different image feature types to indicate the most representative(s) for the discrimination between benign and malignant clusters.

However, there is a specific feature type which has not been under consideration, concerning mainly the location of the cluster inside the breast. Radiologists consider the location of the cluster an important indicator, as it has been shown that the majority of malignant clusters are more often located at the upper outer quadrant of the breast [7]. Few studies have incorporated location features in their CAD_x pipeline. Veldkamp et al. [8] used the relative distance of the cluster from the pectoral muscle and the breast's skin line as measures for the discrimination of clusters of MCs. Russakoff et al [9] generated a set of probabilistic breast cancer atlases to reveal areas of higher risk of breast cancer. However, in the latter study only ROIs including cancerous masses have been used for the generation of the atlases. ROIs including clusters of MCs have not been under consideration. The main reason for the limited use of location features is that their computation requires segmentation of important breast landmarks, such as breast skin line, pectoral muscle and nipple, which of course is not a trivial problem of its own.

In this study, we propose a fully automated framework for the generation of probabilistic atlases based on the relative location of clusters of MCs inside the breast. For the generation of the atlases, we used images from the Digital Database of Screening Mammography (DDSM), which is the largest publicly available database of mammograms [10]. We present briefly the algorithms developed for the segmentation of breast area and the pectoral muscle and we discuss in detail the procedure followed for the atlas generation. We then present the atlases obtained following the aforementioned procedure, which reveal areas of higher occurrence of clusters inside the breast, as well as areas of higher probability of malignancy.

II. Methods and Materials

A. Data Collection

As mentioned above, we used cases provided by the DDSM which contains digitized mammograms of approximately 2600 cases. Almost all cases contain images

Manuscript received September 20, 2013.

I. I. Andreadis is with the the School of Electrical and Computer Engineering, National Technical University of Athens, 15573, Athens, Greece (e-mail: iandr@ biosim.ntua.gr).

G. M. Spyrou is with the Biomedical Research Foundation Academy of Athens, Athens, Greece (e-mail: gspyrou@bioacademy.gr).

P. A. Ligomenides is with the Academy of Athens, Athens, Greece (pligo@academyofathens.gr).

K. S. Nikita is with the School of Electrical and Computer Engineering, National Technical University of Athens, 15573, Athens, Greece (corresponding author phone: +30-2107722968; fax: +30210 7722320; e-mail: knikita@ece.ntua.gr).

978-1-4799-3163-7/13/$31.00 ©2013 IEEE

from both mediolateral (MLO) and craniocaudal (CC) views and information on the boundaries of the annotated regions, where the lesion has been detected and biopsy has been done. We collected all the cases of the database that contained annotated clusters of MCs and we ended in a dataset consisted of 843 mammograms from the CC view and 846 mammograms acquired from the MLO view, containing both benign and malignant clusters. Only few cases have been excluded from the study, such as cases that contained clusters diffused throughout the whole breast area without a local annotation specified by radiologists or cases where segmentation of mammograms was not feasible. Also, certain cases from the folders cancer_02 and cancer_03 of the database, which have been digitized using a DBA digitizer (M2100 ImageClear) at 42 μm, required a preprocessing stage using a median filtering in order to remove the noise that existed on the images.

B. Breast Segmentation

Although a great variety of breast shapes and sizes exists, the task of generating an atlas requires only the determination of breast's edges. As a result, the first step in the proposed framework addresses the task of segmenting the breast area from the background in a mammogram. To this end, we applied a segmentation algorithm presented in [11]. It is a simple and fast algorithm which is mainly based on three steps, including analysis of the mammogram's histogram, application of morphological filters and smoothing of breast's border. The method has been evaluated on cases from the DDSM database and presented promising results. We applied the specific algorithm to segment the breast area on all the available mammograms, both from CC and MLO view. The segmentation performance was satisfactory for the majority of the considered cases. In a few cases only, we had to remove manually some segmented areas that clearly did not belong to the breast.

Apart from the breast's silhouette, another important landmark that has to be located is the pectoral muscle, which is a triangular area appeared in the left upper corner of a mammogram and it is visible exclusively on mammograms acquired from the MLO view. The algorithm proposed is described briefly hereafter: after the proper segmentation of the breast reported previously, we choose an initial line on the upper left area of the image in order to locate on that line the boundary pixel between the pectoral muscle and the breast area. The boundary pixel is considered the pixel which appears the greatest decrease in the average brightness throughout the line. The criteria followed for the choice of the initial line and the choice of the boundary pixel were proposed in [12]. The specific process is repeated from all the lines of the image above and below the initial line, by limiting the search only in neighboring pixels of the boundary pixels defined on previous lines, since the border of the pectoral muscle is usually a straight line or at least a small curve. The process is repeated until all lines in the segmented area are considered or the muscle's border is fully defined when the boundary reaches the left edge of the image. At the end, due to the fact that the muscle's boundary may be coarse with small variations, we apply a least squares method to obtain a smooth and more precise pectoral muscle boundary.

Finally, for the determination of nipple location, since nipple is hardly visible on mammograms, we did not apply a segmentation algorithm. Instead, we used knowledge concerning the geometry of the breast to determine a point from the breast border which should be close to the actual location of the nipple. For the images from CC view, we considered as nipple the point of the breast border with the largest distance to the chest side. On the contrary, in images from the MLO view, the point of the breast border with the largest distance from the pectoral muscle is considered as the nipple.

C. Cluster Position Features

After locating the above landmarks in the mammogram, we proceed on finding the relative location of the cluster inside the breast. The exact annotation and location of the cluster in the mammogram are provided in the files of the DDSM database. Some important location features used to represent the relative location of a cluster inside the breast are presented in fig. 1.

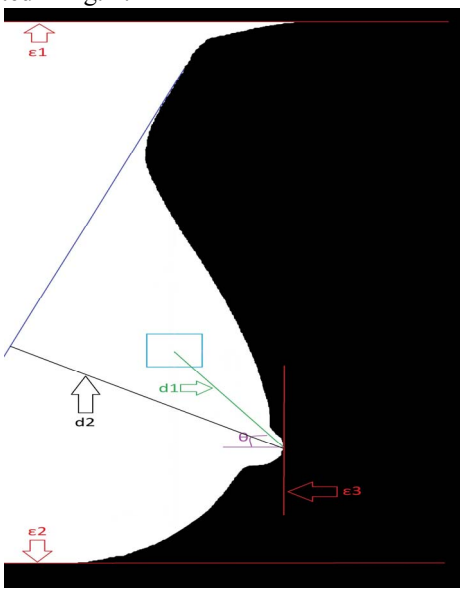

Fig. 1. Description of the relative location of the cluster inside the breast area

The current figure concerns a mammogram from MLO view, but the same principles are followed in the case of CC view. Therefore, for both CC and MLO mammograms we are able to compute breast's length, width and area exploiting the output image of the breast segmentation algorithm (red lines ε1, ε2, ε3). We are also able to compute the distance of the centre of the cluster (represented as a rectangular region in light blue) to the nipple of the breast (d1), as well as the angle that is formed between the current line and the horizontal level (angle θ). For CC mammograms, the distance of the centre of the cluster to the breast is computed relatively to the distance from the chest wall to the nipple, while for MLO mammograms the corresponding distance is taken relatively to the distance

between the nipple and the pectoral muscle (d2). Such features provide the appropriate information for the relative location of the cluster inside the breast.

D. Atlas Generation

In this paragraph, we discuss the main steps proposed for the generation of breast atlases. First of all, we have to define the atlas shape for each different view (CC or MLO). The breast silhouettes of two random cases of the DDSM database have been chosen to be the shape-containers on which the remaining ROIs will be properly scaled and projected. The atlas is simulated as a table of counters, one for each pixel, all initialized to zero. For each mammogram under consideration we performed the following steps:

-- We apply the segmentation algorithm to extract the breast's skin line. In case of mammogram from MLO view, we apply also the pectoral muscle segmentation algorithm.

-- We extract from the files of DDSM the exact location and annotation of the ROI containing the suspicious cluster.

-- We compute the location features discussed in paragraph C of the current section to quantify the relative location of the cluster inside the breast.

-- Using the features above, we project the ROI on the breast atlas on the corresponding location into the atlas shape. For each pixel of the projected ROI, the counter used for the occurrence to the specific location is incremented by one.

-- The values of all the counters of the shape-container are normalized to generate the final atlas.

We have to note here that the projected ROI does not have the same shape as initially annotated by the radiologists. Instead, a circular ROI of standard radius is projected, in order to focus only on the location of the cluster and eliminate the bias created by the size of the ROI as indicated by the radiologists.

III. RESULTS

A. Cluster Occurrence Maps

We followed the procedure discussed in the previous section to generate cluster occurrence atlases for both CC and MLO views. We used all the 843 cases for the former view and the 846 for the latter ending in the cluster occurrence maps presented in fig. 2 and 3 respectively. The atlas shape is superimposed in white, while the red color indicates higher occurrence of clusters, followed by yellow and green color indicating smaller occurrence. The values of occurrence of both figures are normalized to the maximum value of occurrence observed and therefore the red color corresponds to value 1. We have divided the image in four different areas based on the centre of the breast atlas (A, B, C and D). We may notice that there exist indeed areas of the breast with higher occurrence of clusters of MCs. In the case of CC view, the higher values of occurrence are observed in the two upper parts of the atlas (A and B) above the line that connects the nipple to the chest (external quadrant of the breast).

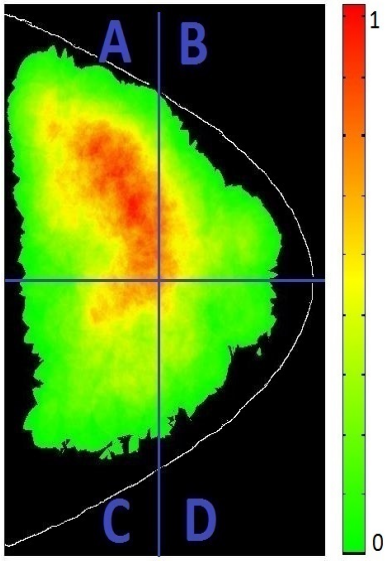

Fig. 2. Occurrence map for CC view

Similarly, for the MLO view, the majority of clusters are projected on the two parts above the line that connects the nipple to the pectoral muscle (upper quadrant of the breast). The specific observations are consistent with the observation that a great percentage of lesions are located on the upper outer quadrant of the breast [7].

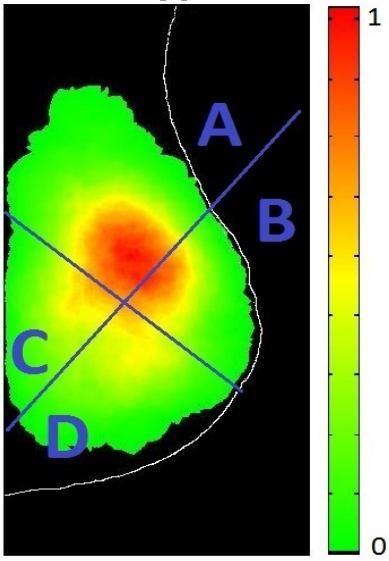

Fig. 3. Occurrence map for MLO view

B. Cluster Discrimination Maps

This time we focus on the discrimination between benign and malignant clusters in order to investigate whether there are areas more susceptible to cancer than others. Therefore, we separated the mammograms from CC view on two groups based on the pathology status of the annotated clusters (425 benign clusters/418 malignant clusters) and we generated an occurrence atlas for each group. Based on the occurrence values of these two maps, we proceeded on the generation of a probabilistic map between benignity and malignancy.

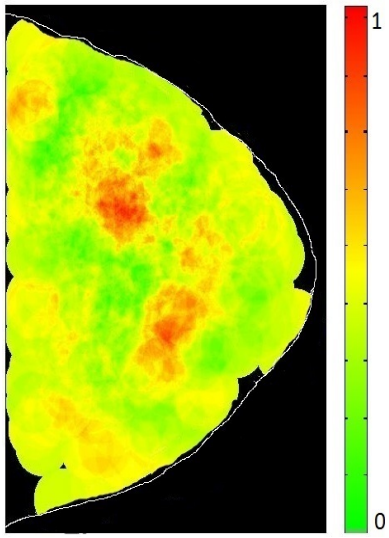

Fig. 4. Probabilistic map for CC view

For each pixel of the occurrence maps, we calculate probabilities by dividing the number of benign or malignant counters to the total occurrence at that pixel (sum of benign and malignant counters). As a result, the probabilistic map presented in fig. 4 is generated for the CC view. Red color (value=1) indicates higher probability of malignancy, while green color (value=0) the opposite. Yellow color implies equal probabilities. We repeated then the same process for the mammograms from the MLO view. We used 437 benign cases and the rest 409 were malignant. Fig. 5 presents the corresponding probabilistic map for the MLO view.

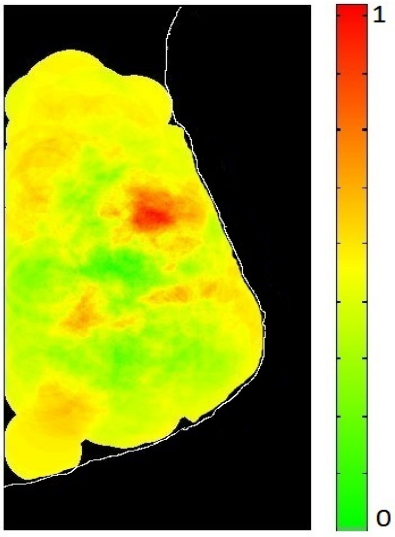

Fig. 5. Probabilistic map for MLO view

In both cases, we may observe that the majority of the breast area is presented in yellow color. This fact indicates that many areas present equal probabilities to host a benign or malignant cluster of MCs. However, in both maps we may notice areas in red color, therefore areas more susceptible to cancer, while there are areas in green color which appear increased probability towards benignity.

IV. Discussion

In this paper, we introduce the use of topological information for cluster of microcalcifications in order to investigate their potential to support diagnosis. We propose a framework including segmentation and determination of important breast landmarks, computation of distances of the centre of the cluster to those landmarks and finally a pipeline leading to the generation of probabilistic maps. This work is introductory to the investigation of location features since we establish the main principles for their use and the generation of the breast atlases. The preliminary results are quite promising, since breast areas of higher occurrence and higher risk for malignancy are revealed. The specific fact indicates that the location of a cluster inside the breast may provide information of high importance for the discrimination of its pathology status. Our future plans include the adoption of location features in the diagnostic process of a CAD_x system, expecting to provide quantified topological information that will optimize diagnosis. The generated maps will be used to enhance the classification results of new considered cases, as a new cluster can be projected on the atlas and a priori probabilities for cancer can be estimated.

References

[1] M. Elter and A. Horsh, "CADx of mammographic masses and clustered microcalcifications: a review," *Med. Phys.*, vol. 36, pp. 2052–2068, 2009.

[2] M. Giger, H. Chan and J. Boone "Anniversary paper: history and status of CAD and quantitative image analysis: the role of medical physics and AAPM," *Med. Phys.*, vol. 35, pp. 5799–5820, 2008.

[3] M. Kallergi "Computer-aided diagnosis of mammographic microcalcification clusters," *Med. Phys.*, vol. 31, pp. 314-326, 2004.

[4] A. Papadopoulos, D. Fotiadis, A. Likas, "Characterization of clustered microcalcifications in digitized mammograms using neural networks and support vector machines," *Artif. Intell. Med.*, vol. 34, pp. 141-50, 2005.

[5] A. Karahaliou, S. Skiadopoulos, I. Boniatis, P. Sakellaropoulos, E. Likaki, G. Panayiotakis, and L. Costaridou, "Texture analysis of tissue surrounding microcalcifications on mammograms for breast cancer diagnosis," *Br. J. Radiol.*, vol. 80, pp. 648-656, 2007.

[6] I. Andreadis, G. Spyrou and K. Nikita, "A comparative study of image features for classification of breast microcalcifications," *IOP Measurement Science and Technology*, vol. 22, no. 11, 2011.

[7] American College of Radiology (ACR), *ACR Breast Imaging Reporting and Data System, Breast Imaging Atlas*, 4th Edition, Reston, VA. USA, 2003.

[8] W.J. Veldkamp, N. Karssemeijer, J. D. Otten, and J.H. Hendricks, "Automated classification of clustered microcalcifications into malignant and benign types," *Med. Phys.*, vol. 27, pp. 2600-2608, 2000.

[9] D.B. Russakoff and A. Hasegawa, "Generation and Application of a Probabilistic Breast Cancer Atlas," In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006, LNCS, 4191, pp. 454-461, 2006.

[10] M. Heath, K. Bowyer, D. Kopand, R. Moore and W. Kegelmeyer. "The Digital Database for Screening Mammography,"in 2001 Proc. of the 5th IWDM, Yaffe M. Medical Physics Publishing, pp. 212-218.

[11] T. Ojala, J. Nappi, and O. Nevalainen, "Accurate segmentation of the breast region from digitized mammograms," *Comput. Med. Imag. Graph.*, vol. 25, pp. 47-59, 2001.

[12] L. Wang, M. Zhu, L. Deng and X. Yuan, "Automatic pectoral muscle boundary detection in mammograms based on Markov chain and active contour model", *J Zhejiang Univ-Sci C*, vol. 11, no. 2, pp. 111-118, 2010.