Neurocomputing 152 (2015) 388–402

Contents lists available at ScienceDirect

Neurocomputing

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

# A new automatic mass detection method for breast cancer with false positive reduction

Xiaoming Liu [a,b,c,d], Zhigang Zeng [a,b,*]

[a] School of Automation, Huazhong University of Science and Technology, Wuhan 430074, China
[b] Key Laboratory of Image Processing and Intelligent Control of Education Ministry of China, Wuhan 430074, China
[c] College of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan 430081, China
[d] Hubei Province Key Laboratory of Intelligent Information Processing and Real-time Industrial System, Wuhan 430081, China

ARTICLE INFO

*Article history:*
Received 19 November 2013
Received in revised form
27 August 2014
Accepted 16 October 2014
Communicated by H. Zhang
Available online 10 November 2014

*Keywords:*
Mass detection
Breast cancer
Level set
Local binary pattern
Support vector machine

ABSTRACT

Mass localization is a crucial problem in computer-aided detection (CAD) system for the diagnosis of suspicious regions in mammograms. In this paper, a new automatic mass detection method for breast cancer in mammographic images is proposed. Firstly, suspicious regions are located with an adaptive region growing method, named multiple concentric layers (MCL) approach. Prior knowledge is utilized by tuning parameters with training data set during the MCL step. Then, the initial regions are further refined with narrow band based active contour (NBAC), which can improve the segmentation accuracy of masses. Texture features and geometry features are extracted from the regions of interest (ROI) containing the segmented suspicious regions and the boundaries of the segmentation. The texture features are computed from gray level co-occurrence matrix (GLCM) and completed local binary pattern (CLBP). Finally, the ROIs are classified by means of support vector machine (SVM), with supervision provided by the radiologist's diagnosis. To deal with the imbalance problem regarding the number of non-masses and masses, supersampling and downsampling are incorporated. The method was evaluated on a dataset with 429 craniocaudal (CC) view images, containing 504 masses. Among them, 219 images containing 260 masses are used to optimize the parameters during MCL step, and are used to train SVM. The remaining 210 images (with 244 masses) are used to test the performance. Masses are detected with 82.4% sensitivity with 5.3 false positives per image (FPsI) with MCL, and after active contour refinement, feature analysis and classification, it obtained 1.48 FPsI at the sensitivity 78.2%. Testing on 164 normal mammographic images showed 5.18 FPsI with MCL and 1.51 FPsI after classification. Experiments on mediolateral oblique (MLO) images have also been performed, the proposed method achieved a sensitivity 75.6% at 1.38 FPsI. The method is also analyzed with free response operating characteristic (FROC) and compared with previous methods. Overall, the proposed method is a promising approach to achieve low FPsI while maintaining a high sensitivity.



## 1. Introduction

Breast cancer is one of the leading causes of cancer death in the world, it is reported that in 2013, the number of new breast cancer patients in USA will increase by 296,980, and about 39,620 among them will die for the disease [1]. Although some imaging modalities, such as magnetic resonance imaging (MRI) and sonography, are currently being investigated to improve sensitivity and specificity of breast cancer diagnosis, X-ray mammography is still the most prevalent imaging procedure for the early detection of breast cancer [2]. Early diagnosis can increase the survival rate.

There are several types of abnormality in mammogram, among them, micro-calcifications and masses are the most common types. A mass typically appears on a mammogram as a relatively dense region, whose properties could be characterized using local density, gradient, texture, and other measures. Mass detection and segmentation is an important step in the analysis of mammograms. The task is not easy since masses are usually embedded and surrounded by different structures, such as parenchyma, muscle, blood vessels, and fibrous tissues.

Researchers have developed several methods for the task, see [3] for a recent review. Some work concentrate on segmenting mass from a small patch of the mammogram, and other work can detect and segment mass in full mammograms. The mass segmentation approaches can be divided into two major categories [3]: supervised segmentation and unsupervised segmentation. The unsupervised approach can be further classified into region-based,

* Corresponding author at: School of Automation, Huazhong University of Science and Technology, Wuhan 430074, China.
*E-mail address:* hustzgzeng@gmail.com (Z. Zeng).

http://dx.doi.org/10.1016/j.neucom.2014.10.040
0925-2312/

contour-based, and clustering-based methods. The supervised approach can be model based method or a classifier is trained to distinguish true masses from non-masses.

Since our method belongs to the supervised approach, we introduced some typical supervised methods below. Bator et al. [4] improved template based mass detection method by using the template with the brightness distribution generated by an evolutionary algorithm. Huo et al. [5] proposed a region growing based lesion extraction method. Their method was performed on the $512 \times 512$ regions which are manually cropped from mammographic images. Abrupt change in size and circularity were used to determine the termination of region growing. Li et al. [6] proposed a two step method for mass segmentation. Regions of interest are first extracted from the images by adaptive thresholding, and the initial segmentation is further refined with a modified Markov random field (MRF) model based method. Iterated Conditional Modes (ICM) was used to solve the MRF estimation problem. The parameter values such as window size used in the initial segmentation need careful tune. Timp et al. [7] utilized both edge information and a prior knowledge about the gray level distribution of the region of interest around the mass, and obtained the contour by dynamic programming optimization technique. Sheshadri et al. [8] detected masses based on morphological watershed transformation. They preprocessed the image to attenuate the curvilinear structures present in a mammogram, such as blood vessels, veins, milk ducts, speculations and fibrous tissue. Rojas et al. [9] presented an automatic mass detection method. They improve the image contrast based on local statistical measures of mammograms firstly, then regions are segmented via thresholding at multiple levels, and features are computed from each segmented region for false positive reduction. Szekely et al. [10] used a three-step hybrid system to detect masses. A global segmentation method is first applied to find regions of interest with decision tree classification and a multiresolution MRF model. Then, three local segmentation methods (radial gradient index method, histogram method and a constrained region growing) were used and combined to extract the masses. Finally, features were extracted and classified to decide whether a patch contains a mass or not. Zheng et al. [11] combined several artificial intelligent techniques with the discrete wavelet transform (DWT) to detect masses. They used fractal dimension analysis to determine the initial locations of the regions suspicious for cancer, and used dogs-and-rabbits clustering algorithm to initiate the segmentation at a subband of the DWT decomposition. Finally, a tree-type classification strategy was applied to determine whether a region contains mass. Eltonsy et al. [12] detected the masses based on the presence of concentric layers surrounding a focal area with suspicious morphological characteristics and low relative incidence in the breast region. Wang et al. [13] investigated structured SVM for mass detection for image blocks. Ramirez et al. [14] investigated mammographic region classification using statistical and multi-resolution features. Wavelet energy measures are extracted from multi-level decomposition, properties of Shannon and Tsallis entropy are compared. The method obtained a sensitivity varying from 86.67% to 91.67%, with specificity varying from 82.50% to 100.00%. Want et al. [15] investigated extreme learning machine for the mass detection. They showed that extreme learning machine has some advantages in training speed and in accuracy compared to SVM.

Among the steps of the mass detection procedure, the accurate segmentation of suspicious mass region is an important step, since it affects the extracted features, and influences the accuracy of the classifier used to separate true masses from non-masses. Active contour models (snakes) [16] have attracted much attention for image segmentation in recent decades. An active contour model minimizes an energy function associated with a deformable contour, which consists of an internal energy and an external energy. The internal energy usually controls the smoothness and elasticity of the contour, and the external energy drives the contour to object boundaries. Recently, a few works have applied this technique to the task of mass segmentation in mammographic images. Brake et al. [17] segmented mass by a discrete active contour with image gradient based external energy. Sahiner et al. [18] used a discrete active contour model with both the gradient and region information for the mass segmentation. In these works, the contour was represented by the vertices of an $N$-point polygon and each vertex was tracked during the process, the representation can not deal with topology merging and splitting. Yuan [19] used a region based geometric active contour model with level set representation to segment masses, which allows the model to handle the topological changes naturally, and proposed a dynamic stopping criterion to terminate the contour evolution. To cope with nonuniformity in the background distribution, Yuan used a background trend correction. A shortcoming of all the above active contour methods is that they are global methods, meaning that they make strong assumptions on the intensity distribution about background (pixels very far away from the mass region are used in background representation), which is not practical for mammogram images. For example, in Chan-Vese model [20], they require all the pixels in the breast except the mass have a constant (or smooth) grayscale, which is certainly a too strict assumption, since the mammographic image consists of several structures, such as parenchyma, muscle, blood vessels, fibrous tissues, and they have different grayscales. In recent years, several approaches have been proposed to overcome the shortcoming of traditional global methods [21,22], but few of them have been applied to segment masses from mammographic images.

In this paper, a new full automatic mass detection method is proposed. The framework of the proposed method is shown in Fig. 1, following steps of a typical supervised mass detection method. For the suspicious location and segmentation step, we integrated the MCL technique [12] and NBAC [23]. The local level set based mass detection approach has the following two advantages: (1) it is easy to explain the result and can utilize prior knowledge since the parameters in the initial segmentation step (modifed MCL) can be adjusted with observations on the given images; (2) it can be applied not only to cropped subimages, but also to the whole mammogram images, due to the localized active contour. For the feature extraction step, both Geometry and texture features are extracted. For the first time, the complete local binary pattern texture feature [24] is introduced for the mass detection problem. With the extracted features, a SVM classifier is trained to separate true mass from non-mass samples. The unbalance problem between mass and non-mass classification in the detection problem is considered, and the proposed approach can obtain better performance than previous works.

## 2. Initial detection and segmentation

MCL [12] detection is an empirically optimized, rule-based algorithm based on observations made on a group of mammograms that contain masses. The hypothesis of MCL technique is that the growth of a mass disrupts the normal breast parenchyma

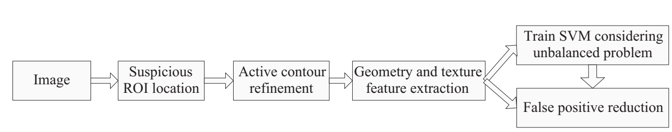


**Fig. 1.** Flowchart of the proposed method.