Expert Systems with Applications 41 (2014) 7361–7369

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Quantitative analysis of morphological techniques for automatic classification of micro-calcifications in digitized mammograms

C.C. Diaz-Huerta [a], E.M. Felipe-Riveron [b,*], L.M. Montaño-Zetina [a]

[a] Research and Higher Studies Center, National Polytechnic Institute, A.P. 14-740, 07000 Mexico City, Mexico
[b] Centro de Investigación en Computación, Instituto Politécnico Nacional, Av. Juan de Dios Batiz w/n and Miguel Othon de Mendizabal, P.O. 07738, Mexico City, Mexico

ARTICLE INFO

Article history:
Available online 10 June 2014

Keywords:
Mammogram analysis
Morphological reconstruction
Digital mammography
Micro-calcification detection
Mathematical Morphology

ABSTRACT

In this paper we present an evaluation of four different algorithms based on Mathematical Morphology, to detect the occurrence of individual micro-calcifications in digitized mammogram images from the mini-MIAS database. A morphological algorithm based on contrast enhancement operator followed by extended maxima thresholding retrieved most of micro-calcifications. In order to reduce the number of false positives produced in that stage, a set of features in the spatial, texture and spectral domains was extracted and used as input in a support vector machine (SVM). Results provided by TMVA (Toolkit for Multivariate Analysis) produced the ranking of features that allowed discrimination between real micro-calcifications and normal tissue. An additional parameter, that we called Signal Efficiency∗Purity (denoted SE∗P), is proposed as a measure of the number of micro-calcifications with the lowest quantity of noise. The SVM with Gaussian kernel was the most suitable for detecting micro-calcifications. Sensitivity was obtained for the three types of breast. For glandular, it detected 137 of 163 (84.0%); for dense tissue, it detected 74 of 85 (87.1%) and for fatty breast, it detected 63 of 71 (88.7%). The overall sensitivity was 85.9%. The system also was tested in normal images, producing an average of false positives per image of 13 in glandular tissue, 11 in dense tissue and 15 in fatty tissue.

1. Introduction

Breast cancer is the most common malignant neoplasm in western women, being the main cause of death in some countries around the world.

Specialists have developed screening programs with the aim of detecting warning signs of cancer, when it is not palpable. One of the main early symptoms is the appearance of small calcium deposits, known as micro-calcifications, whose diameter range from 0.1 to 3 mm. They are seen as bright spots in a mammogram and can be found isolated or forming clusters. Micro-calcifications have been commonly found in women over 50 years old, but recently have been detected in younger women, around 40 years old. In this case, its detection has been more difficult because the glandular breast density is also observed as a bright region in the mammogram, and they could be mistaken in that tissue. In many works, a cluster of micro-calcifications is a group of at least 5 micro-calcifications in a cubic centimeter. Many experts agree that most of them are benign; however, some of them could be malignant and evolve into a carcinoma. Clinical studies revealed that 30–50% of breast cancer cases showed micro-calcifications in mammography images and between 60% and 80% were confirmed by histological examination (Sickles, 1986; Manrique, 1999).

Different methods for automatic detection of micro-calcification clusters in digital or digitized mammograms have been developed in the past (Nishikawa, 2002; Sahiner, Chan, Roubidoux, et al., 2004; Wroblewska, Boninski, Przelaskowski, & Kazubek, 2003). Thangavel, Karnan, Sivakumar, and Mohideen (2005) and Cheng, Cai, Chen, Hu, and Lou (2003) have made an extensive survey of the state-of-the-art on the automatic detection of micro-calcifications in mammograms, which in general follow four steps: first, to reduce the noise and enhance micro-calcifications using image processing techniques; second, to apply a given segmentation technique; third, to extract a group of features of micro-calcifications. Finally, to classify the analyzed objects based on these features in order to determine if the analyzed objects are artifacts or real micro-calcifications.

Due to the many microscopic objects to be identified in a mammogram, there has been the necessity of developing computer algorithms to help physicians to locate them. The creation of an automatic system is intended to provide a "second opinion" in

* Corresponding author. Tel.: +52 55 5729 6000x56515.
E-mail addresses: celia@fis.cinvestav.mx (C.C. Diaz-Huerta), edgardo@cic.ipn.mx (E.M. Felipe-Riveron), lmontano@fis.cinvestav.mx (L.M. Montaño-Zetina).

http://dx.doi.org/10.1016/j.eswa.2014.05.051
0957-4174/

the detection of micro-calcifications. (Cheng et al., 2003; Mohanalin et al., 2010; Papadopoulos, Fotiadis, & Costaridou, 2008; Zyout, Abdel, & Jacobs, 2009). The most recent advances are focused in developing a Computer-aided detection system for clustered micro-calcifications in reconstructed digital breast tomosynthesis volumes. The algorithm is based on analyzing small volumes that enclose the regions of micro-calcifications. It included morphological methods as region growing and iterative thresholding (Sahiner et al., 2012). Other research is concerned on a knowledge base algorithm, proposed by Oliver et al. (2012). It is based on extracting local features that define the morphology of micro-calcifications. Then a boosting scheme is followed in order to create a strong classifier from the sum of weak classifiers.

There are other approaches mainly focused on the feature selection. The research group of Velayutham and Thangavel (2012) began with a simple processing stage based on histogram equalization and watershed algorithm for segmenting the micro-calcifications. Then, a set of features that included Haralick features, Grey Level Co-occurrence Matrix and some modifications of this matrix (for example Gray Level Difference Matrix-GLDM), and Surrounding Region Dependency Matrix (SRDM)) were used. Then, they proposed an unsupervised method for feature selection based on rough set-based entropy measures in order to remove redundant features. They recommended their method when class labels are not known or are incomplete. The method proposed by Yu and Huang (2010) used a wave filter and two thresholds in the stage of image processing. Then a set of textural features based on Markov random field (MRF) and fractal models as well as statistical features were extracted and used as inputs to a BPNN for automatic classification.

Regarding automatic classification of clusters of micro-calcifications, the most common techniques are artificial neural networks (ANN) and support vector machines (SVM). A common problem consisted on the imbalanced distribution of benign and malign instances in the training sets, affecting the performance of classifier. Ren proposed the balanced learning with optimized decision enabling effective learning and then evaluate the performance of both classifiers (Ren, 2012).

The purpose of this work is to analyze the performance of four algorithms reported in the literature for automatic detection of individual micro-calcifications, based on morphological operators. The general procedure in those algorithms consists on the following steps: first, each complete image undergoes a contrast enhancement process in order to highlight micro-calcifications. Then, a thresholding operation is applied, in order to extract size, shape and position of the objects of interest. At this stage, each potential micro-calcification is compared with those detected by an experienced radiologist, who previously set the ground truth by hand. At this point, sensitivity is compared among the four algorithms and the one with the highest value is selected. After evaluation of sensitivity, a set of features in spatial, texture and spectral domains is obtained, which are inputs to a support vector machine (SVM), in order to filter most of the false positives coming from the processing stage, and with different parameters are finally evaluated in a support vector machine to make an automatic classification of micro-calcifications.

The four algorithms are based on Mathematical Morphology, where the digitized mammogram f is considered as a set of pixels that represents a topographical relief and each pixel is considered as an elevation proportional to its intensity. The dark and light structures of the image correspond to valleys and peaks of this relief, respectively. Some of these peaks correspond to micro-calcifications in mammograms surrounded by background tissue.

Recently, different processing algorithms have been reported in the literature (Betal, Roberts, & Whitehouse, 1997; Dengler, Behrens, & Desaga, 1993; Fu et al., 2005; Halkiotis, Botsis, & Rangoussi, 2007; Papadopoulos, Fotiadis, & Likas, 2002; Soille, 2010; Ustymowicz & Nieniewski, 2006). We have implemented them to test their accuracy with different parameter values in order to determine those values that allow the detection of most of the real micro-calcifications.

The paper is organized as follows. Section 2 briefly deals with the materials and methods used in the paper. Section 3 describes the background mainly focused on the morphological operators in which algorithms are based on. Analysis of the algorithms related to the research is included in Section 4. In Section 5 the description of the different types of features to be extracted and evaluated is included. Section 6 explains the support vector machine used for the automatic classification (Diaz-Huerta, Felipe-Riverón, & Montaño-Zetina, 2012). Results and their discussion appear in Section 7 and finally the conclusions are given in Section 8.

2. Materials and methods

Data were obtained from the mini-MIAS database (http://peipa.essex.ac.uk/info/mias.html), which consisted of 322 digitized mammography images owned by the UK National Breast Screening Programme. They were obtained by rescaling MIAS images from 50 to 200 microns pixel size, and consisted of images of 1024×1024 pixels with 8 bits per pixel (0 corresponded to black, and 255 corresponded to white). The database provided a diagnosis for each mammogram, giving 23 images with micro-calcifications (benign or malign). It also provided the coordinates (x, y) at the center of abnormality either in cluster or isolated micro-calcifications and the radius (in pixels) of a circle that contained them. The four algorithms were implemented in MATLAB, using the Image Processing Toolbox morphological functions.

All the mammograms in the mini-MIAS database had labels detailing the orientation and projection of mammography, and notes regarding the patient. We removed that information in the following way: first, we analyzed the gray level in the skin-line region, which was between 14 and 30 (in a scale of 0–255). Pixels with intensity less than 15 were set to zero, because most of them corresponded to the dark region of mammogram and tissue near the surface of the breast. The rest was set to 255, which was associated to breast region, labels and some artifacts related to the X-ray process (lines or bright points due to dust). Then, we applied a morphological opening over this binary image using a disk of radius 30 pixels as a structuring element, in order to separate those labels superimposed to the breast region.

The original breast region was obtained with the process of reconstruction by dilation (Gonzalez & Woods, 2008; Serra, 1982) where the marker f was a black image of size 1024×1024 with an isolated white pixel (with gray level 255) at the center, and the mask g was the original image.

The process of reconstruction by dilation, denoted by $R_g^{\delta}(f)$, consists on successive dilations $\delta_g(f)$ of marker image f (black image with isolated pixel) until the propagation of the resulting image is totally limited by the mask image g (original) (Soille, 2010). After this stage, labels were eliminated.

The four image processing algorithms evaluated in this work were applied on each entire image in order to improve the contrast of micro-calcifications over their surrounding and locate those micro-calcifications. Then the segmentation produced a binary image, with white objects associated to potential micro-calcifications.

3. Background

A mammogram f is considered as a set of points arranged in a matrix of size $M \times N$, as described in Section 2. Each mammogram