2005 WSEAS Int. Conf. on DYNAMICAL SYSTEMS and CONTROL, Venice, Italy, November 2-4, 2005 (pp427-432)

Feature Extraction in Digital Mammography Using LabVIEW

MIHAELA LASCU
DAN LASCU
Department of Measurements and Optical Electronics
Faculty of Electronics and Telecommunications
Bd. Vasile Pârvan no.2
ROMANIA
mihaela.lascu@etc.utt.ro, dan.lascu@etc.utt.ro
http://www.etc.utt.ro

Abstract: - This paper presents some characteristics of mammography images since these will motivate the algorithms for micro calcification detection implemented in LabVIEW (Laboratory Virtual Instrument Engineering Workbench). LabVIEW is a graphical programming language that uses icons instead of lines of text to create programs. We are interested in compression and noise removal of mammography images because these realize a preprocessor for the identification of micro calcification clusters in mammograms. We propose a general strategy for constructing algorithms and implementing them in LabVIEW for extracting micro calcification clusters. A non-linear method is implemented in LabVIEW for performing image enhancement.

Key-Words: - mammograms, compression, noise removal, micro calcification detection, digitized image, non-linear enhancement, LabVIEW, Digital Database for Screening Mammography DDSM, wavelet transform WT.

1 Introduction

The devastating impact of breast cancer in the whole world is well known. These points are necessary for early cancer detection. Mammography is an X-ray imaging procedure for examination of the breast. It is used primarily for the detection and diagnosis of breast cancer but also for preoperative localization of suspicious areas and in the guidance of needle biopsies. Mammograms are complex in appearance and signs of early disease are often small or subtle. Furthermore, the consequences of errors in detection or classification are costly. The number of mammograms generated daily is large and therefore it is very desirable to develop image processing tools which facilitate the handling of mammograms and aid the radiologist in diagnosis. Breast cancer is detected on the basis of four types of signs on the mammogram: a) The characteristic morphology of a tumor mass; b) Certain presentations of mineral deposits as specks called micro calcification; c) Architectural distortion of normal tissue patterns caused by the disease; d) Asymmetry between images of the left and right breasts.

The Digital Database for Screening Mammography (DDSM) is a database of digitized film-screen mammograms with associated ground truth and other information. The purpose of this resource is to provide a large set of mammograms in a digital format that may be used by researchers to evaluate and compare the performance of computer-aided detection (CAD) algorithms.

The evaluation of a CAD algorithm often begins with a retrospective evaluation of cancer cases. A preliminary evaluation of this type is more time and cost effective that a prospective evaluation in a clinical setting. While the data for a retrospective CAD performance evaluation may be obtained by digitizing mammograms from the case files at a mammography center, this task may be time consuming and thus expensive to achieve. This expense may be avoided by using data from the DDSM.

The breast is exposed to a small dose of radiation to produce an image of internal breast tissue. The image of the breast is produced as a result of some of the x-rays being absorbed (attenuation) while others pass through the breast to expose either a film (conventional mammography) or digital image receptor (digital mammography). The exposed film is either placed in a developing machine—producing images much like the negatives from a 35mm camera—or images are digitally stored on computer.

Breast compression is necessary in order to: a) Even out the breast thickness so that all of the tissue can be visualized; b) Spread out the tissue so that small abnormalities won't be obscured by overlying breast tissue; c) Allow the use of a lower x-ray dose

since a thinner amount of breast tissue is being imaged; d) Hold the breast still in order to eliminate blurring of the image caused by motion; d) Reduce x-ray scatter to increase sharpness of picture.

2 Mammograms as Digitized Images

A digitized image is an array of pixel values. For mammography, this array is generally not square and this causes some technical difficulties in wavelet-based image processing algorithms. Also the size of mammograms is large when compared with many other images. However, we shall assume that the digitized mammogram is not only square but of size 2^mx2^m. Typical values are $m = 9, 10$.

Thus, a digitized mammogram, will be an array of nonnegative integers

$$p_k, k = (k_1, k_2), k_1, k_2 \in \{0,...,2^m - 1\} \quad (1)$$

The range of the integer values p_j is related to the scanner and the dynamical range of the film.

The pixel values as obtained from a bivariate function F defined on the unit square $\Omega := [0,1]^2$ by taking cell averages

$$p_k = \int_{Q_k} F(x,y)dxdy, \quad 0 \le k_1, k_2 \le 2^m - 1 \quad (2)$$

where $Q_k := [2^{-m}k_1, 2^{-m}k_1 + 2^{-m}] \times [2^{-m}k_2, 2^{-m}k_2 + 2^{-m}]$ is the square with side length 2^{-m} and lower left vertex $2^{-m}k = 2^{-m}(k_1, k_2)$. Thus, we view the pixel values as samples of the underlying function F.

2.1 Characteristics of mammography images

A typical mammogram is shown in Fig.1.

The micro calcifications which we wish to identify appear as small bright spots in the mammogram. Their diameters are up to .7mm, with an average size of .3mm. Thus, the micro calcifications can be identified with certain dyadic levels j in the wavelet decomposition-namely, those for which 2^j are comparable with the size of the micro calcification. The dyadic levels j = -7, -8, -9 typically correspond to frequencies where micro calcifications occur. It is important not only to retain all micro calcifications but also their shape since these are important in diagnosis.

However, the techniques we put forward may also have application to the detection of lesions. Moreover, the detection of micro calcification clusters has several features in common with other problems of feature classification and automated detection.

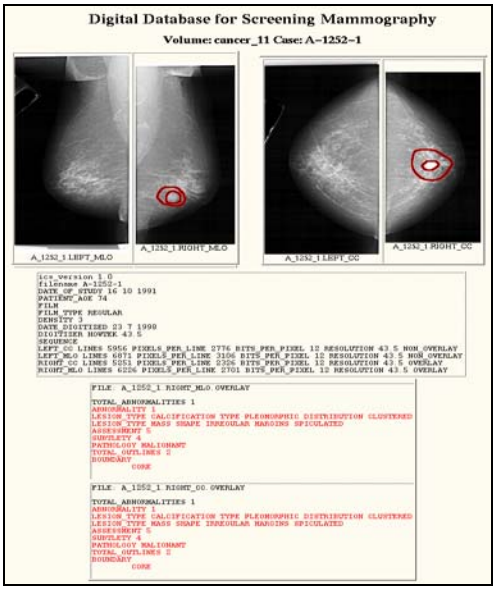

Fig.1 A typical mammography image with micro calcifications.

3 Compression and Noise Removal

While compression and noise removal are important for the storage and transmission of images, in the present paper we are interested in them as a preprocessor for the identification of micro calcification clusters in mammograms.

The main steps in wavelet-based compression are [1]:

Step 1. Computation of wavelet coefficients;
Step 2. Threshold/Quantization;
Step 3. Encoding;
Step 4. Decoding;
Step 5. Computation of pixel values.

The performance of compression algorithms depends on the choice of wavelet basis. It is generally agreed upon that the bi-orthogonal wavelets give the best compression with a fixed wavelet basis (4).

The wavelet transform WT decomposes a signal $f(x)$ by performing inner products with a collection of analysis functions $\{\psi\ (a, b)\}$, which are scaled and translated version of the wavelet ψ;

$$(W_\psi f)(a,b) = \langle f, \psi(a,b) \rangle = \quad (3)$$
$$\int_{-\infty}^{+\infty} f(x)\overline{\psi(a,b)(x)}dx,$$

$$\psi(a,b)(x) = a^{-1/2}\psi\left(\frac{x-b}{a}\right). \quad (4)$$

The distinction between the various types of wavelet transforms depends on the way in which the scale and translation parameters in (4) are discretized. The definition of the continuous wavelet transform assumes that at the most redundant end a and b vary in a continuous fashion. At the other extreme, nonredundant transform, decomposition into wavelet bases only requires the values of the transform at the dyadic scales:

$$a = 2^j,$$

and for translation parameters that are critically sampled:

$$b = k \cdot 2^j.$$

In between those two extremes, there are many varieties of over complete transforms that use a finer sampling of these parameters; for example, wavelet frames and redundant wavelet analyses.

4 Algorithms for the Detection of Micro Calcification Clusters

We shall propose a general strategy for constructing algorithms for extracting micro calcification clusters. We wish to distinguish between two types of algorithms. The first are autonomous [1] and do not involve a computer operator. The second are interactive and allow decisions to be made to improve parameter settings which depend on the image. Both approaches have the same major steps which we now describe.

Step 1. Compression of the digitized image: an algorithm begins by utilizing compression as a preprocessor to the identification of the micro calcification clusters. The purpose of this step is to remove noise and still retain all micro calcifications.

Step 2. Selecting only high frequency terms: the purpose of this step is to retain only those frequency terms that correspond to the size of the micro calcifications.

Step 3. Reconstructing pixel values: this step constructs the image corresponding to the wavelet decomposition.

Step 4. Nonlinear enhancement: The purpose of this step is to remove background and retain only the micro calcifications. Let us denote by $\tilde{p}_j$ the pixel values of the reconstructed image in Step 3. We shall modify these pixel values to obtain the new pixel values p_j^*. The new values p_j^* depend not only on $\tilde{p}_j$ but also on p_j:

$$p_j^* := E\left(p_j, \tilde{p}_j\right) \quad (5)$$

where E is a nonlinear enhancement function.

The properties we want for E are the following. If $\tilde{p}_j$ is small then p_j^* should be set to zero. We would also like to enhance $\tilde{p}_j$ in a way that depends on the background brightness of the original pixel values p_ν, if the difference between ν and j is also large, then we would like p_j^* to be an increase of the value $\tilde{p}_j$. Among all steps in this general algorithm, this step and the next benefit most from interaction.

Step 5. Threshold pixel values: this step thresholds pixel values to retain only the most intense. In the autonomous algorithm this is obtained by retaining only the 300 largest pixel values p_j^*. We denote by $p_j^\#$ the pixel values after this step that has been completed. We call the $p_j^\#$ *hot pixel* values. The hot pixels are indications of micro calcifications. However, further processing is necessary to be sure that the entire micro calcification is retained intact and also to eliminate hot pixels that are not part of micro calcification clusters.

Step 6. Removing isolated pixels. This step is intended to remove any isolated hot pixels. A micro calcification should correspond to several hot pixels. In our autonomous algorithm, we remove isolated hot pixels as follows. For each hot pixel $p_j^\#$, we create a 3x3 square of pixels centered at j. If no pixel in this square other than $p_j^\#$ is hot, then we remove $p_j^\#$ from the hot list.

Step 7. Filling out micro calcifications: The purpose of this step is to fill out micro calcifications. It could happen that the threshold in Step 5 removed pixels which correspond to micro calcifications and we want to restore these. In our autonomous algorithm, we do this as follows. If $p_j^\#$ is a hot pixel, we form a 3x3 square of pixels centered at j and we consider a ν from this square. If p_ν is greater than p_j-5 and if $p_\nu^\#$ was not one of the hot pixel values, then we change the value of $p_\nu^\#$ from zero to $p_j^\#$ and add it to our list of hot pixels.

Step 8. Test for clusters. Finally, we want to test for clusters. Given a hot pixel value $p_j^\#$, we examine all hot pixels in a 61x61 square of pixels centered at j. If $p_\nu^\#$ is a hot pixel from this square and $\left|p_\nu^\# - p_j^\#\right| < 10$, then we say $p_\nu^\#$ is connected to $p_j^\#$ and write $p_\nu^\# \sim p_j^\#$. We say that $p_j^\#$ is related to $p_j^{'\#}$ if there is a sequence $p_{\nu k}^\#$, k=0,…m, of hot pixel values with $\nu_0 = j$ and $\nu_m = j'$ and $p_{\nu k}^\# \sim p_{\nu k+1}^\#$, k=0,…,m-1. Then we choose the smallest rectangle

which contains all pixel values $p_j^{'\#}$, which are related to $p_j^{\#}$. In this way, we obtain boxes of hot pixel values which correspond to boxing our micro calcification clusters.

5 Image processing techniques

NI-IMAQ is a complete and robust application programming interface (API) for image acquisition. NI-IMAQ performs all the computer and board specific tasks for straightforward image acquisition without register-level programming. NI-IMAQ is compatible with NI-DAQ and all other National Instruments driver software for integrating imaging into any National Instruments solution. NI-IMAQ provides the interface path between LabVIEW, Measurement Studio, and other programming environments and the hardware product. NI-IMAQ has both high-level and low-level functions for maximum flexibility and performance. High-level functions include single-shot and continuous-mode image acquisition. Low-level functions include imaging sequence setup. With only three icons in LabVIEW , it is possible to snap and display an image. You can build more complex configurations using this as a foundation.

IMAQ Vision software from National Instruments adds high-level machine vision and image processing to LabVIEW, Measurement Studio, and other programming environments. IMAQ Vision includes an extensive set of MMX-optimized functions for gray-scale, color and binary image display; image processing, including statistics, filtering and geometric transforms; and pattern matching, shape matching, blob analysis, gauging and measurement.

The possibilities and range for image processing and machine vision are numerous, if not over helming. So many algorithms exist for you to select from you might ask which one is right for your application and where to begin.

To start with, in many applications you need a quantitative or statistical description of your image or region of interest. Statistical functions are calculated quickly; you can solve many inspection applications using simple functions such as average and standard deviation. Pattern matching functions are key for machine vision applications for locating features in the image.

In Fig.2 is the block diagram of the virtual instrument (implemented algorithm) for the detection of micro calcification clusters.

Using this graphical program we obtain the following images Fig.3, Fig.4 concerning mammography's with cancer.

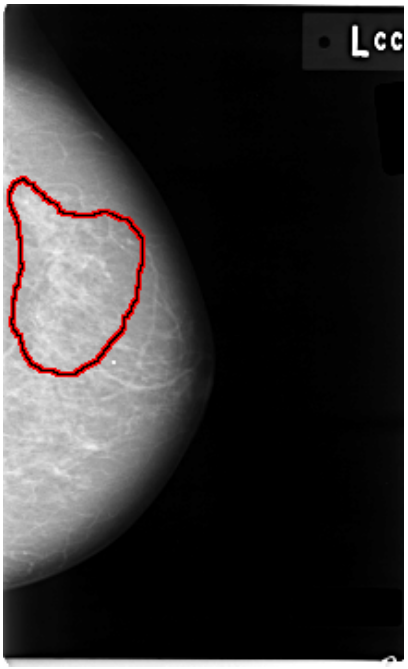

Fig.3 Abnormality 1
Lesion type calcification
amorphous
distribution regional.

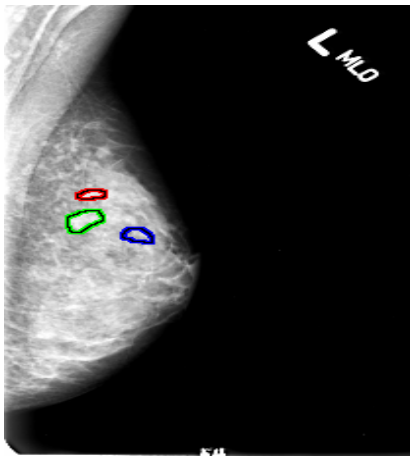

Fig.4 Abnormality 1,2,3
Lesion type calcification
punctuate-amorphous
polymorphic
distribution clustered.

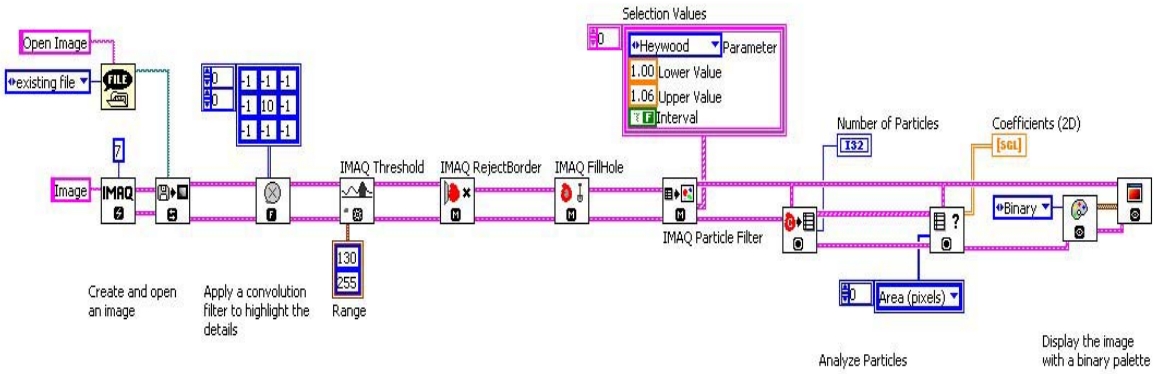

Fig.2 Block diagram for the detection of micro calcification clusters.

The presented program is able to do following activities:1. Display initial image; 2. Smooth the image; 3. Highlight edges; 4. Thresholding the image; 5. Invert the image; 6. Fill holes in particles; 7.Remove border objects; 8. Filter particles. 9. Analyze particles.

Algorithm breadth and accuracy is very important. There is much to consider when choosing vision software – most importantly, whether the software tools can correctly and accurately measure important part or object features down to the subpixel. If the software is not accurate and reliable, then it does not matter how fast your computer is or how many pixels your camera has. Keep in mind that it is much easier to make accurate code faster than to make fast code more accurate. The NI Vision Development Module and Vision Builder AI include hundreds of accurate and reliable vision functions. The four most common machine vision application areas that have been used in our paper are listed below, along with the most popular *algorithms*.

1. Enhancing an image – Use filtering tools to sharpen edges, remove noise, or extract frequency information. Use image calibration tools to remove nonlinear and perspective errors caused by lens distortion and camera placement. You also can use the image calibration tools to apply real-world units to your measurements, so the tools return values in microns, millimeters, or miles instead of pixels.

2. Checking for presence – This is the simplest type of vision inspection. To check for part or feature presence, you can use any of the color, pattern-matching, or histogram tools. A presence checks always results in a yes/no or pass/fail.

3. Locating features – Locating features is important when aligning objects or determining exact object placement, serving as a standard for all subsequent inspections. Edge detection, grey-scale pattern matching, shape matching, geometric matching, and color pattern matching are all tools you can use to locate features. The tools return the object position (X, Y) and rotation angle down to one-tenth of a pixel. Geometric matching is immune to overlapping objects or objects that change in scale.

4. Measuring features – The most common reason to use a vision system is to take a measurement. Typically, you use edge detection, particle analysis, and geometric function tools to measure distance, diameter, total count, angles, and area. Whether you are calculating the total number of cells under a microscope or the angle between two brake-caliper edges, these tools always return a number instead of a location or pass/fail value.

Algorithm Performance - While accuracy and ease of use often are the two most important factors in choosing a vision system, execution speed is a third consideration. No matter how many hundreds of algorithms you have to choose from or how quickly you can build an application with them, if the inspection tools are inefficient and take too long to run, then much of your work goes to waste. NI vision software is highly optimized to squeeze performance from every possible source, resulting in software that rivals the fastest vision software package speed in the world. In fact, when compared to a leading vision software provider, NI vision software is consistently faster in many categories, as shown in the chart below:

Table 1. Vision software package speed.

	NI Vision Software Speed (ms)	Leading Vision Software Speed (ms)	NI Speed Increase
Histogram	0.91	2.03	2.2X
Geometric Transform	3.1	10.3	3.3X
Morphology	1.8	5.9	3.3X
Geometric Matching	93.0	149.8	1.6X
Object Classification	7.5	–	–

In our LabVIEW program, we use the **grayscale morphology.** Morphological transformations extract and alter the structure of particles in an image. You can use grayscale morphology functions to do the following: 1.Filter or smooth the pixel intensities of an image; 2.Alter the shape of regions by expanding bright areas at the expense of dark areas and vice versa; 3.Remove or enhance isolated features, such as bright pixels on a dark background; 4.Smooth gradually varying patterns and increase the contrast in boundary areas

Applications include noise filtering, uneven background correction, and gray-level feature extraction. You can also use grayscale morphological transformations to enhance non-distinct features before thresholding the image in preparation for blob analysis.

Grayscale morphological transformations compare a pixel to those pixels surrounding it. They change the shape of particles by processing each pixel based on its number of neighbors and the values of those neighbors. A neighbor is a pixel whose value affects the values of nearby pixels during certain image processing functions. Morphological transformations use a 2D binary mask called a structuring element to define the size and effect of the neighborhood on each pixel, controlling the effect of the binary morphological functions on the shape and the boundary of a particle.

We also use in our program **thresholding.** Thresholding enables you to select ranges of pixel values in grayscale and color images that separate the objects under consideration from the background. Thresholding converts an image into a binary image, with pixel values of 0 or 1. This process works by setting to 1 all pixels whose value falls within a certain range, called the threshold interval, and setting all other pixel values in the image to 0. Figure 5a shows a grayscale image, and 5b shows the same image after thresholding.

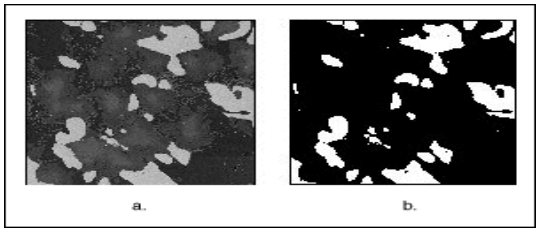

Figure 5. An Image Before and After Thresholding

6 Conclusion

After digitization, mammograms were automatically cropped to remove much of the background (non-breast tissue) area. They were then manually processed to darken (digitally zero) pixels in regions that contained patient identifiers and were stored in files using a truly loss-less compression algorithm.

Another subtle issue is what elements of ground truth to count. That is, should an algorithm aim to find only proven cancers, all abnormalities suspicious enough to merit additional workup, or some other criterion? Furthermore should a mass or micro calcification detection algorithm be evaluated using abnormalities of all types, or just ones that the algorithm was designed to detect? The method used to evaluate the algorithm was as follows. Each detection region was represented by the location of its maximum suspiciousness value. A detection region was scored as false positive if the pixel location of the abnormality fell outside all ground truth regions. A detection that fell inside a ground truth region counted as a true positive detection. In the event where multiple detections fell inside a ground truth region, one counted as a true positive detection, and the others counted as false positive detections. This was done because only one prompt is desired for each ground truth region. Any ground truth region that did not correspond to a detection location was counted as a false positive detection. Finally, all benign and malignant masses (micro calcifications) that were present in the ground truth were used for the performance evaluation.

References:

[1] Akram Aldroubi, Michael Unser, *Wavelets in Medicine and Biology*, Publishing House CRC Press, 1996.

[2] J. Dengler, S.Behrens and J.F. Desaga, Segmentation of micro calcifications in mammograms, *IEEE Trans. Med. Imaging* ***12*** *International Journal of Science and Technology*, 1993, pp. 634-642.

[3] Robert Shroy, Michael Van Lysel, Martin Yaffe, X-Ray equipment, *The Biomedical Engineering Handbook,* 1995, pp.953-989.

[4] Jon Olansen, Eric Rosow, *Virtual Bio-Instrumentation, Biomedical, Clinical and Healthcare applications in LabVIEW,* Publishing House Prentice Hall PTR, 2002.