Proceedings of the 10th WSEAS International Conference on COMPUTERS, Vouliagmeni, Athens, Greece, July 13-15, 2006 (pp671-676)

Compression, Noise Removal and Comparison in Digital Mammography using LabVIEW

MIHAELA LASCU, DAN LASCU
IOAN LIE, MIHAIL TĂNASE
Department of Measurements and Optical Electronics
Faculty of Electronics and Telecommunications
Bd. Vasile Pârvan no.2
ROMANIA

Abstract: - In the present paper we are interested in compression and noise removal as a preprocessor for the identification of microcalcification clusters in mammograms. LabVIEW (Laboratory Virtual Instrument Engineering Workbench) is a graphical programming language that uses icons instead of lines of text to create programs. We propose a general strategy for constructing algorithms and implementing them in LabVIEW for compression, noise removal and extracting microcalcification clusters. The comparison method presented in this paper aims to improve mammogram comparison by estimating the underlying geometric transformation for any mammogram sequence. It takes into consideration the various temporal changes that may occur between successive scans of the same woman and is designed too overcome the inconsistencies of mammogram image formation.

Key-Words: - mammograms, compression, noise removal, microcalcification detection, LabVIEW, Digital Database for Screening Mammography DDSM, wavelet transform WT, biomedical image processing, image registration, image sequence analysis, image matching, blob analysis.

1 Introduction

This paper presents some characteristics of mammographic images since these will motivate the algorithms for *microcalcification* detection implemented in *LabVIEW* (*Laboratory Virtual Instrument Engineering Workbench*) [1].

To count cancerous cells we use a common image processing technique called particle analysis, often referred to as *blob analysis*.

The devastating impact of breast cancer in the whole world is well known. This points to the need 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 microcalcification; 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.

The reliable diagnosis of abnormalities from a single mammogram is an extremely difficult task even for a skilled radiologist, and so it is increasingly the case that pairs of mammograms are compared [4]. These may be, for example, the left and right mammograms taken at the same session. Equally, when mammograms from an earlier time are available, the radiologist will routinely compare the older and more recent images. For this reason alone, the development of mammogram registration is increasingly important for the early detection of

pathology. The potential clinical applications are as follows: 1) for women at high risk of developing breast cancer (e.g., women with family history of breast cancer or genetic susceptibility), usually have more frequent mammograms taken in order to detect a malignancy at as early a stage as possible. Previous ("normal") mammograms are used as a baseline for comparison with recent ones. 2) for postmenopausal women who often decide themselves, or are advised by their gynecologist physician, to undergo *hormone replacement therapy* (HRT). However, there is a suggestion, based on clinical experience, that significant regional increases in tissue density could be an early indicator of breast cancer for women using HRT. For this reason, it is important to be able to register HRT mammogram sequences, aiming at a more effective comparison for early detection of lesions. 3) for retrospective studies that aim to analyze temporal data in order to assess the accuracy and effectiveness of diagnosis in hospitals/screening centers. Such studies aim to define the rate of missed cancers and interval cancers, as well as to further educate clinicians in the important task of early diagnosis.

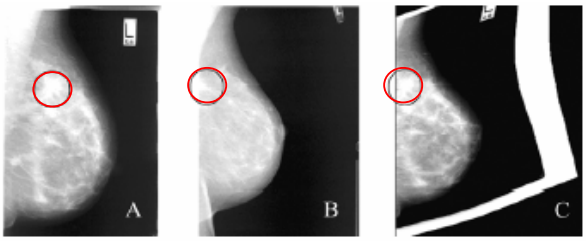

Fig.1. Typical mammography images with microcalcifications of the same woman (A and B) are shown together with C, the registered version of A to B.

The differences in imaging conditions and the temporal changes (5-year interval) render the comparison of A and B difficult, while by comparing B (recent mammogram) and C the physician can be reassured that the encircled region in B is scar tissue from the excision of a previous cancer in C (also encircled in the same location).

2 Compression and Noise Removal

We first review briefly the elements of wavelet-based compression.

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 microcalcification clusters in mammograms.

A schematic for wavelet-based compression is presented in Fig.2.

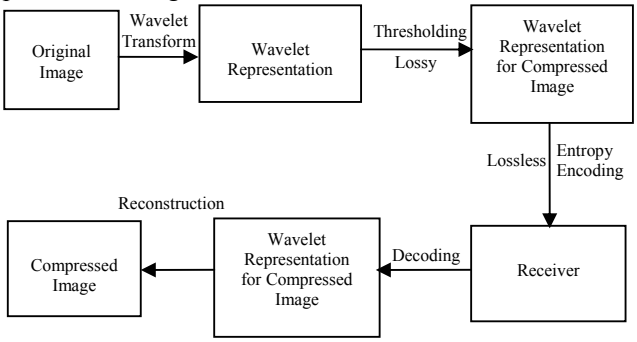

Fig.2. Wavelet based compression

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

Step 1. *Computation of wavelet coefficients*:

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

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

The range of the integer values j_{pg} is related to the scanner and the dynamical range of the film.

The following viewpoint of a digitized image is useful in the analysis that follows. We can view the pixel values as obtained from a bivariate function F defined on the unit square $\Omega:=[0,1]^2$ by taking cell averages:

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

where

$$Q_k := \left[2^{-m}k_1, 2^{-m}k_1+2^{-m}\right] x \left[2^{-m}k_2, 2^{-m}k_2+2^{-m}\right]$$

Thus, we view the pixel values as samples of the underlying function F.

We choose a univariate scaling function φ and from the pixel values (p_k), we create an approximation f to F from the space V_{-m} and represent the mammogram as in

$$F \approx f = \sum_j c(k)\varphi_{-m,k} \tag{3}$$

This is usually accomplished by defining

$$c(k) := p_k, \quad 0 \le k_1,k_2 \le 2^m-1, \tag{4}$$

and defining $c(k)$ for other values of k by some extension strategy.

Step 2. *Thresholding*:

We gain lossy compression of the image by reducing the size of the wavelet coefficient file. We can use the fast wavelet transform (FWT) to convert the representation (3) to the wavelet representation. In this way, we obtain

$$f = \sum_{k\in Z^2} c(k)\varphi_{-m,k} = \sum_{k\in Z^2} c_0(k)\phi_{0,k} + \quad (5)$$
$$+ \sum_{j=-1}^{-m+1}\sum_{\eta\in\Psi}\sum_{k\in Z^2} d_{j,\eta}(k)\eta_{j,k}.$$

In each of these sums, the k can be restricted to those values for which $\eta_{j,k}$ is nonzero on Ω. The wavelet coefficients $d_{j,n}(k)$ can be computed from the pixel values in $N = 2^{2m}$ the original number of pixel values. We call (5) the wavelet representation of the image. It will be convenient to speak only about the coefficients $d_{j,n}(k)$ of $\eta_{j,k}$. Similar statements apply to the coefficients of $\phi_{0,k}$.

There are two essential methods for compression: thresholding and quantization. In this paper we will use thresholding. Thresholding means that we pick a threshold ε_j for each level $j = -m+1,...,-1,0$ and retain only those coefficients whose absolute value exceeds ε_j. Thus, thresholding replaces $d_{j,n}(k)$ by $\tau_j(d_{j,\eta}(k))$ where the function τ_j defined by

$$\tau_j(x) := \begin{cases} x, & |x| \ge \varepsilon_j \\ 0, & |x| < \varepsilon_j. \end{cases} \quad (6)$$

This is called *hard thresholding;* the function τ_j is not continuous. *Soft thresholding* would replace τ_j by the Lipschitz continuous function

$$\tau_j^0(x) := \begin{cases} x, & |x| \ge \varepsilon_j \\ 2(|x| - \varepsilon_j/2)\operatorname{sgn} x, & \varepsilon_j/2 \le |x| < \varepsilon_j \\ 0, & |x| < \varepsilon_j/2. \end{cases} \quad (7)$$

Soft thresholding is numerically stable; a small change in $d_{j,n}(k)$ results in a small change in the output $\tau_j^0(d_{j,\eta}(k))$. Thresholding will replace small wavelet coefficients by zero. This not only has the desired effect of *compression* but also *removes noise*. There are several results which show that soft thresholding, with a proper choice of thresholding parameters, gives an optimal algorithm for removing Gaussian noise.

Step 3. *Encoding*:

For the purposes of storage or transmission of the compressed image, a lossless encoder is applied to the file of wavelet coefficients. While standard arithmetic and runlength encoding can be applied, the best results are obtained with customized encoders [3] which take into account the spatial correlation of the wavelet coefficients. The encoded wavelet coefficient file is our compressed representation of the image and can be stored or transmitted.

Step 4. *Decoding*:

The encoded file is decoded to obtain the compressed wavelet coefficient file. This is the same file as at the end of Step 2.

Step 5. *Computation of pixel values*.

The inverse fast wavelet transform is used to compute the pixel values of the compressed image. This step takes again N operations. These are then the pixel values of the compressed image which is our approximation to the original image. In compression algorithms we have to take into account the choice of wavelet basis, the choice of metric and the level of compression.

3 Image processing techniques

IMAQ Vision Software from National Instruments [2] 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 overwhelming. 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 (ROI). 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.

To count cancerous cells we use a common image processing technique called particle analysis, often referred to as *blob analysis*. Blob analysis is the process of detecting and analyzing distinct two-dimensional shapes within a region of the image. Blob analysis can provide your application with information about the presence or absence, number, location, shape, area, perimeter and orientation of blobs-microcalcifications- within an image.

A simple definition of a blob is a group of connected pixels. In general, blobs are thought of as a group of contiguous pixels that have the same intensity. Image processing operates on these blobs to calculate the area or perimeter or to count the number of distinguishable blobs. Before you can apply blob analysis you must preprocess the image by converting a gray-scale image – an image with 256 levels - to an image with only two gray scales – zeros and ones. The objective is to separate the important microcalcifications, blobs, from the unimportant information, background. A technique called thresholding appropriately separates the microcalcifications (blobs) from the background.

The result of the thresholding process is a *binary image*, which is an image of pixel values of only ones and zeros. The microcalcifications are represented by the connected pixels of ones and the background is represented by the zeros. By binarizing the image into ones and zeros, the task of writing image processing algorithms for microcalcification analysis is made easier. To find the area of a microcalcification, you have to count the pixels with values of one that are connected. Another benefit of binarizing the image for blob analysis is that the blob analysis calculations are fast. Many biomedical imaging applications can be solved using blob analysis, but also this kind of analysis is useful in industrial inspection applications. The steps for counting cancerous cells are the following:

- *Acquiring the image*: configure the board for acquisition using a point-and-click user interface; the configuration software saves the settings for the board to a configuration file, which is then used by the NI-IMAQ driver software in the development environment to simplify acquisition of images;
- *Thresholding to create a binary image*: the threshold function segments an image into two regions, an object region (microcalcification) and a background region; in this process all pixels that fall within the gray-scale interval defined as the threshold interval are given the value one.; all other pixels in the image are set to zero; the result is a binary image that can be processed very rapidly; generally, algorithms to process binary images are faster than algorithms for gray-scale images.
- *Filtering to remove noise and particles on the border of the image*: IMAQ function filters or removes the particles below a certain pixel size; another function removes the particles on the border of the image, because you cannot accurately determine the size of particles on the border of an image;
- *Microcalcifications (blob) analysis to count cells*: IMAQ functions analyze blobs in an image; it is possible to count, label, measure cells and objects; calculate area, perimeter, orientation, location and 49 other parameters; there are many blob calculations such as area, perimeter, moment of inertia, orientation, mean chord, width, height, ellipse axis, elongation factor, circularity factor, type factor, projection, location, bounding rectangle; use morphology functions to erode, dilate, fill holes, convex (fill holes on the edges), reject objects on the border and separate blobs.

In figure 3b is represented the front panel of an acquired cancerous mammography, while the block diagram of the LabVIEW virtual instrument for acquiring the mammography image is in the following figure 3a:

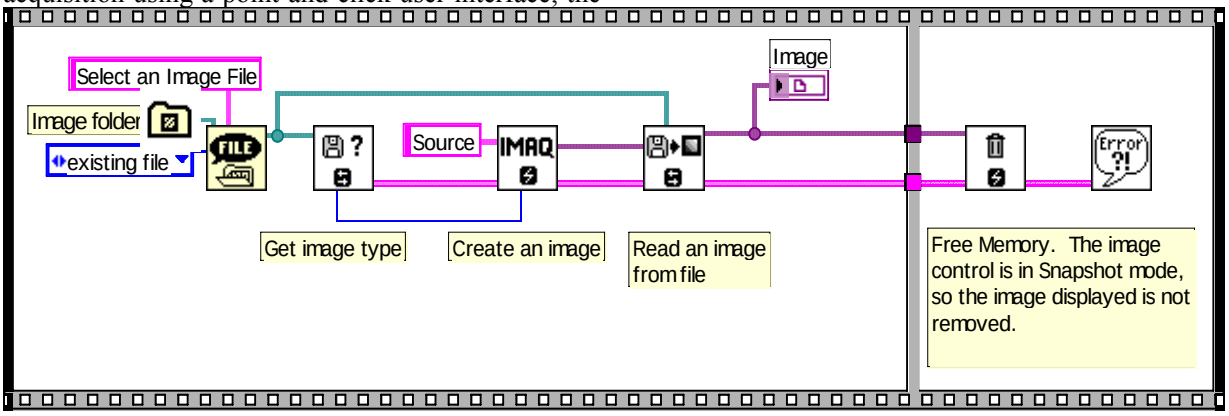

Fig.3a LabVIEW block diagram for getting, creating and reading an image.

Fig.3b LabVIEW front panel for getting, creating and reading an image.

The following program shows how to compute the histogram of an image or a region in the image. This program also shows how to set up and use the display tools to select and work on different regions in the image interactively.

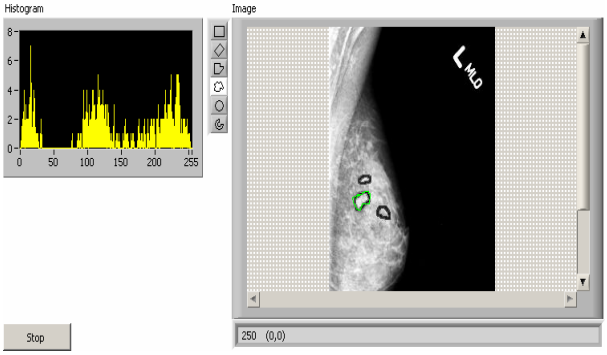

Fig.4. LabVIEW front panel for the histogram of a ROI (one of the three microcalcifications) in a mammography with three microcalcifications.

The following program shows how to use the display tools to draw a line on the window that displays an image and get the pixel values of the image along that line (line profile). You typically use line profiles in inspection tasks to find the edges of an object or part being inspected.

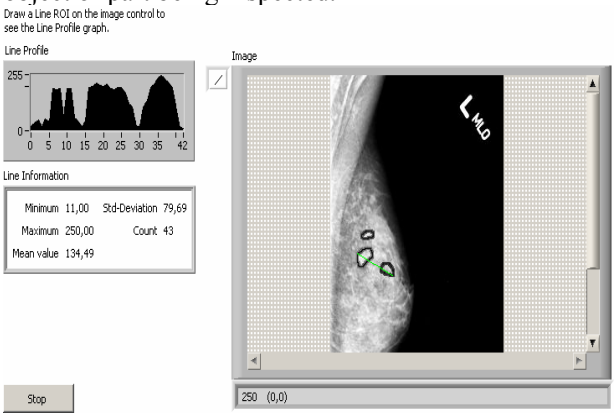

Fig.5. LabVIEW front panel to inspect line profiles.

The blob analysis program performs a series of grayscale filtering, threshold, binary morphology, and particle analysis operations to measure the areas of all the large circular particles of the image.

Fig.6.Filtered image with enhanced edge information.

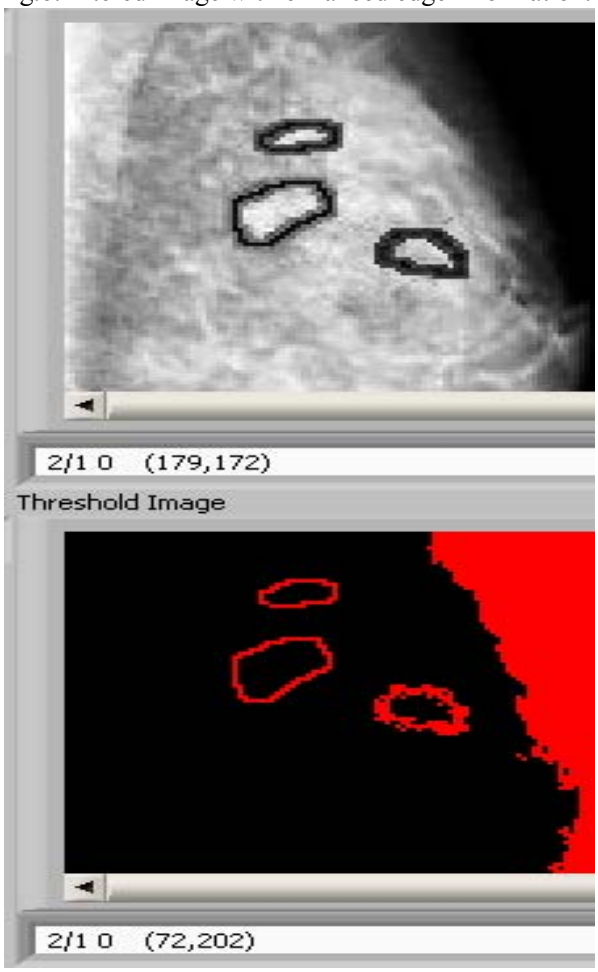

Fig.7. Original image and threshold image.

4. Mammography Comparison

Mammogram registration cannot usually be applied directly for the automatic detection of small abnormalities since the breast is a highly dynamic organ in which numerous normal changes occur regularly.

This way, subtracting registered mammogram sequences may lead to a large number of false positives (in this context defined as normal changes and misregistrations), appearing in the difference image.

Mammogram registration could potentially be used to trace back "missed cancers" in previous mammogram sessions. The novelty of the method presented here derives mainly from the detailed understanding of the temporal changes that can occur between successive acquisitions. Breast compression and imaging condition variability often lead to a nonrigid transform in the image plane (even a small difference in compression can lead to a significant and uneven displacement of the breast structures), and a nonrigid transformation of image intensity, respectively. Temporal changes can also occur due to breast positioning resulting in a rotation of the breast between two mammographic exposures.

The intensities of a mammogram pair can be normalized using the representation of interesting tissue [4] which results in the standard mammogram form (SMF), a standardized representation of a mammogram computed from the image intensities (film or digital) and imaging parameters of the system used to acquire the image. The method described in this [4] deals with the *geometrical alignment* of mammogram sequences. This method incorporates a robust method to select consistent boundary landmarks for the *automatic alignment* of the breast boundary [4].

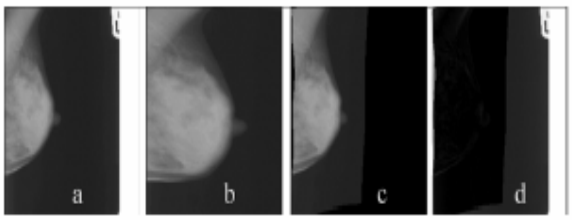

Fig. 8. Mammograms comparison using registrations. (a) The original mammogram. (b) A random mammogram transform is applied and the mammogram is significantly deformed nonrigidly. (c) Mammogram is aligned to its original shape. (d) The difference image shows that the images are almost identical.

The following program shows how to use the pattern matching tools in IMAQ Vision for mammograms comparison.

We used pattern matching in the following three general applications areas: *Alignment* - determine the position and orientation of a known microcalcification by locating features. We use the features as points of reference on the microcalcification; *Gauging*- measure lengths, diameters, angles, and other critical dimensions concerning the microcalcifications; *Inspection* - detect simple flaws, such as missing parts.

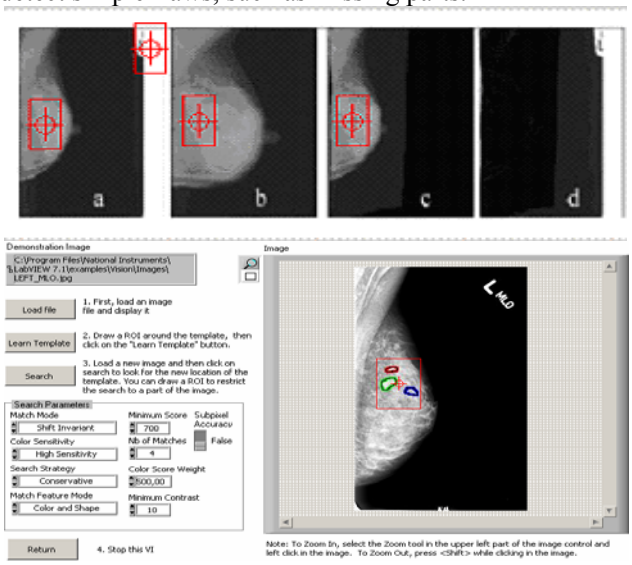

Fig.9. Mammogram sequences comparison.

References:

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

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

[3] Jerome Shapiro, *An embedded hierarchical image coder using zero-trees of wavelet coefficients,* IEEE Computer Society Press, 1993, 214-223.

[4] K. Marias, C. Behrenbuch, S. Parbhoo, A. Seifalian, *A registration framework for the comparison of mammogram sequences* Medical Imaging Trans. Vol24, June 2005 pp: 782 – 790.

[5] Z.Yang,, *Wavelets and Image Compression*, Ph.D. thesis, the University of South Carolina, Columbia,1995.

[6] J.F.Canny, *A computational approach to edge detection*, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.8, No.6, 1986,679-698.