Basic Feature Extractions from Mammograms

Marija Dakovic and Slavoljub Mijovic
University of Montenegro; Faculty of Natural Science and Mathematics
Podgorica, Montenegro
mdakovic987@gmail.com

Abstract—**Brest cancer is the most frequent cause of cancer-induced deaths in women in Europe. Systematic early detection through screening, effective diagnostic pathways and optimal treatment have the ability to substantially lower current breast cancer mortality rate. To produce image of the internal breast structure (mammogram) with adequate quality, each part of the imaging chain must function properly. Nowadays, image processing algorithms play a significant role in enhancements and visibility of specific image details. In this work, digitizing mammograms were analyzed using basic point operators to highlight particular features and to extract quantitative information. Main focus was to the contrast enhancement between "suspicious" breast structures and adjacent tissues. Basic algorithms such as normalization, equalization and thresholding were applied and their actions have been shown in both: the image and its histogram. Matlab as an image processing tool was used. It was shown that basic point operators could be successfully used to help radiologists by increasing probability in early detection of breast cancer, even if the original image was not optimally taken.**

Keywords- mammogram; histogram; gray-scale; Matlab; image processing; normalization; equalization; thresholding;

I. INTRODUCTION

Transmission X-ray radiography, which has been used for over 100 years, is based on the partial absorption of X-rays in material, which depends on thickness (x) and the material-dependent absorption length (λ) through D'Alember's Law (Fig. 1),

$$I(x) = I(0)\exp(-x/\lambda) \qquad (1)$$

which describes the exponential decrease of beam intensity with thickness [1]. An image in medicine represents the spatial distribution of the patient tissue components within the field of view. Visualization of important details requires separation of the "structures of interest" against the "background" (e.g. in mammography (a special kind of breast radiography), micro-calcifications in the breast glandular tissue) [2]. The quality of the various components of the imaging chain (focal spot, imaging geometry, image receptor, video camera and amplifier, image processing software, image display) has also influence on the image signal, obtained at the viewing station.
Nowadays, especially in mammogram's analysis, a significant role play different processing algorithms and computer aided detection (CAD), but their evaluation still makes doubts about their effectiveness.

In this paper, basic feature extractions from mammograms are described, using Matlab software for the image processing.

II. METHOD

Every "image" starts out in analog form and must be converted (digitized) to matrix of pixels, and stored as binary numbers for further computer processing. The need for processing a raw image is to highlight particular features and to extract quantitative information.

In mammography, many cancers escape detection due to the density of surrounding breast tissue. Since contrast between the soft tissues of the breast is inherently low and because relatively minor changes in mammary structure can signify the presence of a malignant breast tumor, the detection is more difficult in mammography than in most other forms of radiography [3]. Moreover, there are different abnormalities as signs for cancer and until now three basic structures can be distinguished [4]: clusters of micro-calcifications, masses and architectural distortion. Calcifications are tiny deposit of calcium which appears as small bright spots on the mammogram. They are characterized by their type and distribution properties. A mass is defined as a space-occupying lesion seen at least two different projection. Masses are described by their shape and margin characteristics. An architectural distortion occurs when normal architecture is distorted with no definite mass visible. A typical example of each of these abnormalities is shown in Fig. 2. It means that effectiveness of any processing algorithm depends on *a priory* information i.e., what we are looking for.

The image quality obtained in mammography can be described and quantified by contrast, sharpness and noise. Furthermore, the derived quantity signal-to-noise ratio (SNR) has be used as a comprehensive image quality parameter. Main focus here is the contrast enhancement. The higher the contrast the stronger the bright and dark image areas will emerge. Sometimes this effect creates the impression of a sharper image as well. Although, there are a number of different definitions for the resulting radiation contrast [2], we used a simple one, defined as

$$C = I_1 / I_2 \qquad (2)$$

where we called C as a contrast ratio, and I_1, I_2 brightness of the "structures of interest" and adjacent "background" tissues respectively.

Different algorithms exist and have advantages and disadvantages for the specific task required for breast imaging-diagnosis and screening. Matlab is the most popular of current tools for image processing. The main advantage of this tool is that one can transpose mathematics pretty well directly and sees how it works.

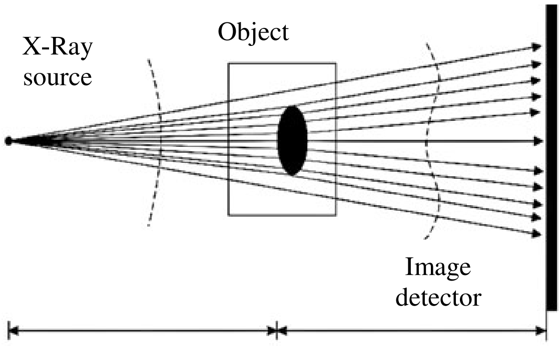

Figure 1. Geometry in radiography imaging

The mammograms, used in this work, are processed using so-called point operations, where each pixel value of the original image is replaced with a new value, using mathematical functions or compute from the image itself. The digitized images are taken from the most easily accessed the Mammographic Image Analysis Society (MIAS) database and the Digital Database for Screening Mammography (DDSM) [5]. The sizes of all pictures are $N \times N$ (1024x1024 pixels) with 200 micron pixel edge and representation of the each pixel with an 8-bit word.

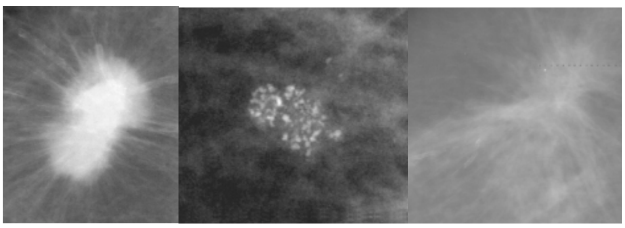

Figure 2. Examples of a mass (left); cluster of micro-calcifications (center), and architectural distortion (right)

III. RESULTS AND DISCUSSIONS

The intensity histogram of an image is a basic parameter which characterizes the image quality. It plots the number of pixels with a particular brightness level against the brightness level. A 8-bit pixel gives brightness levels, ranging between 0 and 255, and cover enough the range of the signal to noise ratio [6].

We did not use whole mammograms but only a part where "suspicious structures of interest" were found. The reason to do so lies in a fact that any mammogram has a lot of black pixels (where X-rays passes directly to the detector without crossing the breast), which do not carry any useful information but complicate its histogram representation. A region of interest (ROI 386x386 pixels) of a mammogram and its histogram is shown in Fig. 3.

A "suspicious" tissue is marked. It is obvious that the image has low contrast and the histogram has not used all available grey levels.

Accordingly, the image could be stretched to use them all, and the image would become clearer. Popular techniques to stretch the range of intensities include histogram (intensity) normalization. Here, the original histogram is stretched and shifted to cover all the 256 available levels (Fig.4). This point operator replaces the brightness at points in the image according to a linear brightness relation,

$$\mathbf{N}_{x,y} = k \cdot \mathbf{O}_{x,y} + l \qquad \forall x, y \in 1, N \tag{3}$$

where are l a level, k a gain, $\mathbf{N}$ the brightness of the points in a new image and $\mathbf{O}$ the brightness of the points in the old image. The level controls overall brightness and is the minimum value of the new image. The gain controls the contrast, or range, and if the gain greater than unity, the output

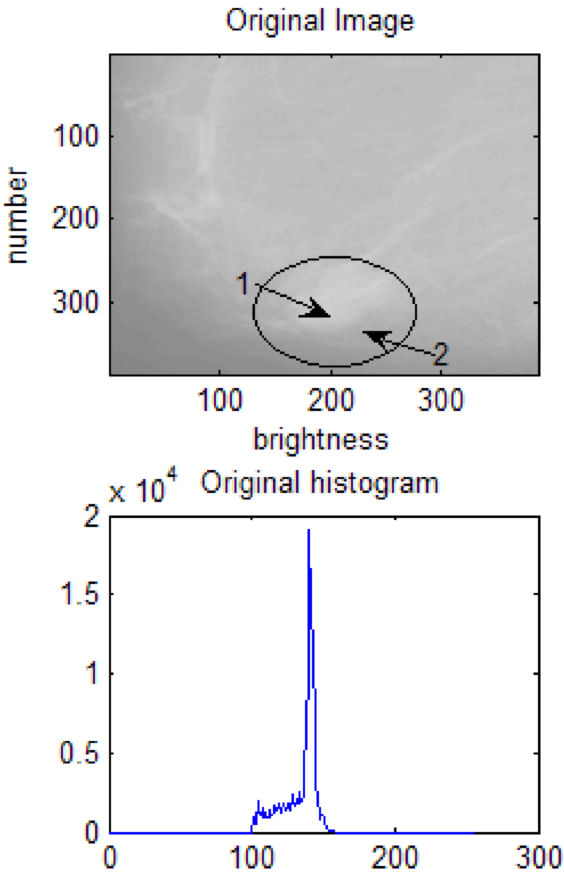

Figure 3. An image of a low quality mammogram and its histogram

range will be increased. Because of a linear transformation, normalization action has no loss of any information from the

This manuscript is done in the framework of the National project "Physics of Ionized Gases and Ionizing radiation" image, and one can return to the original image, just by subtracting the level and taking reciprocal value of the gain.

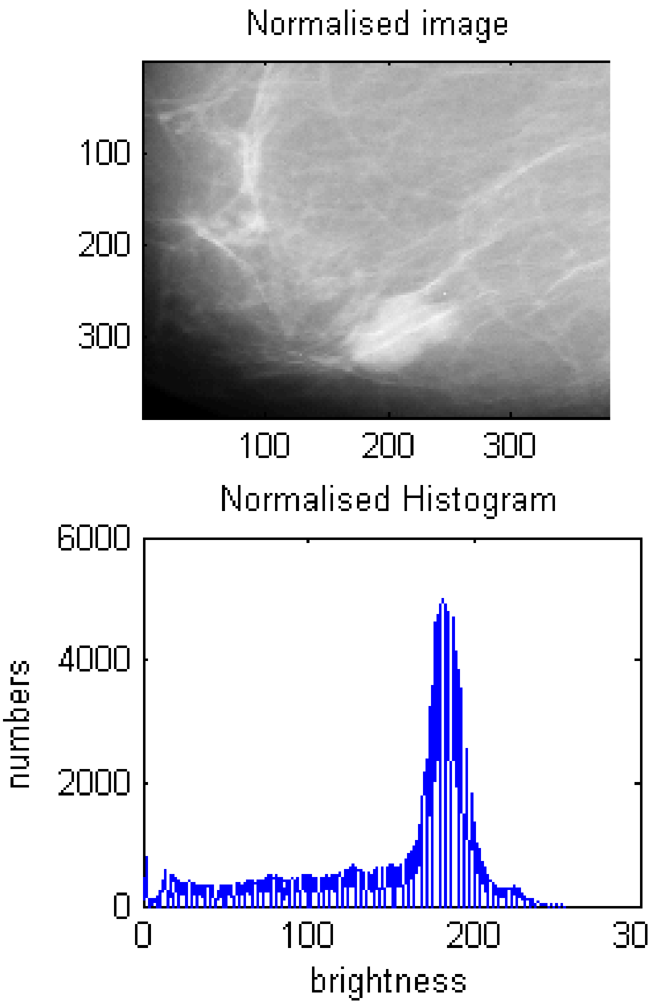

Figure 4. The normalized image and its histogram

Histogram equalization is a non-linear process aimed to highlight image brightness in a way particularly suited to human visual analysis. Histogram equalization aims to change an image in such a way as to produce an image with a flatter histogram, where all levels have equal probability. To develop the operator for equalization, the histogram of the old image is needed. If the old image is a square image, there are N^2 points in the old and new images and if the intensities are in a range of M levels, so the sum of points per level in each should be equal:

$$\sum_{l=0}^{M} \mathbf{O}(l) = \sum_{l=0}^{M} \mathbf{N}(l). \qquad (3)$$

Furthermore, since we are looking for an new image with uniformly flat histogram the condition (3) should be fulfilled for an arbitrarily chosen levels p, as well. Thus, the cumulative histogram up to level p should be transformed to cover up to the level q in the new histogram:

$$\sum_{l=0}^{p} \mathbf{O}(l) = \sum_{l=0}^{q} \mathbf{N}(l). \qquad (4)$$

Since the new histogram is uniformly flat, the cumulative histogram up to level p, should be a fraction of overall sum. This gives a mapping for the output pixels at level q, from the input pixels at level p [6].

Although, this is a non-linear process and is irreversible. Thus, one cannot return to the original image after equalization and, in contrast to intensity normalization, some information is lost. The result of the action of this point operator is shown in Fig 5. The intensity equalized image, has much better defined features than in the original image.

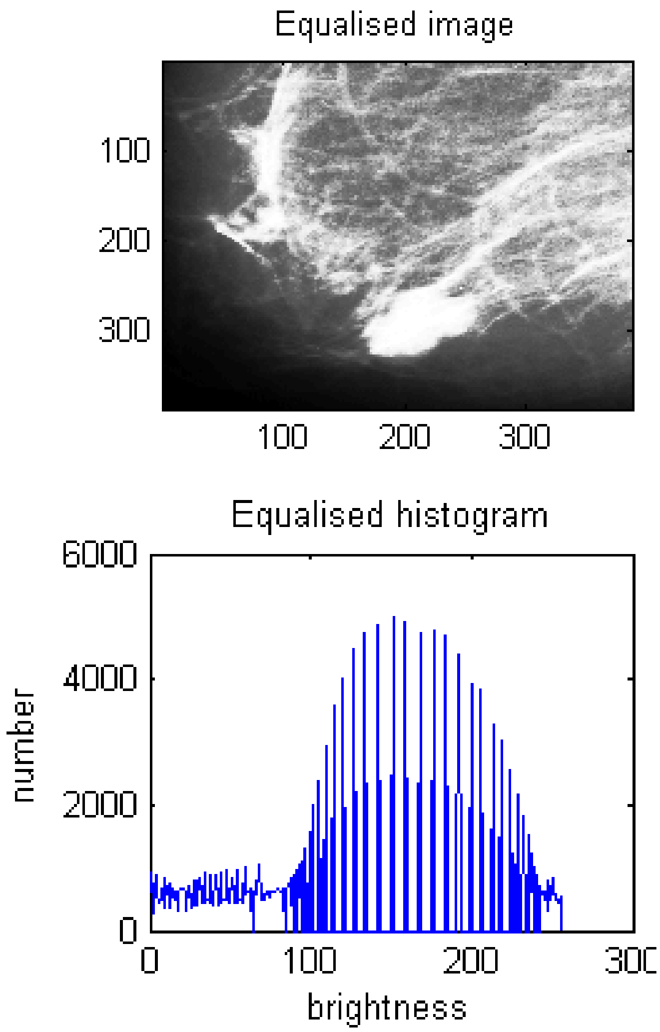

Figure 5. The equalized image and its histogram

The histogram shows the non-linear mapping process whereby white and black are not assigned equal weight, as they were in intensity normalization. Accordingly, more pixels are mapped into the darker region and the brighter intensities become better spread, consistent with the aims of histogram equalization.

Finally, the point operator of major interest is called thresholding. This operator selects pixels that have a particular value, or are within a specific range. It can be used to find objects within an image if their brightness level (or range) is known. Detailed *a priory* information is needed. The result of the action of this point operator is shown in Fig. 6. The image shows a thresholded image where all pixels above 160 brightness levels are set to white, and those below 160 brightness level are set to black. Here, this threshold of 160 is arbitrarily chosen, however the "suspicious" tissue is detected.

Concerning to an analysis of the contrast enhancement in above processing actions, brightness of two points in the "structures of interest" (position 1) and adjacent "background" (position 2) were taken (see Fig. 1).

The results are given in the Table 1. As one can see, significant contrast ratios are increased when normalization or equalization were done. The performance of the equalized image is very convincing since it is well mapped to the properties of human vision. If we replace pixel values with ones computed according to (3), the result of histogram equalization will not change. This is to be expected, since the linear operation of the brightness change does not change the overall shape of the histogram, only the size and position.

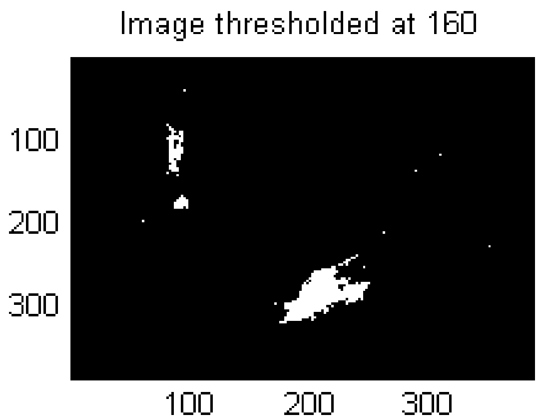

Figure 6. The thesholded image

However, noise in the image acquisition process will affect the shape of the original histogram, and hence the equalized version. So the equalized histogram of an image will not be the same as the equalized histogram of an image with some noise added to it [6].

TABLE I

Image	**Brightness in Position 1** (X:202; Y:323)	**Brightness in Position 2** (X:219; Y:339)	**Contrast ratio *C***
Original	144	125	1.15
Normalized	201	114	1.76
Equalized	238	54	4.41

IV. CONCLUSION

Radiologists visually search mammograms for specific abnormalities. These abnormalities have different shapes, margins and structures. To extract useful information from mammograms the quality of the image plays foremost role. Image processing as a chain to improve this quality contains many algorithms and different approaches. To compare them among themselves with aiming to estimate their relative effectiveness is still complex task. Probably, a unique algorithm does not exist to fulfill all requirements.

Basic feature extraction from mammograms can be done using the point operators and manipulation with their histograms. The histogram representation of an image and manipulation with them, gives better objective insights in the image quality. All these processes, such as normalization, equalization and thresholding are aimed to help radiologist to detect any abnormality in breast tissues. Although, the equalized image is the most suitable for human vision, from the above mention reasons, intensity normalization should be preferred when image's histogram requires manipulation. Further characterization of these lesions in breast tissues requires more sophisticated techniques, knowledge in this field, and *a priory* information.

REFERENCES

[1] Ed by Gordon Fraser, "The New Physics for the twenty-first century," Cambridge University Press 2006.

[2] H. Aichinger, J. Dierker, S. J. Barfuß and M. Säbel, Radiation Exposure and Image Quality in X-Ray Diagnostic Radiology, Springer-Verlag Heilderberg 2004.

[3] I. Brodie R. A. Gutcheck "Radiographic information theory and application to mammography", *Medical Physics* Vol. 9, pp.79 1982.

[4] Ed by Al Bovik, "Handbook of Image and Video Processing", Academic Press, San Diego, 2000.

[5] The Mammographic Image Analysis Society (MIAS) database and the Digital Database for Screening Mammography (DDSM), http://www.mammoimage.org/databases/.

[6] M. Nixon, A. Aguado, "Feature Exstraction & Image Processing", Elsevier Ltd. 2008.