3-23

MVA2007 IAPR Conference on Machine Vision Applications, May 16-18, 2007, Tokyo, JAPAN

Pre-processing Algorithms on Digital Mammograms

H. Mirzaalian[1] M.R Ahmadzadeh[2] S. Sadri[3] M. Jafari[4]

[1,2,3] Department of Electrical and Computer Eng., Isfahan University of Technology, Iran

[1]hengame_m@yahoo.com, [2]ahmadzadeh@cc.iut.ac.ir, [3]sadri@cc.iut.ac.ir

[4]Department of Medical Science, Shahid Beheshti University, Iran, drmehdi@yahoo.com

Abstract

Mammography is the best method for early mass detection. In order to limit the search for abnormalities by Computer Aided Diagnosis systems to the region of the breast without undue influence from the background of the mammogram, extraction of the breast contour and pectoral muscle is necessary. Breast contour helps to find the position of the nipple, which its position is important for mass detection in the next stages and presence of pectoral muscle in the mammogram could bias the detection procedures. So during analysis, the pectoral muscle should preferably be excluded from processing. In this paper we propose one low-pass mask for detecting breast contour and a new method for the identification of the pectoral muscle in most medio-lateral oblique mammograms based on Non-Linear Diffusion algorithm which is an edge preserving smoother. Evaluation of the breast contour and pectoral muscle detected in the mammograms were performed by the Hausdorff Distance Measure (HDM) and also the Mean of Absolute Error Distance Measure (MAEDM) based on a distance transform and image algebra between the edges identified by radiologists and by the proposed methods. Then we compare our results by other segmentation methods. Our proposed algorithms show superior results in comparison.

1 Introduction

Mammography is the most widely used method to screen asymptomatic women for early detection of breast cancer. The large number of mammograms generated by screening of population must be interpreted and diagnosed by relatively few radiologists [1]. In order to improve the accuracy of interpretation, a variety of computer-aided diagnosis (CAD) systems have been proposed [2].

Interpretation of breast images involves two major processes: Computer-Aided Detection (CADe) and Computer-Aided Diagnosis (CADi). Detection is the ability to identify potential abnormalities and classifying regions of a mammogram as positive or negative. Diagnosis is the ability to characterize or classify a detected abnormal entity as being either benign or malignant. However, long before CADe algorithms can perform their task of identifying, pre-processing steps must be taken. These include: mammogram orientation, label and artifact removal, mammogram enhancement and mammogram segmentation [3].

Generally, segmentation in digital mammogram includes two stages: breast contour identification and pectoral muscle segmentation [3].

The pectoral muscle represents a predominant density region in most medio-lateral oblique (MLO) views of mammograms, and can affect the results of image processing [4].

In this work, we propose two preprocessing algorithms, one for the breast contour extraction and the other for pectoral muscle segmentation.

The organization of the paper is as follows: In section 2 proposed algorithm for extraction of breast contour is described. In section 3 Non-Linear Diffusion method is introduced. Section 4 presents the proposed algorithm for pectoral muscle segmentation. Section 5 discuses experimental results and conclusions.

2 Algorithm for extracting the breast contour

The block diagram for extracting breast contour algorithm is shown in Figure 1. Here a short description of each block is given.

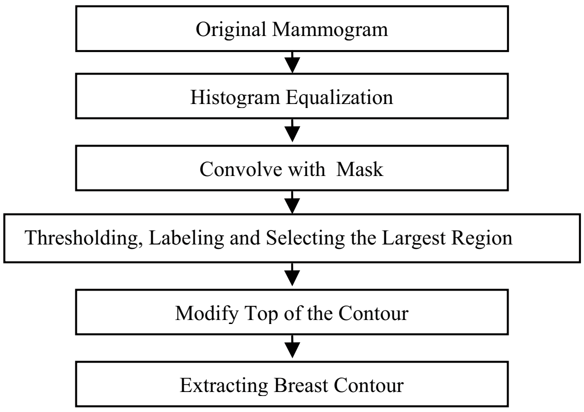


Figure 1. Block diagram of breast contour extraction

2.1 Histogram equalization

Because this algorithm should operate similarly on different mammograms, it is recommended to normalize images before any processing. For having this, at the beginning of the algorithm histogram equalization is used.

2.2 Convolve with mask

In this stage we propose the below mask for convolving with the image.

$$MASK = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (1)$$

This mask is acting like an integrator so it is a low-pass filter. It goes without saying that the size of the mask is dependent on the resolution of the image. In Figure 2 Fourier Transform of one mammogram and the proposed mask is shown.

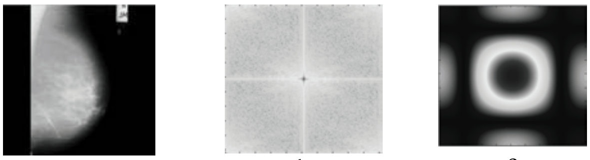


Figure 2:(a) Mammogram mdb004 from MIAS, (b) Fourier-Transform of the mammogram, (c) Fourier-Transform of the proposed mask

Because of that boundary of the breast is related to low frequency, breast contour can be extracted by thresholding the image which is resulted from convolving the mammogram with the mask.

2.3 Thresholding and labeling

One suitable threshold for the image can be achieved from the knee of the cumulative histogram. After having binary image, some morphologic operations such as closing and opening are used to remove small noises. Larger noises are removed by labeling algorithm. The breast region is subsequently identified as the largest non-zero component.

2.4 Modifying ends of breast border

Usually, contrast between breast region and the background in the top of the image is very low, which leads to inaccuracy border detection. To fix this problem we have added additional post-processing stage. Figure 3(a) shows only top part of a mammogram which its detected border usually is very erroneous. Figure 3(b) shows intensity of one row of image 3(a). Experimentally we observed that the real border is located at the main valley of the curve. This point has been circled in the Figure 3(b). To detect this point we have designed the following mask.

Mask=[1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 -2 -2 -2 0 0 0 0 0 0 0 0 0 0 0 0] (2)

This mask is convolved with the top of the mammogram (see Figure 3(c)). Figure 3(d) shows intensity of one row of figure 3(c). By scanning figure 3(d) from left to right, the first nonzero pixel corresponds to the real border (Figure 3(e)). Candidate points finally are fitted by a robust polynomial. The output is shown in Figure 3 (f). This post-processing stage can be used for the bottom part of the image as well.

3 Non-Linear Diffusion

One of the most important problems in image processing is denoising. Usually the procedure used for denoising, is dependent on the features of the image, aim of processing and also post-processing algorithms [5].

Denoising by low-pass filtering not only reduces the noise but also blurs the edges. On the contrary non-linear diffusion is smoother which is edge preserving [6].

The non-linear diffusion is based on an analogy of physical diffusion processes, like the temperature diffusion on a metal bar, or the diffusion between two fluids put together. These physical diffusion processes are modeled by the following differential equation [7]:

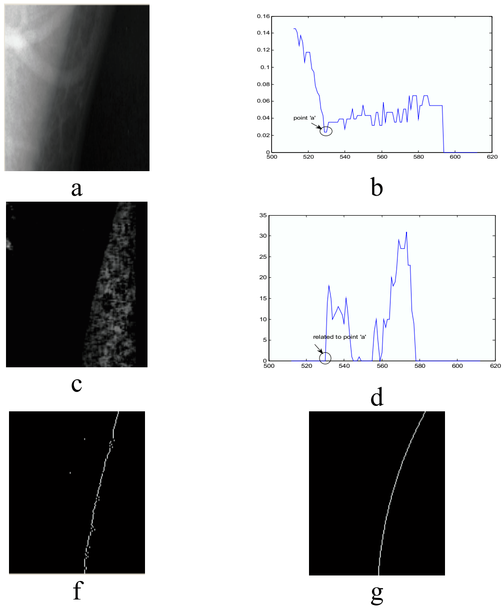


Figure 3. (a) Image of top of the mdb018; (b) values of row 77; (c) convolved image 'a' and proposed mask; (d) values of row 77 of image 'c'; (e) border extracted; (f) polynomial fitted curve (smooth border)

$$\frac{\partial U}{\partial t} = D\nabla.(\nabla U) \qquad (3)$$

where U is the concentration (temperature), constant D is the diffusivity (a conductance), $\nabla.$ is divergence operator and ∇ is gradient operator. Accordingly concentration variation will be faster where the concentration gradient is higher [7].

Relation (3) is linear and the procedure of diffusing acts on noise and edges at the same time. Thus using a linear diffusion would quickly destroy the borders [8].

To overcome this problem, non-linear diffusion is introduced. The nonlinear diffusion makes the diffusivity parameter, D, no longer a constant value, but instead the diffusivity becomes a function of the concentration gradient which decreases for higher gradients as following:

$$\frac{\partial U(x,t)}{\partial t} = \nabla.(D(x,t)\nabla U(x,t)) \qquad (4)$$

This new formulation allows performing an image denoising while preserving the borders.

4 Algorithm for extracting pectoral muscle

In this section the overall method used for pectoral muscle detection is proposed. The flowchart of this method is shown in Figure 4. Here a short description of each block is given.

4.1 Definition of ROI

In order to define an appropriate region of interest (ROI) containing the pectoral muscle, after obtaining breast border approximately, five control points are defined. These control points, *N1* to *N5*, which are used to define ROI are shown in Figure 5.

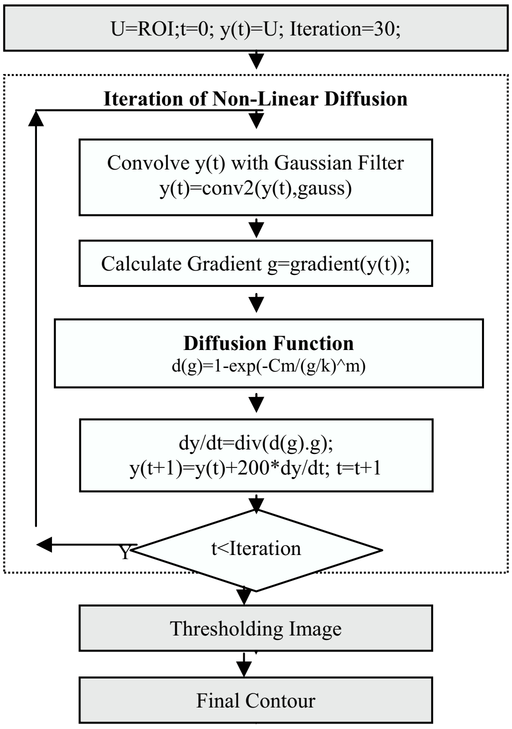


Figure 4: Flow chart of the procedure for identification of the pectoral muscle

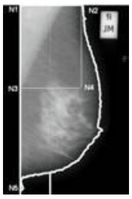


Figure 5: Approximate breast border and control points, *N1* to *N5* used to limit the region of interest (rectangle marked) for the detection of the pectoral muscle

N1: top-left corner pixel of the breast contour.
N2: top-right corner pixel of the breast contour.
N5: lowest pixel on the left edge of the boundary
N3: mid-point between *N1* and *N5*;
N4: the point that completes a rectangle with *N1*, *N2*, and *N3*

4.2 Non-Linear Diffusion

The diffusivity function, $D(x,t)$, that is used in this paper is given by:

$$D(x,t) = 1 - \exp\left(-\frac{C_m}{\left(\frac{\nabla U(x,t)}{\lambda}\right)^m}\right) \qquad (5)$$

The constant C_m is selected to make the flux, $\nabla U(x,t) * D(x,t)$, ascending for $\nabla U(x,t) < \lambda$ and descending for $\nabla U(x,t) > \lambda$, as can be seen in Figure 6.

λ is a contrast parameter (if the gradient is inferior to λ, the flux is increasing with the gradient and if the gradient is larger then λ, the flux is decreasing as the gradient grows).

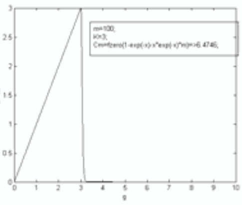


Figure 6: Flux $\nabla U(x,t) * D(x,t)$ for equation 5

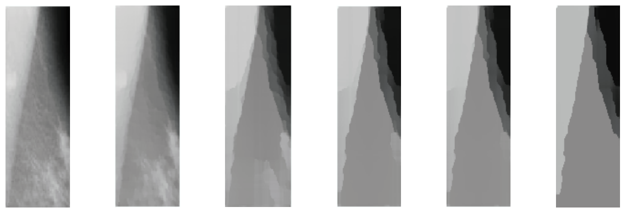
m defines the speed of the diffusivity (and the flux) changes for a variation in the gradient. Big values of m make the flux change quickly [7].

In Non-Linear Diffusion block before calculating gradient of image, it is denoised with Gaussian filter to smooth partially the abrupt changes of pectoral muscle.

Implementing this algorithm for large number of mammogram shows that suitable λ and m for better enhancing of the region which contains pectoral muscle is 3 and 100 respectively.

4.3 Thresholding Image

As it can be seen in figure 7(f), pectoral muscle region can be extracted by thresholding. The threshold is selected as average intensity of the up-left pixels in the image.



a b c d e f

Figure 7: Results of iteration of non-linear diffusion for the detection of the pectoral muscle on ROI for mdb012 from MIAS

5 Experimental Results and Conclusions

The proposed methods are applied to 90 mammograms from Mammography Image Analysis Society (MIAS) database [9]. Figure 8(a) shows the original mammogram mdb028 and Figure 8(b) the convolved image. Figure 8(c) shows the binary image. As it can be seen, border in the top of the image has not been detected properly. Figure 8(d) shows the output image after amending the top of the image by the final post-processing stage.

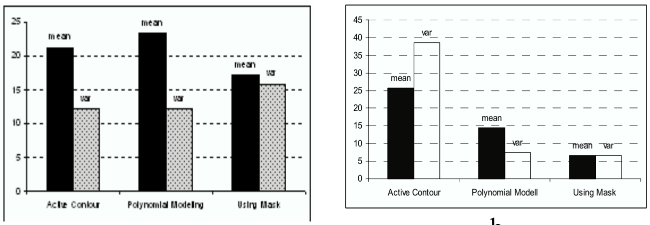
a b c d

Figure 8. Mammogram segmentation results for MIAS image mdb028. (a) Original mammogram; (b) Convolved image; (c) Binary image; (d) Final contour after modifying top of the Image.

We compare our results by hand drawn border lines which are recognized by two expert radiologists.

Evaluation of the breast contour detected in the mammograms was performed by the Hausdorff Distance Measure (HDM) [10] and also the Mean of Absolute Error Distance Measure (MAEDM) based on a distance transform and image algebra between the edges identified by radiologists and by the proposed method.

The comparison of error between our method and Ferrari's method [11] (based on active contour) and Wirth's method [12] (based on polynomial modeling of the background) is shown in Figure 9. As it can be seen our method outperforms two other methods.

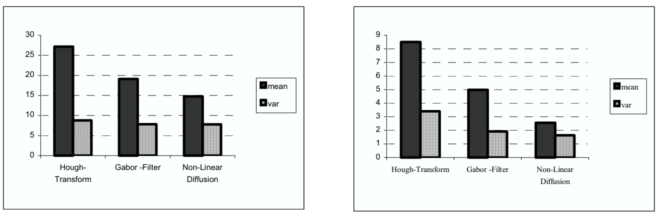


a b

Figure 9. A comparison of error between our method and other methods for 20 mammograms: (a) By HDM, (b) By MAEDM

For evaluating proposed pectoral muscle segmentation method, we compare our results by two other pectoral muscle segmentation methods which are proposed by Karssemeijer et al. [13] and Ferrari et al. [14]. The first one is based on Hough-Transform and the second one is based on Gabor-Filters.

Figure 10 shows the comparison of error between our method, Hough-Transform method and Gabor-Filter method for 20 mammograms with HDM and MAEDM respectively. As it can be seen, our method outperforms both of them.



a b

Figure10: Evaluated methods: (a) HDM, (b) MAEDM

Acknowledgment: We would like to thank Prof. Rangayyan, faculty of electronic and computer department at the University of Calgary and Alberta, who generously made available to us hand drawn border lines and also their results on Hough transform method.

References

[1] M.L. Giger, R.M. Nishikawa, M. Kupinski, U. Bick, M. Zhang, R.A. Schmidt, D.E. Wolverton, C.E. Comstock, J. Papaioannou, S.A. Collins, A.M. Urbas, C.J. Vyborny, and K. Doi, "Computerized Detection of Breast Lesions in Digitized Mammograms and Results with a Clinically-Implemented Intelligent Workstation", *presented at Computer Assisted Radiology and Surgery, Berlin, Germany*, pp.325-330, 1997.

[2] M. A, Wirth, Nonrigid Approach to Medical Image Registration: Matching Images of the Breast, Ph.D. Thesis, RMIT University, Melbourne, Australia, 2000.

[3] M. A. Wirth, A. Stapinski, "Segmentation of the breast region in mammograms using active Contours", in Visual Communications and Image Processing, Vol. 5150, 2003, pp.1995-2006.

[4] S. M. Kwok, R, Chandrasekhar, Y, Attikiouzel, M. T, Rickard, "Automatic Pectoral Muscle Segmentation on Mediolateral Oblique View Mammograms", *IEEE Trans. Med. Imag*, VOL. 23, NO. 9, pp. 232-246, Sept2004.

[5] J. Starck, E. Cand`es, D. Donoho, "The curvelet transform for image denoising", *IEEE Transactions On image processing*, 11, pp. 670-684, 2000.

[6] A. Buades, B. Coll, J. Morel, "A non-local algorithm for image denoising", *IEEE Comp. Soc.*, vol. 2, pp. 60 – 65, 2005

[7] P. Perona, J. Malik, "Scale space and edge detection using anisotropic diffusion", *IEEE Comp. Soc.* , vol. 12, no. 7, pp. 629-639, 1990.

[8] J. Weickert, W. Kropatsch, R. Klette, F. Solina, Theoretical foundations of anisotropic diffusion in image processing, Theoretical foundations of computer vision, Computing Supplement 11 , Springer Wien, pp. 221-236, 1996.

[9] J. Suckling, J. Parker, D. R. Dance, S. Astely, I. Hutt, C. R. M. Boggis, I. Ricketts, E. Stamakis, N. Cerneaz, S. L. Kok, P, Taylor, D. Betal, and J. Savage, "The mammographic image Analysis society digital mammogram database," in Digital Mammography: Proc. of the 2nd International Workshop on Digital Mammography, York, England: Elsevier, pp. 375–378, 1994.

[10] D. P. Huttenlocher, G. A. Klanderman, W. J. Rucklidge, "Comparing images using the Hausdorff distance," IEEE Trans. Pattern Anal. Machine Intell., vol. 15, pp. 850–863, 1993.

[11] R. J. Ferrari, R. M. Rangayyan, J. E. L. Desautels, and A. F. Frere, "Identification of the breast boundary in mammograms using active contour models", Medical and Biological Engineering and Computing, 42: 201-208, 2004.

[12] M. A. Wirth, *A Nonrigid Approach to Medical Image Registration: Matching Images of the Breast,* Ph.D. Thesis, RMIT University, Melbourne, Australia, 2000.

[13] N. Karssemeijer, "Automated classification of parenchymal patterns in mammograms," *Phys. Med. Biol*., vol. 43, no. 2, pp. 365–378, 1998.

[14] R. J. ,Ferrari, R. M. Rangayyan, J. E. L. ,Desautels, R. A. Borges, and , A. F. Frère, "Automatic Identification of the Pectoral Muscle in Mammograms", *IEEE Trans. Med. Imag*, VOL. 23, NO. 2, pp. 232-245, 2004