

Automatic Breast Boundary Segmentation of Mammograms

R. R. Nanayakkara, Y. P. R. D. Yapa, P. B. Hevawithana, P. Wijekoon

Abstract—Accurate breast boundary estimation and segmentation of breast tissue region from the background of the mammogram image is an important pre-processing task in computer-aided diagnosis of breast cancer. This paper presents an automated system to estimate skin-line and breast segmentation. The proposed method is based on automatic seed region selection, modified fast marching algorithm to propagate the seed region and automatic boundary point selection with intensity gradient information to initial boundary estimation and morphological operators to final boundary estimation and breast tissue region segmentation. Performance of the proposed method was tested by using 136 mammogram images with all types of breast tissues taken from mini-MIAS database. The results obtained from the experimental evaluation indicate that the sensitivity of this algorithm is 99.2% of the ground truth breast region and accuracy of the segmentation is 99.0%. By analyzing the results we can conclude that this system is capable of estimate the breast boundary and segment the breast area from background for all three types of breast tissues with high accuracy level.

Keywords— Breast Cancer, Mathematical Morphology, Modified Fast Marching Algorithm

I. INTRODUCTION

According to the statistics about cancer, breast cancer is the most frequent cancer in women worldwide, with nearly 1.67 million new cases diagnosed in 2012 [1]. It is the second leading cancer type in the world and represents 12% of all new cancer cases and 25% of all cancers in women [1]. Early detection of breast cancer increases the prognosis and hence increases the rate of survivals. Mammography is still the widely used breast screening tool for detecting breast cancers. Retrospective studies have shown that the estimated sensitivity of radiologists in breast cancer screening is only about 75% [2]. Therefore many computer vision techniques have been proposed to increase the accuracy of detecting breast cancers. These techniques are divided in to three phases as Pre-processing, Computer Aided Detection or Diagnosis (CAD) and post-processing.

In the pre-processing stage, a given mammographic image is mainly categorized in to three important regions such as breast tissue area, pectoral muscle area and background area including markers.

Manuscript Received on March 2015.

R. R. Nanayakkara, Postgraduate Institute of Science, University of Peradeniya, Sri Lanka.

Y. P. R. D. Yapa, Department of Statistics & Computer Science, Faculty of Science, University of Peradeniya, Sri Lanka.

P. B. Hevawithana, Department of Radiology, Faculty of Medicine, University of Peradeniya, Sri Lanka.

P. Wijekoon, Department of Statistics & Computer Science, Faculty of Science, University of Peradeniya, Sri Lanka.

Accurate breast skin line segmentation is a very important step in enhancing and analyzing of mammograms for computer aided detection of breast cancer. Breast tissue in the breast skin line area is less dense compared to the other adjacent tissues. Because of this reason skin line area in mammogram is very low in grey level contrast [3]. Hence, if the skin line is not detected accurately there is a chance that lesions attached to the skin line are overlooked.

The proposed method mainly consists with four steps Pre-processing, Segmentation, Skin-line estimation and Post processing. In pre processing step, first noise in the mammogram was removed. Attribute morphological operators were used to remove external articles such as identification labels, markers and wedges and to improve the area homogeneity. In the segmentation stage, the seed region was selected automatically by using thresholding and morphological operators. Then the selected seed region was propagated until the skin-line is reached by using modified fast marching algorithm, and an endpoint was selected automatically to stop the region growing. Finally, at the post-processing stage, morphological processing was used to remove false pixels from the skin-line and obtained a smoothed estimate.

II. METHODOLOGY

A. Pre-processing

External artifacts near the breast boundary directly caused to reduce the accuracy of breast skin-line estimation and segmentation results [3]. To increase the accuracy of final estimates, these artifacts must be removed. First, Gaussian smoothing was used to remove the noise. Then the attribute morphological based article suppression algorithm to remove those external articles from the background [4]. Attribute morphological operators, such as area-opening operator can be used to remove such artifacts less than a certain size. Alternating Sequential Filter (ASF) (area closing followed by area opening) was used finally to improve the homogeneity among neighboring regions [5]. Attribute morphology doesn't need a pre-defined structuring element as in standard morphological operators. Morphological filters using standard structuring elements always do not guarantee that it will reconstruct the original structure to be preserved, as it is. Attribute morphology will be a good solution for problems that we need to remove unwanted articles on an image without damaging the shape of the original object.

B. Automated seed region selection

To automatically select a seed region inside the breast tissue

region, first a threshold was applied [6] and then, different morphological operators such as area opening, dilation, erosion were used to smooth the boundary of thresholded breast region (seed contour). The estimated boundary contour was compared with 138 relevant mask images which were validated by medical experts and the seed region was completely inside the actual boundary. Hence the detected region was used as the seed region for the proposed method.

C. Determining breast orientation and automated end point selection

Intensity profile along a mammogram (fig. 1) is an important characteristic used for estimating an end point on the skin-line. However the intensity profile itself does not provide sufficient information to extract the boundary point. Therefore the intensity gradient profile will be used which provides useful information to extract the boundary position and breast orientation. Fig. 1 shows the intensity profile along the selected line on a mammogram and Fig. 2 shows the intensity gradient profile of the same line.

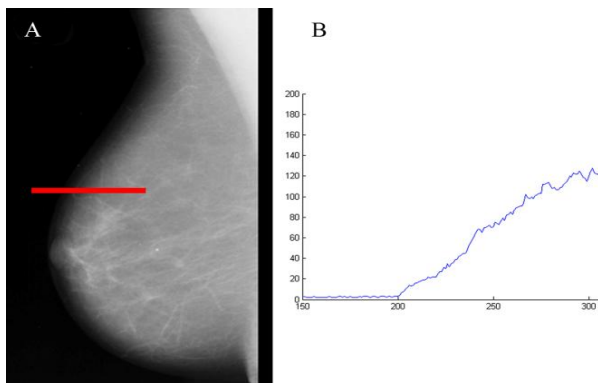


Figure 1: (A) Selected line along a mammogram (B) Intensity profile along the selected line

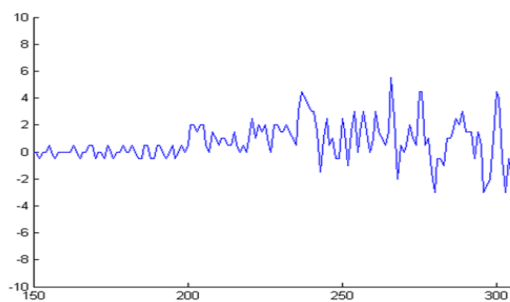


Figure 2: Intensity gradient profile along the selected line
From intensity gradient profile first, local maxima and minima and their coordinate points were extracted [7]. Peak is the highest or lowest point between two valleys or hills on the profile. When comparing all peaks, if maximum of absolute values of peaks belongs to local maxima then it is a Right Oriented mammogram or if it belongs to local minima it is a Left Oriented mammogram. Fig. 3 demonstrates an intensity gradient profile of a Right oriented mammogram. Peaks are shown by asterisks. Fig. 4 demonstrates an intensity gradient profile of a left oriented mammogram.

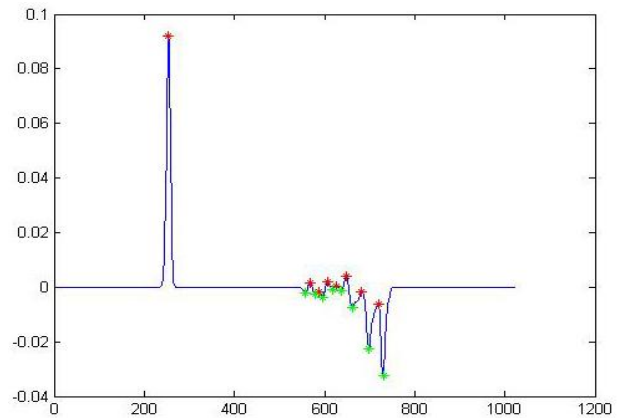


Figure 3: Intensity gradient profile of a right oriented mammogram

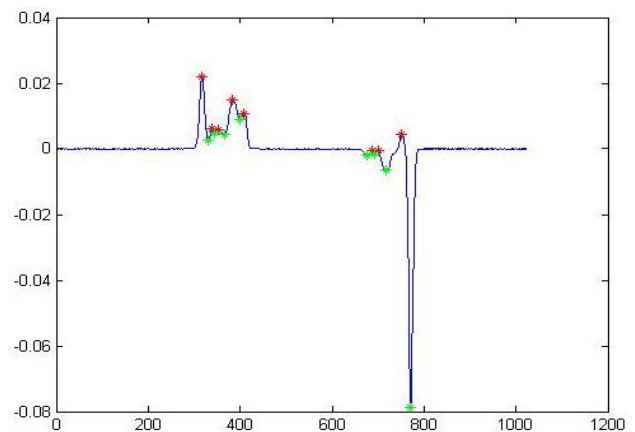


Figure 4: Intensity gradient profile of a left oriented mammogram

To segment the breast region from the background, modified fast marching method was used. After determining the breast orientation, an end point should be selected to stop the propagation of the seed region. Generally, lower and upper area of mammograms contains noise besides the low contrast at breast-edge. This could sometimes introduce undesirable artifacts on lower and upper area of skin-line [3]. Therefore intensity profile line in the middle area of the mammogram was taken to select an end point. When selecting the end point, minimum peaks (local minima) were considered for the right oriented mammograms and maximum peaks (local maxima) were considered for the left oriented mammograms. Simply a minimum or maximum peak of intensity gradient profile could not be selected as an end point. Peaks due to suppressed artifacts and the peak due to real mammogram film edge (mammogram images in MIAS data base have been clipped or padded to make the size of every image to 1024x1024 pixels.) should be avoided. All peaks in the intensity profile which are located far from 80 pixels from seed region were avoided. Then, all peaks which are located inside the seed region were also avoided, and the maximum peak of remaining peaks was selected as an end point. Finally, the modified fast marching method was applied on mammogram to segment the breast region from background [3]. In Fig. 5 the seed region is shown by drawn contour and the selected end point is shown

by asterisks for mammograms with different characteristics.

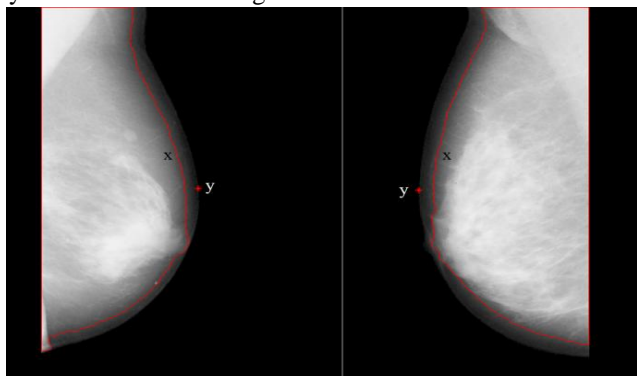


Fig. 5: Automatically selected seed region (x) and automatically selected end-point (y)

Fig. 6 shows the segmented breast region and breast contour superimposed on Original mammogram image, based on the fast marching method.

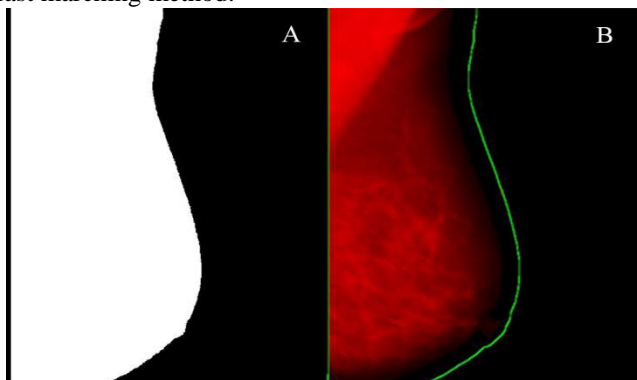


Fig. 6: (A) Segmented breast tissue region (B) Breast contour superimposed on original mammogram image

D. Post processing

After estimating the breast skin-line, morphological filter was applied to suppress inconsistencies on the estimated skin-line. Standard morphological closing-opening (ASF) operator was used to remove small false-positive and false-negative artifacts from lower and upper area of the segmented mammogram image (3). Flow chart 1 shows the different steps of proposed method.

III. PERFORMANCE EVALUATION

In the literature, breast skin line segmentation methods can be classified in to five different methods such as Histogram based methods (8-11), polynomial modeling based methods (12, 13), gradient based methods (14-17), active contour method (18-21) and classifiers (22-25). Almost all algorithms related to mammographic breast boundary detection, have been evaluated qualitatively based on observations made by a radiologist. Most of them have not provided quantitative statistical comparison of measurements. *Completeness* (CM) and *Correctness* (CR) are the two evaluation metrics used to derive the quantitative measurement statistically (26).

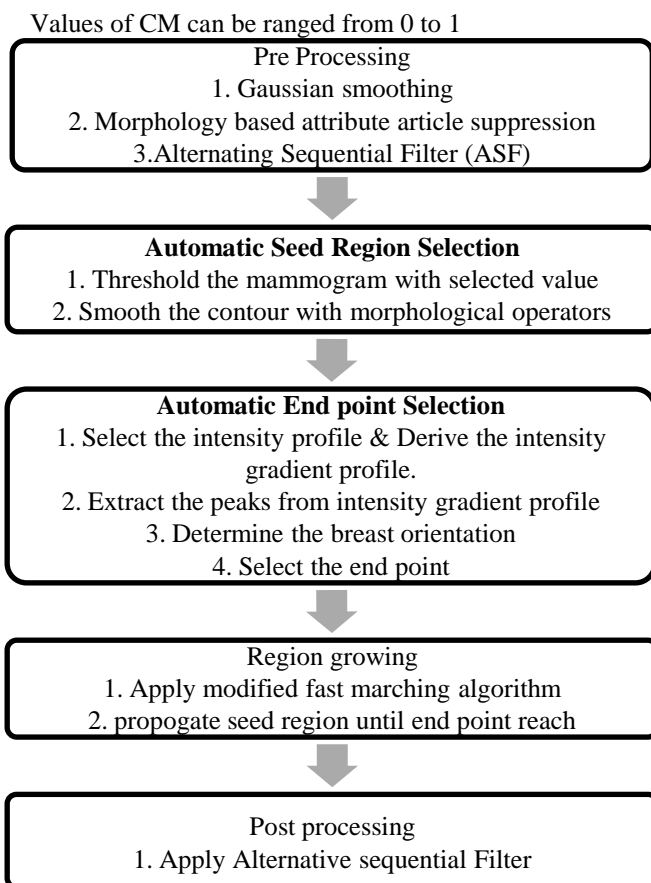
Completeness - percentage of the true breast region (GT) extracted by the segmentation algorithm.

$$Completeness = \frac{TP}{(TP + FN)}$$

Values of CM can be ranged from 0 to 1

Correctness - percentage of correctly extracted breast region.

$$Correctness = \frac{TP}{(TP + FP)}$$



Flow chart 1: Steps of proposed method

IV. EXPERIMENTAL RESULTS

This experiment was done using 136 mammograms taken from mini-MIAS database (27). Mini-MIAS database is an unrestricted database, which is a popular database among mammogram research community. It contains 320 mammograms digitized at 200 micron pixel edge and those images have been clipped or padded so that every image consists of 1024x1024 pixels. The 136 images used for the experiment, consisted all types of mammograms: 36 of images with fatty tissue, 60 of images with fatty-glandular tissue, and 40 of images with dense glandular breast tissues. Different types of breast tissues give different levels of visibility due to different levels of X-ray attenuations (28). Generally, mammograms from breasts with fatty tissues are darker and low in contrast while dense-glandular tissues are brighter. Table 1 shows the results according to breast tissue type.

Table 1: Mean Completeness and mean Correctness according to the breast tissue type

Tissue Type	Mean CM	Mean CR
Fatty	0.99	0.99
Fatty-Glandular	0.99	0.98
Dense-Glandular	0.99	0.99

Mean value for completeness is 99.2% and correctness is 99% for all the 136 mammogram images used for the experiment. Which means only 0.8 % of pixels



from breast region was missed and 1% of pixels were recognized erroneously from the background. Corresponding standard deviations were 1.04% and 2.72 % respectively. This shows that the segmentation algorithm maintains a higher accuracy level. Also the algorithm is robust with respect to the different breast density types. By combining CM and CR measures, the proposed algorithm gives an average quality measure of 98.21%. According to literature, there are very few segmentation algorithms that have been tested broadly. Bick et al. [10] have tested their algorithm on 740 mammogram images, and have visually rated that their algorithm is “acceptable” on 97% of mammogram images and segmentation problems encountered on other 2.9% mammogram images due to digitization artifacts. The contour based algorithm developed by Mendez et al. [29] that have been tested on 156 mammograms and claimed that the breast contour is “accurate” or “nearly accurate” only in 89% of the mammogram images. The algorithm developed by Abdel-Mottaleb et al. [15] that has been tested on 500 mammogram images and achieved “acceptable” boundary in 98% of the images. Chandrasekhar et al. [30] claim that their algorithm estimates breast skin-line with 94% accuracy on mammogram images in MIAS database [27]. Raba et al. [31] have tested their method over 320 images, and have obtained a 98% of “near accurate” results, which includes “accurate” results. Dehghani and Dezfooli [32] have tested their method over 60 images, and have obtained a 99% of “near accurate” results, which include the “accurate” results. According to the results they recognized for their method as robust approach and even they accept the method since it provides useful regions without meaningful information loss. Most of the above proposed methods did not provide quantitative statistical evaluation for the accuracy but visually rated accuracy levels or “acceptable” and “nearly acceptable” categorization evaluation based on observation made by radiologist. Accuracy of most of the above methods is high but they include the near accurate or acceptable results.

Original Image, Segmented mask and result obtained by applying the proposed segmentation algorithm on Fatty, Fatty-Glandular and Dense-Glandular images are shown in figs. 7, 8 and 9 respectively.

Fig. 7 demonstrates the result obtained by applying the algorithm for Fatty image “mdb077” with CM = 0.9998 and CR= 0.9973 which means that the segmented image describes 99.98% of GT and it has correctly extracted 99.73% from GT region.



Fig. 7: Segmentation of Fatty mammogram “mdb077”. (A) Original image, (B) Segmented mask, (C) Estimated skin-line and segmented region superimposed on the original image

Fig. 8 demonstrates the result obtained by applying the algorithm for Fatty-Glandular image “mdb008” with CM =

0.9973 and CR= 0.998 which means that the segmented image describes 99.73% of GT, and it has correctly extracted 99.8% from GT region.

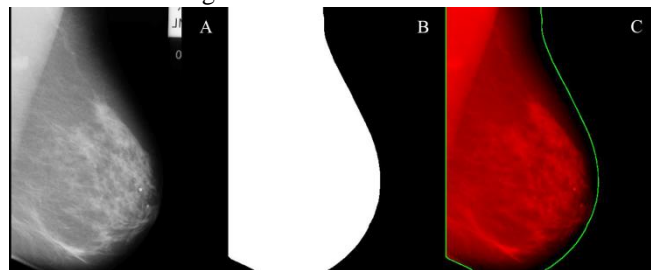


Fig. 8: Segmentation of Fatty-Glandular mammogram “mdb008”. (A) Original image, (B) Segmented mask, (C) Estimated skin-line and segmented region superimposed on the original image

Fig. 9 shows the result obtained for Dense -Glandular image “mdb191” with CM = 0.9919 and CR= 0.9999, which means that the segmented image describes 99.19% of GT and it has correctly extracted 99.99% from GT region.

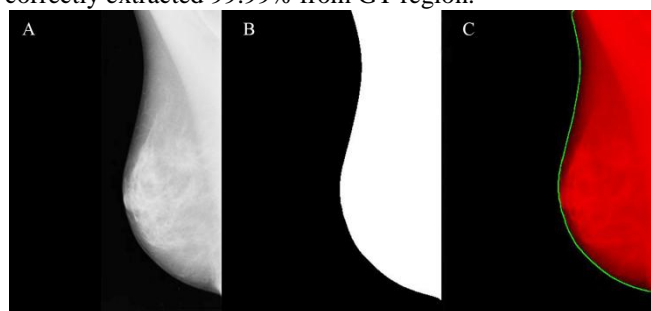


Fig. 9: Segmentation of Dense-Glandular mammogram “mdb191”. (a)Original Image, (b) Segmented mask, (c) Estimated skin-line and segmented region superimposed on the original image.

V. DISCUSSION

Accurate breast boundary estimation and breast segmentation from the background of the image is a significant pre-processing in computer-aided diagnosis of breast cancer. Due to external articles, low contrast in the breast boundary region and noise present in the mammogram, breast skin line segmentation is not an easy task.

By analyzing the results of proposed automated breast boundary segmentation system, we can conclude that this system is capable of estimate the breast boundary and segment the breast area from background for all three types of breast tissues with high accuracy level. Moreover this method is fully automated and can operate independent of image orientation.

The proposed algorithm can provide as a pre-processing component of an intelligent workstation for computer-aided diagnosis in mammography.

ACKNOWLEDGEMENT

R. R. Nanayakkara would likes to thank Board of studies of Statistics and Computer Science, Postgraduate Institute of Science, University of Peradeniya.



REFERENCES

- J. Ferlay, I. Soerjomataram, M. Ervik, R. Dikshit, S. Eser, C. Mathers, M. Rebelo, D.M. Parkin, D. Forman, F. Bray, GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC Cancer Base No. 11 [Internet]. Lyon, France: International Agency for Research on Cancer; 2013. Available from: <http://globocan.iarc.fr>, accessed on 11/2/2015.
- R. Bird, Professional quality assurance for mammography screening programs. *Radiology*. 1990;175(2):587-605. Pub Med PMID: 2217807. DOI: 10.1148/radiology.177.2.2217807.
- R.D. Yapa, K. Harada, Breast skin-line estimation and breast segmentation in mammograms using fast-marching method. *International Journal of Biological, Biomedical and Medical Sciences*. 2008;3(1):54-62.
- M.A. Wirth, J. Lyon, D. Nikitenko, A. Stapinski, editors. Removing radiopaque artifacts from mammograms using area morphology. *Image Processing Proc SPIE 2004: Proceedings of SPIE Medical Imaging; 2004; San Diego, California, USA: SPIE; 2004. p. 1054-65. DOI: 10.1117/12.535372.*
- P. Salembier, J. Serra, Flat zones filtering, connected operators, and filters by reconstruction. *Image Processing, IEEE transactions on*. 1995;4(8):1153-60. DOI: 10.1109/83.403422.
- J. Nagi, S. Abdul Kareem, F. Nagi, S. K. Ahmed, editors. Automated breast profile segmentation for ROI detection using digital mammograms. *Biomedical Engineering and Sciences (IECBES), 2010 IEEE EMBS Conference on; 2010. IEEE. DOI: 10.1109/iecbes.2010.5742205.*
- B. Eli, Eli Billauer's Freelancing Electrical Engineer. Israel: Eli Billauer; [20 July 2012; cited 2013 03]; Available from: <http://billauer.co.il/peakdet.html>.
- T. Ojala, J. Näppi, O. Nevalainen. Accurate segmentation of the breast region from digitized mammograms. *Computerized Medical Imaging and Graphics*. 2001;25(1):47-59. DOI: 10.1016/s0895-6111(00)00036-7.
- T.-K. Lau, W. F. Bischof. Automated detection of breast tumors using the asymmetry approach. *Computers and biomedical research*. 1991;24(3):273-95. DOI: 10.1016/0010-4809(91)90049-3.
- U. Bick, M. L. Giger, R. A. Schmidt, R. M. Nishikawa, D. E. Wolverton, K. Doi. Automated segmentation of digitized mammograms. *Academic Radiology*. 1995;2(1):1-9. PubMed PMID: 9419517. DOI: 10.1016/s1076-6332(05)80239-9.
- J. J. Heine, M. Kallergi, S. M. Chetelat, L. P. Clarke. Multi resolution wavelet approach for separating the breast region from the background in high resolution digital mammography. *Digital Mammography: Springer; 1998. p. 295-8. DOI: 10.1007/978-94-011-5318-8_49.*
- R. Chandrasekhar, Y. Attikiouzel, editors. Gross segmentation of mammograms using a polynomial model. *Engineering in Medicine and Biology Society, 1996 Bridging Disciplines for Biomedicine Proceedings of the 18th Annual International Conference of the IEEE; 1996. IEEE. DOI: 10.1109/iembs.1996.652707.*
- M. M. Goodsitt, H-P. Chan, B. Liu, S. V. Guru, A. R. Morton, S. Keshavmurthy, et al. Classification of compressed breast shapes for the design of equalization filters in x-ray mammography. *Medical Physics*. 1998;25(6):937-48. PubMed PMID: 9650184. DOI: 10.1118/1.598272.
- C. Zhou, H-P. Chan, N. Petrick, M. A. Helvie, M. M. Goodsitt, B. Sahiner, et al. Computerized image analysis: Estimation of breast density on mammograms. *Medical physics*. 2001;28(6):1056-69. PubMed PMID: 11439475. DOI: 10.1118/1.1376640.
- M. Abdel-Mottaleb, C. S. Carman, C. R. Hill, S. Vafai, Skinline detection in digitized mammograms. *Journal of digital imaging*. 1997;10:224-5. PubMed Central PMCID: PMC3452812. DOI: 10.1007/BF03168707.
- A. Morton, H. Chan, M. Goodsitt. Automated model-guided breast segmentation algorithm. *Med Phys*. 1996;23:1107-8.
- N. Karssemeijer, G. Te Brake. Combining single view features and asymmetry for detection of mass lesions. *Digital Mammography: Springer; 1998. p. 95-102. DOI: 10.1007/978-94-011-5318-8_16.*
- R. Ferrari, R. Rangayyan, J. Desautels, A. Frere, Segmentation of mammograms: identification of the skin-air boundary, pectoral muscle, and fibro-glandular disc. *Proceedings of the 5th international workshop on digital Mammography*. 2000:573-9. PubMed PMID: 15125. DOI: 10.1148/radiol.2291032535.
- S. Thiruvenkadam, M. Acharyya, N. Neeba, P. Jhunjhunwala, S. Ranjan, editors. A region-based active contour method for extraction of breast skin-line in mammograms. *Biomedical Imaging: From Nano to Macro, 2010 IEEE International Symposium on; 2010. IEEE. DOI: 10.1109/isbi.2010.5490383.*
- M. A. Wirth, A. Stapinski, editors. Segmentation of the breast region in mammograms using snakes. *Computer and Robot Vision, 2004 Proceedings First Canadian Conference on; 2004. IEEE. DOI: 10.1109/ccrv.2004.1301473.*
- R. Martí, A. Oliver, D. Raba, J. Freixenet, Breast skin-line segmentation using contour growing. *Pattern Recognition and Image Analysis: Springer; 2007. p. 564-71. DOI: 10.1007/978-3-540-72849-8_71.*
- P. K. Saha, J. K. Udupa, E. F. Conant, D. P. Chakraborty, D. Sullivan, Breast tissue density quantification via digitized mammograms. *Medical Imaging, IEEE Transactions on*. 2001;20(8):792-803. PubMed PMID: 11513030. DOI: 10.1109/42.938247.
- H. E. Rickard, G. D. Tourassi, A. S. Elmaghaby, editors. Self-organizing maps for masking mammography images. *Information Technology Applications in Biomedicine, 2003 4th International IEEE EMBS Special Topic Conference on; 2003. IEEE. DOI: 10.1109/itab.2003.1222538.*
- M. Wirth, J. Lyon, D. Nikitenko, editors. A fuzzy approach to segmenting the breast region in mammograms. *Fuzzy Information, 2004 Processing NAFIPS'04 IEEE Annual Meeting of the; 2004. IEEE. DOI: 10.1109/nafigs.2004.1336329.*
- J. M. Gauch. Image segmentation and analysis via multiscale gradient watershed hierarchies. *Image Processing, IEEE Transactions on*. 1999;8(1):69-79. DOI: 10.1109/83.736688.
- W. Wirth, D. Nikitenko, J. Lyon, Segmentation of Breast Region in Mammograms using a Rule-Based Fuzzy Reasoning Algorithm. *International Journal on Graphics, Vision and Image Processing*. 2005;5(2):45-54.
- J. Suckling, The mini-MIAS database of mammograms. In: *Society TMA, editor. Digital Mammography Database ver 1.2.ed.*
- P. C. Johns, M. J. Yaffe. X-ray characterisation of normal and neoplastic breast tissues. *Physics in medicine and biology*. 1987;32(6):675. DOI: 10.1088/0031-9155/32/6/002.
- A. J. Méndez, P. G. Tahoces, M. J. Lado, M. Souto, J. Correa, J. J. Vidal. Automatic detection of breast border and nipple in digital mammograms. *Computer methods and programs in biomedicine*. 1996;49(3):253-62. DOI: 10.1016/0169-2607(96)01724-5.
- R. Chandrasekhar, Y. Attikiouzel, editors. Automatic breast border segmentation by background modeling and subtraction. *5th International Workshop on Digital Mammography (IWDM), (Yaffe M ed), Medical Physics Publishing, Madison, USA; 2000.*
- D. Raba, A. Oliver, J. Martí, M. Peracaula, J. Espunya, Breast segmentation with pectoral muscle suppression on digital mammograms. *Pattern Recognition and Image Analysis: Springer; 2005. p. 471-8. DOI: 10.1007/11492542_58.*
- S. Dehghani, M. Dezfooli, A method for improve preprocessing images mammography. *International Journal of Information and Education Technology*. 2011;1(1):90-3. DOI: 10.7763/ijiet.2011.v1.15.

AUTHORA PROFILE



R. R. Nanayakkara is a M.Phil candidate at Postgraduate Institute of Science, University of Peradeniya, Sri Lanka.



Dr. Y. P. R. D. Yapa is a senior lecturer at Department of Statistics and Computer Science, Faculty of Science, University of Peradeniya Sri Lanka. He has obtained Ph.D (Information Engineering), from Hiroshima University, Japan. He is having about more than 12 years of teaching experience. Database Systems, Image Processing, Statistical Data Mining and Bioinformatics are his research areas.



B. Hewavithana is a Professor at the Department of Radiology, Faculty of Medicine, University of Peradeniya, Sri Lanka and Consultant Radiologist at Teaching Hospital Peradeniya. Imaging in early detection of breast cancer and parenchymal liver disease are her research interests.



Pushpakanthie Wijekoon is a Professor at the Department of Statistics and Computer Science, Faculty of Science, University of Peradeniya, Peradeniya, Sri Lanka. She has obtained a Ph.D (Statistics) from University of Dortmund, Dortmund, Germany. She is having about more than 30 years of teaching experience. Linear Models, Multivariate Techniques, and Theory of Statistics are her research area.