A FULLY AUTOMATED COMPLETE SEGMENTATION SCHEME FOR MAMMOGRAMS

Stylianos Tzikopoulos[1], *Harris Georgiou*[1], *Michael Mavroforakis*[1],
Nikos Dimitropoulos[2] *and Sergios Theodoridis*[1]

[1]National and Kapodistrian University of Athens, Dept. of Informatics and Telecommunications,
Panepistimiopolis, Ilissia, Athens 15784, Greece
[2]Delta Digital Imaging, Semitelou 6, Athens 11528

ABSTRACT

This paper presents a fully automated complete segmentation method for mammographic images. Image preprocessing techniques are first applied to mammograms to remove the noise and then a breast boundary extraction algorithm is implemented, in order to distinguish breast tissue from the background. Next, an improved version of an existing pectoral muscle scheme is performed and a new nipple segmentation technique is applied, detecting the nipple when it is in profile. This improves the estimated breast boundary and serves as a key-point for further processing of the image. This composite method has been implemented and applied to miniMIAS, one of the most well-known mammographic databases. This database consists of 322 mediolateral oblique (MLO) view mammograms, obtained via a digitization procedure. The results are evaluated by an expert radiologist and are very promising. Accordingly, it is expected that this procedure can produce improved results, when applied to high-quality digital mammograms.

Index Terms— image processing, mammogram, automated segmentation, breast boundary, pectoral muscle

1. INTRODUCTION

Breast cancer, i.e., a malignant tumor developed from breast cells, is considered to be one of the major causes for the increase in mortality among women, especially in developed countries. More specifically, breast cancer is the second most common type of cancer and the fifth most common cause of cancer-related death [10].

Mammography has been proved to be the most effective and reliable screening method for early breast cancer detection [12], as it achieves sensitivity and specificity of 62% and 90% respectively, while the corresponding values for the clinical breast examination (CBE) evaluation are 24% and 95% [9]. However, the large number of mammograms generated by population screening must be interpreted and diagnosed by the relatively small number of radiologists. Moreover, when observing a mammographic image, abnormalities are often embedded in and camouflaged by varying densities of breast tissue structures, resulting in high rate of missed breast cancer [16]. In order to reduce the increasing workload and improve the accuracy, a variety of computer-aided diagnosis (CAD) systems, that perform computerized mammographic analysis have been proposed. All of them require, as a first stage, the segmentation of each mammogram into its representative anatomical regions, i.e., the breast border, the pectoral muscle and the nipple.

The breast border extraction is a necessary and cumbersome step for typical CAD systems, as it must identify the breast region independently of the orientation of the breast in the image and of the possible noise, including artifacts. It should also have a fast running time and be adequately precise, so that to improve the accuracy of the overall CAD system.

Automatic segmentation of the pectoral muscle can be useful in many ways [7]. One example is the reduction of the false positives of mass detection procedure, because of the similarity between the pectoral region and the mammographic parenchyma. In addition, pectoral muscle must be excluded in an automated breast tissue density quantification method. Thus, the importance of pectoral muscle detection in a CAD system is clear and profound.

The location of the nipple is of great importance too, as it can serve as a key point for the whole mammographic image. Most CAD systems use the nipple as a registration point for comparison, when trying to detect possible asymmetries between the two breasts of a patient [17]. These automatic methods can also use the nipple as a starting point for cancer detection, as cancer appears in the glandular (not the fatty) tissue of the breast [6]. In addition, radiologists pay specific attention to the nipple, when examining a mammogram [2].

In this work, we propose a fully automated complete segmentation method for mammographic images. We implement the algorithm used in [8] to extract the breast boundary; we modify the algorithm in [7] to improve the pectoral muscle estimation; and we propose a new nipple detection technique for the mammograms, which include the nipple in profile, using the output of the breast boundary extraction procedure. The latter novel algorithm proposed in this work not only lo-

cates the nipple point, but also can serve as an improvement for the existing breast boundary algorithm, which misses the nipple, if it is in profile. The improvement is observed, when updating the breast boundary, in order to include the detected nipple. Our methodology has been tested on all the 322 mediolateral oblique view mammograms of miniMIAS [14], giving promising results, according to specific measures and an expert radiologist's evaluation.

This paper is organized as follows: In section 2 the mammographic image databased used is presented. Section 3 describes the methods implemented, including the preprocessing techniques and the segmentation algorithms. Section 4 presents the results obtained by these algorithms. The discussion and the conclusions are summarized in section 5.

2. DATASET

The previous methodology was implemented on miniMIAS database [14], available freely online for scientific purposes and consisted of 161 pairs of mediolateral oblique (MLO) view mammograms. The images of the database originated as the product of a film-screen mammogram process in the United Kingdom National Breast Screening Program. The films were digitized and the corresponding images were annotated, according to their breast density, by expert radiologists. Any abnormalities were also detected and described as calcifications, well-defined, spiculated or ill-defined masses, architectural distortions or asymmetries. Additionally, the severity of each abnormality is given, meaning benign or malignant cases.

A typical image is shown in figure 1. The precense of high noise is readily observed and this makes the segmentation of the image a demanding task. The original 0.2 mm/pixel images were resized to 0.4 mm/pixel as in [2], in order to reduce the required computational time. The initial bit depth of 8 bits was preserved.

3. METHODS

3.1. Mammogram preprocessing

As mentioned earlier, preprocessing techniques are necessary, in order to find the orientation of the mammogram, remove the noise and enhance the image.

3.1.1. Image orientation

The orientation of the mammogram is determined according to [8]. The image is transformed, so that the chest wall location, i.e., the side of the image containing the pectoral muscle, is on the left side of the image and the pectoral muscle is at the upper-left corner of the image. An example is shown in figure 2a. In order to determine the chest wall location, the decreasing pixel intensity of breast tissue near the skin-air interface

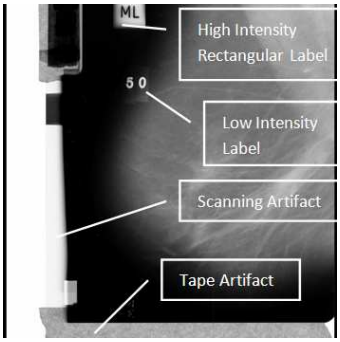


Fig. 1: Different types of noise observed at a mammogram

is used. This tissue is estimated by employing the minimum cross-entropy thresholding technique [1] twice in the original image. By estimating the first derivatives of these pixels using the appropriate masks, we can determine the chest wall location. Next, the top of the image is determined. At first we extract the vertical centroid of the image and then we assume that the asymmetric region closest to the right side of the vertical centroid is the tip of the breast. The image is flipped vertically if needed to place this asymmetric region below the vertical centroid, resulting in an image the right way up.

3.1.2. Noise Estimation

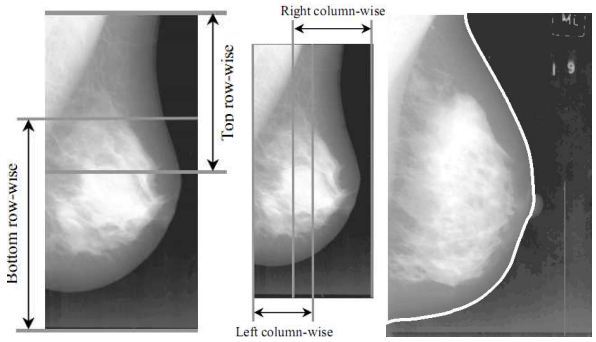
The different types of noise, that appear in miniMIAS images can be observed in figure 1. The algorithm should estimate these regions and exclude them from the remaining process.

High intensity noise is characterized the noise that corresponds to high values of optical densities, such as labels or scanning artifacts (figure 1). The algorithm used to find these regions and extract them from the remaining process is described in detail in [8] and uses the 2-level minimum cross entropy thresholding criterion [1], combined with logical and morphological operations.

Tape artifacts are any markings left by tapes, or other shadows presenting themselves as horizontally running strips, that can be also observed in figure 1. In order to locate a tape artifact, we use the algorithm described in [8]. The property of a tape artifact, that is used for the detection and exclusion from the further processing, is the horizontal line corresponding to their edges. After the high intensity noise is detected and replaced by black pixels and the image orientation is determined, we enhance the horizontal edges of the image using a 3x3 horizontal Sobel mask [13]. Finally, the Radon transform [11] is performed on theleft- half of the image containing the pectoral muscle, in order to detect straight lines.

3.1.3. Image Filtering

Except for the noise already described, the images of the database contain also noise from the digitization process, such as speckle noise. This type of noise should be removed,



(a) Row-wise interfaces (b) Column-wise interfaces (c) Final Interface

Fig. 2: The different interfaces estimated (a,b) and combined to the final one (c)

in order to enhance the quality of the image and make the segmentation task easier. The filtering technique that was selected to remove noise, preserve the edges of the image and help the breast boundary detection algorithm is the median filtering, as described at [5].

3.2. Image Segmentation

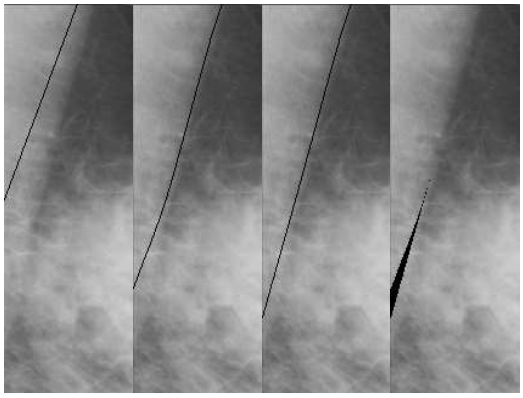
3.2.1. Breast Boundary Detection

The method used to extract the background of the mammograms is described in [8]. In this algorithm the skin-air interface is divided into one obtained from the image rows (row-wise interface) and one obtained from the image columns (column-wise interface), which are combined for the final estimate. Each one of the two interfaces is divided into two parts, as shown in figures 2a and 2b, having at the end four estimates to combine for the final one. Each one of them is transformed into a function with one value for each row or column.

The algorithm relies on the idea that the skin-air interface is the smoothest section of identical pixels near the breast edge. Based on that, we threshold the image using a specific threshold, extract the interface and fit polynomials of degree 5 to 10. Then the square error between the polynomials and the interface is calculated, punishing large values of intensities, so that not to detect contours internal of the breast. This procedure is repeated for several values of the threshold and the final estimate automatically chosen is the one producing the least error when compared with the inherently smooth polynomial.

3.2.2. Pectoral Muscle Detection

Pectoral muscle is a high-intensity, triangular region across the upper posterior margin of the image, appeared only in medio-lateral oblique view mammograms, as it is obvious at



(a) Straight Line (b) Final estimate (c) Improvement (d) Difference pixels

Fig. 3: Pectoral muscle detection. a) Straight line estimation and b) The final estimate of the algorithm of the bibliography, c) Improvement we propose, d) Difference pixels of the two techniques

figure 3a. The technique used to segment it [7] can be divided into:

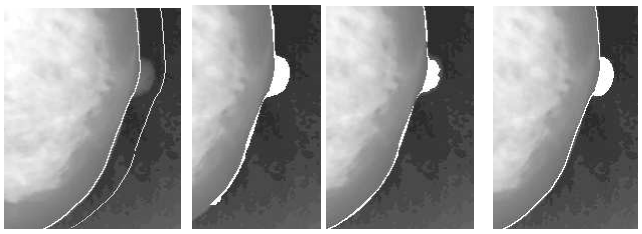
- Straight Line Estimation and Validation: The pectoral muscle is estimated as a straight line (figure 3a) and given as input to the next stage.
- Iterative Cliff Detection: The straight line estimate is iteratively refined to a curve that more accurately depicts the pectoral margin (figure 3b).

Region enclosing is performed at the end of the process, if the bottom end of the estimate is not aligned with the left edge of the image. Then the bottom end is extended by a straight line parallel to the initial straight line estimation. We have improved the existing algorithm, by fitting a straight line parallel to the straight line, which best fits the already detected estimate. The idea behind this is that the region enclosing procedure should use the updated estimate of the pectoral muscle and not the initial straight line estimation. Using the updated approach we get better results, as it can be shown at figure 3c.

3.2.3. Nipple Detection

We propose a new technique for the detection of the nipple. Our motivation was the fact that the breast boundary detection algorithm fails to detect the nipple when it is visible, as figure 2c shows, because the nipple presents some relatively sharp corners. Our starting point is to use the already estimated boundary, in order to detect the nipple, if it is visible.

The algorithm we propose uses the thresholds selected for the Right-Column, Bottom-Row and Top-Row interfaces of the breast boundary detection (Figures 2a, 2b), since it is in these regions that the nipple may be contained. Considering



(a) Search area (b) S_1 mask (c) S_3 mask (d) Final nipple estimate

Fig. 4: a) Defined search area for a nipple, b) S_1, c) S_3 binary masks searched for containing a nipple, d) Final nipple estimate

a threshold value, we assume a search area of 10mm's width right of the already detected breast boundary (of a mammogram facing right), as figure 4a shows and we threshold the search area, after performing a gaussian filter to minimize the noise of the background pixels. For each row of the search area, the first zero pixel (the pixel whose value at initial image is smaller or equal with the threshold value) is detected and all the previous columns are given the value 1, creating a new binary mask S_T, where T is the threshold value. S_T is assumed to be an area that may contain a nipple. The above procedure is repeated for the minimum and maximum values of the thresholds, as well as for the intermediate values, resulting to several binary masks, some of which are shown in figures 4b and 4c.

Considering a binary mask S_T, an ellipse with moving center at each pixel of the boundary and with variable semi-major and semi-minor axis from 2mm's to 10mm's is drawn, trying to model the possible nipple. The major axis is considered to be the tangent of the boundary at the specific point. Note that the smallest value of the the axis is smaller than the one in [2], in order to be able to detect smaller nipples. If the pixels of the circle that correspond right to the boundary have also non-zero values at binary image S_T, then a possible nipple is detected and those pixels are considered as a region of interest (*roi*).

Then we use the area $S_{T_{max}}$, defined as the binary mask obtained by the largest value of the thresholds, in order to avoid detecting possible noise as nipple. The basic idea is that the segmented mask obtained by the largest value of threshold T_{max} should contain at least one pixel of the nipple; otherwise we have detected noise as possible nipple area. Thus a logical AND operator is performed between each region of interest *roi* and $S_{T_{max}}$ and the corresponding *roi* is discarded if the result is a black binary image containing no white pixels.

Repeating the above procedure for all the binary images S_T we obtain several *roi*'s and we consider the largest of them as the possible nipple, as figure 4d shows.

4. RESULTS

4.1. Mammogram preprocessing techniques

The preprocessing techniques were executed for the images of the database. They were successful for all the images, except for the image orientation algorithm, which has the potential to fail in images where the breast is cut off. The noise is successfully detected and the tissue corresponding to tape artefacts is excluded from the further processing of the image.

4.2. Breast Boundary Detection

The fully automatic breast boundary detection algorithm was tested to the images of the database and was compared with the manually segmented images of the database given in Wirth [15]. In order to evaluate the results we used the Tanimoto Coefficient (TC) [4] and Dice Similarity Coefficient (DSC) [3]. Considering two overlapping regions A and B, they can be defined as $TC = \frac{N(A \cap B)}{N(A \cup B)}$ and $DSC = \frac{2N(A \cap B)}{N(A)+N(B)}$, having unity as optimal values. We define a search area of 10 mm's around the ground truth boundary using the morphological operation of dilation and we estimate the above metrics between the ground truth mask and the mask obtained by the fully automatic breast border estimation method, but only at the search area defined above. This way, we consider only the region around the boundary, so that to have more reliable measures. We obtained the mean values of 0.900 and 0.945 for the TC and DSC respectively for the 322 images of the database, whereas the corresponding standard deviations are 0.079 and 0.055. In other words, one can claim that the fully automated segmentation algorithm gives similar results as the manual segmentation method. An example can be shown at figure 5b.

4.3. Pectoral Muscle Detection

The pectoral muscle detection algorithm described above was tested using the images of the database. The results were successful; one of them can be shown at figure 3. It is obvious that the initial straight line estimate (figure 3a) fails to detect the pectoral muscle correctly. But the final estimate (figure 3b) is obviously much better. Finally, the improvements we propose and can be shown at figure 3c gives us the best so far estimate about the pectoral muscle. At figure 3d the difference pixels of the two estimates are shown.

4.4. Nipple Detection

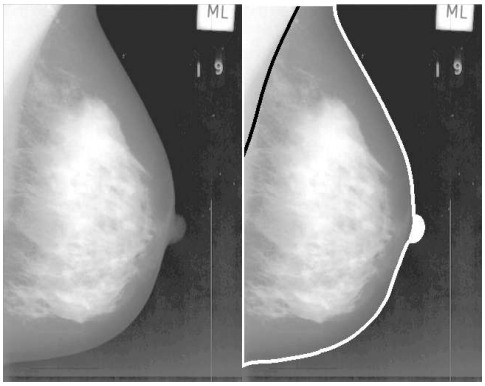
The proposed nipple detection algorithm was executed to all the images of the database. Note that the algorithm should detect a nipple, only if it is in profile and visible. So a manual annotation of the mammograms about the visibility of the nipple and its exact point was completed by an expert radiologist. The method was tested and the corresponding truth table

Nipple	Not Visible	Visible
Not Detected	189	30
Detected	15	88

Table 1: Truth table of the nipple detection algorithm

is the one of the table 1. From the 118 mammograms with a visible nipple, the nipple was correctly detected in the 88 of them, whereas in 30 mammograms was no nipple detected. These 30 mammograms were carefully observed and in 25 of them the nipple was recognized partly in profile, less than 1 mm. At these cases, the already detected breast boundary has succeeded to segment the nipple. From the 204 mammograms with no nipple in profile was a nipple detected to 15 of them. After careful observation of these cases, we conclude that the algorithm fails, due to extremely high level of noise.

An example of the nipple detection algorithm is shown at figure 5b. In order to evaluate the improvement by the nipple detection technique, we estimate the new values of the TC and DSC measures, when including the detected nipple to the breast boundary estimation. They have the values 0.903 and 0.947 with standard deviations 0.078 and 0.055 respectively. Although the increase may not be large, it must be noted that the boundary changes only in cases where the nipple is detected (103 images) and the area of the boundary changes is too small comperated to the whole boundary of the image.



(a) Initial image (b) Breast boundary, pectoral muscle and nipple detected

Fig. 5: a) Initial image, b) Breast boundary, pectoral muscle and nipple detected

5. DISCUSSION AND CONCLUSION

The results of the method, when performed to all the images of the miniMIAS database, were detailed observed and evaluated. The reader should note the high level of noise of the images due mainly to the digitization process, making a hard task the segmentation process .

The proprocessing techniques are characterized as successful except for the breast orientation algorithm, which failed in two images. However in these images, the breast tissue is cut off from the image, resulting to a non-acceptable mammography, according to the radiologists. The implemented breast boundary detection technique is based on a simple inference and gives acceptable results, according to specific measures as well asto careful observation of the detected boundary of the images. However it has the drawback of cutting off the nipple when it is mostly in profile. The boundary derived from the pectoral muscle detection scheme is acceptable and it further improves through the modification we propose. The new nipple detection technique can serve as an improvement for the already known breast boundary, but also as a key point for the further processing of the image, because of the importance of the point of the nipple in a mammographic image. Note that this technique cannot be objectively compared to the algorithms of the bibliography. The most similar one is the work at [2], which uses a small subset of the miniMIAS database and has a different goal than ours. The results are evaluated by an expert radiologist and are promising enough to expect even better results, when applied to high quality digital mammograms.

6. REFERENCES

[1] A. D. Brink and N. E. Pendock. Minimum cross-entropy threshold selection. *Pattern Recognition*, 29(1):179–188, 1996.

[2] R. Chandrasekhar and Y. Attikiouzel. A simple method for automatically locating the nipple on mammograms. *Medical Imaging, IEEE Transactions on*, 16(5):483–494, 1997.

[3] L.R. Dice. Measures of the Amount of Ecologic Association Between Species. *Ecology*, 26(3):297–302, 1945.

[4] R. O. Duda and P. E. Hart. *Pattem Classification and Scene Analysis*. New York Wiley, 1973.

[5] R.C. Gonzalez and R.E. Woods. *Digital Image Processing*. Prentice Hall, 2007.

[6] H. Knauerhase, M. Strietzel, B. Gerber, T. Reimer, and R. Fietkau. Tumor Location, Interval Between Surgery and Radiotherapy, and Boost Technique Influence Local Control After Breast-Conserving Surgery and Radiation: Retrospective Analysis of Monoinstitutional Long-Term Results. *International Journal of Radiation Oncology, Biology, Physics*, 2008.

[7] S.M. Kwok, R. Chandrasekhar, Y. Attikiouzel, and MT Rickard. Automatic pectoral muscle segmentation on mediolateral oblique view mammograms. *Med-*

ical Imaging, IEEE Transactions on, 23(9):1129–1140, 2004.

[8] M. Masek, Electronic University of Western Australia School of Electrical, Computer Engineering, and University of Western Australia Centre for Intelligent Information Processing Systems. *Hierarchical Segmentation of Mammograms Based on Pixel Intensity*. PhD thesis, 2004.

[9] M. E. Mavroforakis. *Geometric Approach to Statistical Learning Theory through Support Vector Machines with application to medical diagnosis*. PhD thesis, School of Science, Department of Informatics and Telecommunications, October 2008.

[10] R.M. Nishikawa. Current status and future directions of computer-aided diagnosis in mammography. *Computerized Medical Imaging and Graphics*, 31:224–235, 2007.

[11] J. Radon. Über die Bestimmung von Funktionen durch ihre Integralwerte längs gewisser Mannigfaltigkeiten. *Berichte Sächsische Akademie der Wissenschaften, Leipzig, Mathematisch-Physikalische Klasse*, 69:262–277, 1917.

[12] MKJ Siddiqui, M. Anand, PK Mehrotra, R. Sarangi, and N. Mathur. Biomonitoring of organochlorines in women with benign and malignant breast disease. *Environmental Research*, 98(2):250–257, 2005.

[13] I. Sobel. Camera models and machine perception AIM-21, 1970.

[14] J. Suckling, J. Parker, D. Dance, S. Astley, I. Hutt, C. Boggis, I. Ricketts, E. Stamatakis, N. Cerneaz, S. Kok, et al. The Mammographic Image Analysis Society Digital Mammogram Database. In *Exerpta Medica. International Congress Series*, pages 375–378, 1994.

[15] M.A. Wirth. *MIAS Mask Database*. University of Guelph, Canada, 2005.

[16] A. Wroblewska, P. Boninski, A. Przelaskowski, and M. Kazubek. Segmentation and feature extraction for reliable classification of microcalcifications in digital mammograms. *Opto-Electronics Review*, 11(3):227–235, 2003.

[17] F.F. Yin, M.L. Giger, K. Doi, C.J. Vyborny, and R.A. Schmidt. Computerized detection of masses in digital mammograms: Automated alignment of breast images and its effect on bilateral-subtraction technique. *Medical Physics*, 21:445, 1994.