Automatic Segmentation of Fibroglandular Tissue

Christina Olsén and Aamir Mukhdoomi

Department of Computing Science,
Umeå University, SE-901 87 Umeå, Sweden
{colsen, c01ami}@cs.umu.se

Abstract. In this paper, a segmentation algorithm is proposed which extracts regions depicting fibroglandular tissue in a mammogram. There has been an increasing need for such algorithms due to several reasons, the majority of which are related to the development of techniques for Computer Aided Diagnosis of breast cancer from mammograms. The proposed algorithm consists of a major phase and a post-processing phase. The purpose of the major phase is to calculate the threshold value that yields a segmentation of glandular tissue which is achievable by thresholding. The method by which we calculate this threshold value is based on the principle of minimizing the cross-entropy between two images. The resulting segmentation is then post-processed to remove artifacts such as noise and other unwanted regions. The algorithm has been implemented and evaluated with promising results. In particular, its performance seems to match that of medical professionals specialized in mammography.

1 Introduction

Breast cancer in females often occur in fibroglandular part of the breast tissue. Due to several reasons, the majority of which are related to the development of techniques for Computer Aided Diagnosis of breast cancer from mammograms, there has been an increasing need for develop an automated segmentation algorithm for extracting the glandular tissue. A first step in that process might be to locate and extract the glandular tissue disc. The next step would be to analyze the different structures within that disc. The location and delineation of the glandular tissue disc is also important, as a quality assurance, in order to determine if the entire glandular tissue disc is depicted in the mammogram. To make this procedure fully automatic is a highly difficult segmentation task since many different tissue types within the tree-dimensional breast superimpose as projected down to a two-dimensional image plane.

Several methods have been proposed previously to perform the segmentation of glandular tissue in digital mammograms. Saha et al. [13] described an automatic method to segment dense tissue regions from fat within breasts using scale-based fuzzy connectivity methods. Bovis and Singh [1] and Karssemeijer [4] have devised segmentation algorithms based on classification methods through

feed-forward Artificial Neural Network and kNN classifier respectively. Various studies found in the literature [12], are also based on the texture analysis of the tissue in mammograms. Different tools are used for analyzing texture such as for instance fractal dimensions and image filters [5].

In addition to the kind of methods mentioned above, a method based on a simple concept of analyzing the histogram of an image is also found in the literature. From this method an optimal threshold value is calculated that can be used to create a distinction between two regions of the image that have different intensity compositions. In this case, the required optimal threshold should create a clear distinction between the fibroglandular tissue and the remaining breast tissue. This method of finding the optimal threshold is called *Minimum Cross-Entropy*, and was used by Masek [8] to segment different regions in digital mammograms. His method however requires manual setting of many parameters for the different types of tissue. The aim of this paper is to develop an objective and fully automated segmentation algorithm for extracting the glandular tissue disc from mammograms. To reach this objective the idea by Masek [8] is used, because of its simplicity, robustness and the fact that the segmentation is independent of the exact location of the glandular tissue disc in a mammogram.

2 The Method

In this paper, a collection of 200 mammograms were randomly chosen from the two databases that are commonly used in image analysis projects related to digital mammography, i.e. *Mammographic Image Analysis Society's* (MIAS) digital mammography database and the *Digital Database for Screening Mammography* (DDSM). These 200 mammograms are part of a bigger research project, where the randomly chosen images were manually examined by five medical experts [10]. Among other tasks, these experts were asked to mark what they perceived as anatomical landmarks in each mammogram. This dataset was divided into 40 randomly chosen training images used during the development and 160 test images used for evaluation of the proposed algorithm. All the images used have dimension 1024×1024 and are 8-bit grayscale images.

2.1 Cross-Entropy Based Thresholding

It has already been established that the region of interest (ROI) for this algorithm is the region of the mammogram that depicts glandular tissue. In the process of extracting the glandular tissue disc it is assumed that the position of the nipple, the breast boundary, the pectoralis muscle and other unwanted regions such as for instance labels are extracted by Olsén's algorithms [9].

In order to use thresholding based segmentation, it is important to examine the images in question to determine if their intensity composition allows segmentation of ROIs by thresholding. However, in these images it can be observed that there are well defined peaks that correspond to the glandular tissue in the histogram of each mammogram. Such a peak has been found to exist in all

the mammograms in varying shapes and sizes. Therefore a robust algorithm is needed that can detect this peak and calculate an acceptable threshold intensity with consistent performance independent of the shape and size of the peak.

Cross-Entropy. The concept of cross-entropy was first introduced by Kullback [6], and has proven to be very useful in different applications as it provides a robust method to measure how good one probability distribution approximates another probability distribution by quantifying the difference between the two.

If P and Q are two discrete probability distributions, such that $P = p_1, \dots, p_n$ and $Q = q_1, \dots, q_n$, then the cross-entropy distance between P and Q is given as:

$$HCE(Q, P) = \sum_{i=1}^n q_i \log \frac{q_i}{p_i} \quad (1)$$

Minimum Cross-Entropy and the LLBP Approach. If a digital image substitutes P in Equation (1) and its thresholded version substitutes Q , then the cross-entropy measure explains how good the thresholded image estimates the original image. Following this line of thought, the principle of minimum cross-entropy states that in order to find the solution (i.e. thresholded image) that best estimates the original image, one should minimize the difference (cross-entropy) between them. In our case this will practically involve minimizing cross-entropy values between an image and its thresholded version over a certain range of threshold values.

Li and Lee [7] and Brink and Pendock [2] have designed image thresholding algorithms based on minimum cross-entropy as well. These two algorithms, along with two other approaches [15,11], have been tested and evaluated on synthetic images as well as actual mammograms by Masek [8]. Masek [8] has combined them into a single method designated as *LLBP approach* that has been shown to give best result when used to segment components of a mammogram. The LLBP approach is the basic idea behind the algorithm proposed below. Formally stated, the main principle in LLBP approach is that, given an image with histogram P , one strives to find a thresholded histogram Q such that distance (cross-entropy) between P and Q is minimized.

2.2 The Complete Segmentation Algorithm

The image shown in Fig. 1(a) will be used as an example of an input image, I_{input} , to explain the different steps of this algorithm. Fig. 1(b) shows the histogram h_{breast} of this image. Black background can and should be exempted from any calculations since this part will not include any tissue depicted in the image, therefore the histogram is calculated on the breast tissue part, I_{breast} , of the image only. This algorithm is divided into two phases. First, an optimal thresholded image is acquired and then the image is post processed (including several image processing operations) to compensate for incorrect segmentations.

Fig. 1. a) Image mdb032 from MIAS database. b) Image histogram of the pixels belonging to breast tissue of mammogram.

Fig. 2. a) Segmentation produced by thresholding in stage 1. b) Segmentation produced by thresholding in stage 2. c) The combined segmentation from stage 1 and stage 2, obtained as the union of the segmentations in a) and b). d) Segmentation from c) marked in the input image from Fig. 1(a).

Phase I. Let $CEM(f(\cdot), t_1, t_2)$ denote a function that performs cross-entropy minimizations on the image as will be explained below. The minimization is performed over a range of intensity values, in which each intensity value is chosen as a threshold and cross-entropy is calculated between the original and the thresholded image until the minimum cross-entropy is reached. Here, $f(\cdot)$ is the image to be thresholded and the parameters t_1 och t_2 define the respective lower and upper bound of this range, as explained later.

Through series of tests it has been observed that the choice of an appropriate range of intensity values, especially the lower bound, for the minimization process is important to produce an optimal result. Experimentation shows that peak A corresponds to the pixels belonging to the glandular tissue (Fig. 1(b)). Empirical tests show that a threshold value near the left end of peak A as an initial value often gives the desired optimal threshold. A reliable and automated method to calculate such an initial value has been devised here that uses the cumulative histogram to detect the ascents and descents of the major peaks in an image histogram. The bends in the cumulative histogram curve correspond to ascends and descends of peak A in the image histogram. To find these bends a line segment, joining the end points of the cumulative histogram curve, is used to calculate the distance between the points on the curve and the line segment.

The point farthest away from the line has shown to be a good initial value for the cross-entropy minimization [3].

Depending on the size and structure of glandular tissue disc, the form of peak A can vary. Kurtosis is a measure useful for describing the histograms relative flatness. Calculations show that if the tip of this peak is flatter in form compared to the tip of a normal distribution i.e. if $kurtosis(h_{breast}) < a$, then the above mentioned method gives a good initial value. However, for images that give $kurtosis(h_{breast}) \geq a$, such an initial value causes the minimization to converge too fast towards a local minimum giving a too low threshold value. In these cases, the intensity value with highest count in the image histogram has proven to be a better initial value. This step of using the maximum intensity count as the initial value instead of using the cumulative histogram is called *the max step*.

With a suitable initial value, cross-entropy minimization is performed to receive a suitable threshold used to create the thresholded image. This is referred to as *stage 1 thresholding* and is illustrated in Fig. 2(a). Let T_{s1} be the threshold value from stage 1 and I_{s1} denote the set of image pixels from I_{breast} that have intensity higher than T_{s1} . Then I_{s1} denotes the glandular tissue disc according to stage 1 thresholding.

In many cases, stage 1 leaves out parts of glandular tissue disc due to too high threshold value. Therefore thresholding through the function *CEM* is repeated in stage 2. The input image is denoted $I_{input_{s2}}$ to make it clear that it is the input image to stage 2. $I_{input_{s2}}$ consists of all the pixels that belong to the breast tissue except those pixels that were segmented in stage 1 thresholding. It has been observed that the set of pixels of $I_{input_{s2}}$ does not require the *max step*, since the minimization does not converge to local minima as in stage 1. Therefore the initial value is chosen using the cumulative histogram method only. The resulting thresholded image from this stage will be referred to as $I_{output_{s2}}$, see Fig. 2(b). The segmentations from stage 1 and stage 2 are combined to create a single segmented image, denoted $I_{output_{s1 \vee s2}}$, as shown in Fig. 2(c).

Phase II. During the training of the proposed algorithm it was observed that in some cases the resulting extraction of the glandular tissue disc did not agree with the markings of the glandular tissue disc done by five experts. The reason was that in these cases regions that have a high probability of not belonging to the actual glandular tissue disc was not properly removed by the steps in Phase I. Therefore, in this phase, the final segmentation from phase I is post-processed in order to remove these regions. These regions are a result of image noise and/or simply the fact that threshold based segmentation always has its limitations. To remove unwanted regions and improve the segmentation, a sequence of image processing operations has been devised, that can be divided into several procedures.

First we want to remove unwanted regions that might belong to fat or muscle tissue. The area of all connected components (clusters of white pixels) in the $I_{output_{s1 \vee s2}}$ is calculate, i.e the number of pixels for each cluster. A limit is estimated such as that A_{limit} is the fraction between the largest area and

Fig. 3. a) Binary image BW_{large} . b) Division of BW_{large} into four quadrants with the centroid of the largest region as point of the origin. c) The final binary segmentation. d) The final segmentation c) marked in the original input image (Fig. 1(a)).

a constant k , $A_{limit} = \frac{A_{largest}}{k}$. Based on these calculations a new binary image, $BW_{largeinitial}$, is created that only contains regions from $I_{output_{s1 \vee s2}}$ that have area greater than or equal to A_{limit} . However, for cases where a elongation representing the muscle is connected to a large cluster in $BW_{largeinitial}$, further post-processing is necessary. This region is detected by analyzing relatively abrupt changes in width of the regions. For this purpose, $BW_{largeinitial}$ is sliced horizontally into l rectangular small images of equal heights. In each sliced image, width of the middle row is calculated to represent the width of the whole slice. If the difference in width between the set of two consecutive slices is greater than a certain limit, m then the change in width is too sharp and thus revealing the presence of a elongated upper region. When this limit is reached the segment is cut off to exclude the unwanted elongation. However, it is important that the limit of differences, m , is calculated such that it has the same relativity to the widest part of the segmentation. To ensure this, the segmented region is resized so that the widest part has width of n pixels. Ratio between height and width of the segmented region is preserved in all cases. The final segmentation of this post-processing procedure, thus containing only the regions with large pixel count, is denoted BW_{large} (Fig. 3(a)).

A shortcoming of removing connected elongation is that too many small regions might be excluded resulting in underdetermined glandular tissue disc. It is the anatomical fact that glandular tissue gets more concentrated and compact as one moves closer to the nipple from inside of the breast. That is, the probability of a small region belonging to glandular tissue disc increases. In addition, all regions that are located in the same quadrant as the nipple are assumed to belong to the glandular tissue. Therefore, BW_{large} is divided into 4 quadrants like a Cartesian coordinate system such that the centroid of the largest region acts as the point of the origin; see Fig. 3(b). Since the position of the nipple is already known, its quadrant is determined. Let BW_{extra} be, $BW_{extra} = I_{output_{s1 \vee s2}} \setminus BW_{large}$, in other words, all the smaller regions that were excluded in the post-processing procedure resulting in the binary image BW_{large} . Centroid for all these regions is calculated and a new binary image is created that contains small regions whose centroid is located in the same quadrant as the nipple. This new binary image is denoted BW_{mam} .

The small regions BW_{small} , is defined as $BW_{small} = BW_{extra} \setminus BW_{mam}$, and is processed separately. For each quadrant a rectangle and a circular sector is calculated. Let x_{max} and y_{max} be respective maximum horizontal and vertical distances achieved by segmented pixels in BW_{large} , then x_{max} and y_{max} are used as width and height of the rectangle. The maximum radial distance from the origin, achieved by segmented pixels in BW_{large} is used as a radius to calculate the sector. All of the small regions, whose centroid lies within the intersection of the rectangle and the sector, are accepted as part of the glandular tissue disc. This procedure is repeated for all the quadrants except the quadrant of the nipple.

Finally, the union of all the binary images, obtained in previous post-processing procedures, is returned as the final segmentation. Morphological closing is done to smooth the boundary of this segmentation. As a last step, any holes present in this segmentation are also filled to insure a closed boundary of the glandular disc. The resultant binary- and intensity images respectively are shown in Fig. 3(c) and Fig. 3(d).

3 Results

3.1 Empirical Evaluation of the Different Parameters

The results from the training data show that the desired result of the cross-entropy minimization procedure in stage 1, explained in Phase I Sect. 2.2, is reached if performed on pixels within intensity range $[max(I_{min}, t_1), min(t_2, I_{max})]$. Here, I_{min} and I_{max} are the minimum and maximum intensity values in the input image respectively and $t_1 = 20$ and $t_2 = 200$. For stage 2, where the cross-entropy minimization procedure is iterated, the parameters t_1 and t_2 , should be 20 and $max(I_{input_{s2}})$ respectively.

Furthermore, for the post-processing procedures, in Phase II Sect. 2.2, empirical tests on the 40 training images show that an optimal post-processing performance is reached if the parameters have certain defined values. For the kurtosis analysis $a = 3$ is shown to provide successful results for estimating the degree of flatness of the image histogram. The optimal constant, k for calculating the area limit, is found to be $k = 5$. Here, the area limit was useful for determining which clusters were large enough to be assumed to represent the depicted glandular tissue disc in the original image.

We saw that for some cases where an elongation representing the muscle is connected to a large cluster it is desirable to remove this part. Therefore the current binary image at that stage was sliced horizontally into a number of rectangular small images of equal heights. The number of rectangular small images needed for this post-processing procedure is $l = 10$. This elongation was automatically detected by a difference analysis of the width between the set of two consecutive rectangular small images and the limit for abrupt changes was empirical found to be $m = 140$. It is also necessary for this part that $n = 500$.

Table 1. Result from an evaluation method proposed by Olsén and Georgsson, [10]. Rank 1 corresponds to the best agreement and rank 6 to the least agreement with the rest of the markings in the ensemble ($A_1 - A_6$). The level of agreement is calculated on a leave one out basis.

	Rank sum Expert	
1	359	A_4
2	365	A_2
3	434	A_5
4	436	A_3
5	480	A_6
6	488	A_1

3.2 The Performance of the Algorithm

A common method of measuring the performance, of an automated segmentation algorithm, is to compare the results with a ground-truth. It is impossible to establish true segmentation of the glandular tissue disc. Hence, for segmentation evaluation, we need to choose a surrogate of true segmentation. In the field of medical image analysis, a ground-truth is often obtained from markings manually created by human-experts. Segmentation evaluation based on ground truth obtained by a group of experts has been studied for a long time. However, the most appropriate way to compare computer generated segmentation to segmentations created by a group of experts is so far unclear [14]. One of the principal reasons for absence of a standard evaluation method is that the inter-expert variations among manually created segmentations are very high. This makes it difficult to create an objective ground-truth.

Fig. 4 shows an example of the inter-expert variations between the manually segmented regions of the glandular tissue disc, in the same mammogram, provided by five experts in radiology. By observing Fig. 4, it can be concluded that an automated segmentation algorithm cannot be tested against a ground-truth from just one expert, because the evaluation would be subjective. Olsén and Georgsson [10] discuss problems involved in using manually created ground-truths and propose a method striving to create an objective performance measure, based on markings provided by experts. The authors have looked at the usage of ensembles of domain experts when assessing the performance of a segmentation algorithm. Olsén and Georgsson [10] estimate a function that calculates the degree of agreement of a given segmentation with the estimated ground-truth. All the five expert markings along with the result from the proposed segmentation algorithm (referred to as the *system segmentation*) are individually ranked by this function. A rank sum is then defined which expresses the overall agreement of expert markings and the system segmentation for the whole set of test images. The rank sum calculated for the test set of 160 images is shown in Table 1, where ranks are arranged in descending order.

Fig. 4. The glandular tissue disc manually outlined in mammogram no. 138 (MIAS) by a panel of five experts ($A_1 - A_5$) in mammography. The inter-expert variations between the manually segmented regions of the same glandular tissue disc in the same mammogram is, as visible, large.

The five expert markings used to estimate the ground-truth are denoted as $A_{1...5}$ and A_6 denotes the system segmentation. This result indicates that proposed algorithm mixes in with the human experts. In other words, the performance of the proposed algorithm seem to match the human professionals in mammography.

4 Discussion

In Sect. 3, inter-expert variations were explained as one of the primary reasons to why it is difficult to make an objective assessment of segmentation algorithms, for fibroglandular tissue in mammograms. Furthermore, an example of evaluation was presented by showing results from a newly developed method [10]. Since this evaluation was based on markings from a certain group of experts, it is possible to get different ranks for a different expert panel. Therefore the results presented in Table 1 can not be considered as exact performance evaluation for the algorithm developed in this paper, but rather an estimation of how good segmentation this algorithm produces on average.

5 Conclusion

A novel approach for extracting the glandular tissue disc has been presented in this paper. This approach was based on global histogram analysis of the glandular regions in mammograms. Based on the arguments in Sect. 3, it can be concluded that this algorithm is able to calculate a good approximation of the location and delineations of the glandular tissue disc. The only truths we can use for segmentation evaluation is medical experts in the field. However, since the most appropriate way to compare computer generated segmentation to segmentations created by a group of experts is so far unclear, evaluation of segmentation algorithms for mammograms is very complex. In fact, in fields where only surrogate truths are available formation of an objective evaluation method is as important as construction of the segmentation routines that are being evaluated.

References

1. Bovis, K., Singh, S.: Classification of mammographic breast density using a combined classifier paradigm. PANN Research, Department of Computer Science, University of Exeter, Exeter, UK. In: 4th International Workshop on Digital Mammography, pp. 177–180 (2002)
2. Brink, A.D., Pendock, N.E.: Minimum Cross-Entropy Threshold Selection. *Pattern Recognition* 29(1), 179–188 (1996)
3. Chandrasekhar, R., Attikiouzel, Y.: Automatic breast border segmentation by background modeling and subtraction. In: M. J. Yaffe, editor, *Digital Mammography, Computational Imaging and Vision*, Medical Physics Publishing, Madison, Wisconsin pp. 560–565 (2000)
4. Karssemeijer, N.: Automate classification of parenchymal patterns in mammograms. *Physics in Medicine and Biology* 43, 365–378 (1998)
5. Kasparis, T., Charalampidis, D., Georgiopoulos, M., Rolland, J.: Segmentation of textured images based on fractals and image filtering. *The Journal of Pattern Recognition Society*, 34 (2001)
6. Kullback, S.: *Information Theory and Statistics*. John Wiley, New York (1959)
7. Li, C.H., Lee, C.K.: Minimum Cross-Entropy Thresholding. *Pattern Recognition* 26(4), 617–625 (1993)
8. Masek, M.: *Hierarchical Segmentation Of Mammograms Based On Pixel Intensity*. PhD thesis, Centre for Intelligent Information Processing Systems, School of Electrical, Electronic, and Computer Engineering. University of Western Australia, Crawley, WA, 6009 (February 2004)
9. Olsén, C.: *Automatic Assessment of Mammogram Adequacy*. Licentiate Thesis, Umeå University, Department of Computing Science (2005)
10. Olsén, C., Georgsson, F.: Assessing ground truth of glandular tissue. In: Astley, S.M., Brady, M., Rose, C., Zwigelaar, R. (eds.) *IWDM 2006*. LNCS, vol. 4046, pp. 10–17. Springer, Heidelberg (2006)
11. Pal, N.R.: On Minimum Cross-Entropy Thresholding. *Pattern Recognition* 29(4), 575–580 (1996)
12. Rangayyan, R.M.: *Biomedical Image Analysis*. In: *Biomedical Engineering Series*, CRC Press LLC, Boca Raton (2005)
13. Saha, P.K., Udupa, J.K., Conant, E.F., Chakraborty, D.P., Sullivan, D.: Breast tissue density quantification via digitized mammograms. *IEEE Transactions On Medical Imaging* 20(792–803), 575–580 (2001)
14. Warfield, S.K., Zou, K.H., Wells, W.M.: Simultaneous truth and performance level estimation (staple): An algorithm for the validation of image segmentation. *IEEE Transaction on Medical Imaging* 23, 903–921 (2004)
15. Xue, Y., Zhang, Y., Lin, X.: Threshold selection using cross-entropy and fuzzy divergence. *Proceedings of the SPIE* 3561, 152–162 (1998)