Breast peripheral area correction in digital mammograms

Meritxell Tortajada∗, Arnau Oliver1, Robert Martí2, Sergi Ganau2, Lidia Tortajada2, Melcior Sentís2, Jordi Freixenet1, and Reyer Zwiggelaar3

1 Dept of Computer Architecture and Technology, University of Girona, Girona, Spain.

2 UDIAT-Centre Diagnòstic, Corporació Parc Taulí, 08208 Sabadell, Spain.

3 Dept of Computer Science, Aberystwyth University, Aberystwyth SY23 3DB, UK.

* Corresponding author. M. Tortajada, Ed. P-IV, Campus Montilivi, University of Girona, 17071 Girona (Spain).

Email: txell@eia.udg.edu; Phone: +34972419812; Fax: +34972418976.
Breast peripheral area correction in digital mammograms

Abstract

Digital mammograms may present an overexposed area in the peripheral part of the breast, which is visually shown as a darker area with lower contrast. This has a direct impact on image quality and affects image visualisation and assessment. This paper presents an automatic method to enhance the overexposed peripheral breast area providing a more homogeneous and improved view of the whole mammogram. The method automatically restores the overexposed area by equalising the image using information from the intensity of non-overexposed neighbour pixels. The correction is based on a multiplicative model and on the computation of the distance map from the breast boundary. A total of 334 digital mammograms were used for evaluation. Mammograms before and after enhancement were evaluated by an expert using visual comparison. In 90.42% of the cases, the enhancement obtained improved visualisation compared to the original image in terms of contrast and detail. Moreover, results show that lesions found in the peripheral area after enhancement presented a more homogeneous intensity distribution. Hence, peripheral enhancement is shown to improve visualisation and will play a role in further development of CAD systems in mammography.

Keywords

Mammography; Image Processing; Radiographic Image Enhancement; Computer-Assisted Diagnosis.

1. Introduction

As a consequence of the current digital revolution, traditional film-based hospitals are converting to digital hospitals, where patient medical records, chart information, and test results are easily available electronically for physicians from anywhere in the hospital and beyond. As such, full-digital mammography is gaining importance compared to conventional film-screen mammography due to the possibility of separating and individually optimising digital acquisition, digital storage, and digital display [1-3].

Digital detectors offer higher quantum efficiency and higher resolution than traditional screen-film receptors [4]. These characteristics translate into both lower dose and improved image quality mammograms. Berns et al. [5] showed that digital mammography acquisition is a highly significant 35% shorter acquisition in time. Once images are acquired, the Digital Imaging and Communications in Medicine (DICOM) [6] standard handles the storage and communication protocol. This allows enabling the integration of the different imaging devices, such as scanner machines, displays, and workstations in a fully digital system, usually referred to as the Picture Archiving and Communication System (PACS). From there, the images are sent to the screening workspace, where one or more experts analyse and diagnose cases.
In contrast with film-screen imaging, in digital imaging experts view images on electronic displays (also called soft-copy displays). These systems offer new opportunities. For instance, there is experimental evidence that alternating current and prior mammograms on the same display allows better evaluation of temporal changes than conventional display of images next to each other [7]. However, a faulty, inadequately calibrated, or improperly set up display can compromise the overall quality of a diagnostic procedure [8].

In order to help radiologists during breast imaging evaluation, different image processing algorithms are being developed to improve the visualisation of digital mammograms. This may be achieved by either enhancing some image features to allow the detection of different types of lesions [9,10] or by improving the quality of the mammograms to compensate for possible acquisition limitations [11]. This paper focuses on the latter, specifically on the correction of the presence of an overexposed boundary area in the majority of mammograms, as shown in Figure 1(a). This effect is due to breast thickness variation during mammographic acquisition, and cannot be solved by modifying the typical contrast parameters that viewers provide (window width and window centre). During acquisition, the breast is compressed with a tilting compression paddle; hence breast thickness is non-uniform across the mammogram, being thinner in the periphery and thus overexposing this area. To compensate for thickness variations in the breast periphery, we propose an automatic peripheral enhancement algorithm.

The proposed enhancement method applies a multiplicative correction factor for each pixel of the overexposed area. This factor depends just on the pixel intensity and its neighbourhood with already corrected intensities. Hence, the procedure starts correcting the overexposed pixels adjacent to the non-overexposed area and iteratively corrects the other pixels. This process stops at the skin-line, which is defined as the boundary between the breast and the dark background. Notice how the use of a multiplicative factor guarantees grey-level continuity. Figure 1(b) shows an example of the application of this algorithm. Note that the correction of the overexposed area also affects the behaviour of the window width and windows centre contrast adjustments. The proposed peripheral enhancement not only represents an improvement in the appearance of the images but also for breast cancer detection [12,13].

The rest of the paper is structured as follows. Section 2 reviews the literature on peripheral breast enhancement. Section 3 describes the proposed method. Results are shown in Section 4. Section 5 shows the benefits of our approach in two different applications. The paper ends with discussion.

2. Background

Several methods have been proposed for overexposed area correction in mammography, which can be classified into non-parametric [14-16] or parametric [17-21] approaches. The former ones try to adjust the intensity of overexposed areas by means of traditional image processing techniques, like segmentation and equalisation. On the other hand, parametric approaches adjust the intensity of the images according to a given model, which may be as
specific as the type of digital detector [17] or as general as a 3D representation of the breast [18-21]. A different approach was proposed by Goodsitt et al. [22,23] who designed physical filters to adjust x-ray beam distribution in order to compensate the tissue thickness.

In this work we focus on a non-parametric approach for image correction. Non-parametric approaches allow improving the quality of a single image without needing extra information. In traditional non-parametric approaches, the correction of the overexposed area follows an additive approach. Thus, a specific factor is added to the pixel intensities appearing too high in order to obtain a more homogeneous intensity distribution throughout the breast. The different approaches may vary in the determination of the correction area and in the factor which is added. Regarding area, algorithms can be either applied to the whole area [14] or just to a region determined by a segmentation algorithm [11]. In the first case, the correction in the inner part of the breast is inappreciable, since the additive factor is assumed to be zero in that part. In the second case, authors segment the breast using intensity based features in order to locate overexposed areas, and the correction algorithm is limited to this part. Regarding the correction factor, Bick et al. [14] generated a curve based on the mean grey level intensity of all points lying at the same distance from the skin line. A second curve was generated by subtracting this intensity curve from the mean intensity value of the image. The intensity value of this second curve at each distance was then added to the corresponding pixels. Karssemeijer and te Brake [16,24] first computed a smoothed version of the mammogram. Subsequently, all pixels below a threshold were corrected by subtracting from the original intensity the smoothed intensity and adding the mean value of the inner part of the breast. A similar approach was also developed by Byng et al. [15]. The underlying assumption of these approaches is that thickness variations are smoother than density variations.

3. Peripheral area correction

Figure 2 depicts our proposal for breast peripheral enhancement. The method can be divided into two main steps: 1) determining overexposed and non-overexposed areas and 2) equalising the mean intensities of both areas. The goal is to enhance the intensities of the overexposed area to make them similar to the ones of the non-overexposed area. Before determining the overexposed area, an initial step is necessary in order to separate the breast region from other areas of the mammogram. Besides, the distance transform of the image is computed to speed-up the overall process. An optional final step is the integration of the pectoral muscle into the final mammogram. In following sections we explain in more detail all these steps.

3.1. Breast area segmentation

In addition to the actual breast, mammograms contain other regions that must be removed before applying the enhancement technique, namely the background (which may contain some labels) and the pectoral muscle. In conventional film-screen mammograms, the background is noisy and inhomogeneous and approaches have been proposed to deal with this problem [25,26]. However, in full-field digital mammograms this region is totally black and uniform, and hence it does not require any pre-processing; a clear advantage over
conventional screen-film. Therefore, the background is composed by all pixels with intensity equal to zero. Additionally, labels can be removed by keeping just the larger region of the image.

Different approaches have been proposed for detecting the pectoral muscle in MLO images [25,27,28]. In this work we implemented the proposal of Kwok et al. [27]. Roughly, initially the pectoral edge is estimated by a straight line obtained by comparing the brighter and darker regions of the mammogram’s upper corner. Subsequently, this straight line is transformed to a curve. This process consists in moving each pixel in the line through a few pixels along its perpendicular line and looking for the best cliff candidate, which is placed as new edge. When all the pixels in the line are moved to each candidate position, the new edge is smoothed by least square fitting. This process is iteratively performed until convergence. Although the algorithm was initially developed for digitised mammograms, we adapted it for digital mammograms.

Therefore, mammograms are divided into breast area, background, and pectoral muscle, and only the breast area is kept for subsequent processing. Figure 3(a) shows the result of this step, where black areas are not considered in the rest of the processing steps.

3.2. Overexposed area determination

The overexposed area is determined by computing the histogram of the breast area. Figure 3(b) shows the histogram of the breast shown in Figure 3(a). Two different areas can be seen in the histogram. The right part of the histogram corresponds to brighter intensities, the ones located in the inner region of the breast. In this region, no single intensity dominates over the rest and the wide peak reflects the inhomogeneous tissues of that region. Conversely, the left part of the histogram shows a narrow high peak, which corresponds to the outer region of the breast that shows a darker intensity, due to the overexposing effect. This is the part that needs to be corrected.

To locate the overexposed part we analysed the histogram of the mammogram, automatically locating the first local peak and subsequently looking for its adjacent local valley. The intensities which are contained between the histogram origin and this valley are the ones that will be corrected. These intensities correspond to the pixels of the overexposed area, as is shown in the binary map of Figure 3(c). This map has been refined using morphological transformations to obtain a smoother contour. In initial experiments, the overexposed area was determined by the automatic Otsu thresholding algorithm. However, in dense breasts, this algorithm tended to merge the overexposed area with fatty tissue.

3.3. Distance transform

Before the calculation of the correction factor, the Euclidean distance map of the mammogram is obtained by computing the minimum distance from each pixel to the breast skin-line. This image helps to speed up the whole process, since the use of the distance map makes possible to deal with all pixels at the same distance at the same time.
The distance transform depends on the metric used, being the Euclidean or Manhattan metrics the most common ones. The latter provides a faster solution to the problem, especially in multidimensional images, at the cost of obtaining only an approximated solution. Rosenfeld et al. [29] provided the roots for fast computation of the Euclidean distance transform, and recently Fabbri et al. [30] reviewed different approaches for computing it, distinguishing among three classes according to their own implementation. In ordered propagation algorithms [31], the algorithm starts from the seeds (0 distance) and progressively transmits the information to other pixels in order to increase the distance. In raster scanning algorithms [32], 2D masks are used to guide the processing of pixels line by line, top to bottom, then bottom to top. Finally, independent scanning schemes process each row of the image, independently of the others, and then process each column of the result (similarly to the implementation of the Fourier transform of an image by a sequence of 1D transforms in orthogonal directions). Independent scanning schemes can be divided into parabola intersection algorithms [33], mathematical morphology based algorithms [34], and algorithms based on Voronoi tessellation [35,36], the latter providing a more general scheme and allowing for parallel computing implementations.

The implementation used in this paper follows the approach of Maurer et al. [35], which allows fast computation of the Euclidean distance transform to multidimensional spaces. At each dimension level, the distance transform is computed by constructing the intersection of the Voronoi diagram (whose sites are the feature voxels) with each row of the image. This construction is performed efficiently by using the distance transform in the next lower dimension. Authors demonstrated that the algorithm has linear time complexity ($O(N)$), while the parallel version of the algorithm being used here runs in $O(N/p)$ time, being $p$ the number of processors.

3.4. Correction factor

Once the overexposed area has been determined, the goal is to correct its intensities by considering the values of the non-overexposed area of the mammogram and following an iterative process. Firstly only the pixels belonging to the inner overexposed area are corrected looking at the intensities of its neighbouring pixels which are not overexposed. Once these pixels have been corrected, the overexposed area mask is reduced and the new inner overexposed pixels are corrected. This process is applied until all pixels have been corrected.

Formally, given a pixel $x$ of breast $B$ with intensity $I(x)$, we define its neighbourhood as:

$$N(x) = \{ t \in B : d(t,S) = d(x,S) \wedge d(t,x) \leq k \}$$

where $t$ is a pixel of the breast, $S$ is the skin-line contour, $d()$ refers to the distance (notice that the first distance is point-to-line while the second one is the common Euclidean distance between two points), and $k$ represents the size of the neighbourhood (discussed in the Section 4). Notice that with this definition, we are considering as a neighbourhood $N(x)$ of pixel $x$ the closest neighbours of $x$ with same distance to the breast skin-line.

We also define $N_{in}(x)$ as the neighbourhood of $x$ located one pixel further inside the breast:
\[ N_{th}(x) = \{ t \in B : d(t,S) = d(x,S) + 1 \land d(t,x) \leq k \} \]

With these definitions and from the furthest overexposed pixel from the skin-line boundary to the closest, the intensity of each pixel is iteratively corrected as:

\[ I'(x) = I(x) \frac{\overline{I_{nm}}(x)}{\overline{I_{n}}(x)} \]

where \( \overline{I(x)} \) refers to the mean intensity of each neighbourhood. The rationale behind this equation is twofold. By dividing by the mean value of the neighbours at the current distance we are normalising the overexposed values without losing information, while when multiplying by the mean of previous values we are giving more weight to the intensity continuity constraints. Notice that the use of the distance map allows us to speed up the process by dealing at the same time with all the pixels which are located at the same distance of the skin-line.

### 3.5. Pectoral muscle integration

The final step of the algorithm is the integration of the pectoral muscle into processed images. This step may be unnecessary for posterior computer-based analysis such as the detection of abnormalities [37,38] or the analysis of the breast density [39,40]. However, we have experimentally observed that radiologists feel more comfortable when the pectoral muscle appears in the mammogram. This is done just adding the result of the pectoral muscle segmentation and the enhanced mammogram, since they are disjoint regions.

### 4. Results

A total of 334 full-field digital mammograms (FFDM) acquired using a Hologic Selenia mammographic system, with resolution equal to 70 micron-pixel, size 4096 x 3328 or 3328 x 2560, and 12-bit depth were used in this work. These images included CC and MLO views of 117 women, 50 full exams (CC and MLO of left and right breasts) and the rest were CC and MLO of either right or left breasts. After double-reading by the experts, 98 images contained masses (10 of them were located in the periphery region).

The only parameter in our algorithm that needs to be tuned is the size of the neighbourhood used for correcting the intensities of the overexposed pixels (parameter \( k \)). We fix this size to be 100 pixels (i.e. 7 mm). Figure 4 shows the resulting image when using smaller, similar, and larger values of \( k \). If \( k \) is too small the neighbourhood is also small, and the new value assigned depends on just few pixels. Hence, the propagated value is always close to the reference pixels and corrected pixels are assigned with similar intensity. On the other hand, when \( k \) is large, the appearance of the mammograms is better, although a dark ribbon appears due to the fact that the new value depends from pixels far from it. The best results were obtained with \( k \) close to 100 pixels. Actually, the tuning of this parameter allows for a certain leeway, since we obtained similar results using values of \( k \) between 70 and 120 pixels (4.9 and 8.4 mm).
Figure 5 shows the result of the enhancement on five different mammograms, three CC and two MLO views (displayed without adjusting the contrast of the images). The first column corresponds to the original mammograms while the second shows corrected images. In MLO views the boundary found automatically between the pectoral muscle and the breast is depicted in black. Comparing the mammograms, the dark overexposed areas have been corrected, obtaining a more homogeneous intensity distribution along the breast area, and obtaining a mammogram with much more visual detail in the peripheral area. Additionally, notice that the structures lying between the overexposed and non-overexposed areas are continuous and have not been altered or disrupted in any place. This means that the tissue shown in the improved area corresponds to the real tissue of that region, and that our approach is not introducing artefacts that could affect the subsequent analysis of the image.

To quantitatively analyse the improvement in terms of image quality, Figure 5(c) and (d) shows the histograms of the breast area (without including the pectoral muscle in MLO images) before and after the enhancement, respectively. As expected, the big peak in the overexposed area is not present in the corrected histogram, and in contrast, the smooth wide peak reflects that the intensities are distributed among all the breast area.

To determine the quality of the processed images, qualitative and quantitative analyses were performed. The qualitative analysis was performed by means of visual assessment by one radiologist with more than 10 years of experience in mammographic images. All original and corrected images were displayed side by side and the observer just labelled the images as correctly processed or not, where correctly processed means that the quality of the images substantially improved after the enhancement. A total of 90.42% of the images were evaluated as correctly processed. In the rest of the cases, the overexposed area was not as dark as in the rest of the images, due to either less radiation or breast physiology, and the correction was actually not necessary. When applying our method to these images, the area selected for correcting was the full mammogram, and the final output was the original image but globally enhanced, i.e. without obtaining any local effect. In any case, the quality of the images decreased after the processing.

A quantitative analysis was also performed by comparing the mean intensity and the histograms of the original and the enhanced mammograms. Regarding mean intensity, since overexposed areas have darker intensities, a “visually better” mammogram should obtain a larger mean intensity value. On the other hand, we also have to define what a “better” histogram is. Histograms from overexposed mammograms present a high peak in dark intensities. This peak is the one that should be reduced when enhancing the mammograms, obtaining a histogram with a well-defined central peak after enhancement. To measure this, we compute the skewness and the kurtosis of the histogram. The skewness measures how asymmetric the histogram is, being 0 a totally symmetric distribution. The kurtosis measures how wide the peak is, being 3 the kurtosis defined by a normal distribution. Figure 6 compares the mean intensity and both the skewness and kurtosis measures for the 334 analysed mammograms summarised in terms of boxplots. As expected, the mean intensity value has increased, almost doubling its value. On the other hand, skewness has been reduced, hence obtaining a more symmetric distribution, while kurtosis is also reduced, obtaining a wider peak
of the breast area. Analysing the results using a paired t-Student test, the improvements in each measure were significant with p-value < 0.0001.

5. Applications of the breast peripheral enhancement

In this section we show two important applications where the use of the peripheral enhancement allows improving the procedures: first, in expert detection of mass routinely performed in clinical practice, and secondly, in a computer aided application based on computing the breast density of the mammogram.

5.1. Mass detection after enhancement

To perform an estimation of the benefits of the approach for manual analysis of the breast we asked to the same experts that diagnosed the mammograms to analyse all the cases again but after the periphery enhancement (the radiologists were different than the one that performed the qualitative evaluation). Notice that there was more than one year of difference between the first and the second analysis, and we can consider that radiologists did not remember the cases.

As expected, radiologists were able to diagnose again all the cases containing masses. After asking their opinion about the enhancement, they reported two main favourably reasons to the enhancement. First of all, masses were easier to detect, allowing a faster analysis of the images. For instance, Figure 7 shows three examples of mammograms with masses located at the overexposed area. Notice that the masses are highlighted after peripheral correction. And secondly, the border of the masses was also clearer compared to without the enhancement. This is an important result, since the border and the shape of the masses are two important features when diagnosing the mass as cancer or benign.

5.2. Automatic breast density estimation after enhancement

Breast density is an important risk factor for breast cancer. It is well known that dense breasts are more likely to develop breast cancer than fatty breast, and besides, the analysis of dense breasts is more difficult, since the own tissue may mask the abnormalities. Breast density estimation allows for the classification of mammograms according to their internal tissue. Consequently, it is a first step to a more personalised analysis.

Oliver et al. [39] developed an automatic method for breast density analysis based on first segmenting the mammograms into two classes: fatty and dense. Subsequently, features from both sets were extracted and used to classify the mammograms. Authors obtained good results using two different digitised databases (kappa = 0.81 and kappa = 0.67, respectively). However, when using this approach in digital mammograms, the method fails to obtain good results due to overexposed area (kappa less than 0.4). As shown in Tortajada et al. [43], in order to overcome this issue and obtain better results, the peripheral enhancement was applied. The results were improved to kappa = 0.88.
6. Discussions and conclusion

Overexposed areas in mammograms are observed as darker areas, and the structures lying in these regions are hardly visible. In this work we presented a novel algorithm that automatically enhances the intensity of overexposed areas to obtain similar intensities to the rest of the mammogram. Our approach was inspired by the additive models that are being used to enhance mammograms [14-16]. However, we claim that our multiplicative model not only provides intensity continuity on the overexposed area (the same as the additive model) but also preserves the intensity relationship between the pixels of the overexposed area.

The presence of the overexposed area cannot be solved by common visualisation contrast tools, due to the fact that most of these tools are based on histogram information. For instance, manually applying one of the most common tools such as the adjustment of window width and centre [41,42] the whole mammogram cannot be properly visualised. Actually, experts should tune the parameters for both overexposed and non-overexposed areas independently, which is not possible. When applying our enhancement approach, we are locally correcting the overexposed area without considering histogram information. Instead, each pixel is corrected using just neighbourhood information. After applying the peripheral enhancement, and using the window width and centre adjustment tool, the full mammogram can be properly visualised, as shown in Figure 1.

The proposed approach presents many benefits, either for manual or automatic analysis of the mammogram. In manual analysis, the proposed enhancement helps uncover different kind of abnormalities in the peripheral zone. Although expert radiologist may actually find them after thorough analysis of the mammogram, abnormalities are easier to detect after the enhancement. Moreover, the physical limits of the mass, which were not visible on the original mammograms, could properly be assessed after enhancement. On the other hand, automatic analysis, such as computer aided detection and breast density measurement systems can also profit from this enhancement [43,44].

We have provided a peripheral enhancement algorithm for digital mammograms. This step is necessary to balance the overexposure of the breast periphery. The algorithm has been tested on a large database of digital mammograms, obtaining visually more comfortable mammogram in 90.42% of the images. In these images, the correction of the overexposed area helps improve the image quality and obtain better visualisation and assessment.

Acknowledgements

This work was partially funded by the Spanish R+D+I grant n. TIN2012-37171-C02-01. M. Tortajada holds a UdG BR-GR10 grant.

Conflict of interest

None declared.
Bibliography


List of figures

Figure 1. Example of the peripheral enhancement: (a) original and (b) enhanced images. The images are obtained after the best manual window width and window centre configuration.

Figure 2. Graphical description for the overexposed area correction.

Figure 3. Example of the thresholding process: (a) breast area segmentation, (b) histogram of the breast area clustered in 64 bins, (c) overexposed area segmentation.

Figure 4. Example of the peripheral enhancement when varying the value of the size of the neighbourhood. (a) k = 20, (b) k = 100, (c) k = 180. The best results were obtained using values of k around 100 pixels.

Figure 5. Example of the peripheral enhancement: (a) original mammogram and (b) corresponding corrected mammogram, where in MLO views the border between pectoral muscle and the breast is shown in black, (c) breast area histogram of the original mammogram, and (d) breast area histogram of the corrected mammogram. Notice that the narrow peak of darker intensities has been removed and the histogram is now more homogeneous.

Figure 6. Quantitative analysis of the original and enhanced histograms in terms of boxplots. (a) mean intensity of the breast area, (b) skewness, and (c) kurtosis.

Figure 7. Example of the peripheral enhancement in mammograms containing masses: (a) original mammogram and (b) corresponding enhanced mammogram. The white arrow indicates the location of the mass in each image, which is zoomed in the small box for a better visualisation. Notice that using the proposed enhancement the contrast of the lesion is now clearer.