Computers in Biology and Medicine 67 (2015) 61–73

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/cbm

Breast image pre-processing for mammographic tissue segmentation

Wenda He [a,*], Peter Hogg [b], Arne Juette [c], Erika R.E. Denton [c], Reyer Zwiggelaar [a,**]

[a] Department of Computer Science, Aberystwyth University, Aberystwyth SY23 3DB, UK
[b] School of Health Sciences, University of Salford, Salford M6 6PU, UK
[c] Department of Radiology, Norfolk & Norwich University Hospital, Norwich NR4 7UY, UK

ARTICLE INFO

Article history:
Received 8 May 2015
Accepted 2 October 2015

Keywords:
Mammographic segmentation
Risk assessment
Density classification
Peripheral enhancement
BI-RADS
Tabár

ABSTRACT

During mammographic image acquisition, a compression paddle is used to even the breast thickness in order to obtain optimal image quality. Clinical observation has indicated that some mammograms may exhibit abrupt intensity change and low visibility of tissue structures in the breast peripheral areas. Such appearance discrepancies can affect image interpretation and may not be desirable for computer aided mammography, leading to incorrect diagnosis and/or detection which can have a negative impact on sensitivity and specificity of screening mammography. This paper describes a novel mammographic image pre-processing method to improve image quality for analysis. An image selection process is incorporated to better target problematic images. The processed images show improved mammographic appearances not only in the breast periphery but also across the mammograms. Mammographic segmentation and risk/density classification were performed to facilitate a quantitative and qualitative evaluation. When using the processed images, the results indicated more anatomically correct segmentation in tissue specific areas, and subsequently better classification accuracies were achieved. Visual assessments were conducted in a clinical environment to determine the quality of the processed images and the resultant segmentation. The developed method has shown promising results. It is expected to be useful in early breast cancer detection, risk-stratified screening, and aiding radiologists in the process of decision making prior to surgery and/or treatment.

1. Introduction

Breast cancer is the most frequently diagnosed cancer in women [1]. To date, the most effective way to overcome the disease is through early detection, precise identification of women at risk, and application of appropriate disease prevention measures [2]. Mammography is the gold standard method in detection of early stage breast cancer before abnormalities become clinically palpable. Within screening mammography, full field digital mammography (FFDM) has become more popular and is gradually replacing screen film mammography (SFM). Many digital mammography units produce images in two forms: 'raw' and 'processed' images. Raw data is often not archived in clinical practice, whilst the appearances of processed (for presentation) images may vary due to different post-processing algorithms applied by mammography manufacturers. A significant amount of dynamic range provided by FFDM systems is redundant after these logarithmic based post-processing. This may result in lower visibility of breast parenchyma in peripheral areas, and large intensity discrepancy between thicker tissue near the chest wall and peripheral areas; see examples in Fig. 1. With such processed images, abnormalities near peripheral areas with less visible structures may be missed during a mammogram examination. When used in computer aided mammography, processed images can lead to less satisfactory breast tissue segmentation, due to inter-fatty/dense tissue intensity variation across the mammograms which jeopardises subsequent analysis in the workflow.

A mammographic pre-processing technique can be developed to enhance the visibility of peripheral areas and improve intensity distribution, in order to facilitate interpretation and benefit follow up analysis. Existing methods in the literature can be categorised into two groups: non-parametric (*e.g.* [3–6]) and parametric (*e.g.* [7,8]) approaches. Most existing methods are intended to be used on 2D mammographic projections. As technology advances, more breast thickness equalisation/correction methods have emerged for 3D volumetric breast density analysis (*e.g.* [9–11]). The proposed approach is in the application domain of 2D mammograms.

An early non-parametric method [3] focused on balancing the mammographic intensity between the breast centre and its peripheral areas, so that the two areas have 'matching' average

* Corresponding author.
** Principal corresponding author.
E-mail addresses: weh@aber.ac.uk (W. He), p.hogg@salford.ac.uk (P. Hogg), arne.juette@nnuh.nhs.uk (A. Juette), erika.denton@nnuh.nhs.uk (E.R.E. Denton), rrz@aber.ac.uk (R. Zwiggelaar).

http://dx.doi.org/10.1016/j.compbiomed.2015.10.002
0010-4825/

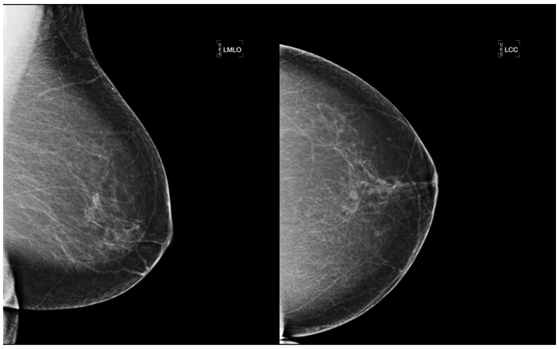

Fig. 1. A compression paddle is used to even out the breast tissue during imaging, however, the peripheral areas may not be fully compressed due to a reduction of breast thickness. This results in air gaps above and beneath the uncompressed areas, leading to a non-uniform exposure and degradation in contrast in these areas.

greylevel values. A log-like-intensity characteristic curve is created based on the average greylevel values that are within the same distances to the skinline, from which a reversal fitted enhancement curve is obtained using a polynomial fit. This fitted enhancement curve defines the necessary correction value for each pixel, which is added to the original pixel value to create the intensity balanced ('equalised') image. Such an approach applies the thickness correction to the entire breast. It works well with a homogeneous fatty/dense breast but localised artefacts can be seen when a breast exhibits large density variation across the mammogram. To better identify breast peripheral areas requiring correction, a large Gaussian filter can be used to blur a mammographic image isotropically first [4], before obtaining a representation of tissue thickness differences with smoother variations, assuming that the breast thickness variations are smoother than tissue density variations. The thickness correction is only applied in the breast periphery determined by a local threshold at the boundary of the compressed and uncompressed part of the breast. To ensure intensity continuity, a locally determined correction factor is used to multiply with the original pixel values, to derive the corrected pixel values in the breast periphery. It is an effective method to correct pixels in the peripheral areas using neighbouring pixels, however, the corrections can be over emphasised with breasts, which have intricate parencymal structures in the periphery areas, and less desirable corrections can be associated with ringing artefacts. To better reflect breast thickness differences, a mammogram is iteratively segmented into fatty and dense areas prior to the correction, and followed by a linear interpolation to replace all the dense tissue with nearby fatty tissue [5]. An alternative anisotropic diffusion filter based approach (direction parallel to the skin edge) was investigated to facilitate the breast thickness estimation. The method processes the entire breast but only adds correction terms to the pixel values in the peripheral areas. Results showed improved peripheral texture appearances for those structural texture with orientations (*e.g.* blood vessels), however, the interior part of the breast may display higher contrast after correction. The method critically depends on accurate iterative segmentation of the dense breast tissue, which can be problematic when the breast exhibits heterogeneous dense tissue. The aforementioned studies used pixel intensity values as correlation factors to estimate the breast thickness, which may not be an accurate estimation/close approximation. Note that true breast thickness may not be attainable retrospectively.

A parametric method as proposed in [7] used a geometric model of the three-dimensional shape of the breast. The breast interior is modelled by two non-parametric planes which requires three degrees of freedom: one for the onset and two for the slopes. The peripheral area is modelled by bands of semi-circles, determined by the breast outline and interior model. Once the parameters of the geometric model are obtained, dense tissue is separated and interpolated with fatty tissue, similar to [5]. Therefore, the breast can be modelled with the original and interpolated fatty tissue. The subsequent correction process is performed by adding a fatty tissue component in the periphery which fills in the air gap between the fitted planes and breast. As in [5], the approach is critically dependent on the accuracy of iterative dense breast tissue segmentation and fatty tissue interpolation. Note that the approach is designed for unprocessed digital mammograms with a linear relationship between exposure and greylevel value, therefore, it cannot be applied to processed FFDM nor SFM with unknown calibration data, which has a non-linear relationship between exposure and greylevel value.

We propose a mammographic pre-processing technique which has the following key novelty aspects: (1) modelling a breast thickness based on its shape outline derived from mediolateral oblique (MLO) and cranio-caudal (CC) views, instead of using an assumed correlation between smoothed pixels and breast thickness; (2) using a selective approach to target specific mammograms more accurately; and (3) both breast interior and exterior are enhanced simultaneously, in order to achieve intensity balancing across the mammogram and increasing breast tissue visibility in the peripheral areas. Mammographic segmentation and risk/density classification were conducted to determine the usefulness of the developed approach and the results were evaluated in a clinical environment.

With respect to mammoraphic risk assessment, Tabár et al. [12] proposed a model based on a mixture of four mammographic building blocks representing the normal breast anatomy, five mammographic risk categorises were identified based on these building blocks (i.e. [nodular%, linear%, homogeneous%, radiolucent%]): T_I [25%, 15%, 35%, 25%], $T_{II/III}$ [2%, 14%, 2%, 82%], T_{IV} [49%, 19%, 15%, 17%], and T_V [2%, 2%, 89%, 7%] [12]. Alternatively, the American College of Radiology's Breast Imaging Reporting and Data System (BI-RADS) [13] was developed, with four breast dense tissue compositions categorised as; B_1 the breast is almost entirely fat ($<25\%$ glandular), B_2 the breast has scattered fibroglandular densities (25–50%), B_3 the breast consists of heterogeneously dense breast tissue (51–75%), and B_4 the breast is extremely dense ($>75\%$ glandular).

2. Data and method

A Hologic Selenia Dimensions 2D FFDM system was used to obtain a total of 360 digital mammograms (i.e. 180 CC and 180 MLO views), processed for optimal visual appearance for radiologists. Two consultant radiologists[1] provided Tabár risk classifications and BI-RADS density ratings for all the mammograms as 'ground truth'. To model mammographic building blocks (breast tissue), a collection of patches were cropped from randomly selected images from the dataset, consisting of examples of (139) nodular, (224) linear structure, (87) homogeneous, and (89) radiolucent tissue; see Fig. 2 for examples.

An overall workflow of the developed approach can be found in Fig. 3.

[1] One radiologist has over 5 years mammographic reading experience, the other has over 10 years mammographic reading experience.