A REGION-BASED ACTIVE CONTOUR METHOD FOR EXTRACTION OF BREAST SKIN-LINE IN MAMMOGRAMS

S. Thiruvenkadam, M. Acharyya, Neeba NV, P. Jhunjhunwala and S. Ranjan

Imaging Technology, GE Global Research
Bangalore, India

ABSTRACT

In this work, we present a novel region-based active contour technique to extract the breast boundary in mammograms. Skin-line extraction in mammograms is non-trivial due to the presence of noise and scanning artifacts. Also, weak contrast near the skin-line boundary, especially in the case of low-density breasts, poses a lot of difficulty in its extraction. Here, we represent the breast boundary by a smooth, parametric curve. The region based data term of the active contour energy is based on the assumption that in a small neighborhood around the skin boundary, the intensity is piece-wise constant. Further, to achieve a good initial guess, we make use of the assumption that at a large scale, the breast image is binary with a fuzzy boundary. A PDE based energy is employed to get the fuzzy membership function. The smoothness term of the above energy allows us to handle noise/scanning artifacts in mammograms, and allows easy extraction of a single contour, which is then input to the active contour energy. Our experiments on the MIAS database gave promising results as validated by an experienced radiologist.

Index Terms— Breast Skin-line, region-based active contour, PDE based fuzzy segmentation

1. INTRODUCTION

Delineation of breast skin-line is useful in many ways in the analysis of mammograms. Information of the skin-line minimizes background noise and artifacts in a computer-aided diagnosis (CAD) system of breast cancer. Moreover, the skin-line boundary profile is also vital information for density correction of peripheral breast tissue [1, 2], and in locating the nipple (frequently used as a reference point for registering mammograms at multiple time points) [3]. Breast contour has also been used for registration between left and right mammograms for detecting asymmetry [4, 5].

Skin-line extraction is non-trivial due to the presence of noise and scanning artifacts. Also, missing/weak contrast near the skin-line boundary, especially in the case of low-density breasts, poses additional difficulty in its extraction. Further, intensity inhomogeneity in mammograms (due to varied tissue densities) makes it hard for simple thresholding schemes to accurately delineate the boundary.

A considerable amount of work has been reported on skin-line extraction algorithms over the past decade. We briefly mention a few works here. Wavelet-based thresholding [6] was adapted to identify the boundary of tissue and non-tissue region (skin-line). In [7], low-level preprocessing steps like histogram equalization, convolution with a low-pass mask, thresholding and connected component analysis have been used. Morphological pre-processing to suppress artifacts and to emphasize the breast region, followed by a fuzzy rule-based algorithm was used in [8]. An algorithm based histogram thresholding, morphological filtering and contour modeling have been illustrated in [9]. In [10], a rough binary mask thresholded using Poisson approximation is refined using the greedy snake algorithm.

Active contour based algorithms have also been used in Ferrari et. al. [11] and Wirth et. al. [12]. In [11], a contrast enhanced algorithm followed by a threshold based binarization and a chain-code was used to find an approximate breast contour; finally, the true breast boundary was detected by using an edge based active contour algorithm. In [12], the initial skin-line boundary was extracted using adaptive thresholding then using uniform Euclidean distances as a constraint was propagated to the upper and lower breast regions to obtain the upper and lower skin-line estimates, using the so-called curve extrapolation method [13].

In this work, we propose a local region-based active contour technique for automatic skin-line segmentation. Region Based active contours as against edge based methods offer robustness to initial curve placement and noise. However, methods using global statistics [14] are usually not ideal for segmenting objects with multi modal intensity as in the case of Mammo images commonly with large shading effects. Recent works [15, 16] have addressed this issue by estimating statistics in a local neighborhood around each pixel, though at a higher computational expense. In our work, we assume that in a small neighborhood around the skin boundary, the intensity is piece-wise constant, which is used in simplifying the above local region based models for computational feasibility. Secondly, the topology of breast region is used to represent the breast boundary by a smooth parametric curve, with a flexibility of imposing additional constraints on curvature, shape etc. to enhance accuracy. Additionally, we assume that on a large scale, the breast image is close to a binary image with mixed intensity values around the skin boundary region. Thus, for initialization to the active contour step, we use a PDE based fuzzy segmentation technique similar to [17]. The fuzzy segmentation energy, is input scale of the object we are looking for, and its smoothness term rejects smaller scaled objects as noise. This property allows us to handle noise/scanning artifacts in mammograms, and allows easy extraction of a single contour close to the actual boundary in the presence of shading/low contrast effects.

To summarize, we use a cascade of fuzzy segmentation based initialization, followed by local region based active contour refinement for skin line segmentation. The combination of the above energies is shown to be effective in handling shading effects, and challenges of noise, artifacts, and multi modal intensity.

2. RELATED WORK

Fuzzy based segmentation works as in [8] do not have any explicit smoothness terms to distinguish between noise/object regions. They
have to resort to either pre/post processing e.g. through morphology, the result of which may not corroborate with the fuzzy labelling step, thus giving an in-accurate and unreliable skin boundary. Also for many images, the multi modal nature of the region close to the skin edge is not accounted for, which makes a subsequent local refinement step (e.g. using Active contours) necessary for such data.

The work in [11] uses gradient based information within active contours, which is known to be susceptible to high levels of noise (e.g. digitized images in Fig. 4). Secondly, initial contour placement could be an issue in cases of shading effects close to the boundary. In such cases, common initialization techniques are likely to give a contour placement quite far away from the boundary (Fig. 4), and gradient based methods using only local information are likely to get drawn to local minima; region based methods have demonstrated robustness to such issues [14].

3. METHODOLOGY
The proposed algorithm works as follows. A preprocessing step eliminates patient labels, characters and artifacts that appear in the image so as to improve the accuracy of detection of the breast boundary. This image is then utilized for skin-line detection using the fuzzy segmentation (FS) energy and the active contour (AC) technique. The overall algorithm flow is described in Fig 2 (a). In the subsequent discussion, we elaborate on each of the steps.

3.1. Label Detection and Removal
The following pre processing steps are employed for label/text extraction (Fig 1). Firstly, histogram based thresholding is adapted to remove the background, and connected components are used for identifying coherent regions. Then, connected components of a specific scale, as determined through learning, are identified as text. The learning process involves running a simple threshold on the images and manually marking the text on about 100 training images. The following pre processing steps are employed for label/text extraction. The work in [11] uses gradient based information within active contours, which is known to be susceptible to high levels of noise (e.g. digitized images in Fig. 4). Secondly, initial contour placement could be an issue in cases of shading effects close to the boundary. In such cases, common initialization techniques are likely to give a contour placement quite far away from the boundary (Fig. 4), and gradient based methods using only local information are likely to get drawn to local minima; region based methods have demonstrated robustness to such issues [14].

3.2. Region-based Active Contour Formulation
Let \( I : \Omega \rightarrow R \) be the image with labels extracted; the problem is to find the boundary that separates \( \Omega \) into the object region and background. We assume that close to the boundary, the image is piece-wise constant. That is, we seek an open curve \( C \) that separates local neighborhoods into two regions, each with Gaussian statistics. Let us decompose \( \Omega \) into disjoint rectangles \( B_i \), \( \Omega = \bigcup_{i=1}^{K} B_i \), where we expect the intensity to be bi-modal gaussian. Let \( \mu^o, \sigma^o \) and \( \mu^b, \sigma^b \) be the distribution in \( B_i \) and \( B_j \), respectively. Similarly, \( \mu^g, \sigma^g \) denotes the background region that lies within a distance \( d \) from \( C \), similarly \( \mu^b, \sigma^b \) gives the distribution in the regions \( C^o \cap B_i \), similarly \( \mu^b, \sigma^b \) gives the distribution in \( C^o \cap B_i \).

We minimize the following energy (1) over the space of open, smooth parametric curves \( C : [0, 1] \rightarrow \Omega \), distributions \( \mu^o, \sigma^o, \mu^b, \sigma^b \):

\[
E[C, (\mu^o, \sigma^o), (\mu^b, \sigma^b)] = \\
\sum_{i=1}^{K} \int_{B_i} \chi_{B_i} \left( \frac{(I - \mu^o)^2}{\sigma^o} + \ln(\sigma^o) \right) dx \\
+ \int_{C^o} \chi_{C^o} \left( \frac{(I - \mu^b)^2}{\sigma^b} + \ln(\sigma^b) \right) dx + \lambda \int ds
\]  

Given the estimates \( \mu^o, \sigma^o, \mu^b, \sigma^b \), the first and second terms of the energy drives \( C \) to partition the rectangle \( B_i \) into two regions; \( C^o \cap B_i \) and \( C^o \cap B_i \), where the distribution is close to \( (\mu^o, \sigma^o) \) and \( (\mu^b, \sigma^b) \). The last term minimizes the length of solutions \( C \), governed by parameter \( \lambda \).

Given an initial guess \( C^o \), we use the Euler Lagrange equations of (1) to iteratively solve for \( (\mu^o, \sigma^o, \mu^b, \sigma^b) \), and \( C \), using an explicit finite difference scheme. Since only rectangles \( B_i \) that intersect with \( C \) affect the update equations, we compute distributions
only at such rectangles. Further, for optimal number of rectangles used, and for better distribution estimates, rectangles are centered at $C_k^i = [x_k^i, y_k^i]$ for uniformly placed, discrete curve points $k$, and iteration $n$.

### 3.3. Fuzzy Segmentation

Although region based AC are fairly robust to initial contour placement, in our case due to the presence of a number of smooth edges of similar scale in the breast image, a reasonable initial guess $C^0$ is important for correct solution. Once the initial contour is placed close to the piecewise constant location near the skin boundary, the AC energy can be relied upon to find the final contour.

To get $C^0$, we make use of a phase field based approach similar to the energy proposed in [17] to first get a fuzzy membership function. Specifically, we use the observation that at a large scale, the breast image is binary with a fuzzy boundary. Hence we look at the following minimization problem (2) over functions $u : \Omega \rightarrow [0, 1]$, $\mu_o$ and $\mu_b$ are the class means.

$$
E_f[u, \mu_o, \mu_b] = 
\int_\Omega u^2 (I - \mu_o)^2 + (1 - u)(I - \mu_b)^2 dx
+ \beta \int_\Omega u^2 (1 - u)^2 dx + \lambda_f \int_\Omega |\nabla u|^2 dx
$$

(2)

$u$ acts as a smooth indicator function that gives pixels a degree of membership to either the object/background regions. The second term penalizes solutions $u$ that deviate from 0 or 1 in considerably sized regions as specified by $\beta$. The third term acts as a smoothness term for $u$. We look for objects with large scale by setting $\lambda_f$ to a high value. The smoothness term rejects smaller scaled objects as noise, thus handles noise/scanning artifacts in mammograms. For numerical implementation, we use steepest descent on the above energy’s Euler Lagrange equations, discretized using a semi-implicit finite difference scheme [17], given initial inputs for class means $\mu_o, \mu_b$. The solution $u$ of the above energy is thresholded to get to the AC initial guess $C^0$.

### 4. RESULTS AND DISCUSSIONS

Experimental data set consisted of 322 images from the MIAS database [18]. Parameters for the fuzzy segmentation and active contour steps were set based on experimentation on 100 images. Validation of the proposed technique has been done on the basis of visual feedback and comments on what is clinically acceptable by an experienced radiologist. The accuracy rating protocol defined by the radiologist is given in Table 1. The last column shows the breakup of our algorithm’s performance on the above dataset.

Based on the above rating protocol we have achieved 92% success within the range of clinically acceptable skin-line detection (rating [1,3]) on the 322 images. The rest are under-segmented and clinically not acceptable.

Figure 3 shows results on some images (histogram equalized for display purposes) taken from the above dataset. The images illustrate the challenges due to shading effects, and multi modal intensity/weak contrast close to the skin boundary. The dotted cyan contour shows the initial guess input from the fuzzy initialization step. The yellow contour shows the Active contour output. The fuzzy segmentation step gets the initial guess close to the boundary in the presence of shading effects/weak contrast. Here, since two-class clustering is used in energy (2), the fuzzy output contour is accurate for bi-modal data (Figure 3(a) & (b)), and away from the boundary (Figure 3(c)-(f)) for data with multi modal intensity near the boundary. Figure 3(g) is an example of accuracy rating 3, which is clinically acceptable with minor undersegmentation near the arm pit. Figure 3(h) is an example of accuracy rating 4, which is unacceptable with undersegmentation in clinically important areas, while Figure 3(i) shows a failure case where due to presence of a large vertical artifact at the image border, the initial fuzzy segmentation step fails to detect a proper object and background.

In many cases (e.g Fig 4), the initial guess is off by as much as 50-100 pixels. The active contour step, using a local region based term is effective in handling local refinement in the above scenario as demonstrated in the images. Secondly, noise and artifacts are even more pronounced for Digitized images (e.g. Fig 4). The smoothness terms in both energies (1) and (2) give robustness to higher noise levels, weak contrast, and ability to reject artifacts below specified scale (given by parameters $\lambda, \lambda_f$ and $\beta$).

### 5. CONCLUSION

Mammo skin line segmentation is challenging due to varied levels of shading effects, multi modal nature of the data, noise and artifacts. Noise and artifacts are more pronounced for Digitized images (e.g. Fig 4) compared to Digital images (e.g. MIAS database). Our approach provides a novel solution to the above issues through a cascade of: Pre-Process image for labels, PDE based fuzzy initialization, and local region based active contour. As shown in the results, fuzzy segmentation handles shading effects and gets the initial guess close to the boundary. Active contours using a local region term (to handle data multi modality), refines the initial guess to the actual boundary. Smoothness terms within both energies give robustness to noise/artifacts. It would be difficult for a standalone Fuzzy Segmentation or Active contour based method mentioned in Section 2 to handle above issues by itself. Our current work is to add shape based priors to our active contour energy to enhance robustness to boundary gaps that could occur in Digitized images.

### 6. REFERENCES


Fig. 3. Digital Data (MIAS) with labels extracted: Detected breast skin-line contours (Yellow) with computed Fuzzy Initial contours (Cyan dotted).


Fig. 4. Noisy Digitized Images: Column I: Histogram Equalized Image. Column II: Result of Fuzzy Initialization Column III: Detected breast skin-line contours (White)