2010 Digital Image Computing: Techniques and Applications

Mammographic Mass Detection with Statistical Region Merging

Mariusz Bajger*, Fei Ma†, Simon Williams* and Murk Bottema*
*School of Computer Science, Engineering and Mathematics
Flinders University, Bedford Park, SA, 5042, Australia
Email: Mariusz.Bajger@flinders.edu.au
Simon.Williams@flinders.edu.au
Murk.Bottema@flinders.edu.au
†Symbion Pharmacy
Email: Kenny.Ma@symbionpharmacy.com

***Abstract*—An automatic method for detection of mammographic masses is presented which utilizes statistical region merging for segmentation (SRM) and linear discriminant analysis (LDA) for classification. The performance of the scheme was evaluated on 36 images selected from the local database of mammograms and on 48 images taken from the Digital Database for Screening Mammography (DDSM). The Az value (area under the ROC curve) for classifying each region was 0.90 for the local dataset and 0.96 for the images from DDSM. Results indicate that SRM segmentation can form part of an robust and efficient basis for analysis of mammograms.**

***Keywords*-segmentation; mammography; mass detection; statistical region merging;**

I. Introduction

Computer-aided detection (CAD) systems are developed to help radiologists detect lesions in mammograms. Currently, such systems have detection rates comparable to radiologists (high sensitivity) but report too many false lesions (low specificity). Unfortunately mammographic image analysis is difficult because of large variability in the appearance of lesions, large variability in the appearance on normal tissue and the subtle difference between normal and diseased tissue. In addition, breast abnormalities are often superimposed on dense breast tissue in which case the detection problem is exacerbated for both human readers and CAD systems. As a result, more research is needed in the detection of all categories of masses ([1]). Recent comprehensive reviews of methods used in CAD for breast cancer are presented in [2] and [3].

As is the case in most object recognition problems, the two crucial steps in automatic detection of lesions are image segmentation and the choice of features used to distinguish cancer from normal tissue. The direction taken in this study is to base segmentation on sound statistical image analysis and then, and in acknowledging that segmentation will not be perfect, features for classifying lesions are selected so as to be robust to errors in the assignment of the lesion boundary.

In particular, this study investigates the use of statistical region merging (SRM) [4] to provide segmentation. SRM is an example of a region growing and merging process but differs from other algorithms in this class by viewing the segmentation process as an inference problem. Advantages of this method are; (1) no prior statistical models are needed for diseased and normal tissue; (2) the only parameter needed to implement the algorithm is one that controls the complexity of the segmentation; and (3) with high probability, the resulting segmentation is close to optimal with respect to overmerging error [4].

The criterion for feature selection in this study is based on the fact that boundaries of lesions are inherently difficult to assign reliably. This is true because the path of x-ray beams through a lesion becomes smaller near the edge of the lesion and so the projection image of the edge is poorly defined. In addition, poor contrast differences between lesion and background and variation in the appearance of normal tissue compounds the problem. Hence the boundary of a lesion cannot be expected to be delineated exactly and so lesion features whose values depend critically on finding the exact boundary of the lesion are not reliable. The benefit of focussing on features that are robust to details of the lesion boundary has been noted previously ([5], [6]).

II. Statistical Region Merging

The statistical region merging technique proposed in [4] considers image segmentation as an inference problem. The image itself is considered as an observed instance of some unknown perfectly segmented image. The true (or statistical) regions of the perfect image are to be reconstructed.

The method was tested in [4] on natural scene images in comparison to the Minimum Spanning Tree (MST) based segmentation proposed in [7] and was found significantly better in handling noise. The SRM technique is well backed up by probabilistic concentration theory and as a result, as we will show further, its outcome can be fairly well estimated prior to the experiment. This is a big advantage of the method since many costly experiments can be avoided.

The following introduction to SMR focuses on single color channel images since our interest is in segmenting mammograms.

978-0-7695-4271-3/10 $26.00 © 2010 IEEE
DOI 10.1109/DICTA.2010.14

Let I be an observation of a true image I^*. Suppose that each pixel in I^* is represented by a family of distributions from which the observed intensity is sampled. The optimal (statistical) regions in I^* possess the homogeneity property: *all pixels have the same expectation across the region and the expectations of adjacent regions differ.* Thus, I is obtained from I^* by sampling statistical pixels for the observed intensity. More precisely, the intensity of each pixel in I is realized as a sum of Q independent random variables, each taking values in $[0, g/Q]$, where g is the number of image intensity levels. The observation was made in [4] that the parameter Q can be seen as a measure of statistical complexity of the image I^*. Higher values of Q result in a finer segmentation.

Similarly to MST segmentation (and any other region-growing algorithms), SRM comprises of two components: a merging predicate and the order of testing the predicate for growing regions. To develop the predicate authors of [4] prove the following: *For any fixed couple (R, R') of regions of I and any fixed $0 < \delta \leq 1$, the probability is no more than δ that*

$$\left|(\bar{R} - \bar{R}') - \mathbb{E}(\bar{R} - \bar{R}')\right| \geq g\sqrt{\frac{1}{2Q}\left(\frac{1}{|R|} + \frac{1}{|R'|}\right)\ln\frac{2}{\delta}}, \tag{1}$$

where $\bar{R}$ denotes the average intensity across the region R and $\mathbb{E}(R)$ is the expectation over all corresponding statistical pixels of I^* of their sum of expectations of their Q random variables for their intensity values. $|\cdot|$ denotes cardinality. Assuming that regions R, R' should be merged if $\mathbb{E}(\bar{R} - \bar{R}') = 0$ formula (1) yields the merging predicate

$$P(R, R') = \begin{cases} true & \text{if} \quad |\bar{R} - \bar{R}'| \leq \sqrt{b^2(R) + b^2(R')} \\ false & \text{otherwise} \end{cases} \tag{2}$$

where

$$b(R) = g\sqrt{\frac{1}{2Q|R|}\ln\frac{2}{\delta}}. \tag{3}$$

Note that in the setting of 4-connectivity (which is utilized in this paper) the number of merging tests N for adjacent regions is bounded above, that is, $N < 2|I|$ for an image I. Thus the predicate will be satisfied with the high probability $p \geq 1 - N\delta$ for N merging tests assuming δ is sufficiently small (in the sequel we follow [4] and use the value $\delta = \frac{1}{6|I|^2}$).

The order of merging satisfies the invariant which implies that if two parts of the true regions are tested then all tests inside each of those regions have already being done. Let S_I be a set containing all pairs of adjacent pixels in I (based on 4-connectivity) and let $R(p)$ be the region containing pixel p. The algorithm first sorts those pairs in increasing order according to a function $f(p, p')$. Then the order is traversed one time with the merging performed for regions $R(p)$ and $R(p')$ if the predicate $P(R(p), R(p'))$ holds true. A common choice for function f is to use the pixel intensity difference

$$f(p, p') = |p_{int} - p'_{int}|. \tag{4}$$

III. Datasets

Two mammographic databases were used to test our schema: the publicly available Digital Database for Screening Mammography (DDSM) [8], and a local database of mammograms (LSM). For the project we selected 48 images from volume `cancer_09` of the DDSM collection. The only criterion for selection was that each image contained an abnormal lesion of mass type. The images in this set had been digitized with Lumisys at 50 μm per pixel spatial resolution and 12 bit depth resolution. The sizes of images differed from 2384-4216 width and 4608-5968 for the height.

All LSM images were digitized using a Vidar Diagnostic Pro Advantage digitizer at 48 μm per pixel spatial resolution and 12 bit depth resolution. The size of each image was the same 5296×3478. For the project at hand, 36 images were selected based on only one criterion that they contained an abnormal (biopsy proven) mass type lesion.

It is important to emphasize that only subsets of the local set LSM were used to train/test some of the algorithms used in the mass detection schema (details will follow in the relevant sections). None of DDSM images were used for training or testing purposes at any stage of the development.

IV. SRM on mammograms

Despite the solid mathematical base and clarity of the condition, the predicate (2) was found to result in overmerging in [4] and replaced by a more sophisticated version for natural scene images. Our preliminary experiments on a set of 12 mammograms from LSM revealed that the original version (2) worked better for mammograms. Moreover, its simplicity allows for insight into the connection between the statistical complexity of the image and the size of segmented objects.

As with other region-growing segmentation methods using parameters there is a question of choosing the best (optimal) value of the parameter. In this case, the parameter Q needs to be selected in a way that small masses, that is, those having 3-5mm diameter, can be detected. Using the value of spatial resolution of images, one can estimate that the size of those objects varies approximately between 110 - 140 pixels. This finding was assumed valid for both DDSM and LSM because the spatial resolution of images in these datasets is very similar. To estimate the right value for Q, let $T_0 = \left|\bar{R} - \bar{R}'\right|$ and assume that the merging predicate (2) is true for the regions R, R'. Solving (2) for Q yields

$$Q \leq \frac{g^2}{2T_0^2}\left(\frac{1}{|R|} + \frac{1}{|R'|}\right)\ln\frac{2}{\delta}. \tag{5}$$

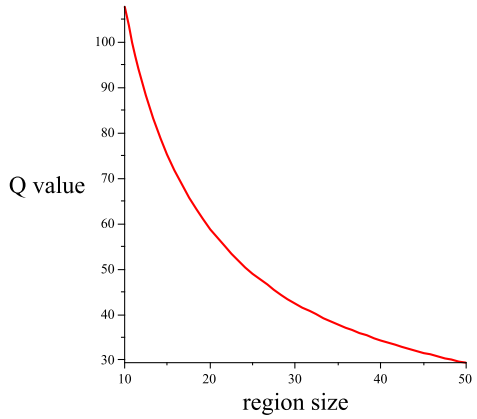

Figure 1. Estimation of the parameter Q for SRM method.

This gives us an upper bound on the value of Q such that two regions with average intensity difference T_0 (or less) will merge. Thus, to prevent merging of regions R, R' such that $|R \cup R'| \approx 140$ and $T_0 = 30$, we get from (5), $Q \approx 30$.

The graph (1) shows the behaviour of Q as a function of size of one of the regions, with the other region size fixed to 100 pixels, and with the threshold T_0 set to 30. The total size of 140 for the union of regions is obtained for $Q = 32$. Note that bigger Q values will also prevent merging but for the price of more regions needed to be processed. Thus, we are most interested in the smallest Q value allowing for mass separation and so $Q = 32$ was picked up as an optimal value for the parameter. The validity of the threshold value $T_0 = 30$, and as a consequence, validity of the Q value was initially confirmed on a training set of 12 images selected from the local dataset LSM of 36 images and it was then used for all images from both DDSM and LSM. The same set of 12 images was also used to test the choice of the function f governing the order of merging (see (4)). To reduce local noise, the single pixel intensity p_{int}, in (4), was replaced by an average over the neighbourhood of p with radius set to 1 (4-connectivity setting).

Note that the connection between T_0 and Q can be seen as a relationship between the saliency of the lesion (as a minimum average intensity difference between the lesion and the surrounding tissue) and the statistical complexity of the image.

V. Mass detection schema

A. Preprocessing

All images from DDSM and LSM were downsampled by a factor of $8 \times 8 \rightarrow 1$ and denoised with the neutrosophic filter proposed in [9]. This particular denoising method was chosen based on research reported in [5] where the filter was tested on mammograms and outperformed Gaussian and anisotropic filtering. For all images the neutrosophic filter was used with the same values of parameters: $\alpha = 0.85, w = 4, \epsilon = 0.05$. Then, the images had contrast adjusted through localized histogram equalization with block sizes 20×14.

Table I
Summary of region features.

Property	Description
Solidity	solidity
Axis ratio	ratio of minor and major axis
Area	number of pixels
RelInt	relative intensity
Radi	average radial distance
C2	the second contrast feature
C3	the third contrast feature
IntEnt	entropy of the intensity distribution
Anisotropy	anisotropy
$m_1 - m_5$	five low-order invariant moments

For each image we also created a binary contour template using the rational wavelet filtering method described in [10]. The templates were used to filter out regions outside the breast area to increase computational efficiency of the schema and to calculate some of the region properties described in the next section.

B. Feature extraction

For each of the segmented regions we constructed a feature vector comprising 14 intensity, shape and texture properties listed in Table (I):

Solidity is the area of the lesion divided by the area of the convex hull of the lesion. By *axis ratio* is meant the ratio of lengths of the major and minor axis of the ellipse having the same normalized second central moment as the region. The *RelInt* is the mean intensity of the lesion relative to the mean intensity of the entire breast region. *Radi* denotes the standard deviation of the Euclidean distance of edge points to the centroid of the region. *C2* and *C3* are two contrast measures defined in [11]. The first one is defined as the square of the difference between the mean gray value of the inside and outside region, divided by the sum of the standard deviations of both areas. The second one calculates the distance between histograms of the inside and outside region. By *anisotropy* we understand the distance between the geometric center and the center of luminosity of the region. Finally, we also used five low-order central invariant moments $m_1, \ldots, m_5$ to characterize regularity of region shapes. The moments are defined for region I as follows ([12], [13]). Let

$$m_{pq} = \sum_{i,j} i^p j^q I(i,j), \quad \text{where} \quad p, q = 0, 1, 2, \ldots. \tag{6}$$

and define the coordinates of the region centre by $\bar{x} = m_{10}/m_{00}$, $\bar{y} = m_{01}/m_{00}$. Then the five moments are

$$m_1 = u_{20} + u_{02} \tag{7}$$

$$m_2 = (u_{20} - u_{02})^2 + 4u_{11}^2, \tag{8}$$

$$m_3 = (u_{30} - 3u_{12})^2 + (3u_{21} - u_{03})^2, \tag{9}$$

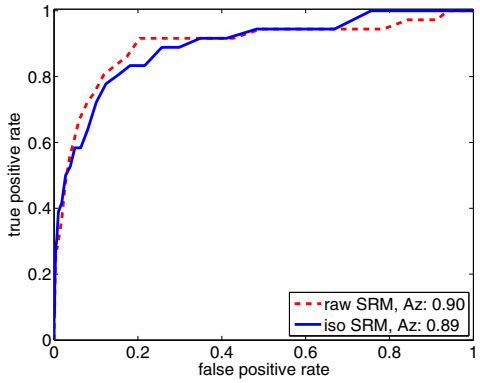

Figure 2. ROC curves for raw SRM segmentation and with isolable-contour analysis for the local LSM dataset.

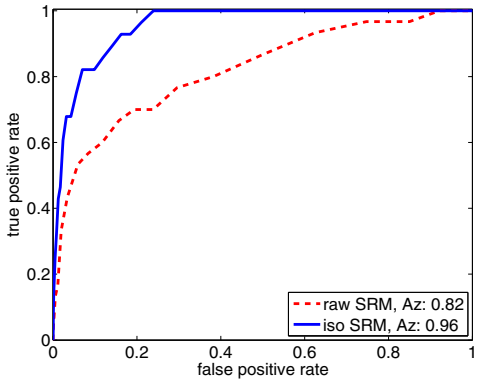

Figure 3. ROC curves for raw SRM segmentation and with isolable-contour analysis for the DDSM dataset.

$$m_4 = (u_{30} + u_{12})^2 + (u_{21} + u_{03})^2, \quad (10)$$

$$m_5 = (u_{30} - 3u_{12})(u_{30} + u_{12}) \times \left[(u_{30} + u_{12})^2 - 3(u_{21} + u_{03})\right] + (3u_{21} - u_{03})(u_{21} + u_{03}) \times \left[3(u_{30} + u_{12})^2 - (u_{21} + u_{03})\right], \quad (11)$$

where

$$u_{pq} = \sum_{i,j} (i - \bar{x})^p (j - \bar{y})^q I(i,j). \quad (12)$$

C. Detection

The accuracy of mass detection was evaluated by using ROC curve analysis. First the experiment was performed on raw regions, as obtained directly from the SRM segmentation, and again, on the same regions refined using the isolable-contour analysis method with SRM regions as seeds. The isolabel technique has been described previously ([14], [15]), and its seed-based implementation which we employed in the study, is detailed in [16]. In both cases the features were calculated for each of the segmented regions and Fisher's linear discriminant function (LDA) was used to assign a mass-like score to each region. The overall performance of the scheme on the local dataset LSM of 36 images and on 48 images selected from DDSM is shown in Figures 2 and 3 and is measured by the area under the ROC curve (the A_Z score).

The high A_Z score of 90 for LSM can be partly attributed to the fact that 12 images from the set were used for training of the SRM segmentation algorithm. However, none of the algorithms was trained or tested on any of the DDSM images, yet the final A_Z score was very high (0.96). It seems that SRM reached its upper limit on the LSM and the isolable-contour analysis could not improve the outcome (in fact, a slight deterioration of A_Z at third decimal place was observed).

Figure 4 shows the only example from the DDSM set where the isolable-contour analysis failed (image

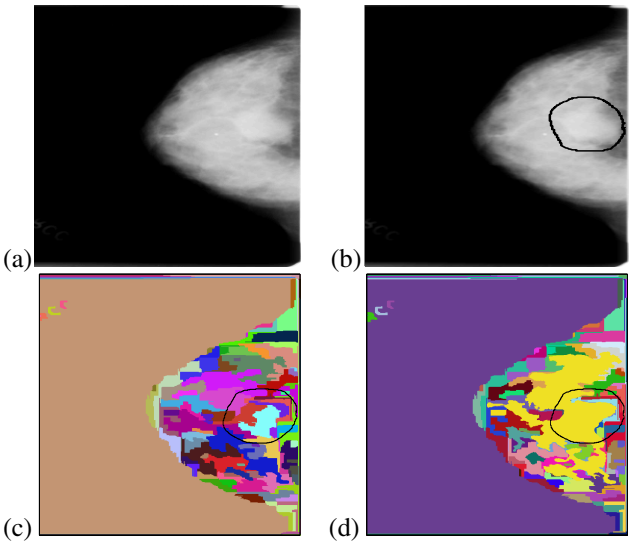

Figure 4. A difficult case from DDSM. (a) The original mammogram. (b) Image with the radiologist annotation. (c) SRM segmentation. (d) The outcome of the isolable-contour analysis

`C_0166_1.RIGHT.CC`). This can be attributed to the particularly high density of the breast. Notably, SRM still produced a very reasonable outcome. In case of our local dataset SRM failed in three cases producing significantly over-segmented mass regions. Of the three failures one mass was partially off the edge of the film and the other two were very low contrast masses. In one of these cases, isolable-contour technique managed to recover from over-segmentation and produced a reasonable mass contour.

VI. Comparison with MST segmentation

As mentioned in Section (II), the study in [4] revealed superiority of SRM over MST in handling noise on natural scene images. Since MST segmentation was also used to segment mammograms ([17], [5]) we looked at some of the differences between the two methods in the context of mammogram segmentation.

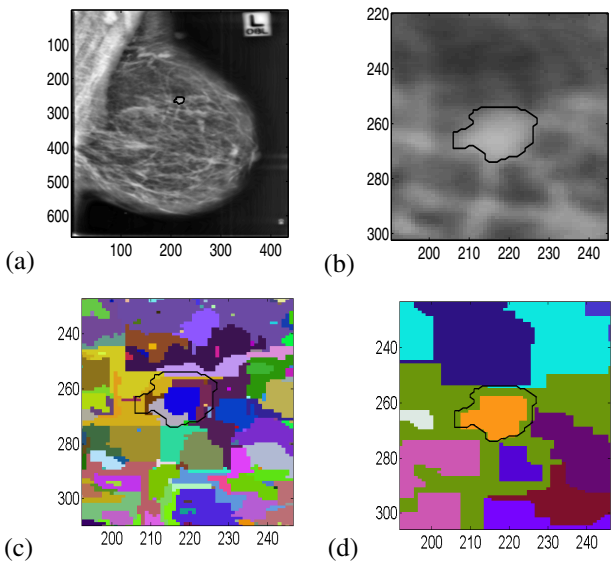

Figure 5. Comparison of segmentation of a small mass with MST and SRM. (a) The original preprocessed mammogram with mass contour marked. (b) The mass region. (c) MST segmentation of the mass region. (d) SRM segmentation of the same region.

We applied MST segmentation (with parameters chosen as suggested in [5] to avoid significant overmerging and, as a consequence, possible mass losses) on the set of 36 images from the local database LSM. The same set was then segmented with SRM using the parameter value estimated in Section IV. To compare segmentations, approximate mass contours were drawn for each image based on the radiologist annotated regions, and the overlap $\mathcal{O}$ of the contour and the set of regions produced by the segmentation was measured as in [18].

$$\mathcal{O} = \frac{area(\mathcal{S} \cap \mathcal{C})}{area(\mathcal{S} \cup \mathcal{C})}, \tag{13}$$

where $\mathcal{S}$ is the set returned from the segmentation algorithm and $\mathcal{C}$ is the set delineated by the drawn contour. As in [5], for $\mathcal{S}$ we took the union of all eligible regions (those with at least half of its area residing within the contour). The outcome is shown on Figure 6(a). MST segmentation produced much finer segmentation (4500-6500 regions per image) than SRM method (150-350 regions per image). As a result the contours were better delineated by the unions of MST regions and fewer massess were lost than for SRM. However, the task of merging the regions was very difficult and computationally infeasible due to their large number. Figure 5 shows an example of segmented small mass with both methods. The total number of segmented regions for this mammogram was 5844 for the MST segmentation, and 223 for the SRM method.

On the other hand, minimal or no merging is needed if the biggest eligible segmented component corresponds well to the mass area. Taking as $\mathcal{S}$ the biggest eligible region we

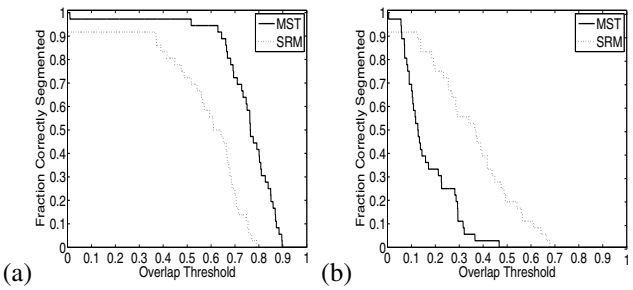

Figure 6. Comparison of MST and SRM performance on LSM. (a) For the union of eligible regions. (b) For the biggest eligible region.

again looked at the proportion of correctly segmented masses as a function of $\mathcal{O}$. Figure 6(b) shows that SRM significantly outperformed MST in this case. For example, it demonstrates that about 40% of masses had 40% or more area segmented correctly with SRM (as a single segmented region), while MST achieved the 40% overlap threshold for less than 5% of cases. Thus the use of MST segmented regions for mass detection can only be sensible if further post-processing like isolable-contour analysis or some merging technique can be efficently performed.

VII. Conclusion

The study shows that statistical region merging segmentation is applicable to mammograms. The regions produced by the method were accurate enough to be used as seed components for other region refinement or merging methods, such as the isolable-contour technique. Mass regions were successfully detected as suspicious in a proposed mass-detection scheme using lesion features robust to errors in assigning boundaries. These results are an improvement for these reported in [5] and [16].

We also showed that the SRM segmentation technique allows for an analytic rigorous judgment of the size of segmented lesions based on image characteristic. By estimating the saliency of the lesions (here measured by the average intensity across the lesion) one can determine the segmentation parameter ensuring that none of the important lesions is lost in the process due to overmerging.

Finally the paper shows that for practical applications SRM outperforms the MST segmentation since it produces significantly smaller number of regions and still preserving the quality of each lesion representation and overall segmentation. SRM segmentation can be well refined in an computationally efficient way for mass detection or other image analysis tasks.

References

[1] M. P. Sampat, M. K. Markey, and A. C. Bovik, *Computer-Aided Detection and Diagnosis in Mammography*, ser. The Handbook of Image and Video Processing. New York: New York Academic Press, 2005, pp. 1195–1217.

[2] H. D. Cheng, X. J. Shi, R. Min, L. M. Hu, X. P. Cai, and H. N. Du, "Approaches for automated detection and classification of masses in mammograms," *Pattern Recognition*, vol. 39, no. 4, pp. 646–668, 2006.

[3] R. M. Rangayan, F. J. Ayres, and J. E. Desautels, "A review of computer-aided diagnosis of breast cancer: Toward the detection of subtle signs," *Journal of the Franklin Institute*, vol. 344, pp. 312–348, 2007.

[4] R. Nock and F. Nielsen, "Statistical region merging," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 11, pp. 1452–1458, 2004.

[5] M. Bajger, F. Ma, and M. J. Bottema, "Automatic tuning of MST segmentation of mammograms for registration and mass detection algorithms," in *Digital Image Computing Techniques and Applications (DICTA2009)*, Melbourne, 2009, pp. 400–407.

[6] A. R. Dominguez and A. K. Nandi, "Toward breast cancer diagnosis based on automated segmentation of massess in mammograms," *Pattern Recognition*, vol. 42, pp. 1138–1148, 2009.

[7] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *Int. J. Comput. Vision*, vol. 59, no. 2, pp. 167–181, 2004.

[8] M. K. Heath, K. Bowyer, D. Kopans, P. K. Jr., R. Moore, K. Chang, and S. Munishkumaran, *Current status of the digital database for screening mammography*, ser. Digital Mammography. Dordrecht, The Netherlands: Kluwer Academic Publishers, 1998, pp. 457–460.

[9] Y. Guo and H. D. Cheng, "New neutrosophic approach to image segmentation," *Pattern Recognition*, vol. 42, pp. 587–595, 2009.

[10] L. Yu, F. Ma, A. Jayasuriya, M. Sigelle, and S. Perreau, "A new contour detection approach in mammogram using rational wavelet filtering and MRF smoothing," in *Digital Image Computing Techniques and Applications (DICTA2007)*, 2007, pp. 106–111.

[11] S. Timp and N. Karssemeijer, "Interval change analysis to improve computer aided detection in mammography," *Medical Image Analysis*, vol. 10, pp. 82–95, 2006.

[12] V. H. M. Sonka and R. Boyle, *Image Processing, Analysis, and Machine Vision*. Thomson, 2008.

[13] R. M. Rangayan, N. M. Faramawy, J. E. Desautels, and O. A. Alim, "Measures of acutance and shape for classification of breast tumors," *IEEE Transactions on Medical Imaging*, vol. 16, no. 6, pp. 799–810, 1997.

[14] S. Shiffman, G. D. Rubin, and S. Napel, "Medical image segmentation using analysis of isolable-contour maps," *IEEE Trans. Med. Imaging*, vol. 19, no. 11, pp. 1064–1074, 2000.

[15] B. W. Hong and M. Brady, "A topographic representation for mammogram segmentation," in *MICCAI 2003*, ser. LNCS, R. E. Ellis and T. M. Peters, Eds., vol. 2879, 2003, pp. 730–737.

[16] F. Ma, M. Bajger, and M. J. Bottema, "Automatic mass segmentation based on adaptive pyramid and sublevel set analysis," in *Digital Image Computing Techniques and Applications (DICTA2009)*, Melbourne, 2009, pp. 236–241.

[17] F. Ma, M. Bajger, J. P. Slavotinek, and M. J. Bottema, "Two graph theory based image segmentation methods for identifying pectoral muscle in mammograms," *Pattern Recognition*, vol. 40, pp. 2592–2602, 2007.

[18] M. Kupinski and M. L. Giger, "Automated seeded lession segmentation on digital mammograms," *IEEE Transactions on Medical Imaging*, vol. 17, no. 4, pp. 510–517, 1998.