Seed selection criteria for breast lesion segmentation in Ultra-Sound images

Joan Massich[1] *, Fabrice Meriaudeau[2], Elsa Pérez[3], Robert Martí[1], Arnau Oliver[1], and Joan Martí[1]

[1] Computer Vision and Robotics Group, University of Girona, Spain,
jmassich@atc.udg.edu,
[2] University of Burgundy, Le2i Laboratory CNRS UMR 5158, Esplanade Erasme, 21000, Dijon, France
[3] Dept of Radiology, Hospital Josep Trueta of Girona, Spain.

Abstract. Purpose : Segmentation plays a central role in medical imaging, though is not a trivial task to perform in some screening modalities such as Ultra-Sound images. This paper addresses the role of automatic seed placement when segmenting breast lesions in B-mode Ultra-Sound images, and proposes a new algorithm to automatically locate seed regions for further region growing expansion.
Methods : In this work some state-of-the-art methodologies for seed placement are reviewed and a new method basing its region selection on assigning a probability of belonging to a lesion for every pixel depending on intensity, texture and geometrical constraints of the pixel is proposed.
Results : The proposed algorithm has been evaluated using a set of sonographic breast images with accompanying expert-provided ground truth, and successfully compared to other existing algorithms.
Conclusions : The experimental results show the performance and robustness of the method when placing seed regions in noisy environments.

Keywords: seed placement, ultra-sound, segmentation, breast cancer

1 Introduction

Breast cancer constitutes one of the leading causes of death for women in developed countries, and is most effectively treated when diagnosed at an early stage [11].

Taking this into account, Digital Mammography is still the most powerful screening tool for breast cancer [6]. However, some studies [12] have shown in a recent past that Ultra-Sound (US) images of the breast can provide useful complementary information in cases where the patients present dense glandular breast tissue, and a tumor presence can be shielded when using mammogram screening. In addition, US images is a non-expensive and non-invasive technique with no side effects, thus rendering sonography an attractive complement to

* We gratefully acknowledge the help of Dr. Gururajan and Dr. Sari-Sarraf from Texas Tech University.This research was partially supported by the Spanish Government MEC grant nb. TIN2007-60553 and the University of Girona BR grant nb. 09/22

digital mammography and leading to a re-emergence of interest in image segmentation applied to ultrasound data [7] due to the segmentation's clinical value. Despite this, performing automatic segmentation in US images is a challenge because they often suffer from poor quality. US imaging tends to generate artifacts like weak edges produced by acoustic similarity between adjacent tissues, shadows presence when the signal gets completely attenuated preventing to screen any further, low contrast as a consequence of the US wave attenuation by the tissue media, or, speckle which is an unwanted collateral artifact coming from coherent interface of scatterers that appear as a granular structure superimposed on the image.

Due to segmentation's clinical value, the literature reports several techniques proposals for both guided and automatic segmentation of lesions in US images that try to overcome all the US screening inconveniences. Among those, region growing procedures that expand a seed accordingly to some criteria, have been reported to be suitable for US image segmentation [5, 4, 7]. However if seeds are not properly selected, the final segmentation results would be definitely incorrect.

This work uses an already stated framework for segmenting breast lesions in US images [5] in order to study the seed placement influence when segmenting and compare several seed selection procedures that can be plugged within such framework. This work also proposes a novel procedure combining texture and intensity features with geometric constrains. The framework has been tested with different seed selection procedures, against a data-set of sonographic images with accompanying expert-provided ground truth.

2 Seed Placement Background

Determining seed points on an US image that lead to a proper segmentation of breast lesion is not a trivial task, basically due to the noisy nature of the US images and the presence of other structures rather than lesions with similar acoustic properties (e.g. in some screening conditions, subcutaneous fat can be mistaken as a lesion). To achieve a fully automatic procedure, seeded segmentation methods require an automatic seed placement as well. The remaining of this section reviews some of those automatic procedures for selecting good seeds.

Pixel Rewarding method to select seed points (PR) Madabushi and Metaxas [4] proposed a method which rewards each pixel according to its position, intensity and texture using an assessment function. The main advantage of this pixel rewarding proposal remains in its spatially constrained seed rewarding along with the fact that the lesion's appearance is obtained by means of a learning step. On the other hand, its major disadvantage remains in choosing an appropriated neighborhood for the term representing the probability mean of the surrounding pixels when calculating the pixel reward. If the neighborhood used is too small, it might incorrectly reward a noisy region; otherwise, if the used neighborhood is too large, a proper seed can be hidden due to its neighbors' low recall.

Gradient-Based method to select seed points (GB) Drukker et al. [1] investigated the use of Radial Gradient Index (RGI) filtering technique to automatize Horsch et al. segmentation proposal [2] by adding automatic seed placement. Such seed placement uses the gradient as the only feature to select seeds by computing the maximum RGI [3] for all the pixels of the input image. RGI is a measure similar to Average Radial Derivative (ARD) coefficient that is used to drive the segmentation by Horsch et al. [2]. Summing up, every pixel is proposed as a potential lesion in order to determine which pixel would have the best reward, so the seed selection is deeply coupled to the segmentation procedure. Clearly, the main drawback of this seed selection is its computational cost, which was partially solved by means of subsampling techniques. However, due to the comprehensive nature of the seed determination, the method remains unadvisable for anything but offine applications.

Intensity Binarized ranked Regions method to select seed points (IBRR) Shan et al. [10] find candidate lesion regions based on intensity and rank them using the region properties; once the region is chosen, a seed point is determined within the selected region.

3 ITG: a novel seed region selection methodology

Both intensity and texture have been stated as a high specificity features when charaterizing breast lesions in US images [12]. In addition, the tendency of centering the lesions when acquiring the images by the radiologists has also been stated [4]. Figure 1 shows the proposed methodology which makes use of Intensity, Texture and Geometric constrains (ITG) and takes advantage of the mentioned statements in order to select a seed region for further region growing expansion. The proposal, combines the probability of a pixel being part of a lesion depending on its intensity, texture and position to generate a joint probability or total probability plane. Then the selection criterium selects the largest region of the connected pixels that satisfy a confidence level of being a lesion. So for selecting the best candidate regions, the probability plane gets thresholded in order to split the image with foreground and background. This thresholding has been empirically set at 0.8 as a good tradeoff between large foreground regions and low lesion belonging recall. Once determined the regions, the largest one gets selected as seed region.

Equation 1 illustrates the Joint probability calculation, where $\tau(x, y)$ indicates the total probability for a pixel (x, y) of being part of a lesion depending on its intensity i, texture t and position (x, y). Since intensity, texture and location

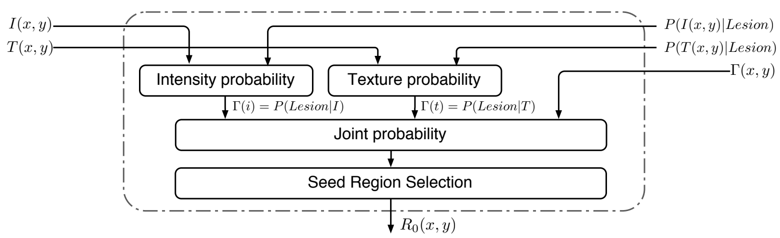


Fig. 1: Block diagram describing the seed region selection proposal.

features can be assumed Independent and Identically Distributed (IID) [4], the total probability corresponds to the three features' probability product. The intensity probability $\Gamma(i)$ is computed from a Probability Density Function (PDF) determined during a training step necessary to compute $P(i|Lesion)$. $\Gamma(i)$ can be computed from the intensity PDF by the assumption of a Bayesian framework. $\Gamma(t)$ has the same nature as $\Gamma(i)$ and the texture PDF also needs to be determined during a training step. The seed location constrain $\Gamma(x, y)$ corresponds to a bivariate Gaussian function where the variances have been visually assessed and fixed at $\frac{1}{3}$ of the image size.

$$\tau(x, y) = \Gamma(i) \cdot \Gamma(t) \cdot \Gamma(x, y) \tag{1}$$

The texture measure used is given by equation 2 and corresponds to the difference between the pixel intensity $I(x, y)$ and the intensity mean of its N nearest neighbors (here an eight pixel neighborhood is used).

$$T(x, y) = I(x, y) - \frac{1}{N}\sum_{\delta=0}^{N-1} I_{\delta}(x, y) \tag{2}$$

In summary, the proposed methodology uses five inputs to automatically determine a seed region: the intensity image, the texture image, the intensity and texture PDFs, and the seed location prior; along with a fixed parameter to split the probability plane. Figure 2 illustrates all the steps involved during the course of action, where the upper row represents the procedure inputs (intensity, texture, geometrical constrains and learned PDFs), and the lower row shows the probability image for the intensity feature (e), for the texture feature (f), the total probability (g) and the final seed region selection (h). The final selected region (the largest) is depicted in magenta and the region candidates obtained when thresholding the probability plane are shown in cyan.

4 Experimental setup

The Gaussian Constraining Segmentation framework proposed by Massich et al. [5] has been used to test and evaluate the seed placement approach. Although such segmentation framework allows different user interation levels, as figure 3 depicts, only the fully automatic procedure has been used in this work. First, an initial region $R_0(x, y)$ is determined and then grown into a preliminary lesion delineation $R(x, y)$ that is used to obtain a multivariate Gaussian function describing the shape, position and orientation of the lesion ($G_{\mu\Sigma}(x, y)$). Finally the Gaussian Constraining Segmentation (GCS) procedure refines the segmentation by thresholding an intensity dependent function $\Psi(x, y)$ constrained by the multivariate Gaussian describing the lesion.

In order to evaluate the segmentations, Massich et al. [5] propose to use Simultaneous Truth and Performance Level Estimation (STAPLE) algorithm [13] to obtain the Hidden Ground Truth (HGT) from multiple expert delineations. Then use the μ-coefficient proposed as a variance of the True-Positive Ratio (TPR) or Jaccard coefficient that takes into accound the experts agreement by means of the HGT.

ISBN: 978-87-981270-9-3

Workshop on Breast Image Analysis
In conjunction with MICCAI 2011

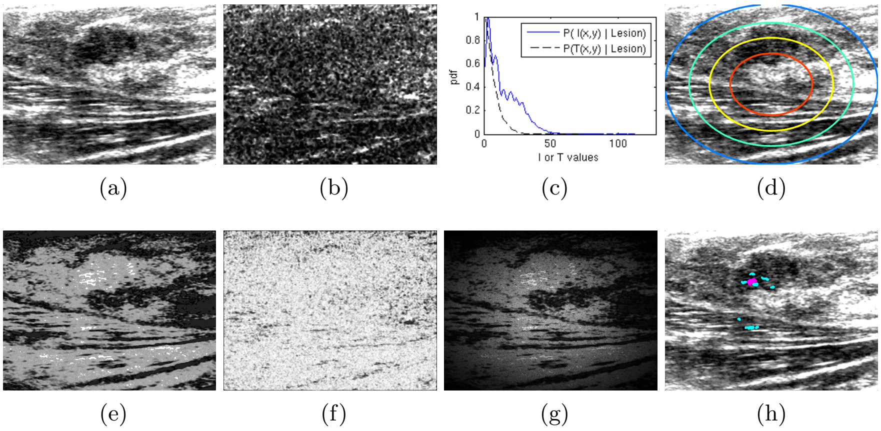


Fig. 2: Seed region selection illustration. (a) pre-processed intensity image (b) texture image (c) intensity and textureProbability Density Functions (d) seed location prior colored as overlay (e) $\Gamma(i)$ (f) $\Gamma(t)$ (g) total joint probability $\tau(x, y)$ (h) candidate regions (in cyan) and the final selected region (in magenta).

For evaluating purposes, a set of 25 sonograms were acquired in the *Hospital Dr. Josep Trueta* of Girona. Each image has seven ground truth delineations provided by different radiology experts. The training and testing of the data is obtained using a leave-one-out methodology.

4.1 Seed region location

When evaluating the seed selection, a key issue is to determine what defines a good seed in terms of the initial seed position. Figure 4a illustrates the ten Areas-of-Interest used in this case of study to test the influence of the lesion center distance and orientation. The Areas-of-Interest have been selected as: out

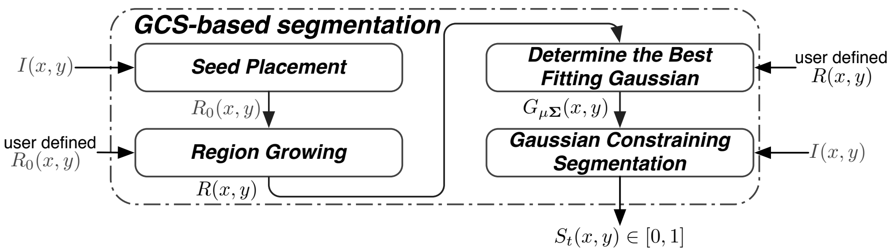
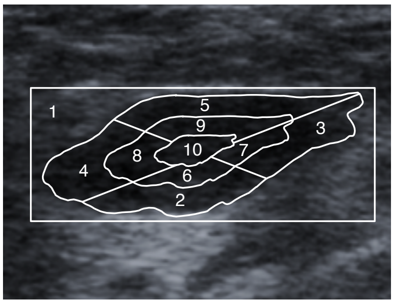
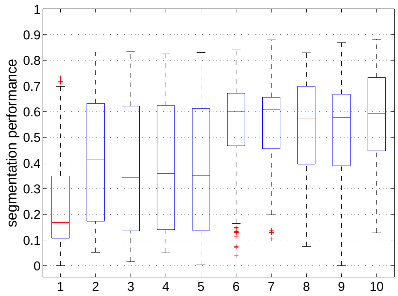


Fig. 3: Methodology block diagram. When user interaction is used (only for semi-automatic segmentation), it overwrites the previous input.



(a)



(b)

Fig. 4: First experiment: (a) distribution of the seeds' regions, and (b) segmentation results in terms of μ-coefficient.

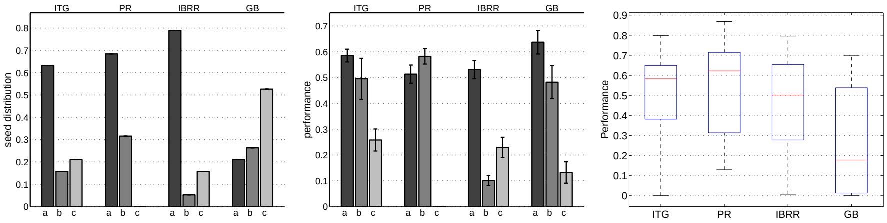
of the lesion (1), inside the lesion close to the boundaries (2-5), inside the lesion slightly shifted from the central part (6-9) and, central part of the lesion (10). Figure 4b shows the segmentation results for each Area-of-Interest according to the μ value. Each of the ten Areas-of-Interest has been randomly sampled with 15 seed regions. The boxplots clearly shows that achieving good segmentation results highly depends on locating the seed regions within the lesion (Areas-of-Interest 2 to 10). The figure also shows that the regions can be clustered in three main classes a to c: (a) Areas-of-Interest 6-10 that correspond to the inner lesion area, (b) 2-5 boundary area, and, (c) 1 anything outside the lesion. The results indicates that the better segmentation results are achieved when the seed is placed in (a), but not necessarily in the most inner region.

4.2 Methodology evaluation

As well to determine the role of the seed region location in terms of segmentation results, the proposed seed selection method has been evaluated by comparing to the methods referred in section 2: PR, GB and IBRR. Figure 5 illustrates the obtained results when comparing methodologies. The first plot (fig. 5a) shows the location of the selected seed regions along three areas based on the first experiment. The second plot (fig. 5b) illustrates the mean and variation of the final segmentation results for each methodology and area. Finally, figure 5c illustrates the performance distribution of each methodology regardless of which area the seed regions are placed. Notice that figure 5a is expressed in terms of the seeds distribution within the three groups of Areas-of-Interest while 5b and 5c refers to the μ coefficient to assess the final segmentation results.

Altough the PR and IBRR methods place more seeds in the central area than the proposed ITG method, the latter has the best performance in terms of final segmentation results when the seed is placed in the central area (a). The GB performance is not significant since its ability to place the seed regions within the (a) area is quite low, and most of the seeds fall outside the lesion.

ISBN: 978-87-981270-9-3



(a) (b) (c)

Fig. 5: Second experiment: comparison between the proposed method (ITG) and the PR, GB and IBRR methods. (a) seed region location (b) μ values depending on the seed location (c) global μ values.

4.3 Noise degradation

This experiment has been devoted to observe the seed placement evolution in challenging noisy scenario by repeating the segmentation on artificially degraded US images. The noise in US images mainly comes from scattering and reflection [9]. The structures present on a sonogram produce an adaptative coherent scatter comonly regarded as a Rician PDF [9], whereas the incoherent or diffuse scattering normally modeled as a Rayleigh PDF [9] leads to speckle artifacts. In order to obtain a fairly realistic US degradation, a percentage of the pixels from the image have been modified by a random walk of aleatory number of steps. For every step within the random walk, an amplitude and phase of the scatter have been simulated. Finally in order to eliminate the impulse nature of the added noise, a spacial correlation has been carried out [8]. From 0% to 100% of the pixels within the images have been alterated following the mentioned scheme by steps of 5%. Accordingly to the doctors, although some images have way more noise than the acceptable for diagnose purposes, the images seem to have been acquired with lower performing US imaging equipment. Figure 6e shows the ratio of seed regions placed within the 75% of the most inner area of the lesion between the ITG method and the PR and IBRR methods. Observe that for the proposed ITG method, values are mainly higher than 0.6 as for the IBRR method. Notice that previous experiments recalled that when the seed is properly placed ITG performs better than IBRR.

5 Conclusions

In this work, the importance of a good seed selection for a region growing like procedure has been stated. Some state-of-the-art seed placement procedures have been implemented, discussed and compared to a novel seed region selection proposal based on assessing the probability of belonging to a lesion for every pixel in the image depending on its intensity, texture and location and selecting the largest area obtained. The location of good seed regions on noisy environments has also been addressed, thus validating the performance and robustness of the methodology when placing seed regions in such noisy environments.

Workshop on Breast Image Analysis
In conjunction with MICCAI 2011

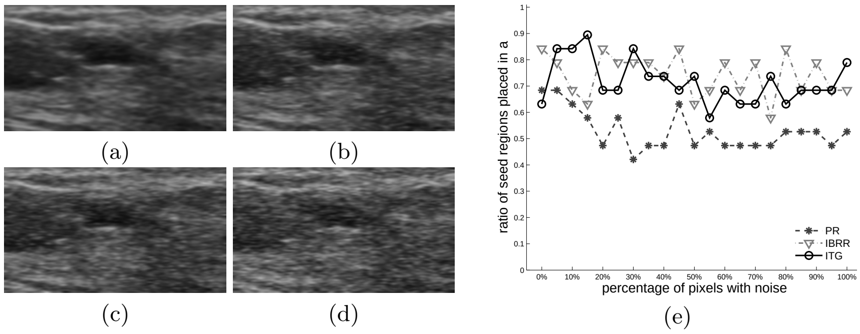


Fig. 6: Third experiment: seed placement performance on noisy enviroment. (a-d) same image with different amount $\{0, 20, 40, 80\}\%$ of added noise (e) ratio of correctly placed seeds at different noise levels.

References

1. Karen Drukker, Maryellen L Giger, Karla Horsch, Matthew A Kupinski, Carl J Vyborny, and Ellen B Mendelson. Computerized lesion detection on breast ultrasound. *Medical Physics*, 29(7):1438–46, Jul 2002.
2. K. Horsch, M.L. Giger, L.A. Venta, and C.J. Vyborny. Automatic segmentation of breast lesions on ultrasound. *Medical Physics*, 28:1652–1659, 2001.
3. M.A. Kupinski and M.L. Giger. Automated seeded lesion segmentation on digital mammograms. *IEEE Trans. Med. Imaging*, 17(4):510–517, 1998.
4. A. Madabhushi and D. Metaxas. Automatic boundary extraction of ultrasonic breast lesions. In *2002 IEEE International Symposium on Biomedical Imaging, 2002. Proceedings*, pages 601–604, 2002.
5. J. Massich, F. Meriaudeau, E. Pérez, R. Martí, A. Oliver, and J. Martí. Lesion segmentation in breast sonography. In *Digital Mammography*, volume 6136 of *LNCS*, pages 39–45, 2010.
6. SK Moore. Better breast cancer detection. *IEEE Spectrum*, 38(5):50–54, 2001.
7. J.A. Noble and D. Boukerroui. Ultrasound image segmentation: A survey. *IEEE Transactions on medical imaging*, 25(8):987–1010, 2006.
8. D Jones R Czerwinski and W. O'Brien. Edge detection in ultrasound speckle noise. *Image Processing*, Jan 1994.
9. J.M. Sandrick H. Lopez R.F. Wagner, S.W. Smith. Statistics of speckle in ultrsound b-scans. volume 30. IEEE Transactions on Sonics and Ultrasonics, 3 1983.
10. Juan Shan, H.D. Cheng, and Yuxuan Wang. A novel automatic seed point selection algorithm for breast ultrasound images. In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, dec 2008.
11. R. Sivaramakrishna, K.A. Powell, M.L. Lieber, W.A. Chilcote, and R. Shekhar. Texture analysis of lesions in breast ultrasound images. *Computerized medical imaging and graphics*, 26(5):303–307, 2002.
12. A.T. Stavros, C.L. Rapp, and S.H. Parker. *Breast ultrasound*. Lippincott Williams & Wilkins, 2004.
13. S.K. Warfield, K.H. Zou, and W.M. Wells. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Transactions on Medical Imaging*, 23(7):903–921, 2004.