An Optimization Approach to Segment Breast Lesions in Ultra-Sound Images using Clinically Validated Visual Cues

Joan Massich, Guillaume Lemaître, Mojdeh Rastgoo, Fabrice Mériaudeau
Université de Bourgogne Franche-Comté
12 rue de la Fonderie, 71200 Le Creusot

Anke Meyer-Baese
Department of Scientific Computing, Florida State University
400 Dirac Science Library Tallahassee, FL 32306, US

Joan Martí
Universitat de Girona
Campus Montilivi, Edifici P4, 17071 Girona

Abstract

As long as breast cancer remains the leading cause of cancer deaths among female population world wide, developing tools to assist radiologists during the diagnosis process is necessary. However, most of the technologies developed in the imaging laboratories are rarely integrated in this assessing process, as they are based on information cues differing from those used by clinicians. In order to grant Computer Aided Diagnosis (CAD) systems with these information cues when performing non-aided diagnosis, better segmentation strategies are needed to automatically produce accurate delineations of the breast structures. This paper proposes a highly modular and flexible framework for segmenting breast tissues and lesions present in Breast Ultra-Sound (BUS) images. This framework relies on an optimization strategy and high-level descriptors designed analogously to the visual cues used by radiologists. The methodology is comprehensively compared to other sixteen published methodologies developed for segmenting lesions in BUS images. The proposed methodology achieves similar results than reported in the state-of-the-art.

1. Introduction

Breast cancer is the second most common cancer. In terms of mortality, breast cancer is the fifth most common cause of cancer death. However, it is ranked as the leading cause of cancer deaths among females in both western and economically developing countries [1].

Medical imaging contributes to its early detection through screening programs, non-invasive diagnosis, follow-up, and similar procedures. Despite Breast Ultra-Sound (BUS) imaging not being the imaging modality of reference for breast cancer screening [2], Ultra-Sound (US) imaging has more discriminative power when compared with other image modalities to visually differentiate benign from malignant solid lesions [3]. In this manner, US screening is estimated to be able to reduce between $65 \sim 85\%$ of unnecessary biopsies, in favour of a less traumatic short-term screening follow-up using BUS images. As the standard for assessing these BUS images, the American College of Radiology (ACR) proposes the Breast Imaging-Reporting and Data System (BI-RADS) lexicon for BUS images [4]. This US BI-RADS lexicon is a set of standard markers that characterizes the lesions encoding the visual cues found in BUS images and facilitates their analysis. Further details regarding the US BI-RADS lexicon descriptors proposed by the ACR, can be found in Sect. 3, where visual cues of BUS images and breast structures are discussed to define feature descriptors.

The incorporation of US in screening policies and the emergence of clinical standards to assess image like the US BI-RADS lexicon, encourage the development of Computer Aided Diagnosis (CAD) systems using US to be applied to breast cancer diagnosis. However, this clinical assessment using lexicon is not directly applicable to CAD systems. Shortcomings like the location and explicit delineation of the lesions need to be addressed, since those tasks are intrinsically carried out by the radiologists during their visual assessment of the images to infer the lexicon representation of the lesions. Therefore, developing accurate segmentation methodologies for breast lesions and structures is crucial to
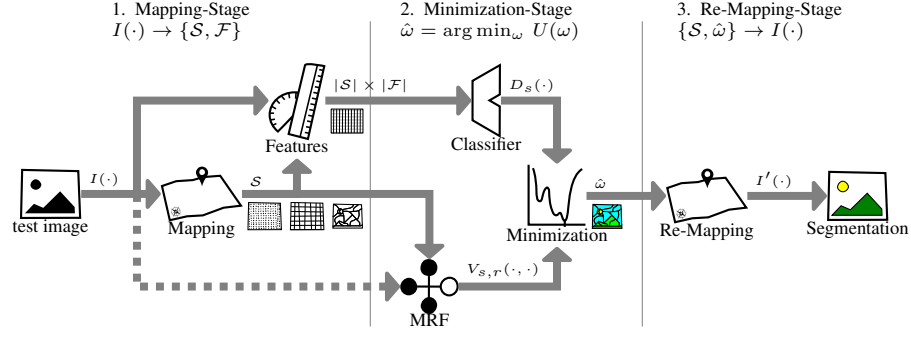
Figure 1: Conceptual block representation of the segmentation methodology.

take advantage of this already validated clinical tools.

2. Description of the segmentation methodology

Optimization methodologies offer a standardized manner to approach segmentation by minimizing an application-driven cost function [5]. Figure 1 illustrates a generic representation of the segmentation strategy, concrete examples of its terms, applied to BUS, can be found in section 3. The overall segmentation can be seen as a three-steps strategy: (1) a mapping of the image into a discrete set of elements $\mathcal{S}$, (2) the optimization stage which is formulated as a *metric labelling* problem, and (3) a re-mapping the labels obtained from the previous stage to produce the final delineation.

To formulate segmentation as a metric labelling problem, the image is conceived as a discrete set of elements $\mathcal{S}$ that need to be labelled using a label l from the labelling set $\mathcal{L}$. Let $\mathcal{W}$ be all the possible labelling configurations of the set $\mathcal{S}$, given $\mathcal{L}$. Let $U(\cdot)$ be a cost function encoding the goodness of the labelling configuration $\omega \in \mathcal{W}$ based on the appearance of the elements in $\mathcal{S}$, their inner relation and some designing constraints. Then, the desired segmentation $\hat{\omega}$ corresponds to the labelling configuration that minimizes this cost function, as described in Eq. (1).

$$\hat{\omega} = \arg \min_{\omega} U(\omega) \qquad (1)$$

This goodness measure $U(\cdot)$ must be defined to take into account the appearance of the target region, its relation with other regions, and other designing constraints. Equation (2) describes this cost function as the combination of two independent costs that need to be simultaneously minimized as a whole.

$$U(\omega) = \sum_{s \in \mathcal{S}} D_s(\omega_s) + \sum_{s \in \mathcal{S}} \sum_{r \in \mathcal{N}_s} V_{s,r}(\omega_s, \omega_r) \qquad (2)$$

Where, the left-hand side of the expression integrates the so-called *data* term, while the right-hand side integrates the

pairwise term, which is also referred to as the *smoothing* term. Both terms are shaped by $\mathcal{S}$ and evaluated in the labelling space $\mathcal{W}$. In our quest to optimize the cost function $U(\cdot)$, it is required to define a representation for the set $\mathcal{S}$, a data term $D(\cdot)$, a pairwise term $V(\cdot)$, and a proper minimization methodology.

The set $\mathcal{S}$ can be, in general, any discrete set representing the image (i.e. pixels, overlapping or non overlapping windows, super-pixels, etc.).

The data term $D(\cdot)$, given a label configuration $\omega \in \mathcal{W}$, penalizes the labelling of a particular image element or site ($\omega_s = l$) based on the data associated to s. In this manner, $D_s(\omega_s = l_{\checkmark}) << D_s(\omega_s = l_{\textbf{x}})$. Figure 2b illustrates the data cost associated to some arbitrary labelling configurations to clarify the desired effect (or behaviour) of this data term. Designing an obscure heuristic to comply with the desired behaviour of $D(\cdot)$ out of the box, is rather a complicated task. Therefore, an easier and cleaner approach is to design this data term $D(\cdot)$ with the help of Machine Learning (ML) because it provides a systematic process that is flexible enough to encode any desired behaviour based on a training stage. This concept is in fact depicted in the upper row in Fig. 1. For each site $s \in \mathcal{S}$, features describing s are designed. Then, different optional steps can be applied to this set of features: (i) features normalization, (ii) features selection or (iii) features extraction. Finally, the data term $D(\cdot)$ is encoded based on ML classifiers, the features and a training step. Thus, the data term $D(\cdot)$ can be seen as a distance or goodness measure reflecting the likelihood for s to belong to class l.

The pairwise term $V(\cdot, \cdot)$ represents the cost associated to ω_s taking into account the labels of its neighboring sites, ω_r, $r \in \mathcal{N}_s$. This term is usualy modeled using Markov Random Fields or Conditional Random Fields. The typical form of this term, given in Eq. (3), is called homogenization
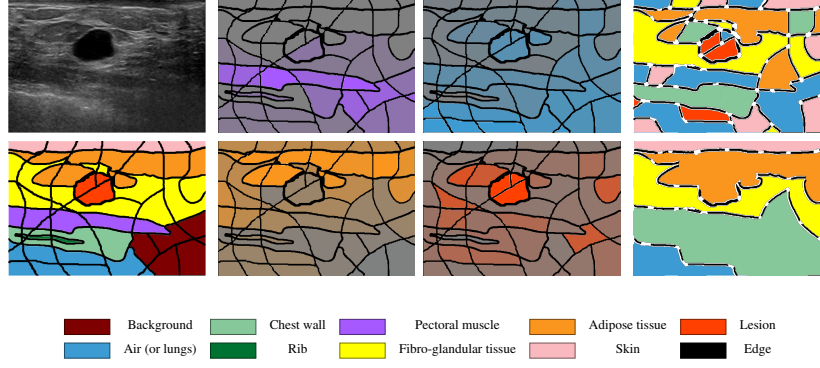
Figure 2: Methodology Highlights. (a) BUS image example. (b) Superpixels' representation coloured using dataset's accompanying multi-label GT. (c) GT color code. (d) Data term: cost of labelling all sites as pectoral, lungs, adipose tissue or lesion. For illustration purposes, highly saturated colour indicates a low data cost - i.e., high confidence to assign the label associated with the color. (e) Pairwise term: labelling configurations with more boundaries produce higher pairwise term cost.
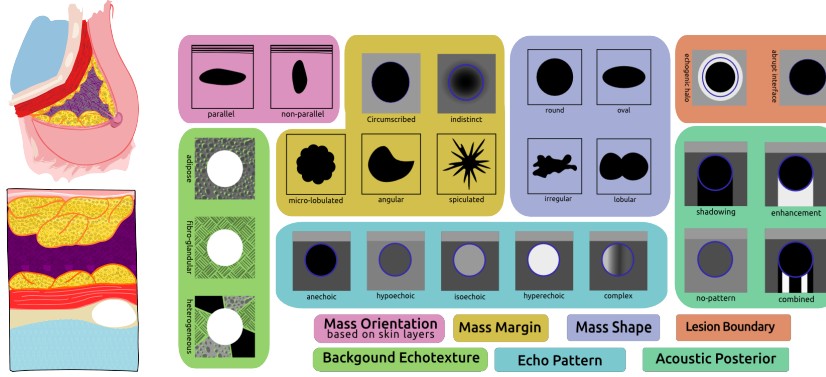

Figure 3: Visual reference: (a) breast structures, (b) US BI-RADS lexicon

which acts as a regularization factor favouring configurations that have a coherent labelling.

$$V_{s,r}(\omega_s, \omega_r) = \begin{cases} \beta, & \text{if } \omega_s \neq \omega_r \\ 0, & \text{otherwise} \end{cases} \qquad (3)$$

Figure 2c shows a visual interpretation of this cost. The more fragmented is the segmentation ω, the higher is the overall pairwise term, since every boundary brings a penalization β to the total cost $U(\omega)$. In this manner, the regularization term can be seen as a post-processing or denoising stage as some sites will flip their labelling if the cost of fragmenting the regions is larger than the cost of adopting their neighbour's label.

The minimization strategy is determined by the nature of $U(\cdot)$ and $\mathcal{W}$, since not all the minimization strategies are applicable or adequate to find $\hat{\omega}$. The size of the labelling space $|\mathcal{W}| = |\mathcal{L}|^{|\mathcal{S}|}$, discontinuities in $U(\omega)$ along $\mathcal{W}$ or the problem of local minima, additionally all the particular of

all the different minimization. Need to be taken into account while choosing the most desirable minimization strategy.

3. BUS images segmentation using optimization

In this section, the problem of delineating structures in BUS images is defined as an optimization problem that can be solved by applying the framework presented in Sect. 2. The segmentation here proposed aims at tying a label $l \in \mathcal{L}$ (*i.e.* {lesion, $\overline{\text{lesion}}$} or {chest wall, lungs, . . . , lesion}) to each element of $\mathcal{S}$ by simultaneously optimizing the data and pairwise terms as illustrated in Fig. 2. Choices made regarding different elements: the representation $\mathcal{S}$, the data term $D(\cdot)$, the pairwise term $V(.)$, and the optimizer choice are summarized in Table 1 and justified thereafter (see Fig. 1 for reference).

$\mathcal{S}$ is considered the result from an over-segmentation of the image using Quick-shift super-pixel [6]. The structures of the breast and their rendering when using a hand-held 2D US probe are sketched in Fig. 3a. Figure 3b illustrates

the lexicon proposed by the ACR [4] and used by clinicians to perform their diagnosis. Thus, our aim is to generate a set of computer vision features which is able to encode the characteristic described in the lexicon. The selected features are as follows:

Appearance Based on the multi-labelled GT, a Median Absolute Deviation histogram model for every tissue label is built. The Appearance feature is computed as the χ^2 distance between a histogram of s and the models generated.

Atlas Based on the multi-labelled GT, an atlas is built to encode the labels likelihood based on the location of s.

Brightness Intensity descriptors are computed based on statistics of s (*i.e:* mean, median, mode) and are compared with some intensity markers of the set $\mathcal{S}$ such as the minimum intensity value, the maximum, its mean, etc.

Self-Invariant Feature Transform (SIFT)-Back-of-Features s is described as a histogram of visual words based on SIFT [7]. The dictionary is built with 36 words.

The relationships between the lexicon and the descriptors described previously are depicted in Table 1. More precisely, we highlight the corresponding elements of the lexicon which is encoded by each feature. A choice regarding the encoding of the data term $D(\cdot)$ has to be made by using a ML classifier. Support Vector Machines classifier with Radial Basis Function kernel is selected to determine the data model during the training stage. The pairwise term in our framework was defined as in Eq. (3). The optimization method used as a solver to minimize our cost function $U(\cdot)$ is Graph-Cuts (GC). GC, where appropriately applied, allows to rapidly find a strong local minima guaranteeing that no other minimum with lower energies can be found [8]. GC is applicable if, and only if, the pairwise term favours coherent labelling configurations and penalizes labelling configurations where neighbours' labels differ such as in our case, given by Eq. (3).

4. Method evaluation and comparison

The proposed methodology is evaluated using a dataset of 16 BUS images presenting a single lesion of variable extension. The size of the lesions ranges from under $1/100$ to over $1/5$ of the image size. The dataset is composed of cysts, Fibro-Adenomas, Ductal Inflating Carcinomas and Inflating Lobular Carcinomas. Every image has accompanying multi-label GT delineating all the depicted structures. This dataset is now publicly available at `http://visor.udg.edu/dataset/#breast`

The lack of publicly available data and source code, limits the comparison between the different methods. For this study, the results published by the other authors have been collected and expressed in terms of AOV in order to share a common metric. Further details can be found in [10] and summarized in Fig. 5.

Figure 5 is divided into three main parts: (i) a table on the top summarizes the core stages of each study framework, (ii) a legend box on the right side informs our testing setup and, (iii) a comparison of the different metrics in a radial manner. An extra element is also represented in this radial representation: a blue swatch delimited by two blue dashed lines. The boundaries of this swatch correspond to the performance of some expert radiologists based on an inter- and intra-observer experiments carried out by Pons et al. [9]. It is interesting to note that some methodologies outperform this swatch. A publicly available dataset should allow a better comparison in that regard.

The results point out the inherent capabilities of ML to cope with data scalability and variability, induce its usage in conjunction with larger datasets. Whereas, Active Contour Model (ACM) methodologies show its effectiveness to model the boundary in a natural manner.

For our proposed framework, the performance in terms of AOV lies within the state-of-the-art despite its final delineation limited by the capacity of the super-pixels to snap the desired boundary. Figure 4 shows some qualitative results where there are limitations of labeling super-pixels when compared with hand-drawn GT. Figure 4 also illustrates the influence of the pair-wise term.

5. Conclusions

This work presents a segmentation strategy to delineate lesions in BUS images using an optimization framework that takes advantage of all the facilities available when using ML techniques. Despite the limitation that the final segmentation is subject to the super-pixels' boundaries, the AOV results reported here are similar to those reported by other methodologies in the literature. A higher AOV result can be achieved by refining the delineation resulting from our proposed framework by post-processing it with an ACM. In this manner, the contour constraints could be applied to achieve a more natural delineation.

References

[1] A. Jemal *et al.*, "Global cancer statistics," *CA: A Cancer Journal for Clinicians*, vol. 61, 2011 (cit. on p. 1).

[2] R. A. Smith *et al.*, "American cancer society guidelines for breast cancer screening: update 2003," *CA: a cancer journal for clinicians*, vol. 53, no. 3, pp. 141–169, 2003 (cit. on p. 1).

Table 1: Design choices summary

$\mathcal{S}$	Quick-Shift super-pixels
	Background Echotexture: encoded in Appearance and SIFT-BoW
$D(\cdot)$	Echo Pattern: encoded in Appearance, Atlas and Brightness
	Acoustic Posterior: encoded in Atlas and Brightness
$V(\cdot, \cdot)$	homogeneity as Eq. (3)
$\arg \min U(\cdot)$	Graph-Cuts

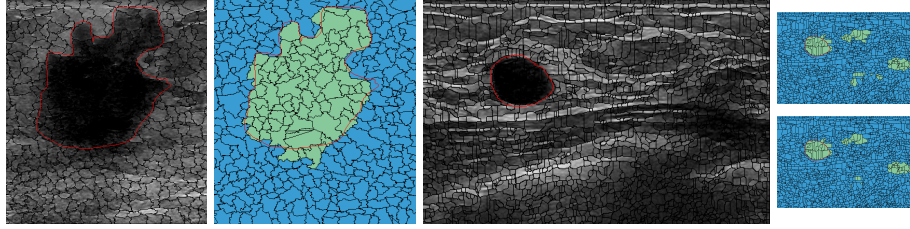Figure 4: Qualitative results. (a) orignal image, super-pixels' delineations and GT. (b) Differences between GT and the delination resulting from super-pixels' boundary. (c) Ex. 2. (d) weak $V(\cdot, \cdot)$ (e) strong $V(\cdot, \cdot)$
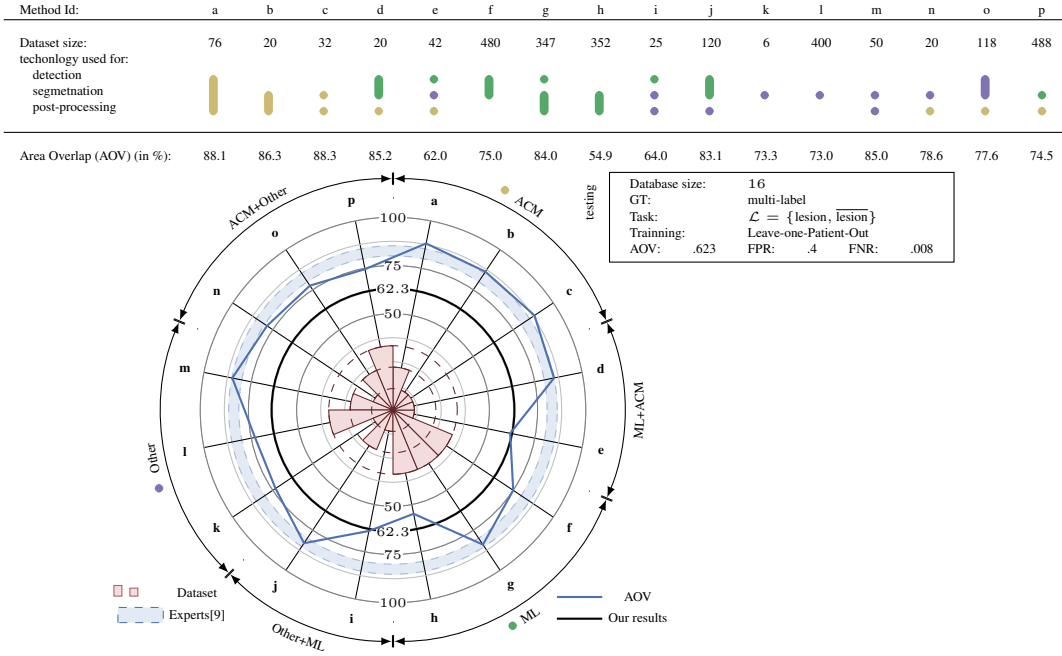

Figure 5: Quantitative results compilation and comparison

[3] A. T. Stavros, D. Thickman, C. L. Rapp, M. A. Dennis, S. H. Parker, and G. A. Sisney, "Solid breast nodules: Use of sonography to distinguish between benign and malignant lesions," *Radiology*, vol. 196, no. 1, pp. 123–34, 1995 (cit. on p. 1).

[4] E. Mendelson, J. Baum, B. WA, *et al.*, *BI-RADS: Ultrasound, 1st edition in: D'Orsi CJ, Mendelson EB, Ikeda DM, et al: Breast Imaging Reporting and Data System: ACR BIRADS – Breast Imaging Atlas*. American College of Radiology, 2003 (cit. on pp. 1, 4).

[5] D. Cremers, M. Rousson, and R. Deriche, "A review of statistical approaches to level set segmentation: integrating color, texture, motion and shape," *International journal of computer vision*, vol. 72, no. 2, 2007 (cit. on p. 2).

[6] R. Achanta *et al.*, "SLIC superpixels compared to state-of-the-art superpixel methods," 2012 (cit. on p. 3).

[7] J. Massich *et al.*, "Sift texture description for understanding breast ultrasound images," in *Breast*

Imaging, ser. Lecture Notes in Computer Science, vol. 8539, Springer International Publishing, 2014, pp. 681–688, ISBN: 978-3-319-07886-1 (cit. on p. 4).

[8]   A. Delong, A. Osokin, H. N. Isack, and Y. Boykov, "Fast approximate energy minimization with label costs," *International Journal of Computer Vision*, vol. 96, no. 1, pp. 1–27, 2012 (cit. on p. 4).

[9]   G. Pons, J. Mart, R. Mart, S. Ganau, J. Vilanova, and J. Noble, "Evaluating lesion segmentation in breast ultrasound images related to lesion typology," *Journal of Ultrasound in Medicine*, 2013 (cit. on pp. 4, 5).

[10] J. Massich, "Deformable object segmentation in ultra-sound images," PhD thesis, Universitat de Girona, Girona (Catalunya), Dec. 2013 (cit. on p. 4).