Breast Ultra-Sound image segmentation: an optimization approach based on super-pixels and high-level descriptors

Joan Massich[a] and Guillaume Lemaître[a,b] and Joan Martí[b] and Fabrice Mériaudeau[a]

[a]LE2I-UMR CNRS 6306, Université de Bourgogne, 12 rue de la Fonderie, 71200 Le Creusot, France;
[b]ViCOROB, Universitat de Girona, Campus Montilivi, Edifici P4, 17071 Girona, Spain

ABSTRACT

This document shows the desired format and appearance of a manuscript prepared for the Proceedings of the SPIE. It contains general formatting instructions and hints about how to use LaTeX. The LaTeX source file that produced this document, `article.tex` (Version 3.3), provides a template, used in conjunction with `spie.cls` (Version 3.3).

Keywords: Breast Ultra-Sound, Machine-Learning based Segmentation

1. INTRODUCTION

Breast cancer is the second most common cancer (1.4 million cases per year, 10.9% of diagnosed cancers) after lung cancer, followed by colorectal, stomach, prostate and liver cancers.[1] 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 death among females in both western countries and economically developing countries.[2]

Medical imaging plays an important role in breast cancer mortality reduction, contributing to its early detection through screening, diagnosis, image-guided biopsy, treatment follow-up and suchlike procedures.[3] Although Digital Mammography (DM) remains the reference imaging modality for breast cancer screening, Ultra-Sound (US) imaging has proven to be a successful adjunct image modality.[3,4] The main advantage of US imaging, opposed to other image modalities, is its discriminative power for visually differentiate solid lesions as benign and malignant.[5] US screening contributes to reduce the amount of unnecessary biopsies,[6] which is estimated to be between $65 \sim 85\%$ of the prescribed biopsies,[7] in favour of a less traumatic short-term US screening follow-up.[8]

For all these reasons, there is a growing interest in the medical community to incorporate US screening as part of the standard procedure,[9] which encourages the development of Computer Aided Diagnose (CAD) systems applied to BUS images.

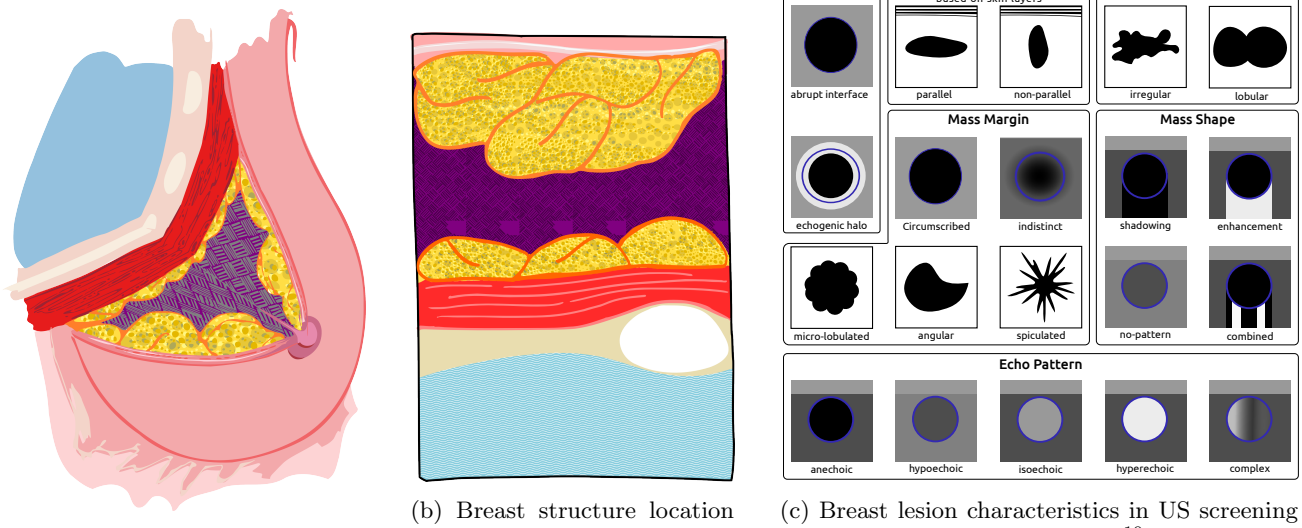
Figure 1 offers a compact, brief and visual idea of breast's structure, their render in a 2D BUS image, and which markers are recommended to study in order to produce a diagnosis.[9] All these markers either describe the lesion's delineation, or describe the relation between the lesion and the surrounding tissue, which also requires a delineation of the lesion to differentiate between the two. *

When radiologists read an image to produce a diagnosis by analysing the markers, there is no need for an explicit delineation of the tissues or lesions present in the image since this is intrinsic to the reading process. However, developing accurate segmentation methodologies breast lesions and structures is crucial for developing CAD systems.

This article proposes a highly modular and flexible framework for segmenting lesions and tissues present in BUS images. The proposal takes advantage of an energy-based strategy to perform segmentations based on discrete optimization using super-pixels and a set of novel features analogous to the elements present in fig. 1.

Further author information: (Send correspondence to J.M.)
J.M.: E-mail: joan.massich@u-bourgogne.fr

*More details regarding the visual cues used by radiologist would be present in the final manuscript as a building block for feature extraction



(a) Breast structure

(b) Breast structure location when USscreening

(c) Breast lesion characteristics in US screening influencing clinical management[10]

Figure 1: Visual reference of breast structures and visual cues used for standard BUS image assessment and diagnosis.

2. SEGMENTATION METHODOLOGY DESCRIPTION

Optimization methodologies offer a standardized manner to approach segmentation by minimizing an application-driven cost function.[11] Figure 2 illustrates a generic representation of the segmentation strategy here adopted to delineate breast tissues or lesions in US images. The overall segmentation can be seen as a three steps strategy: (1) a mapping or encoding of the image into a discrete set of elements $\mathcal{S}$, (2) the optimization stage which is formulated as *metric labelling* problem, and (3) re-mapping the labels obtained from the previous stage to produce the final delineation.

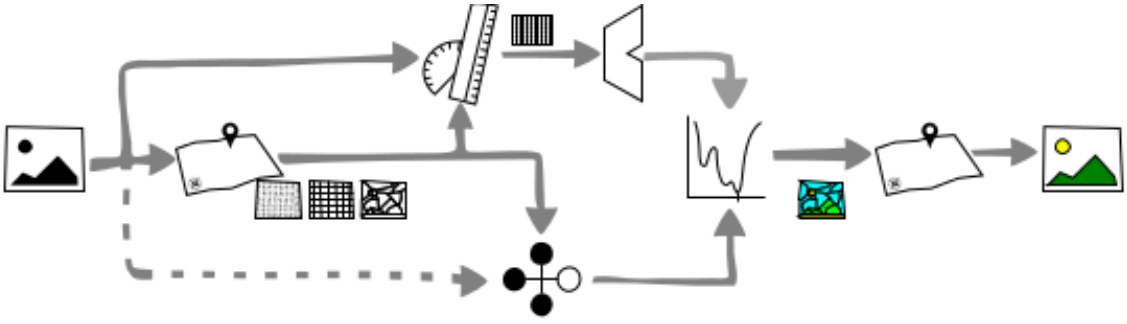
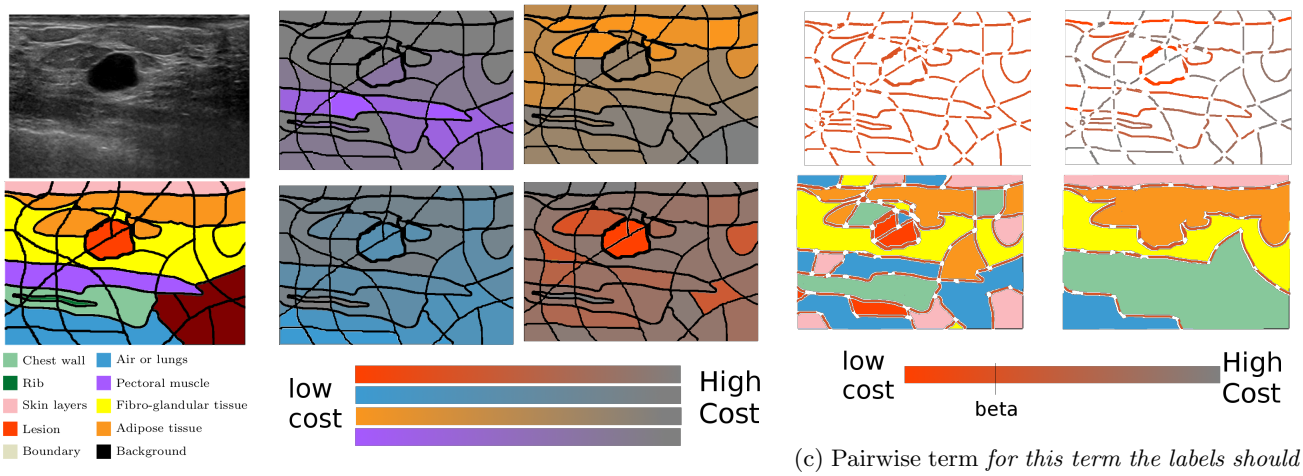


Figure 2: Conceptual block representation of the segmentation methodology

In order to formulate the segmentation like 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 $\mathcal{L}$ (i.e. $l \in \{\text{lesion}, \overline{\text{lesion}}\}$ or $l \in \{\text{lungs}, \text{fat}, \cdots, \text{lesion}\}$). Let $\mathcal{W}$ be all the possible labelling configurations of the set $\mathcal{S}$ given $\mathcal{L}$; and, let $U(\cdot)$ be a cost function encoding how good is a labelling configuration $\omega \in \mathcal{W}$ based on the appearance of the elements in $\mathcal{S}$, their inner relation and some desig constraints. Then, the desired segmentation $\hat{\omega}$ corresponds to the labelling configuration that minimize this cost function, as described in eq. (1).



(a) Problem definition (b) Data term (c) Pairwise term *for this term the labels should be patterns and the cost be the only colour*

Figure 3: Methodology terms interpretation

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

The nature of $U(\cdot)$ and $\mathcal{W}$ dictates which minimization strategies should be applied to find ω sine not every strategy is suitable or desirable.

Equation (2) determines this cost function as the combination of two independent costs that need to be simultaneously minimized as a whole. 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 as the *smoothing* term. Both terms are shaped by $\mathcal{S}$ and evaluated in $\mathcal{W}$.

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

Figure 3 illustrates the terms and elements relating the framework's outline, presented in fig. 2, with its formulation in eq. (2). The problem of delineating the tissues present in BUS image has been used here for this illustrative purposes.

In general, $\mathcal{S}$ can be any discrete set representing the image (i.e. pixels, overlapping or non overlapping windows, etc.). For this work $\mathcal{S}$ is chosen to be a super-pixels representation of the image,[12] though. Figure 3a shows a BUS image example and a its associated super-pixels representation $\mathcal{S}$ coloured according to the image's Ground Truth (GT).

Bear in mind that given an unseen BUS image, the ultimate goal is to represent the image as a set of super-pixels and infer the appropriated labelling for each of them. To do so using the strategy here proposed, it requires to define: a data term, a pairwise term, and a proper minimization methodology.

2.1 Data term

Given a label configuration $\omega \in \mathcal{W}$, the data term penalizes the labelling of a particular image element or site $(\omega_s = l)$ based on the data associated to s. In this manner $D_s(\omega_s = l_{\checkmark}) << D_s(\omega_s = l_{\times})$. To perceive the effect or behaviour of this data term, fig. 3b shows some labelling configurations $\omega^{'}$ where all the sites share the same label, $\omega' \in \{\omega_s = l,\ \forall s \in \mathcal{S}\}$

Designing $D(\cdot)$ that accomplish the desired behaviour by defining an obscure heuristic, is rather complicated task to achieve out of the box. Therefore, an easier and cleaner approach is to take advantage of Machine Learning (ML) techniques to design this data cost in a systematic manner based on a training stage. The idea

is to generate image or data model for each class from training samples, and let $D(\cdot)$ be a distance or goodness measure reflecting the likelihood for s to to belong to class l. Defining the data term in this manner allows for great flexibility while offering a systematic approach towards its design. $D(\cdot)$ is fully customizable through the features, through the construction of the model where several classifiers and training techniques can be applied; or through definition of the relation between the testing sample and the data models.

This type of data term is incorporated to our framework as represented at the upper side of the diagram in fig. 2. Each site s is treated as a sample and the features to describe it are extracted from the original image. For the work here reported, a Support Vector Machine (SVM) classifier is used to determine the data model during the training stage. During testing stage $D_s(\omega_s = l)$ corresponds to the distance between the testing sample and the model associated to l as the SVM classification reward.

Further discussion regarding the feature choices can be found in **??**, whereas other designing choices regarding ML are out of the scope for this work.

2.2 Pairwise or smoothing term

The pairwise term represents the cost of the assignation ω_s taking into account the labels of its neighbour sites, ω_r, $r \in \mathcal{N}_s$. This term models a Markov Random Fields (MRFs) or a Conditional Random Fields (CRFs). The typical form of this term, given in eq. (3), is called homogenization which acts as a regularization factor favouring configurations that have coherent labelling.

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

Figure 3c offers a visual interpretation of this cost. If the resulting segmentation associated to the current labelling configuration ω has a boundary segment, this 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 since some sites will flip their labelling if the cost of producing and edge is larger than the cost of adopting the neighbour's label.

More sophisticated smoothing terms where boundaries have different penalization based not only on site relations in $\mathcal{S}$ but also based on image information would (see fig. 3c) be found in the final version of the document in **??**.

2.3 Searching the best labelling configuration

Once defined $U(\omega)$ so that the cost for a particular labelling configuration ω can be computed, the problem of finding $\hat{\omega}$ corresponding to the global minimum of the space $\mathcal{W}$ of all possible labelling configurations needs to be faced.

This problem falls into the category of NP-hard problems. More over, due to limitations in building $U(\cdot)$ such as noise, training policies, etc. there are no guarantees that the global minimum $\hat{\omega}$ corresponds to the true labelling.

Nevertheless, there is a large body of literature proposing methodologies to find suboptimal solutions to the problem trading-off between time of convergence and accuracy of the solution reached. Szeliski et al.[13] conducted an exhaustive review in terms of solution quality and runtime of the most common energy minimization algorithms used in Computer Vision (CV), such as Iterated Conditional Modes (ICM), Simulate Anealing (SA) or Graph-Cuts (GC).

The minimization strategy used for this work is GC. This technique was initially introduced to solve CV applications by Boykov et al.[14] Soon after its introduction, it become the minimization technique of choice for CV problems. Since, when GC is applicable, it allows to rapidly find a strong local minima guaranteeing that no other minimum with with lower energy can be found.[15] GC is applicable if, and only if, the pairwise term favours coherent labelling configurations and penalizes labelling configurations where neighbours labels differs. Such is our case, given eq. (3).

3. RESULTS

A lack of public data in order to perform fair comparison between methodologies is a recurrent problem in medical imaging. BUS images are no exception.[16] Therefore segmentation can only be compared for the lesions segmentation case, and only against the results reported in the bibliography.

Figure 4 compares our segmentation strategy against the state-of-the-art assuming the following limitations: (a) the other methodologies' performance has been collected from the literature. (b) Since all the segmentation results are reported using different metrics, those have been translated to () as a common evaluation metric. (c) Evaluation datasets not only differ in image acquisition but also their sizes suffer a large variation.

Every methodology from the literature is represented in a uncommon dataset, different levels of user assistance to perform the segmentation,

A fair comparison with other segmentation methodologies is impossible. Trying to assess a novel segmentation strategy for delineating tissues in BUS images is a challenge by itself. Cheng:2009p10580, However, the lack of a common dataset to test all the methodologies with makes impossible a fair comparison between methods. This is easily observed when comparing the segmentation results reported from automatic methodologies those outperform manual segmentations done by trained expert radiologists.

Pons et al.[17] analyzed the inter- and intra-observer variability of manual segmentations of breast lesions in US images. In the experiment, a subset of 50 images is segmented by an expert radiologist and 5 expert biomedical engineers with deep knowledge of a breast lesion appearance in US data. The experiment reported an rate between 0.8 and 0.852 for the 6 actors. This demonstrates the large variability between GT delineations; a fact that needs to be taken into account in order to draw proper conclusions about the performance of a segmentation methodology. However, having multiple GT delineations to better assess the segmentations performance is not always possible. When possible, several strategies have been used to incorporate such information.

lthough the new segmentation technique achieves results are comparable only to some of the results published in the bibliography (see fig. **??**), the proposed methodology has large room for improvement compared to our previous proposal which was pretty tuned up already (see section **??**). In this last affirmation it needs to be taken into account such methodologies against the rest of the methodologies in the literature is unfeasible due to the lacking common dataset.

In order to facilitate the comparison between the proposed methodology and the methodologies reviewed in chapter **??**, despite the bias of being tested in different datasets, the figure **??**a is replicated here in figure **??** this time showing an extra ring in black at 0.623 representing the best performance in fig. **??**a, so that it can be easily compared to the previously reviewed methodologies.

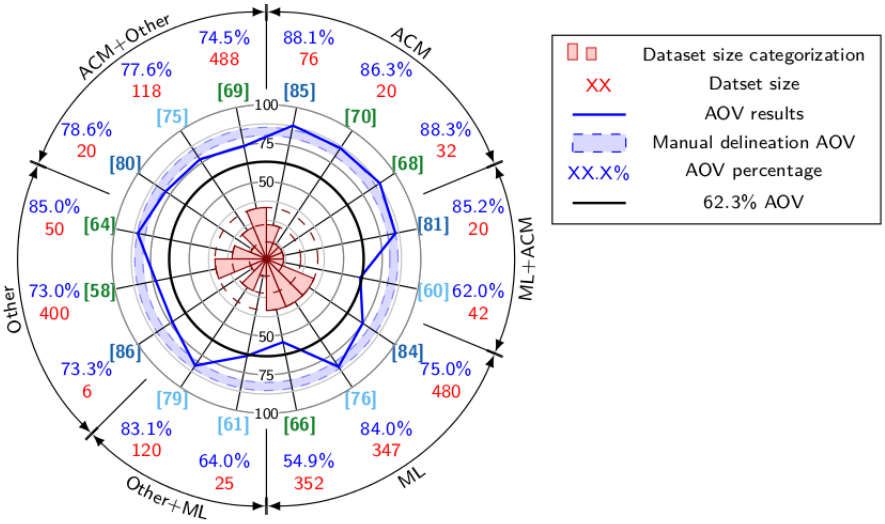


Figure 4: Quantitative AOV results

ACKNOWLEDGMENTS

This unnumbered section is used to identify those who have aided the authors in understanding or accomplishing the work presented and to acknowledge sources of funding.

REFERENCES

1. J. Ferlay, H.-R. Shin, F. Bray, D. Forman, C. Mathers, and D. M. Parkin, "Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008," *International Journal of Cancer* **127**(12), pp. 2893–2917, 2010.
2. A. Jemal, F. Bray, M. M. Center, J. Ferlay, E. Ward, and D. Forman, "Global cancer statistics," *CA: A Cancer Journal for Clinicians* **61**(2), pp. 69–90, 2011.
3. R. A. Smith, D. Saslow, K. A. Sawyer, W. Burke, M. E. Costanza, W. Evans, R. S. Foster, E. Hendrick, H. J. Eyre, and S. Sener, "American cancer society guidelines for breast cancer screening: update 2003," *CA: a cancer journal for clinicians* **53**(3), pp. 141–169, 2003.
4. W. A. Berg, L. Gutierrez, M. S. NessAiver, W. B. Carter, M. Bhargavan, R. S. Lewis, and O. B. Ioffe, "Diagnostic accuracy of mammography, clinical examination, US, and MR imaging in preoperative assessment of breast cancer," *Radiology* **233**(3), pp. 830–849, 2004.
5. 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* **196**(1), pp. 123–34, 1995.
6. S. Ciatto, M. Rosselli del Turco, S. Catarzi, D. Morrone, *et al.*, "The contribution of ultrasonography to the differential diagnosis of breast cancer.," *Neoplasma* **41**(6), p. 341, 1994.
7. Y. Yuan, M. L. Giger, H. Li, N. Bhooshan, and C. A. Sennett, "Multimodality computer-aided breast cancer diagnosis with ffdm and dce-mri.," *Academic radiology* **17**(9), p. 1158, 2010.
8. P. B. Gordon and S. L. Goldenberg, "Malignant breast masses detected only by ultrasound. A retrospective review," *Cancer* **76**(4), pp. 626–630, 1995.
9. 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.
10. S. Raza, A. L. Goldkamp, S. A. Chikarmane, and R. L. B irdwell, "US of breast masses categorized as BI-RADS 3, 4, and 5: Pictor ial review of factors influencing clinical management," *Radiographics* **30**(5), pp. 1199–1213, 2010.
11. 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* **72**(2), pp. 195–215, 2007.
12. R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," 2012.
13. R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, and C. Rother, "A comparative study of energy minimization methods for markov random fields with smoothness-based priors," *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **30**(6), pp. 1068–1080, 2008.
14. Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **23**(11), pp. 1222–1239, 2001.
15. A. Delong, A. Osokin, H. N. Isack, and Y. Boykov, "Fast approximate energy minimization with label costs," *International Journal of Computer Vision* **96**(1), pp. 1–27, 2012.
16. H. D. Cheng, J. Shan, W. Ju, Y. Guo, and L. Zhang, "Automated breast cancer detection and classification using ultrasound images: A survey," *Pattern Recognition* **43**(1), pp. 299–317, 2009.
17. G. Pons, J. Martí, R. Martí, S. Ganau, J. Vilanova, and J. Noble, "Evaluating lesion segmentation in breast ultrasound images related to lesion typology," *Journal of Ultrasound in Medicine* , 2013.