Breast Lesion Segmentation in Ultrasound Images

Deanna Hood, Gerard Bahi and Chalikonda Prabu Kumar

***Abstract*— This paper outlines a segmentation algorithm for breast lesions in ultrasound images. Supervised machine learning has been used to train a model of the lesions vs background in the images using features which take attributes of over-segmented regions of the images such as texture, grayscale intensity, and spatial information into account. The spatial information is further used in post-processing to reduce the likelihood of classifying segments of the image in areas far from the determined likely lesion center. The results obtained with a training database of only 14 images are a mean Dice similarity coefficient in the leave-one-out validation of 0.59, with half of the Dice similarity coefficients greater than 0.57. With an increase in the database size to 20 images, the mean and median Dice similarity coefficients achieved respectively rise to 0.64 and 0.68.**

I. INTRODUCTION

Lesions in ultrasound images can be challenging to segment for reasons such as low contrast, speckle noise, poorly defined edges [1]. The aim of this paper is to outline an algorithm which can successfully overcome these challenges to segment the lesions in ultrasounds of the breast, which is effective: maximising evaluation measures such as the Dice similarity coefficient; fast, with reasonable computational time; and automatic, with no intervention needed to segment a variety of images. The structure of the paper is as follows. Section 2 describes the design of the algorithm, with Section 3 the implementation. Section 4 evaluates the results of the segmentation and Sections 5 and 6 provide future work and concluding remarks.

II. ALGORITHM

A. Overview

The use of supervised machine learning has long been prevalent in lesion segmentation [2] to train a classifier using a database of training images with the ground truth provided, so that the segmentation of lesions in test images with no such ground truth may be predicted. During the training phase, features are extracted from the training images and provided, with the information of their ground truth, to the training model of the classifier in order to make a relationship between the features and their class. The testing of an image is performed by extracting the same features from the image and passing them to the trained classifier, which then provides a prediction of the class, and/or a probability of the matching to a class. In the context of segmentation of lesion vs background in an image, the classification is performed on the different pixels/regions in the image. Correspondingly, the ground truth in the training images also varies over the regions on the image. As such, the features used for training/classification must be extracted from the corresponding regions of the image. It seems reasonable to take these regions of the image as the superpixels generated in the pre-processing step as opposed to arbitrarily-shaped regions, such that a) the segmentation labels assigned are more likely to match the border of the lesion, and b) the features extracted from the regions are over a region which is approximately uniform in its features.

The segmentation algorithm developed has the structure outlined in Figure 1, where the images are first over-segmented into images formed with superpixels, each made of a cluster of normal pixels, on which a variety of features are extracted, and are used to train a classifier if the superpixels are from training set images, or used for classification if they are from images to be segmented. The results of the classification step are refined with some post-processing, before the final segmentation is determined. Finally, if the ground truth information is available, evaluation on the performance of the obtained segmentation is calculated. Each of the steps are outlined in further detail in the following sections.

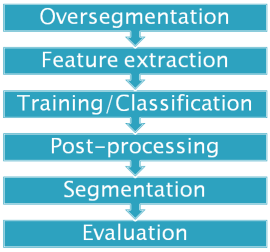

Fig. 1. Pipeline of the lesion segmentation algorithm proposed.

B. Pre-processing

Ultrasound images are known to suffer from speckle noise, which can be overcome with a variety of approaches [1]. This may be appropriate for algorithms with a high reliance on the intensity of pixels, such as region growing algorithms. For classification algorithms, however, the speckle noise observed in the parts of the image with and without the lesion can different and therefore leveraged in the classification process. As such, pre-processing to remove the speckle noise from the ultrasound images was not implemented in this paper.

The use of superpixels has been seen in the literature for over a decade [3], where clusters of pixels with similar spatial and colour information are grouped into over-segmented regions of an image, referred to as superpixels. Superpixels have been used for a variety of uses in computer vision,

including the top-performing submissions to the multi-class object segmentation PASCAL VOC Challenge, because of their ability to capture image redundancy and reduce the required complexity of the image processing tasks which follow, and provide appropriate regions for extracting local features from throughout the image [4]. Because of these advantages, the use of superpixels has been implemented for this paper as a form of pre-processing of the images.

Two of the superpixel algorithms which were investigated were SLIC and Quick Shift. SLIC is known to produce superpixels of a regularized size and shape, which - while showing impressive performance in natural images such as those in the PASCAL VOC Challenge - was not found to be particularly appropriate for use in ultrasound images. Quick Shift, however, does not have these constraints and moreover allows the user to assign more weight to the spatial or colour information during the clustering, as necessary. With emphasis placed on the spatial information, the Quick Shift algorithm was able to adapt more appropriately to the shape of the tissues in the ultrasound and in particular the lesion boundaries, as seen in Figure 2. For some of the images

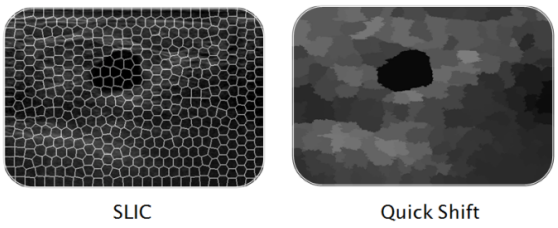

Fig. 2. Comparison of superpixel over-segmentation algorithms on an ultrasound image. Left: SLIC, with regularised superpixel sizes and shapes. Right: Quick Shift, with the mean of each superpixel shown.

to be segmented, the Quick Shift superpixels produced a result which could have simply been thresholded directly to obtain the final segmentation of the lesion. However, in order to appropriately segment all of the test images, machine learning of a classifier was determined as appropriate, in order to predict each superpixels probability of being a lesion.

C. Feature extraction

Features extracted from each superpixel in the training set of images were used with their ground truth information to train the classifier on the characteristics of lesion superpixels. The same features were extracted from superpixels in the images to be tested, to be passed to the classifier for an prediction of their probability of being a lesion. A variety of features were extracted from each superpixel, chosen to capture the range of information contained within each superpixel.

The final features used were:

1) Intensity information: Intensity information contained in each superpixel is captured by extracting features from the histogram such as the mean, median, and standard deviation of the superpixels values. Some of the features which were tested but not used in the final implementation as they did not improve the results are the minimum and maximum intensity in the superpixel.

The histogram information of each superpixel is captured by calculating the Chi-squared distance of the superpixels normalized histogram to each of 50 (empirical value) representative histograms. The representative histograms were determined in pre-training by clustering a matrix of Chi-squared distances of each superpixels histogram to every superpixel in the training sets histogram, with the representative histogram calculated as the mean histogram of the clusters. This replaced the use of the superpixel's normalized histogram as a feature.

The SIFT descriptor of the centroid of each superpixel is calculated by performing dense SIFT feature extraction on a small window around the centroid, with a step size that forced only the center point to be selected. This method was used to ensure that each superpixel had the same number of SIFT descriptors calculated for them, such that the feature vectors were all the same length as required for the classifier.

2) Spatial information: Spatial information of the superpixels is captured by calculating euclidean-distance of superpixels normalised centroid to the top center of the image (0.5,0), normalized between 0 and 1. Originally the distance to the center of the image was used, however this gave equivalent values for superpixels in the top and bottom of the images, which was undesired. The use of area of each superpixel, normalized by the image size, was tested but not used as a feature in the final implementation.

3) Texture information: The texture information of the superpixels is captured by calculating the contrast, correlation, homogeneity and energy statistics of 15-level gray-level co-occurrence matrix (GLCM). The GLCM is calculated on the image region of the bounding box of the superpixel, with the intensity value of the pixels in this region but not contained in the superpixel set to the default, unlikely value, of 1. Because of the 1s in the calculation of the GLCM, the last row and column of the GLCM matrix, where the artefacts of the 1s occur, were ignored for the calculation of the GLCM statistics. Different combinations of the offset (angle and distance) used between pixels for the calculation of the GLCM have been tested, before finally settling on distances of 1 to 6 for each of the angles of 0, 45, 90 and 135 degrees.

D. Classification

Different classifiers were trained with supervised machine learning, passing the binary (lesion vs background) ground truth of the superpixels as the classes to be learned. Some of the classifiers tested were K-nearest neighbours, support vector machine (with best convergence obtained with Radial Basis Function kernel), and random forests. The best performance was obtained with random forests, one of the leading ensemble learning techniques which has proven to be competitive with SVM for classification [5]. In addition, the random forests classification algorithm also has the advantages of being able to be trained in independent parts with different subsets of the training data with the results then

concatenated for the final classification tree, a feature which may prove useful when expanding the database of training images; and it has the advantage of providing a continuous probability output for each superpixel being classified as in either the lesion or background class, as opposed to a discrete classification output by other classifiers, which provides more freedom in the post-processing step which follows.

E. Post-processing

The output of the classification step yields a continuous probability between 0 and 1 for each superpixel belonging to the lesion class. It was observed, however, that there were often superpixels being assigned a high lesion probability because they had one or more similar properties to that of the true lesions. For example, pixels with a similar texture and intensity of the lesions such as those which in the region of the lungs or lesions' shadow in the ultrasound can cause significant misclassification. This problem was

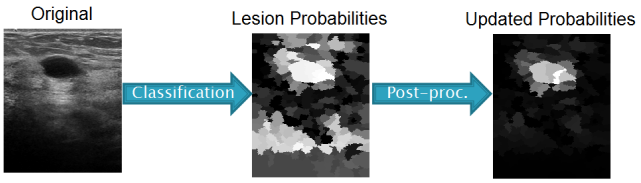

Fig. 3. Aim of the post-processing step: to make the lesion probability map easier to segment.

mitigated with the use of the spatial information of the superpixels, and more specifically their distance to the likely lesion center, which is determined across all of the training images. This was motivated by the observation that the lesion is usually in the top center of the ultrasound image. The probabilities of superpixels located far from this center are penalised by being scaled by

$$min[e^{\frac{-|x-x_0|^2}{2\sigma_x^2}}, e^{\frac{-|y-y_0|^2}{2\sigma_y^2}}]$$

where [x,y] is the superpixels centroid, and $[x_0, y_0]$ is the likely lesion center. The parameters σ_x and σ_y can be interpreted as specifying at what values of $|x - x_0|$ and $|y - y_0|$, respectively, the scaling begins to have a significant effect. A visualisation of the x- and y-scaling factors may be seen in the elliptical bell-curve with [σ_x, σ_y]=[0.3,0.25] and $[x_0, y_0]$=[0.5,0.3] in Figure 4. The minimum function was introduced to combine the x- and y-scaling factors to enforce that superpixels in the bottom center are significantly penalised, according to their large y-distance from the likely lesion center.

F. Segmentation

Following the post-processing step, the updated probability map of each image, containing the probability of each superpixel to belong to the lesion class, is able to be thresholded to determine the final binary lesion segmentation. Due to the variety in appropriate thresholds for each image, rather than fix a threshold to be used for all images, it

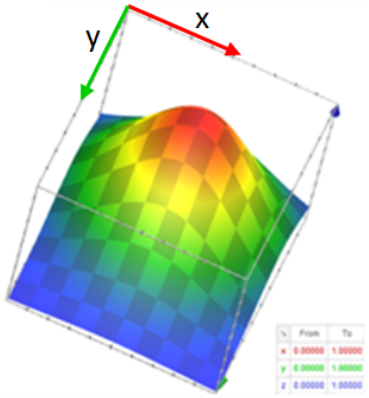

Fig. 4. Elliptical bell curve used to extract the penalties of superpixels with respect to their distance to the likely lesion center, with [σ_x, σ_y]=[0.3,0.25] and $[x_0, y_0]$=[0.5,0.3]. (Image source: Google Search plotting utilities).

was determined appropriate to instead fix parameters in an algebraic expression to determine each image's threshold based on some of the characteristics of the image. Namely, the parameters a, b, c and d such that

$$thresh_i = min(max(probMap_i) - a \cdot std(probMap_i), b) \cdot c + d$$

G. Limitations

The algorithm which has been described, as with all algorithms, has its limitations. This section highlights some of the most relevant limitations of the algorithm for the reader's consideration.

One of the advantages of using the random forest classifier is that additional training data can be added to the system by simply appending the classification model trained with the additional data to the existing model, without having to re-compute the model for all of the available data. However, the average histogram of the clusters of superpixels used to calculate one of the features for the classifier should be determined from all of the training superpixels, and so if at some point the information of the cluster histograms needs to be updated then also all the features need to be re-computed and the model re-trained. Depending on the increase in the database size, it may be feasible to skip the re-calculation of the clusters and therefore re-calculation of the features and therefore re-calculation of the training, and only append the result of training with the new images to the existing model.

When training the classifier with a reasonably small dataset, it is reasonable to use all of the available information for training. However, it is reasonable to expect that in future applications a much larger database would be used for training, as the performance of machine learning algorithms tend to rely on the size of the database. Therefore, it is important for the algorithm to be able to cope with large amounts of data, and so it may be feasible to reduce the redundancy in the training matrix by removing highly correlated data.

The post-processing stage described uses the likely lesion center as determined from all of the training images. If some of the ultrasound images in the training set have a different orientation to the majority of images, this functionality of

this post-processing will be reduced, leading to an increase in the number of false positive results. In addition, if the test image given is oriented contrary to the images used to determine the likely lesion center, the post-processing step will be ineffective. However, this problem is easy to overcome if the consistent orientation of the images is given as a precondition before adding any image to the database.

III. IMPLEMENTATION

A. External libraries

The vl_feat library is an award-winning opensource library which contains C-implementation with a MATLAB interface for popular computer vision algorithms including Quick Shift and SLIC superpixels, and SIFT feature extraction [6]. The MATLAB interface of these implementations were leveraged during the pre-processing and feature extraction stage of the implementation of the algorithm.

B. Parameter selection

It can be difficult to tune the parameters in an algorithm manually when there are tradeoffs to be made across a range of outputs: for example to reduce the number of false positives for one image while preserving the true positives of another. Because of this, an optimiser was used to determine the most appropriate parameters (such as the σ_x and σ_y and likely lesion center for the post-processing, and the threshold for the final segmentation of the images) to yield the optimal objective function, which can be, for example, the mean plus median of the Dice coefficients on the leave-one-out testing on the training database.

The optimiser used was Simulated Annealing, due to its ability to converge to a global optimum for the aforementioned objective function which is continuous and non-convex, as opposed to other optimisers such as Pattern Search which only converged to local optima.

IV. RESULTS

The images in Figures 5, 6 and 7 show several examples of the segmentation results achieved: the original image, the probability map of the superpixels probabilities to be classified as a lesion by the classifier, the refined probability map after the post-processing, and the final segmentation after thresholding. The red border corresponds to the boundary of the ground truth of the image. In all images it can be appreciated that the Quick Shift superpixels adapt quite well to the boundary of the lesion.

In Figure 6 it can be seen that the shadow in the ultrasound caused by the lesion gives high probabilities on the classifier output. However, after applying the post-processing, these probabilities are reduced allowing a good final segmentation result from to be obtained from the thresholding.

In Figure 7 it can be seen that the false positives at the output of the classifier are created by the presence of the lungs in the ultrasound. The impact of the texture features may be appreciated in this figure, as the superpixels at the very bottom of the image are void of texture and less likely to be interpreted as lesion segments than those with some speckle noise. These high probabilities are reduced drastically by the post-processing, providing a probability map which is less difficult to segment with thresholding.

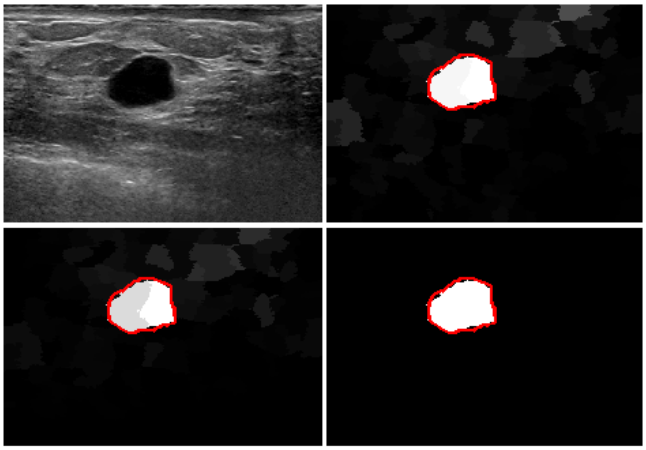

Fig. 5. Results of lesion 000002.png. Top-left: Original image. Top-right: Probability map given by the classifier. Bottom-left: Probability map after post-processing. Bottom-right: Final segmentation with thresholding.

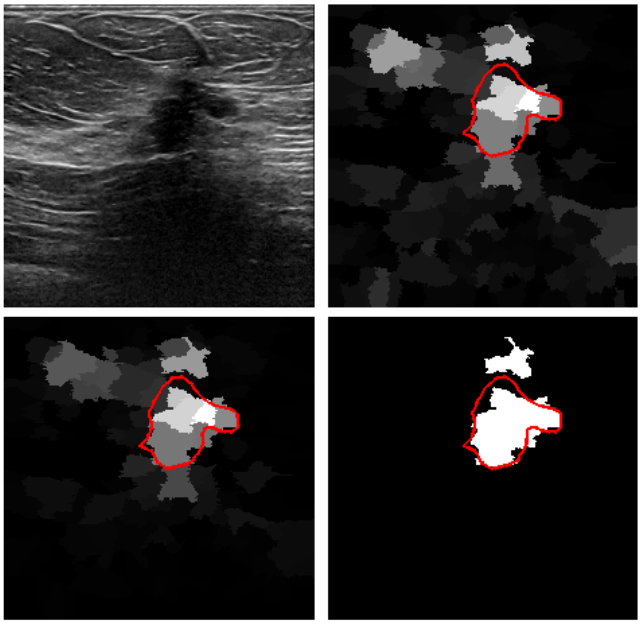

Fig. 6. Results of lesion 000031.png. Top-left: Original image. Top-right: Probability map given by the classifier. Bottom-left: Probability map after post-processing. Bottom-right: Final segmentation with thresholding.

The concept of the post-processing making the final segmentation less difficult may be further illustrated with the use of the ROC curves for the image, an example of which is shown in Figure 8. Not only can it be observed that the ROC curve of the image after post-processing has an increased area underneath over the image before post-processing, but it is evident that the number of thresholds which are close to the optimal position of the ROC curve of (0,1) has increased. This point is further demonstrated in Figure 9, and it can be concluded that after the post-processing, achieving a

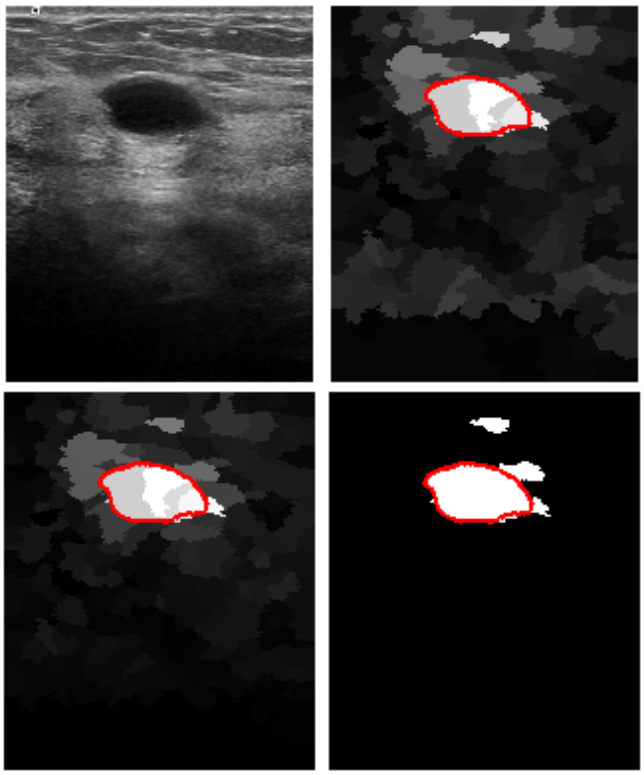

Fig. 7. Results of lesion 000031.png. Top-left: Original image. Top-right: Probability map given by the classifier. Bottom-left: Probability map after post-processing. Bottom-right: Final segmentation with thresholding.

successful final segmentation is not only more likely, but less dependent on the value of the threshold. This makes the algorithm more robust and moreover makes the process of finding a threshold which works for all the images more feasible.

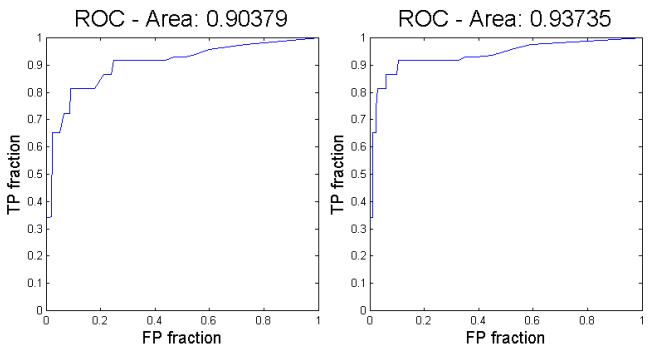

Fig. 8. ROC curves of file 00061.png. The left image represents the ROC cuve of the output probability without using the post processing. The right image represents the ROC curve of the final probability map after the post processing.

Leave-one-out validation was performed on the database of 14 ultrasound images with ground truths available, yielding a mean Dice coefficient of 0.59. A box-and-whisker plot of the Dice coefficients obtained may be seen to the right, from which it may be observed that the majority of

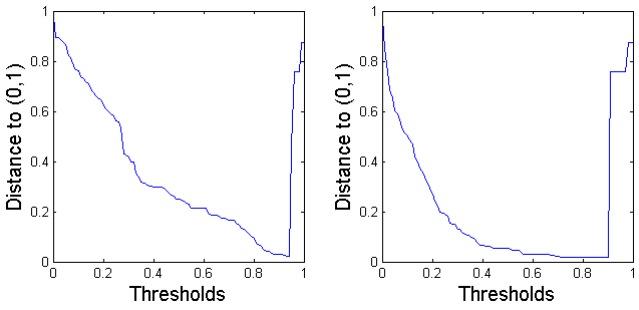

Fig. 9. The graphs display the distance of each threshold used in the ROC curve from the point (0,1) of the ROC curve. The lower the distance the more optimal the threshold is. After post-processing (right), a wider range of thresholds yield optimal results than before (left).

the images yielded a Dice coefficient of above 0.55, with a quarter of all Dice coefficients above 0.87.

The algorithm proposed in this paper is based on machine learning, and so one of the factors for obtaining good results is the database used for training the algorithm. Figure Y presents the results obtained using leave-one-out validation on a database of 20 images, where the mean and median Dice coefficients have been increased to 0.64 and 0.68 respectively. This illustrates the benefit of using machine learning for the algorithm implementation, as the results are able to be improved by having a better database, as opposed to other algorithms where performance is often fixed at a particular level. The advantage of using the random forests classifier which can have additional training concatenated to the model comes into particular use during the times of expanding the database size.

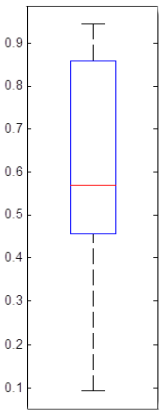

Fig. 10. Box plot of the Dice similarity coefficients obtained using leave-one-out validation on a training set of 14 images. Mean: 0.59. Median: 0.57. Max: 0.94.

Matlab parallelization which uses CPU threads was leveraged for the computation of this algorithm, and as such describing the computational time is not straightforward. The CPU of the machine used to run the algorithm was an Intel core i7 Ivy Bridge with 4 cores with hyper thread. The segmentation of 6 images took a computing time of approximately 5 minutes and highlights that once the training

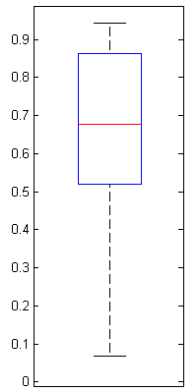

Fig. 11. Box plot of the Dice similarity coefficients obtained using leave-one-out validation on a training set of 20 images. Mean: 0.64. Median: 0.68. Max: 0.94.

of the classifier has been computed off-line, the test images can be tested with sufficient speed for on-line applications.

V. FUTURE WORK

The following recommendations are made for future work on the algorithm proposed in this paper:

- Investigate the use of more sophisticated techniques to segment the post-processed probability map.
- Use only low-correlated data for training the classifier, such that the volume of data that the trainer has to handle is lower but the subset of data will still be representative of the set. This may be particularly useful when dealing with larger datasets.
- Use an optimizer to determine the optimal feature combination, not just to improve the segmentation results but also to determine which features contribute more to successful segmentation which may give insight into yet more appropriate features.
- Investigate potential convex objective functions which may be used for parameter optimization such that better convergence can be obtained with a variety of optimizers.

VI. CONCLUSIONS

The following conclusions can be drawn from the work in this paper:

- In ultrasound images, the Quick Shift over-segmentation method has shown to perform better than SLIC, because of its reduced emphasis on regularised segment shape and size and its ability to place more emphasis when forming segments on their spatial information instead of color.
- It can be difficult to tune a set of algorithm parameters manually to get the best overall performance, however the use of objective function optimizers has proven a useful tool for determining the appropriate parameters for the algorithm automatically.
- Breast lesions in ultrasound images do not have a straightforward model of shape, texture or intensities which can be used for the segmentation process, however using an appropriate combination of these features can enable successful classification.
- False positives can occur from regions of the ultrasound image with similar features to the lesion, such as lesion shadows and lungs, however this can be overcome with the use of spatial information of the regions relative to the likely legion center, making the final segmentation more robust to the threshold selected.

VII. ACKNOWLEDGMENTS

The authors gratefully acknowledge the contribution of Joan Massich and thank him for his guidance.

REFERENCES

[1] A.A. Mahmoud, S. El Rabaie, T.E. Taha, O. Zahran, F.E.A. El-Samie, and W. Al-Nauimy. Comparative study between different denoising filters for speckle noise reduction in ultrasonic b-mode images. In *Computer Engineering Conference (ICENCO), 2012 8th International*, pages 30–36, 2012.

[2] Constantine Kotropoulos and Ioannis Pitas. Segmentation of ultrasonic images using Support Vector Machines. *Pattern Recognition Letters*, 24(4–5):715–727, 2003.

[3] D. Comaniciu and P. Meer. Mean shift: a robust approach toward feature space analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(5):603–619, 2002.

[4] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk. SLIC Superpixels Compared to State-of-the-Art Superpixel Methods. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34(11):2274–2282, 2012.

[5] Chesner Désir, Simon Bernard, Caroline Petitjean, and Laurent Heutte. A Random Forest Based Approach for One Class Classification in Medical Imaging. In Fei Wang, Dinggang Shen, Pingkun Yan, and Kenji Suzuki, editors, *Machine Learning in Medical Imaging*, volume 7588 of *Lecture Notes in Computer Science*, pages 250–257. Springer Berlin Heidelberg, 2012.

[6] A. Vedaldi and B. Fulkerson. VLFeat: An Open and Portable Library of Computer Vision Algorithms. http://www.vlfeat.org/, 2008.