A comparison of automated versus manual segmentation of breast UST transmission images to measure breast volume and sound speed

Mark Sak[1], Neb Duric[1], Peter Littrup[1], Katelyn Westerberg[1]

[1]Delphinus Medical Technologies, 45525 Grand River, Novi, MI, 48374

ABSTRACT

Ultrasound tomography (UST) is an emerging breast imaging modality that can be used to quantitatively measure breast density. However, the sound speed images that are used in this analysis must first be segmented in order to accurately parse any quantitative information. Previously, this segmentation has been done manually, but this is time consuming, especially when dealing with a large number of images that must be masked. An automated masking algorithm has been developed that applies thresholding and morphological operators to UST attenuation images to automatically create masks that separate the breast tissue from the water bath. An initial set of images was tested using this algorithm to fine tune settings and very good agreement was achieved. However, when the optimized settings were applied to a larger dataset of 286 images, the robustness of the algorithm was tested. The manual masks measured a larger volume (921 cm^3) than the automated masks (713 cm^3), but fortunately, the difference in mean sound speed was much smaller (1449 m/s versus 1448 m/s). A majority of the automated masks (72.7%) had a measured Dice similarity coefficient (DSC) of greater than 0.8 which indicates that there was good to great overlap in the volumes of tissue created by the automated method. This algorithm shows promise to be used as a tool to quickly and effectively measure breast density.

1. INTRODUCTION

Ultrasound tomography (UST) is an emerging breast imaging modality[1, 2] that can create quantitative measurements of breast density using sound speed images[3, 4]. However, in order to produce these measurements, the imaged breast tissue must first be segmented from the background water bath. Unfortunately, the sound speed of water is intermediate to the range of sound speeds of breast tissue, so a simple thresholding cannot be used to easily separate the breast tissue to be used in quantitative analysis.

Previous work[3-5] involving UST sound speed measurements relied on a manual masking method. This method required a user to define which slices in the stack corresponded to the breast tissue and then required the user to manually select points that corresponded to the boundary between the breast tissue and the water bath. An ellipse would then be fit to these points to separate the two regions. This method is time consuming, especially when dealing with a large number of breast images. Previous attempts at automating a masking procedure have produced mixed results[6].

However, UST transmission imaging produces both sound speed images and attenuation images. The acoustic attenuation of breast tissue is greater than that of the background water bath and is therefore much better suited for the creation of automated masking algorithms. Since attenuation images are created using the same transmission signals that are used to create the sound speed images, masks that are created from the attenuation images could also be applied to the sound speed images. The automated attenuation masks would therefore remove the most time consuming step from making quantitative UST breast density measurements.

2. METHODS AND MATERIALS

2.1 Manual Masking Method

Two different masking methods were compared here. The initial manual method of masking that is performed on the sound speed images was compared to the automated method that is performed on the attenuation images. The method for manual masking has been described before[5], but will be briefly reviewed here. An algorithm built in ImageJ requires the user to select points on the image that correspond to the boundary between the breast tissue and the water bath. An ellipse is then fit to the selected points and a mask is created where all voxels located outside the mask are ignored. Only the voxels inside the ellipse are considered to be a part of the breast tissue and from these voxels, the quantitative measurements could be made.

2.2 Automated Masking Algorithm

The masking algorithm was created in MATLAB using the Image Processing Toolbox. The mask was created on a slice by slice basis. Since the attenuation of the breast tissue is generally much higher than that of the background water bath, an initial mask was first estimated by applying a threshold. All values above this user defined threshold were given a value of 1 and all values below it were given a value of 0. The threshold value was kept constant for all images. The thresholding produced an intermediate image that roughly corresponded to a finished mask. A series of morphological operators with variable settings were then applied to the thresholded image to finish the mask. The image was first closed using a disk of a fixed size and then all interior holes were filled. Finally, to smooth the edges and boundary of the mask, the image was eroded and then opened using a disk with a user-defined radius. An example of the steps of the algorithm is shown in Figure 1.

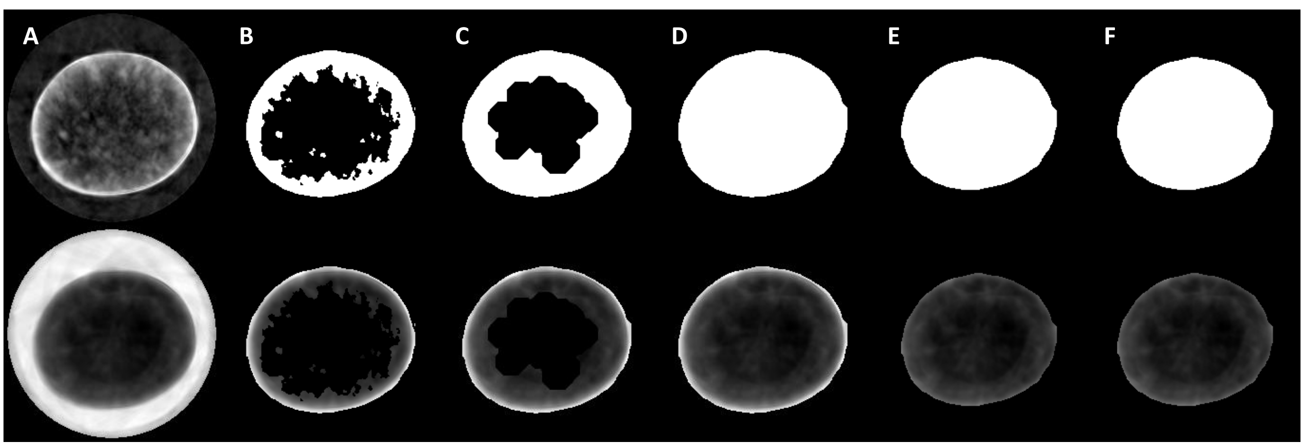

Figure 1 – The top row shows the original attenuation image and the series of steps undertaken to create the final mask. The bottom row shows the sound speed image with the mask generated at each step applied to the image. B) The initial thresholding step creates a ring near the breast boundary. C) The closing of the initial mask. D) The interior of the mask is then filled. E) The mask is then eroded using a disk of a various radius to remove traces of the background water bath. F) The final mask is created after it is opened.

The algorithm creates a mask for every slice in the stack. However, some slices can contain chest wall or the nipple which must be removed to obtain ideal quantitative measures. The algorithm cannot parse which slices contain either of these structures, but these slices were identified when masks were created manually. Therefore, even though the automated algorithm created extra masks, only the slices which corresponded to the actual breast tissue, as determined manually, were compared quantitatively. Quantitative measurements will likely need to be monitored by a user to ensure the proper regions of breast are measured. Fortunately, when compared to manually masking images, visually identifying which slices to exclude is far less time consuming.

2.3 Initial Tests to Set Parameter

To set the parameters for the algorithm, a subset of 15 UST scans were tested with both automated and manual masking. The radius of the disks used for the erosion and opening of the mask were adjusted as these operations would likely have the most effect on the final image. The automated masks that were created were then compared to the manual masks to identify which settings were closest to recreating the measurements made from the manual methods. The Dice similarity coefficient (DSC)[7] was calculated for each automated mask and quantitative measures such as total breast volume and average sound speed were then compared to the manual method as well. The DSC was calculated using the following formula:

$$DSC = \frac{2 \times V_O}{V_M + V_A}$$

where V_M is the volume of the manual mask, V_A is the volume of the automated mask and V_O is the volume of the overlapping volume between both masks. These volumes could be easily calculated using ImageJ.

2.4 Full Data Set Comparison

As part of an unrelated study[8] involving comparison of UST sound speed measurements with mammographic density measurements, a set of 286 UST sound speed and attenuation images were available. These UST images already had manual masks created. The results of the initial set mentioned above allowed for the parameters of the algorithm to be set and tested on this larger data set. The automated algorithm was run on these images and once again the same quantitative measures were compared. The larger data set allowed for a wider variety of cases to be tested to ensure the robustness of the algorithm.

3. RESULTS AND DISCUSSION

3.1 Initial Data Set Results

The initial pilot study tested the various parameter settings for the automated algorithm. In particular, two different settings were adjusted that affected the masking. The two settings were the erode radius and the open radius. Various settings for each parameter were adjusted and masks were created for each setting. Table 1 shows the results for the various parameters and compares them to the manual method using a paired t-test.

Table 1 – Results of Algorithm Parameter Testing on Pilot Data Set

Open Radius (pixels)	Erode Radius (pixels)	Mean Volume (cm^3)	p-value	Mean Sound Speed (m/s)	p-value	Mean DSC
20	5	768.3	< 0.001	1447.2	< 0.001	0.916
20	8	689.2	0.362	1443.7	0.725	0.920
20	10	626.1	0.002	1441.8	< 0.001	0.898
25	8	675.7	0.075	1442.8	0.015	0.919
25	10	611.8	< 0.001	1441.1	< 0.001	0.891
30	10	587.2	< 0.001	1440.5	< 0.001	0.870
35	10	550.8	< 0.001	1440.1	0.003	0.817
Manual Mask Results		702.5		1443.8		

These results show that the optimal settings occur when the disk radius for the open operation is set to 20 pixels and the radius for erosion is set to 8 pixels. The volume measurements for these settings (689 cm^3) are the closest to the manual measurements (703 cm^3) and show no statistically significant difference. This slight difference in volume does not give rise to a statistically significant difference in mean sound speed as the automated mean value of 1443.7 m/s is essentially identical to the mean sound speed of the manual method (1443.8 m/s). Finally, the mean DSC was the highest for any of the settings at 0.920 indicating that the automated masks measure 92% of the same volume as the manual masks. These settings were then applied and used on the larger data set.

3.2 Full Data Set Results

The optimized automated algorithm settings were then applied to the larger 286 scan data set. The mean DSC along with the mean volume and mean sound speed measured for both masking methods are shown below in Table 2. A paired t-test was also performed and the p-value is listed as well.

Table 2 – Quantitative Measurement Comparison between Automated and Manual Masks for Large Data Set

	Volume (cm^3)	Mean Sound Speed (m/s)
Manual Masks	921.4	1449.1
Automated Masks	713.2	1448.1
p-value	< 0.001	0.004
Average DSC	0.818	

These results shows that for the larger dataset, there was more of a noticeably difference in the masks. The volume measured by each method is noticeably different by almost 200 cm^3 which leads to a lower average DSC of 0.818. However, this noticeable volume difference only leads to a small, but statistically significant difference in the mean sound speed of only 1 m/s.

While a lower average DSC value seems to indicate a relatively poorer performance than for the pilot set, it is driven by a small percentage of poor outliers. Table 3 shows the breakdown of all 286 data sets grouped by their DSC value. For each range of DSC values, the mean volume and sound speed for each masking method are also shown.

Table 3 – Quantitative Measurements Distributed According to DSC Value

DSC Range	Count	Manual Volume (cm3)	Automated Volume (cm^3)	Manual Mean Sound Speed (m/s)	Automated Mean Sound Speed (m/s)
< 0.4	16	1025.4	135.1	1446.9	1447.1
0.4-0.5	12	1007.2	363.0	1444.6	1447.1
0.5-0.6	5	1203.4	463.8	1445.4	1442.6
0.6-0.7	19	1139.1	563.1	1444.4	1441.3
0.7-0.8	26	784.8	620.7	1452.9	1456.4
0.8-0.9	40	543.1	439.5	1454.6	1454.0
0.9-1.0	168	983.6	897.1	1448.3	1446.5

These results show that a majority of the scans had good to great overlap, with 208 scans (72.7%) of scans having a DSC greater than 0.8 and 168 scans (58.7%) having a DSC greater than 0.9. As the DSC decreases, the differences in the mean volumes between the two methods increases. For the scans with the lowest DSC values, the automated volumes tended to be much smaller than the manual masking methods.

However, despite the large differences in the measured volumes, the mean sound speed values were much less sensitive to changes in the masking method. The smaller volumes measured by the automated method are a subset of the larger volumes measured manually. This result shows that the mean sound speed of the breast can be sampled somewhat accurately with even a small representative volume of breast tissue. It may not be necessary to perfectly segment the breast tissue to be able to estimate mean sound speed and breast density. Figure 2 shows plots of the automated volume versus the manual volume for every scan as well as the automated sound speed versus the manual sound speed. It shows that there are many scans that had excellent automated segmentation. However, although there is some noise in the volume plot, the noise is much less pronounced for the sound speed plot. Once again, this shows that the sound speed is much less sensitive to masking than the measured volume. There are far fewer outliers in the sound speed plot and the Spearman correlation coefficient is much higher for sound speed than for volume.

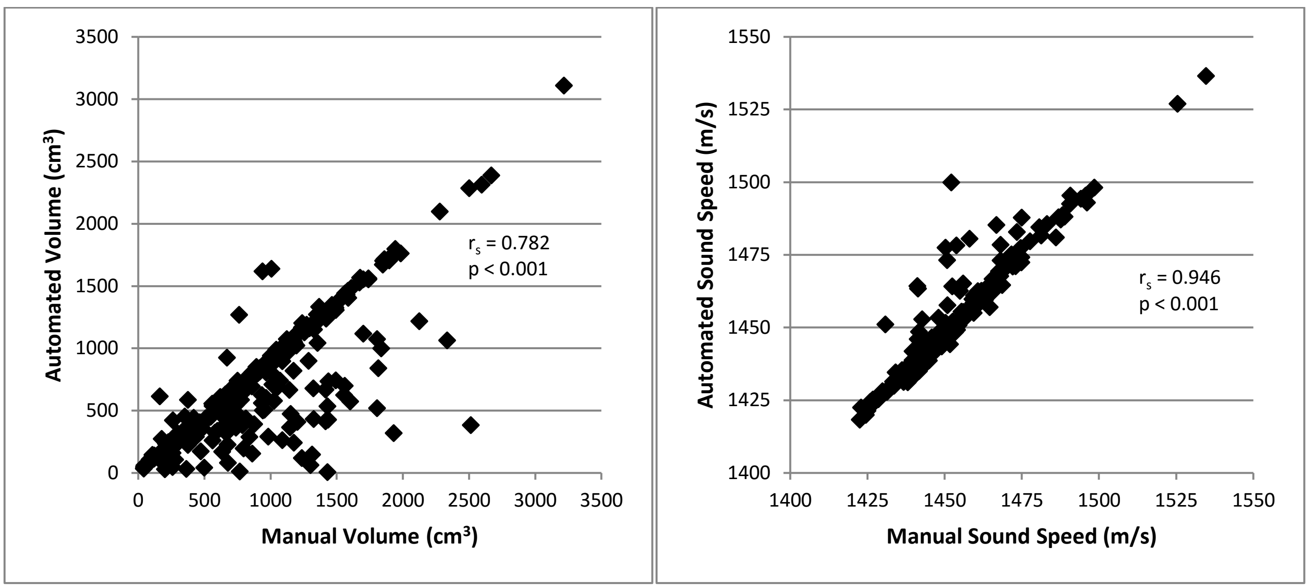

Figure 2 – (Left) Plot of the measured volume of the automated masks versus the volume of the manual masks. (Right) Plot of the mean sound speed measured using the automated masks versus the mean sound speed measured using the manual masks. The noise in the sound speed plot is lower than for the volume plot indicating that measured sound speed is less sensitive to changes in volume.

3.3 Mask Outlier Analysis

For most slices, the automated masking algorithm behaves well and produces a mask that closely mimics the manual mask. However, for some slices the automated masks start to drastically differ from the manual masks. This usually occurs in slices near the chest wall and especially when the breast either fills the field of view or at least has portions that touch the ring transducer. For these slices the algorithm had difficulties in either detecting any breast tissue or accurately defining the breast boundary (Figure 3).

The most likely reason for this misalignment was that the initial threshold was set at an incorrect value to accurately segment the breast from the water bath. For most images, the attenuation of the breast tissue is much higher than the water bath and offers a clear separation between the two. However, there are some images where the breast attenuation is much lower and large regions of tissue can fall below the threshold value. This algorithm then would ignore those regions when choosing the initial mask which results in the final automated mask having a poor DSC. The reason for some images to have low attenuation isn't clear, but some image artifacts may have an effect. For some slices, especially those closer to the chest wall, the breast tissue can touch the ring transducer. When this happens, the image can become full of artifacts that cause the attenuation of the breast to decrease while simultaneously raising the attenuation of the water bath. This would cause issues with thresholding.

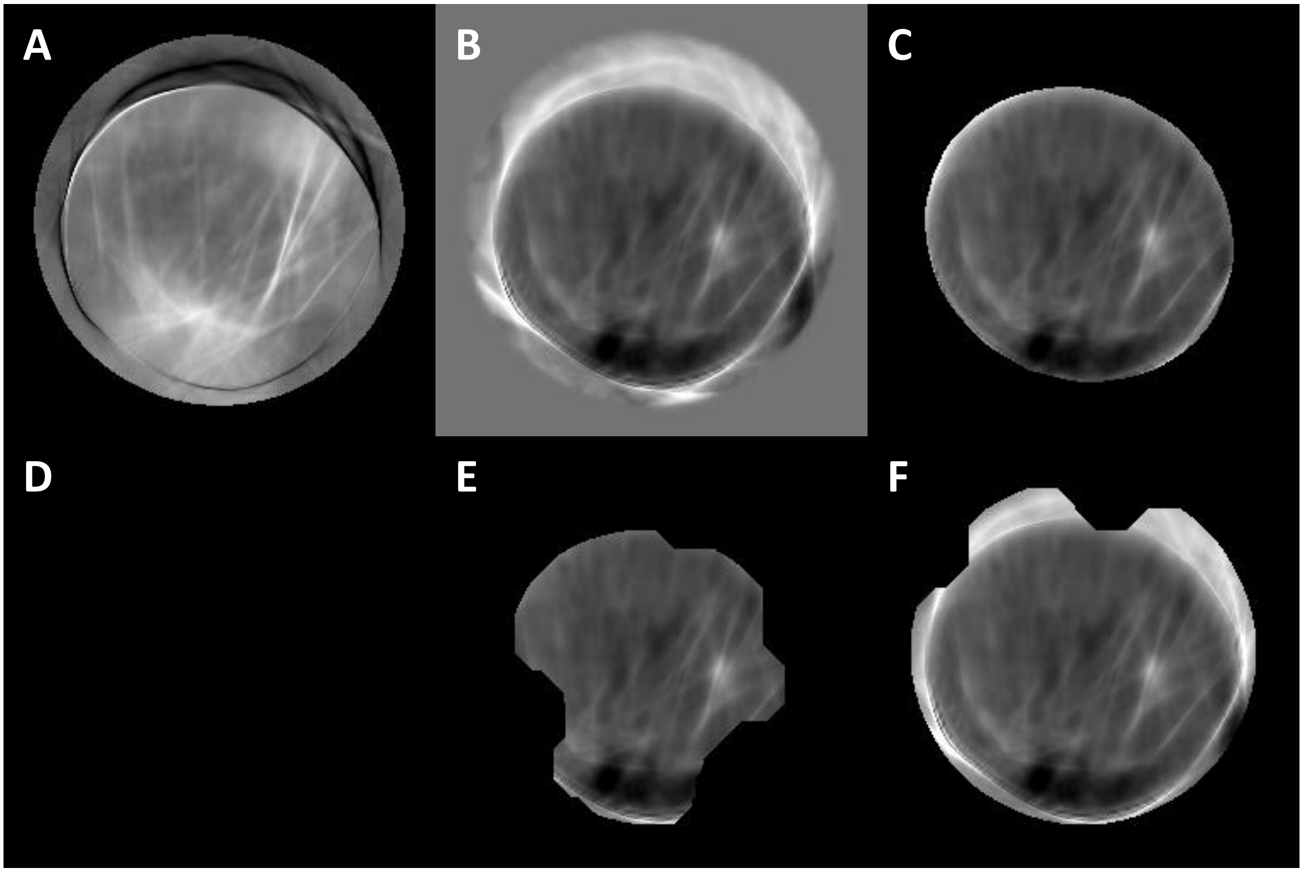

Figure 3 – An example of an image with a poor overall average DSC. A) The original attenuation image that was used to make the mask. B) The original sound speed images used to make the manual mask. C) The sound speed image with the manual mask applied. D) The automated mask created using the default attenuation threshold value. For this slice, these settings could not identify any breast tissue at all. E) The automated mask created using a slightly lower threshold value. Here, the algorithm identified some of the breast tissue, although not a good match, there is some improvement. F) The automated mask created using an even lower threshold value. Here, the breast tissue is completely identified, but the background water bath is also included. The optimal threshold setting should lie somewhere between situation E) and F) but would need to be adjusted for other images as well. The results seen in this figure were representative of the other cases that were examined.

A brief qualitative review of several cases with a low DSC was performed. The algorithm was rerun with a lower threshold value and there was some improvement in the quality of the masks (Figure 3). However, each image needed to have its threshold value selected individually. A careful selection of the initial threshold value is therefore critical to ensuring that there is a good agreement with breast volume. Fortunately, a large disagreement in volume does not necessarily lead to a difference in average sound speed. The images that were the most susceptible to these image artifacts tended to be those that were scanned early on when the UST device was still new. There were some initial issues with the quality of the scans that have since been overcome and UST images that are now being scanned are consistently of a higher quality.

The automated algorithm creates masks for every image, but not every image is required to obtain a quantitative measure of breast density. Some slices need to be removed because they contain the chest wall or the nipple. A user is still therefore required to manually input this information to accurately define the breast tissue. However, this step is far less time consuming than manually masking the images so the automated algorithm still presents dramatic time savings. Therefore, these results show that this automated method shows promise as a tool to quickly and effectively measure breast density. Additional work is still required to expand the robustness of the algorithm.

4. CONCLUSION

An algorithm to automate the creation of masks to be used with UST sound speed images was created. The algorithm applies thresholding along with various morphological operators to UST attenuation images to separate the breast tissue from the water bath. For most cases where the UST images were well behaved, the algorithm works well at effectively segmenting breast tissue. However, a minority of cases require additional parameters to be adjusted to accurately match masks that were created manually. Also, to obtain accurate quantitative measurements, a user will also still be required to remove slices that contain the chest wall or nipple. Fortunately, the measurement of mean UST sound speed is not as sensitive to variations in the volume of breast tissue measured. With additional work, the robustness of the algorithm can easily be expanded to significantly cut down on the time required to produce quantitative results of UST sound speed images.

5. REFERENCES

1. Duric, N., N. Boyd, P. Littrup, M. Sak, L. Myc, C. Li, E. West, S. Minkin, L. Martin, M. Yaffe, S. Schmidt, M. Faiz, J. Shen, O. Melnichouk, Q. Li, and T. Albrecht, *Breast density measurements with ultrasound tomography: A comparison with film and digital mammography.* Medical Physics, 2013. **40**(1): p. 013501-12.
2. Duric, N., P. Littrup, S. Schmidt, C. Li, O. Roy, L. Bey-Knight, R. Janer, D. Kunz, X. Chen, J. Goll, A. Wallen, F. Zafar, V. Allada, E. West, I. Jovanovic, K. Li, and W. Greenway. "Breast Imaging with the SoftVue Imaging system: First results," Medical Imaging 2013: Ultrasonic Imaging, Tomography, and Therapy, 8675-19 (2013).
3. Sak, M., N. Duric, P. Littrup, L. Bey-Knight, M. Krycia, M.E. Sherman, N. Boyd, and G.L. Gierach. "Comparison of breast density measurements made using ultrasound tomography and mammography," Proc SPIE 9419 Medical Imaging 2015: Ultrasonic Imaging, Tomography, and Therapy, 94190R (2015).
4. Sak, M., N. Duric, P. Littrup, M.E. Sherman, and G.L. Gierach. "Using ultrasound tomography to identify the distributions of density throughout the breast," Proc. SPIE 9790 Medical Imaging 2016: Ultrasonic Imaging, Tomography and Therapy, 979019-979019-11 (2016).
5. Sak, M.A., *The role of tissue sound speed as a surrogate marker of breast density*, 2013, Wayne State University: Ann Arbor. p. 225.
6. Balic, I., P. Goyal, O. Roy, and N. Duric. "Breast boundary detection with active contours," Proc. SPIE 9040 Medical Imaging 2014: Ultrasonic Imaging, Tomography, and Therapy, 90400D-8 (2014).
7. Dice, L.R., *Measures of the Amount of Ecologic Association Between Species.* Ecology, 1945. **26**(3): p. 297-302.
8. Sak, M., N. Duric, P. Littrup, C. Li, L. Bey-Knight, M. Sherman, N.F. Boyd, and G. Gierach. "Breast density measurements using ultrasound tomography for patients undergoing tamoxifen treatment," Medical Imaging 2013: Ultrasonic Imaging, Tomography, and Therapy, 8675 (2013).