Comparison of 3D and 2D Breast Density Estimation from Synthetic Ultrasound Tomography Images and Digital Mammograms of Anthropomorphic Software Breast Phantoms

Predrag R. Bakic, Cuiping Li*, Erik West*, Mark Sak*, Sara C. Gavenonis, Nebojsa Duric*, and Andrew D.A. Maidment
University of Pennsylvania, Department of Radiology, 3400 Spruce St., Philadelphia PA 19104
* Karmanos Cancer Institute, Wayne State University, 4100 John R. Street, 5 HWCRC, Detroit, MI 48201

Predrag.Bakic | Sara.Gavenonis | Andrew.Maidment @uphs.upenn.edu
LiC | WestE | SakM | Duric @karmanos.org

ABSTRACT

Breast density descriptors were estimated from ultrasound tomography (UST) and digital mammogram (DM) images of 46 anthropomorphic software breast phantoms. Each phantom simulated a 450 ml or 700 ml breast with volumetric percent density (PD) values between 10% and 50%. The UST based volumetric breast density (VBD) estimates were calculated by thresholding the reconstructed UST images. Percent density (PD) values from DM images were estimated interactively by a clinical breast radiologist using Cumulus software. Such obtained UST VBD and Cumulus PD estimates were compared with the ground truth VBD values available from phantoms. The UST VBD values showed a high correlation with the ground truth, as evidenced by the Pearson correlation coefficient of $r=0.93$. The Cumulus PD values also showed a high correlation with the ground truth ($r=0.84$), as well as with the UST VBD values ($r=0.78$). The consistency in measuring the UST VBD and Cumulus PD values was analyzed using the standard error of the estimation by linear regression ($\sigma_E$). The $\sigma_E$ value for Cumulus PD was 1.5 times higher compared to the UST VBD (6.54 vs. 4.21). The $\sigma_E$ calculated from two repeated Cumulus estimation sessions ($\sigma_E=4.66$) was comparable with the UST. Potential sources of the observed errors in density measurement are the use of global thresholding and (for Cumulus) the human observer variability. This preliminary study of simulated phantom UST images showed promise for non-invasive estimation of breast density.

Keywords: Digital mammography, ultrasound tomography, anthropomorphic breast phantom, breast density analysis, validation.

1. INTRODUCTION

Previously, the correlation between volumetric breast density (VBD) estimates from clinical ultrasound tomography (UST) images and the Cumulus based percent density (PD) estimates from clinical digital mammogram (DM) images have been studied at the Karmanos Cancer Institute. In that study, a Pearson correlation coefficient of 0.75 was observed. Such studies have a practical limitation since the UST images represent a reconstructed 3D image of the breast, while DM represents a 2D projection through the breast. In addition, no information is available clinically about the ground truth breast density values. The work described in this paper has been motivated by our desire to overcome the limitation of using correlation with clinical data alone. To achieve that, we have analyzed synthetic images generated with an anthropomorphic software breast phantom, developed at the University of Pennsylvania. The phantom simulates the arrangement of breast tissues based upon the analysis of histological and radiological images. The ground truth information about the simulated breast tissues provided by the phantom allows calculating the absolute error in breast density measurements and determining potential sources of that error.

The anthropomorphic phantom developed at the University of Pennsylvania has been used previously for validation of breast density estimation methods. Recently, phantom images were used in optimization of digital breast tomosynthesis.
reconstruction for density estimation. In that study distinguishing between fatty and dense tissues was treated as a binary decision task quantified using the ROC analysis. Histograms of the reconstructed pixels corresponding to simulated adipose and dense tissues were analyzed separately. By varying the pixel intensity threshold used for breast density estimation, ROC curves were generated, thus allowing the optimization of reconstruction parameters to maximize the area under the ROC curve. Other applications of this anthropomorphic phantom include estimation of observer performance in tomosynthesis, validation of geometric accuracy in tomosynthesis reconstruction, and power spectral analysis of simulated x-ray images. In addition, a physical version of the anthropomorphic phantom has been recently manufactured.

Design of the anthropomorphic phantom used in this study allows for simulating multimodality breast images. Examples of simulated phantom mammograms, tomosynthesis, MRI, and UST images are shown in Figure 1. This study is a result of the collaboration between the University of Pennsylvania and the Karmanos Cancer Institute. The Karmanos Cancer Institute has developed a method for ultrasound tomosynthesis (UST) imaging of the breast, which is currently undergoing clinical trials. Based upon the use of non-ionizing ultrasound waves to reconstruct 3D images, the UST has the potential to overcome limitations of mammography in the measurement of breast density. In this study we have presented the assessment of methods for estimating absolute breast density values from simulated clinical data. Specifically, we have focused on the assessment of the VBD estimates calculated by thresholding simulated phantom UST images and the PD values estimated from simulated phantom mammograms interactively by a clinical radiologist using Cumulus software.

![Fig. 1: Examples of simulated images of an anthropomorphic software breast phantom. Shown: (a) A vertical slice through the phantom; (b) A simulated mammographic projection through the phantom; (c) A slice through the reconstructed phantom digital breast tomosynthesis volume; (d) A slice through the simulated phantom MRI data; and (e) A slice through the reconstructed phantom ultrasound tomography (UST) volume.](http://spiedigitallibrary.org/)

Proc. of SPIE Vol. 7961 79610Z-2
2. METHODS

2.1. Anthropomorphic software breast phantom

The anthropomorphic software breast phantom used in this study has been previously developed based upon a detailed analysis of breast anatomy visualization by clinical images and sub-gross pathology. The algorithm for phantom generation is based upon geometric primitives, and allows storing the phantoms as a 3D array of voxels with a user-specified spatial resolution. Each phantom voxel belongs to a unique tissue structure, and is characterized by physical properties of the corresponding tissue (e.g., linear x-ray attenuation, tissue elasticity, etc.)

The phantom offers great flexibility in simulating various breast size, glandularity, and internal composition. Starting from a realistic skin surface, the phantom interior includes simulated tissue structures, as adipose compartments, Cooper’s ligaments, and glandular tissue. The adipose compartments and Cooper’s ligaments are simulated using a region-growing algorithm. The region-growing algorithm includes heuristic rules governing selection of seed points and growth parameters. Figure 3(a) shows a coronal section through the software breast phantom used for the UST image simulation. Figure 4(a) shows a vertical section through another phantom realization, with simulated deformation due to mammographic breast compression. The compressed phantom was generated using a finite element model of breast tissue deformation.

For the purpose of this study we have generated 46 phantoms. Phantoms were generated with isotropic voxel resolution of (500 μm)³. The phantoms simulated breasts with two different volumes: 450 ml (26 phantoms) and 700 ml (20 phantoms). Phantom VBD values were calculated as the fraction of the phantom volume occupied by the fibroglandular tissue, Cooper’s ligaments, and the skin. Simulated VBD values were selected to roughly follow the distribution clinically estimated from over 2800 women using digital mammograms and breast CT images. Figure 2 shows a histogram of simulated VBD values.

Fig. 2: Histogram of phantom VBD values used in this study.

2.2. VBD estimation from UST phantom images

UST images of the phantom were simulated based upon the UST prototype developed at the Karmanos Cancer Institute. By the use of the non-ionizing ultrasound waves, the UST allows to create sound speed images of the breast anatomy. The longitudinal sound speed of any material is given by

\[ v = \frac{C}{\sqrt{\rho}} \]
where $C$ is the bulk modulus of the material and $\rho$ is the density. Studies have shown\textsuperscript{17} that for breast tissue, the bulk modulus scales with the cube of density. This leaves a direct relationship between sound speed and tissue density. For the simulation of UST images, each phantom voxel was converted into the corresponding sound speed value scale using a transformation that is linear with respect to tissue density. We applied 5-slice moving averaging and further mean filtering to the converted sound speed images to generate ground truth for the propagating medium. Due to the implicit asymptotic wave theory assumption in the eikonal equation, the above smoothing is necessary for accurate forward modeling. In forward modeling, we simulated acoustic wave propagation based on the eikonal equation using a 42 cm diameter ring acquisition array with 512 evenly distributed elements. The synthetic data were inverted with an iterative bent-ray inversion program\textsuperscript{18} to generate sound speed images of all phantom slices.

The breast density descriptors were then estimated by applying a 2-level clustering method on the sound speed images with the final output being percent density. The clustering was constrained with sound speed threshold values. Two values of the threshold, 1.48 and 1.49 km/sec, were found to provide the optimal results, as shown below. The calculations were performed on each slice separately and the results averaged to yield a volume based percent density.

### 2.3. Cumulus based PD estimation from mammographic phantom images

Synthetic DM images used in this study were generated by simulating x-ray projections through the software phantom deformed to model mammographic compression. The phantom deformation was calculated based upon a finite element model of 50% reduction in compressed phantom thickness. The phantom projections were simulated using a model of mono-energetic x-ray acquisition without scatter. Values of the linear x-ray attenuation coefficients were selected as 0.456 cm\textsuperscript{-1} for adipose tissue, 0.802 cm\textsuperscript{-1} for glandular and connective tissue and skin, and 0.94 \times 10\textsuperscript{-3} cm\textsuperscript{-1} for air, assuming x-ray energy of 20keV. The simulated DM acquisition geometry was selected based upon the clinical Hologic DIMENSIONS (Hologic, Bedford, MA). The simulated acquisition geometry assumed the focal spot – detector distance of 70 cm. The digital x-ray detector model assumed 70 μm/pixel spatial resolution with 25 cm × 28 cm field-of-view. The effect of quantum noise was simulated by adding random pixel variations with Poisson distribution. Such simulated images were postprocessed using a commercially available medical image processing tool (AdaraView\textsuperscript{TM}, Real-Time Tomography, LLC, Villanova, PA) for peripheral equalization similar to what is used clinically.

PD was estimated from phantom DM images by a clinical radiologist (with over 5 years of experience in mammography) using Cumulus 4.0, an interactive software package developed at the University of Toronto and validated in many studies. Cumulus is based on manual exclusion of the pectoral muscle and interactive selection of thresholds for segmenting the breast outline and the regions of dense tissue. PD value is computed as the ratio of the area corresponding to the dense tissue in a mammogram and the total mammographic breast area.

### 2.4. Data analysis

We analyzed the agreement between breast density descriptors (VBD and PD) estimated from simulated phantom images (UST and DM, respectively), and the ground truth VBD values available from the phantom by computing the Pearson correlation coefficient. To identify potential sources of error, we also tested the consistency of VBD and PD measurements by calculating standard error of the estimate using linear regressions of the estimates. The standard error of the estimate is calculated as

$$
\sigma_E = \sqrt{\frac{1}{N-2} \sum_{i=1}^{N} (D_i - D_i^{\text{LinReg}})^2},
$$

where $D_i^{\text{LinReg}}$ represents the linear regression corresponding to the $i$-th density estimate $D_i$. 


3. RESULTS

3.1. Synthetic images

Figures 3 and 4 illustrate synthetic phantom images that have been used for breast density estimation. A coronal slice of a 700 ml phantom with VBD of 20% is shown in Figure 3(a) along with the corresponding slice through the synthetic reconstructed UST volume of the same phantom (Figure 3(b)). For the sake of visual comparison, the same smoothing process as used prior to the UST image simulation was applied to the phantom slice and shown in Figure 3(c). The internal structures in Fig. 3(b) and 3(c) correlate to each other very well. The artifacts in the UST image, especially near the phantom edge, are primarily due to the finite data coverage by simulated realistic size of the transducer array; these artifacts could be avoided by simulating larger transducer array.

![Fig. 3:](image)

(a) (b) (c)

Figure 3: (a) A coronal slice of an anthropomorphic software breast phantom simulating a 700 ml breast with the true VBD of 20%. (b) The reconstructed synthetic UST image corresponding to the phantom slice shown in (a). (c) A smoothed version of the phantom slice from (a), obtained by applying the same averaging method as used in preprocessing of clinical UST data.

Figure 4(a) shows a vertical slice through a 450 ml phantom with VBD of 20%, along with the synthetic mammographic projection through the same phantom (Figure 4(b)). During the Cumulus PD estimation sessions we have asked the clinical radiologist who operated Cumulus to record her comments on the quality and realism of simulated DM images. She recorded her comments for 18 out of 46 analyzed images. Her comments include the following statements: “Images appear fairly realistic”, “Easier to segment more dense breasts – similar to in clinical images”, “Ligaments prominent with lower density – similar to clinical images”, “Denser phantoms appear qualitatively more realistic”, “Difficult to segment diffuse dense breasts”, “Good mix of fibroglandular and fatty equivalent. As expected, more fatty around periphery. Cooper’s ligaments quite linear, make it seem a phantom.”
Fig. 4: (a) A vertical slice of an anthropomorphic software breast phantom simulating a 450 ml breast with the true VBD of 40%. (b) The corresponding synthetic mammographic projection.

3.2. Comparison between the estimated and ground truth PD values

Figure 5 shows a scatter plot of the VBD values estimated from synthetic UST images as a function of the ground truth VBD values available from the phantoms. Shown are the results obtained using two different threshold values used for the segmentation of UST images. The VBD values estimated from UST images using threshold Threshold 1 show a good agreement with the ground truth VBD values which include the simulated skin, with the corresponding Pearson correlation of $r=0.93$. On the other hand, the UST VBD values estimated using threshold Threshold 2 agree well with the ground truth VBD values calculated without the skin, with the corresponding Pearson correlation of $r=0.90$.

Figure 6 shows a scatter plot of the PD values estimated from synthetic DM images using Cumulus software as a function of the ground truth phantom VBD values. The PD values estimated from phantom DM images show a better agreement with the ground truth VBD values calculated without the simulated skin (Pearson correlation $r=0.86$), compared with the ground truth VBD values including skin ($r=0.84$). Because of this observation, in Section 3.3 we compared the Cumulus PD estimates with the UST VBD values estimated using Threshold 2 which have also shown a good agreement with the ground truth excluding skin.
Fig. 5: Comparison of VBD estimates from UST phantom images vs. the ground truth VBD values available from phantoms. VBD estimates were obtained by thresholding the reconstructed UST images, using two different threshold values: (a) Threshold 1 corresponding to the sound speed of 1480 m/s, and (b) Threshold 2 corresponding to the sound speed of 1490 m/s. The UST VBD estimates have been compared with the ground truth VBD values which include (black) or exclude (gray) simulated skin. Shown are also linear regressions of the both graphs.

Fig. 6: Comparison of PD estimates from phantom DM images vs. the ground truth VBD values available from phantoms. PD estimates were obtained using Cumulus operated by a clinical radiologist. PD estimates have been compared with the ground truth VBD values which include (black) or exclude (gray) simulated skin. Shown are also linear regressions of the both graphs.
3.3. Analysis of the sources of error in PD estimation

Figure 7 shows a scatter plot of the PD estimates from phantom DM images and the VBD estimates from phantom UST images. The corresponding Pearson correlation is \( r = 0.78 \).

\[
y = 1.21x - 3.70 \\
R^2 = 0.61
\]

Fig. 7: Comparison of breast density estimates from phantom DM and UST images. The PD estimates from phantom DM images were obtained using Cumulus operated by a clinical radiologist. The VBD estimates from phantom UST images were obtained by thresholding the reconstructed UST images using the threshold value corresponding to the sound speed of 1490 m/s (Threshold 2 in Fig. 5). Shown is also the corresponding linear regression.

Figure 7 suggests that the variation in Cumulus PD estimates may be larger compared to the variation in UST VBD estimates. To test this observation, we calculated the standard error of the estimated Cumulus PD values, using the linear regression from Figure 7; this standard error was equal to \( \sigma_E = 6.54 \) (in percent PD units). Similarly, we calculated the standard error of the UST VBD estimates; this standard error was equal to \( \sigma_E = 4.21 \) (in percent VBD units).

We have also analyzed the intra-reader variability, by comparing PD values from two repeated Cumulus sessions for a subset of 20 phantom DM images. The two estimation sessions were repeated by the same radiologist six months apart (in August 2010 and January 2011). The January 2011 session included analysis of synthetic images from all 46 phantoms, including the 20 phantoms analyzed in August 2010. Results from the repeated Cumulus analysis were shown in Figure 8. The corresponding Pearson correlation value is \( r = 0.94 \) and the standard error of the estimate \( \sigma_E = 4.66 \) percent PD units.
4. DISCUSSION

The results in this study indicated that our phantom based approach has been successfully applied to the analysis of breast density estimation from synthetic images. Based upon its multimodality design, the phantom has been appropriate for the simulation of both UST and DM images. The phantom allowed for simulating a realistic range of VBD values. The visual appearance of phantom DM images was similar to clinical mammograms, as evidenced from the comments by the radiologist observer participating in this study. Importantly, the use of phantom images provided the ground truth information about the simulated tissues, which was used to validate performance of the Cumulus and UST based density estimation methods.

Selection of threshold values used for the UST image segmentation has a noticeable effect on the VBD estimation performance. As seen in Figure 5(a), VBD estimated using Threshold 1 seems to agree better with the phantom ground truth including the simulated skin; the offset value closer to zero and the $R^2$ value higher than for the ground truth with the skin excluded. On the other hand, the estimates obtained using Threshold 2 (Figure 5(b)) seem to agree somewhat better with the phantom ground truth excluding the skin; the slope is closer to 1 compared to the ground truth calculated with the skin included (while offsets and $R^2$ values are comparable). The observed effect suggests that a refinement in UST segmentation might further improve the agreement with the ground truth, e.g., using an iterative determination of the optimal threshold. The observed effect of the UST image thresholding suggests that there might be a room to further improve the accuracy of the UST VBD estimation. An iterative comparison between the phantom ground truth VBD values and the UST VBD estimates may be used to determine the optimal threshold value.

Figures 5 and 6 show that breast density estimates from phantom DM and UST images have relatively consistent values. Cumulus has slightly overestimated the phantom ground truth, as indicated by the linear regression slope of 1.27 in Figure 4. On the other hand, the UST based approach has slightly underestimated the phantom ground truth, with the corresponding regression slope of 0.80 in Figure 5(b). Figure 7 indicated that the Cumulus based density estimation has larger variations compared to the UST; the standard error in the estimate calculated for the Cumulus results are 1.5 times larger compared to the UST based approach. This is probably due to the variations in the performance of human observers in Cumulus analysis, as well as to the use of global thresholding by both Cumulus and the UST based estimation. Such a hypothesis may be supported by the observed relationship between the $\sigma_E$ values calculated from breast density estimates. Since both Cumulus and UST use global threshold it is a plausible assumption...
that the effect of thresholding might result in a similar variation, i.e., (σ_E)_{THR} ≈ (σ_E)_{UST}. = 4.21 (from Figure 7). We may interpret the standard error in the estimate calculated from the repeated Cumulus estimations as describing the observer variations, i.e., (σ_E)_{OBS} ≈ 4.66 (from Figure 8). Assuming that these two sources of variations are independent, they would add up in quadrature, so the Cumulus σ_E value might be approximated by:

\[(σ_E)^2_{Cumul} = (σ_E)^2_{THR} + (σ_E)^2_{OBS} ≈ (4.21)^2 + (4.66)^2 = 39.44,\]

which agrees well with the variation in Cumulus estimates calculated from Figure 7 as (σ_E)^2_{Cumul} = 42.76. The observed effect of global thresholding on the variation in density measurements may suggest that better estimation methods need to include non-global, adaptive analysis.

The effects of observer variation on Cumulus results are comparable with our previous work. In our study of breast density estimates from DM and DBT projection images,\(^{19}\) we compared the Cumulus PD values estimated by three medical physicists in two sessions each two months apart. We observed an average intra-reader Spearman correlation coefficient of 0.86. Our current study yielded a higher intra-reader correlation (r=0.94). Such a result is supported by recent reports on the effect of observer clinical training, which observed higher agreement in Cumulus estimates between clinical radiologists compared to medical physicists.\(^{20}\) The effect of human observer variation justifies the efforts in developing automated density estimation, including the automated analysis of UST images described here.

5. CONCLUSION

The breast density descriptors based upon the reconstructed UST images showed high correlation with the ground truth 3D percent density. Efforts to improve UST VBD estimation should focus on the optimization of the UST image segmentation. This result supports further research on using the UST for non-invasive 3D analysis of breast density. Use of an anthropomorphic software breast phantom has proven practically useful in comparison of percent density descriptors from synthetic UST and DM images.

ACKNOWLEDGEMENT

PRB, SCG, and ADAM acknowledge a support by a Partnership Training Award from the Departement of Defense Breast Cancer Research Program. ND, EW, CL and MS acknowledge that this work was supported by a grant from the Susan G. Komen Foundation (KG100100). The authors are grateful to Real-Time Tomography, LLC, for the use of their commercial breast image preprocessing software AdaraView\textsuperscript{TM}. The authors are grateful to Dr. Ben Rao and Dr. Andy Smith of Hologic for providing description of DM acquisition geometry.

REFERENCES


