Objective models of compressed breast shapes undergoing mammography

Steve Si Jia Feng

*Department of Biomedical Engineering, Georgia Institute of Technology and Emory University
and Department of Radiology and Imaging Sciences, Emory University, 1701 Uppergate Drive Northeast,
Suite 5018, Atlanta, Georgia 30322*

Bhavika Patel

*Department of Radiology and Imaging Sciences, Emory University, 1701 Uppergate Drive Northeast,
Suite 5018, Atlanta, Georgia 30322*

Ioannis Sechopoulos^{a)}

*Departments of Radiology and Imaging Sciences, Hematology and Medical Oncology and Winship Cancer
Institute, Emory University, 1701 Uppergate Drive Northeast, Suite 5018, Atlanta, Georgia 30322*

(Received 8 September 2012; revised 10 December 2012; accepted for publication 14 January 2013;
published 11 February 2013)

Purpose: To develop models of compressed breasts undergoing mammography based on objective analysis, that are capable of accurately representing breast shapes in acquired clinical images and generating new, clinically realistic shapes.

Methods: An automated edge detection algorithm was used to catalogue the breast shapes of clinically acquired cranio-caudal (CC) and medio-lateral oblique (MLO) view mammograms from a large database of digital mammography images. Principal component analysis (PCA) was performed on these shapes to reduce the information contained within the shapes to a small number of linearly independent variables. The breast shape models, one of each view, were developed from the identified principal components, and their ability to reproduce the shape of breasts from an independent set of mammograms not used in the PCA, was assessed both visually and quantitatively by calculating the average distance error (ADE).

Results: The PCA breast shape models of the CC and MLO mammographic views based on six principal components, in which 99.2% and 98.0%, respectively, of the total variance of the dataset is contained, were found to be able to reproduce breast shapes with strong fidelity (CC view mean ADE = 0.90 mm, MLO view mean ADE = 1.43 mm) and to generate new clinically realistic shapes. The PCA models based on fewer principal components were also successful, but to a lesser degree, as the two-component model exhibited a mean ADE = 2.99 mm for the CC view, and a mean ADE = 4.63 mm for the MLO view. The four-component models exhibited a mean ADE = 1.47 mm for the CC view and a mean ADE = 2.14 mm for the MLO view. Paired t-tests of the ADE values of each image between models showed that these differences were statistically significant (max p-value = 0.0247). Visual examination of modeled breast shapes confirmed these results. Histograms of the PCA parameters associated with the six principal components were fitted with Gaussian distributions. The six-component model was also used to generate CC and MLO view mammogram breast shapes, using the mean PCA parameter values of these distributions and randomly generated values based on the fitted Gaussian distributions, which resemble clinically encountered breasts. A spreadsheet with the data necessary to apply this model is provided as the supplementary material.

Conclusions: Our PCA models of breast shapes in both mammographic views successfully reproduce analyzed breast shapes and generate new clinically relevant shapes. This work can aid in research applications which incorporate breast shape modeling, such as x-ray scatter correction, dosimetry, and image registration. © 2013 American Association of Physicists in Medicine. [<http://dx.doi.org/10.1118/1.4789579>]

Key words: breast cancer, mammography, principal component analysis

I. INTRODUCTION

According to the American Cancer Society, over 287 000 new cases of breast cancer were estimated to be diagnosed in 2011.¹ Currently, screening mammography is regarded as the most effective tool for early breast cancer detection and diagnosis. Standard mammography is a low-cost, fast, noninvasive x-ray study that involves relatively low doses of ionizing radiation, but has room for improvement in

performance, demonstrating a sensitivity and specificity of 83.5% and 90.9%, respectively.²

A large amount of the continued research being performed to improve the clinical performance of breast imaging requires the simulation of realistic patient breast models, and this has become an entire research area in itself.^{3–10} Research on breast imaging dosimetry,^{11–20} lesion detectability,^{21–23} technique optimization,^{24,25} image

registration,²⁶ image texture,²⁷ and x-ray scatter correction²⁸ has been performed using software breast models, representing both uncompressed and compressed breasts. In most studies involving a model of the compressed breast, only the cranio-caudal (CC) view has been simulated, with the breast represented as a semi-circle or semi-ellipse. In a few studies, models of the breast compressed for the medio-lateral oblique (MLO) view have been used, but the models were subjectively derived.^{18,19,29–31} Although these models have proved appropriate, the development of more sophisticated models that better represent not only the “average” breast shape but also attempt to include the span of clinically realistic breast shape variations would be useful in many of the areas of research listed above that use software breast models.

Here we present models for the breast shape when compressed for both the CC and MLO views of mammography that leverage principal component analysis (PCA) to accurately reproduce breast shapes and to generate new, clinically realistic shapes.

II. MATERIALS AND METHODS

The proposed models of mammogram breast shape were developed in three distinct steps. First we assembled a large database of clinically acquired CC and MLO view mammograms and systematically catalogued the breast shapes. Then we performed PCA on these shapes to condense the information down to a reduced number of linearly independent variables. And finally we tested the models’ ability to estimate the shape of breasts from mammograms that were not used in the PCA.

II.A. Mammogram database

We have an anonymized database of 477 paired CC and MLO view digital mammograms. This database was previously compiled in our institution and as the anonymization process is irreversible, no further IRB oversight is necessary. These were acquired using a Senographe 2000D digital mammography system (GE Healthcare, UK) that utilizes an x-ray tube with molybdenum and rhodium targets with added molybdenum and rhodium filtration. The system has a 19.14 × 22.94 cm indirect detector with a pixel pitch of 100 μm.

This database was divided into 377 images of each view to be used in the formation of the PCA model, and 100 images of each view to be used for testing.

II.B. Automated edge detection

Each mammogram was automatically thresholded to obtain a binary image, Y_b , separating tissue from open field.³² The extraneous tissue, i.e., tissue not part of the breast under examination, was removed to isolate only the breast tissue in the image via the extraction of connected components.³² This was accomplished by repeated iterations of image dilation with a dilation structure B and image intersection with Y_b :

$$Y_k = (Y_{k-1} \oplus B) \cap Y_b. \quad (1)$$

We chose the midpoint of the length along the chest wall of the mammogram as a starting point for a 5×5 pixel initial guess and a 10×10 structure was chosen for B . The extraction was considered complete when $Y_k = Y_{k-1}$. Using this isolated binary image of the breast, the edge was then located by using Roberts cross gradient operators, and edge thinning.^{32–34} The edge is extended to the chest wall, if needed, to form a set of points that depict the breast shape in the mammogram. This set was then uniformly sampled to generate a subset of 100 (x,y) coordinate points, in terms of distance rather than pixel index, for use in the PCA. This is very similar to a process we previously used to register breast shapes for digital tomosynthesis images.²⁸ For the purposes of this study, the automation of the breast tissue isolation and edge detection allows for many images to be processed quickly without requiring user input and subjective analysis of the images.

To validate the automated edge detection, a radiologist enrolled in a mammography fellowship evaluated the generated edges of 100 CC view mammograms and 100 MLO view mammograms, all of which were randomly selected from the database. Each mammogram was shown side by side with a view of the generated edge overlaid onto the mammogram, allowing the radiologist to simultaneously view the edge and the original image. For each image, the observer was asked if the generated edge correctly represented the breast edge.

II.C. Principal component analysis

The subset of points that delineate the shape of the breast in the mammogram can be defined as 100 correlated observations which we can reduce using PCA to a smaller set of linearly independent variables that encapsulates most of the information contained within the datapoints.^{35,36} PCA has been used in innumerable applications, including in image registration.^{37–39} To perform the analysis, we formed the vector $\mathbf{x}_i$ as follows for each image i :

$$\mathbf{x}_i = [x_1, x_2, x_3, \dots, x_{100}, y_1, y_2, y_3, \dots, y_{100}]^T, \quad (2)$$

where the (x,y) pairs are the pixel coordinates that define the breast shape, in cm. We then formed a square covariance matrix, $\text{cov}(\mathbf{X})$, from the mean-centered $\mathbf{x}$ vectors:

$$\text{cov}(\mathbf{X}) = \frac{\sum_{i=1}^N (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T}{N - 1}, \quad (3)$$

where N is the total number of images processed and $\bar{\mathbf{x}}$ is the mean of the $\mathbf{x}$ vectors. The principal components of the dataset are then represented by the eigenvectors $\mathbf{e}_j$, which can be obtained by solving the eigenvalue problem:

$$\text{cov}(\mathbf{X})\mathbf{e}_j = v_j\mathbf{e}_j, \quad (4)$$

where the associated eigenvalues (v_j) also represent the variance of each principal component. For each view (CC and MLO), the n principal components with the greatest v_j were selected and their respective $\mathbf{e}_j$ assembled into a $100 \times n$ principal component matrix $\mathbf{E}$. This matrix can be used to describe, to a high degree of accuracy, any of the breast edges $\mathbf{x}_i$ that were used in the PCA and forms the basis of our breast

shape model. The PCA parameter vector $\mathbf{r}_i$ for image i can be calculated as follows:

$$\mathbf{r}_i = (\mathbf{E}^T \mathbf{E})^{-1} \mathbf{E}^T (\mathbf{x}_i - \bar{\mathbf{x}}). \quad (5)$$

Initial validation of the model was done by visually comparing the original breast edge of several mammograms from the foundational database to that estimated from its PCA parameters in $\mathbf{r}_i$. An estimate of the original breast edge, $\tilde{\mathbf{x}}_i$, was calculated as follows:

$$\tilde{\mathbf{x}}_i = \bar{\mathbf{x}} + \sum_{j=1}^n \mathbf{r}_{i,j} \mathbf{e}_j, \quad (6)$$

where n is the number of principal components chosen for $\mathbf{E}$.

Histograms of the first six PCA parameters obtained for each of the breast shapes in the foundational database were generated, and each one was fitted with a Gaussian distribution.

All PCA analysis was performed using MATLAB® R2011b (MathWorks, Natick, MA).

II.D. PCA model testing

To test the PCA models of breast shape, we utilized 100 CC view and 100 MLO view mammograms that were not included in the PCA. Each test mammogram m underwent automated edge detection as described above to generate a subset of points that depict a test edge $\mathbf{t}_m$. Using the view-specific principal component matrix, $\mathbf{E}_{CC}$ or $\mathbf{E}_{MLO}$, we calculated the PCA parameter vector, $\mathbf{r}_m$, of each test edge, and the modeled edge $\tilde{\mathbf{t}}_m$ as follows:

$$\mathbf{r}_m = (\mathbf{E}^T \mathbf{E})^{-1} \mathbf{E}^T (\mathbf{t}_m - \bar{\mathbf{x}}), \quad (7)$$

$$\tilde{\mathbf{t}}_m = \bar{\mathbf{x}} + \sum_{j=1}^n \mathbf{r}_{m,j} \mathbf{e}_j. \quad (8)$$

The accuracy of $\tilde{\mathbf{t}}_m$ was assessed by calculating the average distance error (ADE) between it and the original edge, $\mathbf{r}_m$. ADE was calculated as follows:

$$\text{ADE} = \frac{(T_b \cup Y_b) - (T_b \cap Y_b)}{0.5(L_r + L_t)}, \quad (9)$$

where T_b is a binary image, depicting the PCA modeled breast shape, created by shape filling an image of $\tilde{\mathbf{t}}_m$, Y_b is the binary image of the original mammogram, and L_r and L_t are the lengths of the modeled edge and the original edge, respectively. Thus, this metric is formed by dividing the area that lies in between the edges formed by the modeled and original breast shapes by the average of the two edge lengths (Fig. 1). A paired t-test was performed on the ADE values of each tested image between the PCA models generated with differing number of principal components.

II.E. PCA parameter characterization

To understand how the value of each PCA parameter affects the breast shape in the six-component model, we generated novel shapes using selected values based on the mean and standard deviation of the fitted Gaussian probability distributions of the parameter histograms. For this, each parameter

FIG. 1. Average distance error calculation—the automatically generated breast edge (dark) and the PCA modeled edge (white) are shown here on the left for a CC view mammogram. The image shown on the right represents $(T_b \cup Y_b) - (T_b \cap Y_b)$, where T_b and Y_b are binary images depicting the PCA modeled breast shape and the original mammogram, respectively. The area of the region in white shown in the right image is divided by the average of the length of the two edges to obtain the ADE.

was varied individually, with the selection of five values ranging from its mean minus two standard deviations ($\mu - 2\sigma$) to $\mu + 2\sigma$ while the other parameters were held constant at their mean values. In this way, the effect of each parameter when its values span approximately 95% of the modeled clinically relevant range can be visualized.

II.F. PCA parameters and breast size

We examined the relationship between breast size in the mammogram, both in terms of breast area and compressed breast thickness, and each of the calculated PCA parameter values from both $\mathbf{r}_m$ and $\mathbf{r}_i$. The thickness was retrieved from the mammogram database, and the breast shape area was calculated from the binary image, Y_b , of the original mammogram.

III. RESULTS

III.A. Automated edge detection

Our automated edge detection algorithm proved very successful in isolating the breast tissue under examination from extraneous tissue present in the mammogram, and delineating a set of points that form the shape of the breast. Two examples each of CC view and MLO view breast shapes from the mammogram database (that used for the PCA) can be seen in Fig. 2. The delineated breast edges are overlaid on the images in white. The radiologist enlisted in the validation study determined that all 200 automatically generated edges presented, evenly distributed between the two views, are successful representations of the breast shape edge.

FIG. 2. Breast edges—shown here, in white, are edges generated by our automated edge detection algorithm on CC view (top) and MLO view mammograms (bottom). The extraneous tissue seen at the bottom of the CC view mammograms was excluded by our algorithm.

III.B. Principal component analysis

Our PCA of the CC and MLO view mammograms revealed that 93.3% and 85.4%, respectively, of the variance in the database images is contained in the first two principal components, that 99.2% and 98.0% of the variance is contained in the first six principal components (Table I). We have included the eigenvectors of the first six principal components of both models in the supplementary material.⁴⁰

TABLE I. Percentage of total variance contained in the first six principal components—shown here is the variance contained within each individual component and the cumulative variance contained within each component and its predecessors.

Principal component	Percentage of total variance	
	Cranio-caudal view/cumulative (%)	Medio-lateral oblique view/cumulative (%)
α	74.8/74.8	62.3/62.3
β	18.5/93.3	23.1/85.4
γ	3.54/96.8	6.72/92.1
δ	1.03/97.9	3.13/95.2
ϵ	0.72/98.6	1.83/97.1
ζ	0.54/99.2	0.87/98.0

III.C. PCA model testing

The two-component models produced reasonable recreations of breast shapes of mammograms from the testing set; three examples of breast shapes of each view are shown in Fig. 3 (white outline), along with their ADEs. The upper left image depicts a CC-view mammogram for which the two-component model generates a breast shape with a very low ADE that is similar to that of the six-component model (0.88 mm to 0.79 mm, respectively). However, as can be seen in the upper right image, for some cases the two-component model does not recreate the breast shape accurately, resulting in a higher ADE (7.15 mm). Visual inspection and the ADE values of the four-component and six-component PCA models for each view revealed that such models generate more accurate representations that include intricacies of some breast shapes. Test edges generated by the six-component models can also be seen in Fig. 3 (dark outline). Some breast shapes are recaptured to a much greater degree by the six-component model, such as the middle and right images shown in both rows in Fig. 3. Table II shows the mean and maximum ADE for the breast shapes generated with the two-, four-, and six-component models for both the CC and MLO views.

We also compared the performance of the models with differing number of components by examining a box-and-whisker plot of the ADEs of each model (Fig. 4). As can be seen, for every percentile mark (10th, 25th, median/50th, 75th, and 90th), models with more components exhibited lower ADEs than those of the same view with fewer components. Thus, as expected, models with more components consistently perform better. For each view, paired t-tests were performed between the ADE values for each PCA model, and all p-values were found to be less than 0.025.

Histograms of the six PCA parameters for each view were fitted with Gaussian distributions and are shown in Fig. 5, along with the mean and standard deviation of each fit.

III.D. Novel breast shape generation

The six-component objective PCA breast shape models presented here were also used to generate novel breast shapes, i.e., shapes that are clinically relevant but which were not derived from a clinical mammogram. This was achieved by using the view-specific principal component matrix, $\mathbf{E}_{CC}$ or $\mathbf{E}_{MLO}$, and selected values for the PCA parameter vector $\mathbf{r}_m = [\alpha, \beta, \gamma, \delta, \epsilon, \zeta]$. First we generated the models' average breast shapes by using the mean values of the PCA parameters (Fig. 6, top). It should be noted that this "average" breast shape in the CC view results in a shape that is similar to, but not quite, a semi-ellipse, with a chest wall-to-nipple distance of 10.1 cm and a distance along the chest wall of 20.2 cm, and a calculated area of 144 cm². This size corresponds well with the mean equivalent radius of 10 cm and the mean equivalent area of 157.3 cm² determined by Boone *et al.* from 82 patient CC mammograms.⁴¹

We also generated a random PCA parameter set, using the fitted Gaussian probability distributions of the parameter histograms, and generated a random clinically relevant breast

FIG. 3. Modeled breast shapes using two and six principal components—shown here are breast shapes generated from six-component PCA models (dark) of the CC (top) and MLO (bottom) views overlaid on the shapes generated by two-component models (white) and their respective ADEs. In some cases, the two-component model generates adequate breast shape representations (left, top, and bottom). However, in other cases (middle and right) a model with a greater number of components is needed to recapture the intricacies of the breast shapes.

shape for each view (Fig. 6, bottom). The fitted Gaussian distributions from Fig. 5 are included in the supplementary material and can be used to generate random values for the PCA parameters.

III.E. PCA parameter characterization

Figure 7 (CC view) and Fig. 8 (MLO view) show how the breast shape model varies when the six PCA parameters are varied individually. As can be seen, the first PCA parameter, α , is the greatest determinant of the overall size and shape of the breast for both the CC and MLO view models, while

the subsequent five parameters ($\beta, \gamma, \delta, \epsilon, \zeta$) adjust different detailed shape features of the breast.

III.F. PCA parameters and breast size

A strong relationship was found between the first PCA parameter, α , for both views and the breast area (Fig. 9, left). However, there was no apparent relationship between the α parameter and the compressed breast thicknesses in either view (Fig. 9, right). None of the other PCA parameters were found to correlate strongly with either measure of breast size in either view.

TABLE II. Average distance error—the mean ADEs of the two-component PCA model breast shapes was 2.99 mm for the CC view and 4.63 mm for the MLO view, but the six-component model produced more accurate shapes on average and in the worst cases (maximum).

Average distance error (ADE)						
View	Two-component model (mm)		Four-component model (mm)		Six-component model (mm)	
	Mean	Maximum	Mean	Maximum	Mean	Maximum
CC	2.99	7.80	1.47	3.92	0.90	1.94
MLO	4.63	13.6	2.14	6.25	1.43	4.49

FIG. 4. Box-Whisker plot of average distance error—PCA models with more components consistently performed better than those of the same view with fewer components, exhibiting not only lower means but also lower ADEs for the 10th, 25th, 50th/median, 75th, and 90th percentiles. Shown here are the models of the CC-view (left) and MLO-view (right) with components numbering from 2 to 6. All differences in paired ADE values of the PCA models with different number of principal components were statistically significant ($p < 0.025$).

FIG. 5. PCA parameter distributions—The distributions of the first six PCA parameters ($\alpha, \beta, \gamma, \delta, \epsilon, \zeta$) of the PCA breast shape models for both the CC (left) and MLO (right) views can be fitted with Gaussian distributions. The mean and standard deviation of each fit is also shown for each parameter.

FIG. 6. Novel breast shapes generated by the six-component PCA model—These breast shapes were generated by the selection of a set of values for the PCA data vector r_m . The top CC and MLO view modeled breasts were generated using the mean values of the Gaussian fits for each principal component, so they represent an “average” breast shape. The bottom breast shapes were generated using randomly generated values from the Gaussian distributions for each principal component. Axes are in units of centimeters.

IV. DISCUSSION

In this study, we developed objective models of the compressed breast shape in the CC and MLO mammographic views using PCA and tested the models’ ability to both recreate breast shapes that were not part of the original analysis and to create novel clinically relevant breast shapes. Visual examination and quantitative analysis in the form of the average distance error (ADE) revealed that the 2–6 principal component

models of breast shape presented here show great promise and could become useful tools for breast imaging applications, both for analyzing acquired images and for generating realistic breast shapes. However, the number of components to include in the PCA model depends on the application. In the generation of realistic breast shapes, the use of a high number of components should not be onerous and it would provide the greatest variability and detail in clinically relevant shapes, so all six components could be included. However, for other

FIG. 7. Varying PCA parameters of the six-component CC view model—shown here are CC view breast shapes generated by varying each of the six PCA parameters (α , β , γ , δ , ϵ , ζ) individually. Each parameter was set to the mean (μ) value from the fitted Gaussian distributions shown in Fig. 5, as well as $\mu \pm$ one standard deviation (σ), and $\mu \pm 2\sigma$. The arrows indicate the overall direction in which that portion of the breast shape shifts as each parameter is increased. Within each panel, the average breast (in which all the values are set to their mean) is shown in solid black. Axes are in units of centimeters.

FIG. 8. Varying PCA parameters of the six-component MLO view model—shown here are MLO view breast shapes generated by varying each of the six PCA parameters (α , β , γ , δ , ϵ , ζ) individually. The process by which this was accomplished is the same as in Fig. 7. The arrows indicate the overall direction in which that portion of the breast shape shifts as each parameter is increased. Within each panel, the average breast (in which all the values are set to their mean) is shown in solid black. Axes are in units of centimeters.

FIG. 9. Relationship between α and breast size—for both the CC (top) and MLO (bottom) views, a strong relationship was found between α and the breast area (left). However, for the compressed breast thickness (right), there was no strong correlation found.

applications it is possible that the inclusion of more components could introduce high complexity, making the use of a greater number of components unfeasible. For these cases, using only a three- or four-component model could still result in acceptable representations of the breast, given the low ADE values found (Table II, Fig. 4). In fact, the two-component models generate a shape that resembles a breast undergoing mammography (Fig. 3, white) in both the CC and MLO view, and recaptures the general curvature and closed edge of the breast, which might be sufficient for some applications.

Evaluation of each individual PCA parameter (Figs. 7 and 8) shows that the first PCA parameter, α , determines to a great extent the size of the breast shape in the modeled mammogram. We also found that α is strongly correlated with the breast shape area, but not compressed breast thickness. The relationships shown in Fig. 9 can be a starting point for the user to generate new, clinically relevant breast shapes with control of the size of the generated breasts.

We have used the models presented here to characterize the breast shapes seen in almost 500 mammograms each of the CC and MLO views. As revealed by the ADE results (Table II, Fig. 4), our models have reduced the complexities of breast shapes to a small number of parameters; using just 2 to 6 values, we can describe to a great extent the breast shapes in the original mammograms. The use of the presented models to generate novel, clinically relevant breast shapes could

aid research in 3D breast software phantoms by simplifying the mathematical representation of the breast shape. In addition, we have previously developed a software-based scatter correction algorithm for digital breast tomosynthesis imaging (DBT) that utilizes Monte Carlo x-ray simulations of standardized breast shapes, so the PCA models developed here facilitate the scatter correction algorithm's application to clinically encountered breasts.²⁸ We have performed preliminary testing with the central angle x-ray projection images of DBT imaging, and results have also been promising but a larger database of clinical images is required for further testing. And finally, dosimetric characterization of breast imaging systems that are also based on Monte Carlo simulations could leverage our PCA model's ability to randomly generate breasts of different clinically relevant shapes quickly.^{19,42,43}

It is also important to note that the PCA models developed here, and the methodology we have employed, can be used independently of our automated edge detection algorithm. Any user-defined breast edge representation can be used in conjunction with our model to expand the basis of the model or to characterize a breast shape by condensing information down to a few principal components. There have been several other methods proposed to automatically segment or detect the breast boundaries in mammograms.^{44–47}

Limitations of this study include the lack of including an automated nipple detection and removal algorithm, so that the

model of the breast shape would not be affected by its presence. In addition, inclusion of an automated pectoralis detection algorithm, especially in the MLO view, such as that proposed by Kwok *et al.*,⁴⁸ and its inclusion as another edge in the (x,y) coordinate pair vector for the PCA could allow for this image feature to also be included. Other future work would include the expansion of the PCA with more mammograms, allowing for the study of PCA parameter variability with patient characteristics (e.g., age, glandular density), and for the investigation of constraining their values to generate corresponding CC and MLO view mammograms.

V. CONCLUSION

The objective PCA models of breast shape presented here show great promise and could become useful tools for breast imaging applications, both for analyzing acquired images and for generating realistic breast shapes. Testing of the models' ability to replicate clinical breast shapes encountered in an independent database were successful, and the ease of generating random, clinically realistic breast shapes with the models was shown.

ACKNOWLEDGMENTS

The authors would like to thank Carl J. D'Orsi, M.D., for providing the digital mammography database. Supported in part by Grant No. R01CA163746 from the National Cancer Institute, National Institutes of Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute or the National Institutes of Health.

^{a)} Author to whom correspondence should be addressed. Electronic mail: ise-chop@emory.edu

¹ American Cancer Society, "Breast Cancer Facts and Figures 2011-2012" (American Cancer Society, Atlanta, GA, 2012).

² NCI-funded Breast Cancer Surveillance Consortium co-operative agreement, "Performance measures for 1960 150 screening mammography examinations from 2002 to 2006 by age—based on BCSC data as of 2009," 2011, available at http://breastscreening.cancer.gov/data/performance/screening/2009/perf_age.html.

³ P. R. Bakic, M. Albert, D. Brzakovic, and A. D. A. Maidment, "Mammogram synthesis using a 3D simulation. I. Breast tissue model and image acquisition simulation," *Med. Phys.* **29**, 2131–2139 (2002).

⁴ P. R. Bakic, M. Albert, D. Brzakovic, and A. D. A. Maidment, "Mammogram synthesis using a 3D simulation. II. Evaluation of synthetic mammogram texture," *Med. Phys.* **29**, 2140–2151 (2002).

⁵ P. R. Bakic, M. Albert, D. Brzakovic, and A. D. A. Maidment, "Mammogram synthesis using a three-dimensional simulation. III. Modeling and evaluation of the breast ductal network," *Med. Phys.* **30**, 1914–1925 (2003).

⁶ P. R. Bakic, C. Zhang, and A. D. Maidment, "Development and characterization of an anthropomorphic breast software phantom based upon region-growing algorithm," *Med. Phys.* **38**, 3165–3176 (2011).

⁷ B. A. Lau, I. Reiser, R. M. Nishikawa, and P. R. Bakic, "A statistically defined anthropomorphic software breast phantom," *Med. Phys.* **39**, 3375–3385 (2012).

⁸ I. Reiser, S. Lee, and R. M. Nishikawa, "On the orientation of mammographic structure," *Med. Phys.* **38**, 5303–5306 (2011).

⁹ K. Bliznakova, I. Sechopoulos, I. Buliev, and N. Pallikarakis, "BreastSimulator: A software platform for breast x-ray imaging research," *J. Biomed. Graph. Comput.* **2**, 1–14 (2012).

- ¹⁰ C. M. Li, W. P. Segars, G. D. Tourassi, J. M. Boone, J. T. Dobbins III, "Methodology for generating a 3D computerized breast phantom from empirical data," *Med. Phys.* **36**, 3122–3131 (2009).
- ¹¹ D. R. Dance, "Monte Carlo calculation of conversion factors for the estimation of mean glandular breast dose," *Phys. Med. Biol.* **35**, 1211–1219 (1990).
- ¹² D. R. Dance, C. L. Skinner, K. C. Young, J. R. Beckett, and C. J. Kotre, "Additional factors for the estimation of mean glandular breast dose using the UK mammography dosimetry protocol," *Phys. Med. Biol.* **45**, 3225–3240 (2000).
- ¹³ D. R. Dance, K. C. Young, and R. E. van Engen, "Estimation of mean glandular dose for breast tomosynthesis: Factors for use with the UK, European, and IAEA breast dosimetry protocols," *Phys. Med. Biol.* **56**, 453–471 (2011).
- ¹⁴ X. Wu, G. T. Barnes, and D. M. Tucker, "Spectral dependence of glandular tissue dose in screen-film mammography," *Radiology* **179**, 143–148 (1991).
- ¹⁵ X. Wu, E. L. Gingold, G. T. Barnes, and D. M. Tucker, "Normalized average glandular dose in molybdenum target-rhodium filter and rhodium target-rhodium filter mammography," *Radiology* **193**, 83–89 (1994).
- ¹⁶ J. M. Boone, "Normalized glandular dose (DgN) coefficients for arbitrary X-ray spectra in mammography: Computer-fit values of Monte Carlo derived data," *Med. Phys.* **29**, 869–875 (2002).
- ¹⁷ J. M. Boone, "Glandular breast dose for monoenergetic and high-energy X-ray beams: Monte Carlo assessment," *Radiology* **213**, 23–37 (1999).
- ¹⁸ I. Sechopoulos and C. J. D'Orsi, "Glandular radiation dose in tomosynthesis of the breast using tungsten targets," *J. Appl. Clin. Med. Phys.* **9**, 161–171 (2008).
- ¹⁹ I. Sechopoulos, S. Suryanarayanan, S. Vedantham, C. D'Orsi, and A. Karellas, "Computation of the glandular radiation dose in digital tomosynthesis of the breast," *Med. Phys.* **34**, 221–232 (2007).
- ²⁰ D. R. Dance, R. A. Hunt, P. R. Bakic, A. D. A. Maidment, M. Sandborg, G. Ullman, and G. Alm Carlsson, "Breast dosimetry using high-resolution voxel phantoms," *Radiat. Prot. Dosim.* **114**, 359–363 (2005).
- ²¹ I. Reiser and R. M. Nishikawa, "Task-based assessment of breast tomosynthesis: Effect of acquisition parameters and quantum noise," *Med. Phys.* **37**, 1591–1600 (2010).
- ²² X. Gong, S. J. Glick, B. Liu, A. A. Vedula, and S. Thacker, "A computer simulation study comparing lesion detection accuracy with digital mammography, breast tomosynthesis, and cone-beam CT breast imaging," *Med. Phys.* **33**, 1041–1052 (2006).
- ²³ L. Zhou, J. Oldan, P. Fisher, and G. Gindi, "Low-contrast lesion detection in tomosynthetic breast imaging using a realistic breast phantom," *Proc. SPIE* **6142**, 61425A–61412 (2006).
- ²⁴ R. A. Hunt, D. R. Dance, P. R. Bakic, A. D. A. Maidment, M. Sandborg, G. Ullman, and G. Alm Carlsson, "Calculation of the properties of digital mammograms using a computer simulation," *Radiat. Prot. Dosim.* **114**, 395–398 (2005).
- ²⁵ K. Bliznakova, Z. Kolitsi, and N. Pallikarakis, "Dual-energy mammography: Simulation studies," *Phys. Med. Biol.* **51**, 4497–4515 (2006).
- ²⁶ F. J. P. Richard, P. R. Bakic, and A. D. A. Maidment, "Mammogram registration: A phantom-based evaluation of compressed breast thickness variation effects," *IEEE Trans. Med. Imaging* **25**, 188–197 (2006).
- ²⁷ D. Kontos, C. Zhang, N. Ruiter, P. Bakic, and A. Maidment, "Evaluating the effect of tomosynthesis acquisition parameters on image texture: A study based on an anthropomorphic breast tissue software model," in *the Proceedings of 9th International Workshop on Digital Mammography* (Springer, Berlin/Heidelberg, Tucson, AZ, 2008), pp. 491–498.
- ²⁸ S. S. J. Feng and I. Sechopoulos, "A software-based x-ray scatter correction method for breast tomosynthesis," *Med. Phys.* **38**, 6643–6653 (2011).
- ²⁹ I. Sechopoulos, S. Suryanarayanan, S. Vedantham, C. J. D'Orsi, and A. Karellas, "Scatter radiation in digital tomosynthesis of the breast," *Med. Phys.* **34**, 564–576 (2007).
- ³⁰ A. K. W. Ma, D. G. Darambara, A. Stewart, S. Gunn, and E. Bullard, "Mean glandular dose estimation using MCNPX for a digital breast tomosynthesis system with tungsten/aluminum and tungsten/aluminum + silver x-ray anode-filter combinations," *Med. Phys.* **35**, 5278–5289 (2008).
- ³¹ N. Buls, I. Wathion, L. Mommaerts, C. Breucq, and J. de Mey, "Development of a multi-spectral, multi-geometry computational model for X-ray breast imaging," *Radiat. Prot. Dosim.* **139**, 343–347 (2010).
- ³² R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 2nd ed. (Prentice Hall, Upper Saddle River, N.J., 2002).

- ³³T. Pavlidis, *Algorithms for Graphics and Image Processing* (Computer Science, Rockville, MD, 1982).
- ³⁴R. Chandrasekhar and Y. Attikiouzel, "A simple method for automatically locating the nipple on mammograms," *IEEE Trans. Med. Imaging* **16**, 483–494 (1997).
- ³⁵T. F. Cootes and C. J. Taylor, "Statistical models of appearance for medical image analysis and computer vision," *Proc. SPIE* **4322**, 236–248 (2001).
- ³⁶I. T. Jolliffe, *Principal Component Analysis*, 2nd ed. (Springer, New York, 2002).
- ³⁷D. C. Barber, "The use of principal components in the quantitative analysis of gamma camera dynamic studies," *Phys. Med. Biol.* **25**, 283–292 (1980).
- ³⁸F. Pedersen, M. Bergströme, E. Bengtsson, and B. Långström, "Principal component analysis of dynamic positron emission tomography images," *Eur. J. Nucl. Med. Mol. Imaging* **21**, 1285–1292 (1994).
- ³⁹D. Rueckert, A. Frangi, and J. Schnabel, "Automatic construction of 3D statistical deformation models using nonrigid registration," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI* (Utrecht, The Netherlands, 2001), pp. 77–84.
- ⁴⁰See supplementary material at <http://dx.doi.org/10.1118/1.4789579> for the necessary data to use the developed models.
- ⁴¹J. M. Boone, K. K. Lindfors, V. N. Cooper III, and J. A. Seibert, "Scatter/primary in mammography: Comprehensive results," *Med. Phys.* **27**, 2408–2416 (2000).
- ⁴²I. Sechopoulos, S. S. J. Feng, and C. J. D'Orsi, "Dosimetric characterization of a dedicated breast computed tomography clinical prototype," *Med. Phys.* **37**, 4110–4120 (2010).
- ⁴³S. S. J. Feng and I. Sechopoulos, "Clinical digital breast tomosynthesis system: Dosimetric characterization," *Radiology* **263**, 35–42 (2012).
- ⁴⁴T. Ojala, J. Nappi, and O. Nevalainen, "Accurate segmentation of the breast region from digitized mammograms," *Comput. Med. Imaging Graph.* **25**, 47–59 (2001).
- ⁴⁵C. Zhou, H. P. Chan, N. Petrick, M. A. Helvie, M. M. Goodsitt, B. Sahiner, and L. M. Hadjiiski, "Computerized image analysis: Estimation of breast density on mammograms," *Med. Phys.* **28**, 1056–1069 (2001).
- ⁴⁶M. M. Goodsitt, H. P. Chan, B. Liu, S. V. Guru, A. R. Morton, S. Keshav-murthy, and N. Petrick, "Classification of compressed breast shapes for the design of equalization filters in x-ray mammography," *Med. Phys.* **25**, 937–948 (1998).
- ⁴⁷R. J. Ferrari, R. M. Rangayyan, J. E. L. Desautels, R. A. Borges, and A. F. Frere, "Identification of the breast boundary in mammograms using active contour models," *Med. Biol. Eng. Comput.* **42**, 201–208 (2004).
- ⁴⁸S. M. Kwok, R. Chandrasekhar, Y. Attikiouzel, and M. T. Rickard, "Automatic pectoral muscle segmentation on mediolateral oblique view mammograms," *IEEE Trans. Med. Imaging* **23**, 1129–1140 (2004).