

Journal of Medical Imaging

MedicalImaging.SPIEDigitalLibrary.org

Evaluation of non-Gaussian statistical properties in virtual breast phantoms

Craig K. Abbey
Predrag R. Bakic
David D. Pokrajac
Andrew D. A. Maidment
Miguel P. Eckstein
John M. Boone

Evaluation of non-Gaussian statistical properties in virtual breast phantoms

Craig K. Abbey,^{a,*} Predrag R. Bakic,^b David D. Pokrajac,^c Andrew D. A. Maidment,^b Miguel P. Eckstein,^a and John M. Boone^d

^aUniversity of California, Santa Barbara, Department of Psychological and Brain Sciences, Santa Barbara, California, United States

^bUniversity of Pennsylvania, Department of Radiology, Philadelphia, Pennsylvania, United States

^cDelaware State University, Department of Computer and Information Sciences, Dover, Delaware, United States

^dUniversity of California at Davis, Department of Radiology, Sacramento, California, United States

Abstract. Images derived from a “virtual phantom” can be useful in characterizing the performance of imaging systems. This has driven the development of virtual breast phantoms implemented in simulation environments. In breast imaging, several such phantoms have been proposed. We analyze the non-Gaussian statistical properties from three classes of virtual breast phantoms and compare them to similar statistics from a database of breast images. These include clustered-blob lumpy backgrounds (CBLBs), truncated binary textures, and the UPenn virtual breast phantoms. We use Laplacian fractional entropy (LFE) as a measure of the non-Gaussian statistical properties of each simulation procedure. Our results show that, despite similar power spectra, the simulation approaches differ considerably in LFE with very low scores for the CBLB to high values for the UPenn phantom at certain frequencies. These results suggest that LFE may have value in developing and tuning virtual phantom simulation procedures. © The Authors. Published by SPIE under a Creative Commons Attribution 4.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.JMI.6.2.025502](https://doi.org/10.1117/1.JMI.6.2.025502)]

Keywords: Laplacian fractional entropy; breast phantoms; image statistics; natural scene statistics.

Paper 19014R received Jan. 11, 2019; accepted for publication May 20, 2019; published online Jun. 14, 2019.

1 Introduction

Virtual breast phantoms have many appealing qualities for evaluating breast imaging technology. These phantoms are the output of simulations of breast tissue, often with the intent of capturing the effect of anatomical variability on the performance of imaging systems. As such, the “phantom” may be considered as the ensemble of image backgrounds produced by the simulation procedure, with additional acquisition noise simulated as part of a specific implementation. Virtual phantoms have advantages that arise from known ground truth about the object being imaged and the ability to evaluate a not-yet-existent (or otherwise inaccessible) imaging system through a simulation of the imaging process.

However, in order to be effective, a virtual phantom must accurately capture the effects of patient-structured images. We refer to this objective as phantom realism, with the assumption that a more realistic phantom will more fully capture anatomical effects than a less realistic phantom. But it is not clear at this point in time how to validate such a comparison or how such a validation might depend on the task that drives development of the imaging system.

The approach we explore in this work is based on the idea that realism can be characterized by comparing statistical properties of the phantom ensemble (or the subsequent images) and actual patient images in cases where such images are available. Traditionally, statistical properties have been limited to the mean and power spectrum,^{1–3} which characterize first- and second-order statistical properties (e.g., mean, variance, covariance), under the assumption of stationarity. It has been well established that the distribution of soft tissue in breast images has multiscale

characteristics^{4–7} that lead to a power law governing the power spectrum of anatomical variability. The power spectrum of breast images is well approximated as the sum of this power law with the power spectrum of quantum noise generated during the process of image acquisition. This relationship was first observed in film-screen mammography⁸ and later extended to full-field digital mammography (FFDM)⁵ and other breast imaging modalities, such as digital breast tomosynthesis^{9,10} and dedicated breast computed tomography (CT),^{11–13} albeit with different exponents, demonstrating the ubiquity of the power-law formalism. As a result, the power-law spectrum has served as a useful constraint for stochastic simulations of breast anatomy in virtual breast phantoms.

However, it is also known that the power spectrum does not fully capture the appearance of most medical images.¹⁴ For example, it is generally easy to differentiate between a real patient mammogram and a Gaussian process that has matched the power spectrum of the mammograms. Figure 1 gives examples of each of these processes. The top row of the figure shows small regions of x-ray projection mammograms from clinical exams that have been log-converted to density. Note that we have not applied the usual “for-display” processing to these images since we are interested in the properties of the objects being imaged and not necessarily the final displayed image. We contend that textural components of the mammograms make them readily discriminable from the Gaussian textures in the bottom row of the figure.

Figure 1 suggests that if the statistical-properties approach to evaluating phantom realism is going to be most effective, we will likely need to evaluate properties beyond the mean and power spectrum that fully characterize a stationary Gaussian process. We consider these to be “non-Gaussian” statistics that assess higher-order structure in the images. We investigate one

*Address all correspondence to Craig K. Abbey, E-mail: ckabbey@ucsb.edu

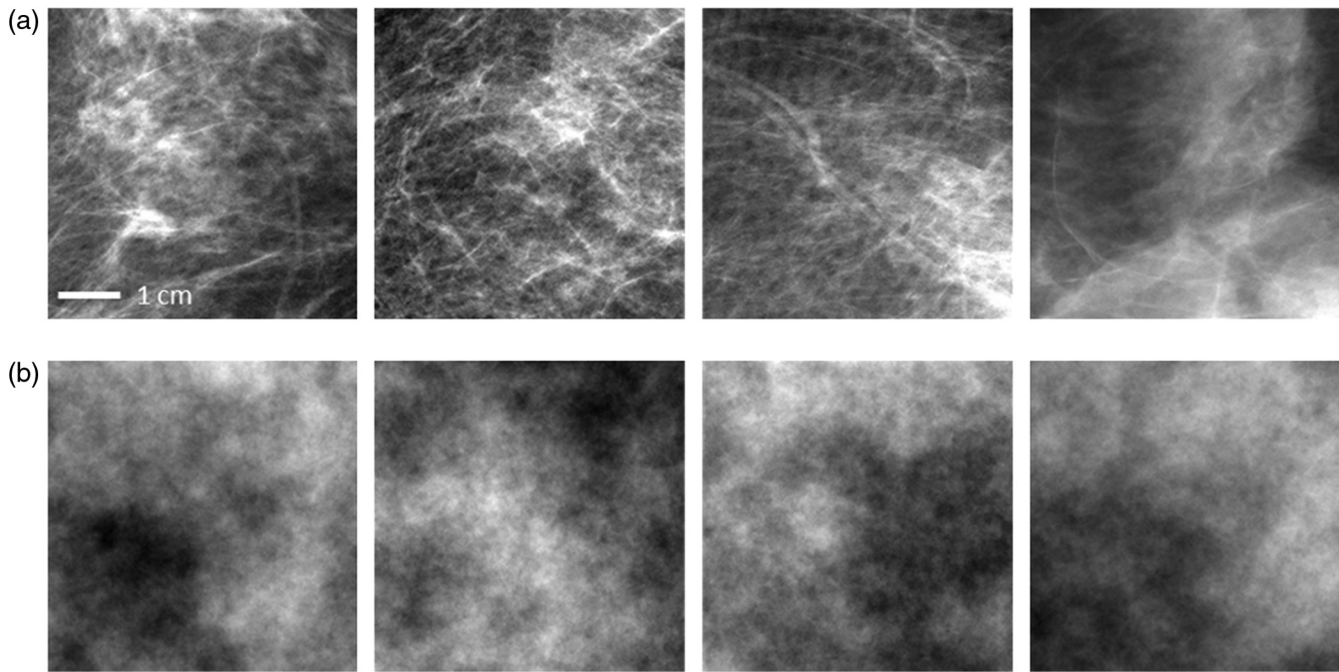


Fig. 1 Mammographic backgrounds and Gaussian stochastic textures. (a) The $5\text{ cm} \times 5\text{ cm}$ patches of mammographic density images. (b) Gaussian stochastic textures that match the mean and power spectrum of mammographic densities in the breast interior.

such measure, known as Laplacian fractional entropy (LFE), for this purpose.¹⁴ LFE is based on the response histograms of Gabor filters, which are intended to represent receptive fields in early visual cortex.^{15–17} Much like visual receptive fields, the measure can be tuned to different center frequencies and orientations. The LFE measure computes the entropy of a response histogram relative to the expected histogram from a Gaussian process matched for mean and variance. Relative entropy will be zero when the two histograms fully correspond, which occurs when the response histogram is consistent with a Gaussian distribution. This makes the measure insensitive to Gaussian statistics. The Laplacian distribution (two-sided exponential) is used as a yardstick for measuring how much a response histogram deviates from Gaussian form. An LFE value of 100% is interpreted as being as “non-Gaussian” as a Laplacian distribution.

The question we would like to address in this work is whether a higher-order statistical property, LFE in this case, provides additional information over the power spectrum regarding phantom realism. For simplicity, we use FFDM images as a test case for this hypothesis. We examine LFE in three classes of virtual breast phantoms comprising nine different simulations of FFDM. These include virtual breast phantoms based on clustered-blob lumpy backgrounds (CBLBs),^{18,19} binary truncation (BT) processes,^{10,20} and the UPenn virtual breast phantom.^{21–23} For comparison, we also have patient images from a set of 35 clinical breast exams that we use to estimate the anatomical noise in FFDM images of patients. In addition, we also evaluate a stationary Gaussian process with a power-law power spectrum at the two-pixel sizes used in this work (70 and 100 μm). Our approach is different than what was used in our preliminary conference report,²⁴ in that we determine the acquisition noise-power spectrum in each simulation by matching the power spectrum of the FFDM images. This allows us to see what additional information is learned by evaluating the non-Gaussian statistics.

2 Materials and Methods

2.1 Mammograms

The mammograms we use were obtained from the UC Davis Medical Center and the University of Pennsylvania Medical Center under IRB-approved human-subject protocols that included deidentifying all image data. Images from 26 patients came from the UC Davis Medical Center and images from 9 patients came from the University of Pennsylvania Medical Center. Patient images were selected on the basis of having complete data, which consisted of two standard views (CC and MLO) of each breast for a total of 140 images. All of the mammograms were acquired on Hologic Dimensions systems (Hologic Inc., Bedford, Massachusetts) as part of breast cancer-screening exams. However, patient images in our sample were collected retrospectively on the basis of a recommendation for biopsy, and so they should not be considered a representative sample of the screening population.

A subset of 26 of these exams (acquired at UC Davis) was used previously in the original LFE paper.¹⁴ All breast image data were saved at the “for-processing” stage, after detector calibration had been performed, but before any display processing had been applied. This step was done to avoid any non-Gaussian statistics that might be introduced by transformations in the “for-display” processing.¹⁴ The images were log-converted from transmission to density before the analysis was performed. Pixel size in the mammograms was 70 μm .

2.2 Clustered-Blob Lumpy Backgrounds

The CBLBs have been developed by Bochud et al.^{18,19} as an extension of Rolland’s and Barrett’s type I lumpy backgrounds²⁵ for applications in x-ray projection mammography.

The idea behind lumpy backgrounds was that a background could be generated as a superposition of lump profiles placed at different locations throughout the background area. The CBLB extended this by allowing the lump profile to be a localized probability density for a cluster made up from smaller lumps.

A total of five previously published CBLB simulation procedures are considered here. These include the process described in the original publication¹⁸ and referred to as “OpEx99,” as well as four virtual breast phantoms developed in a subsequent publication.¹⁹ These are all sampled on a 100- μm -pixel size that was used in the publications. Two virtual phantoms based on oriented and isotropic processes are referred to as “SimpOri” and “SimpIso,” respectively, with “Simp” indicating that a single CBLB process is used. The remaining two phantoms both sum two CBLB processes modeling glandular and fibrous structures, respectively. These also incorporate oriented and isotropic processes and are referred to as “DoubOri” and “DoubIso,” with “Doub” indicating that a double CBLB process is used. The parameters for these processes are all taken directly from the original publications, with the only change being to scale the number of clusters to match the larger image size (1024 \times 2048 instead of 256 \times 256) tested. For each phantom, 50 independent backgrounds are generated for analysis. Figure 2 gives examples of each of these processes [Figs. 2(a)–2(e)] along with a power-law Gaussian texture [Fig. 2(h)], sampled on a 100- μm -pixel grid, for reference.

It should be noted that the CBLB phantoms directly simulate a two-dimensional (2-D) image. This is considerably simpler than the methods below, which simulate a three-dimensional (3-D) breast and then project the 3-D simulation onto a 2-D detector.

2.3 Binary Truncation Processes

The BT phantoms are based on the idea that the breast can be well modeled as mixture of adipose and glandular tissues, each characterized by its respective attenuation coefficient. These models of soft tissue contrast have typically neglected calcifications. The spatial distribution of the attenuation coefficients is generated by thresholding a 3-D random process, initially developed by Gong et al.²⁶ The approach of Abbey and Boone²⁰ (referred to as AB08) uses a 3-D Gaussian process truncated to $\sim 30\%$ volume glandular fraction. Reiser and Nishikawa¹⁰ (referred to as RN10) used phase randomization followed by thresholding to generate the distribution of adipose and glandular tissues and targeted a 75% glandular fraction.

It should be noted that both of these approaches have higher glandular fractions than the 19% average volume glandular fraction of the breast, as measured from breast CT images.²⁷ Both methods could be easily modified to meet this average glandular fraction by increasing the truncation threshold. There are other differences between the two processes (how they generate and truncate the 3-D random field) that can affect LFE. However, our interest is in evaluating whether LFE changes substantially between the different simulation procedures, and so we have adhered as closely as possible to the simulation procedures used in the publications.^{10,20}

Binary 3-D backgrounds were generated with an isotropic sampling of 100 μm in x , y , and z . The resulting grid was 1024 \times 2048 \times 512, and thus approximately captured the 5-cm average thickness of a compressed breast in a screening mammography exam while achieving the size needed for a comparable region of interest (ROI) of the breast interior relative to FFDM images. Phantoms were projected onto the detector by summing the z -dimension of the 3-D phantom for a

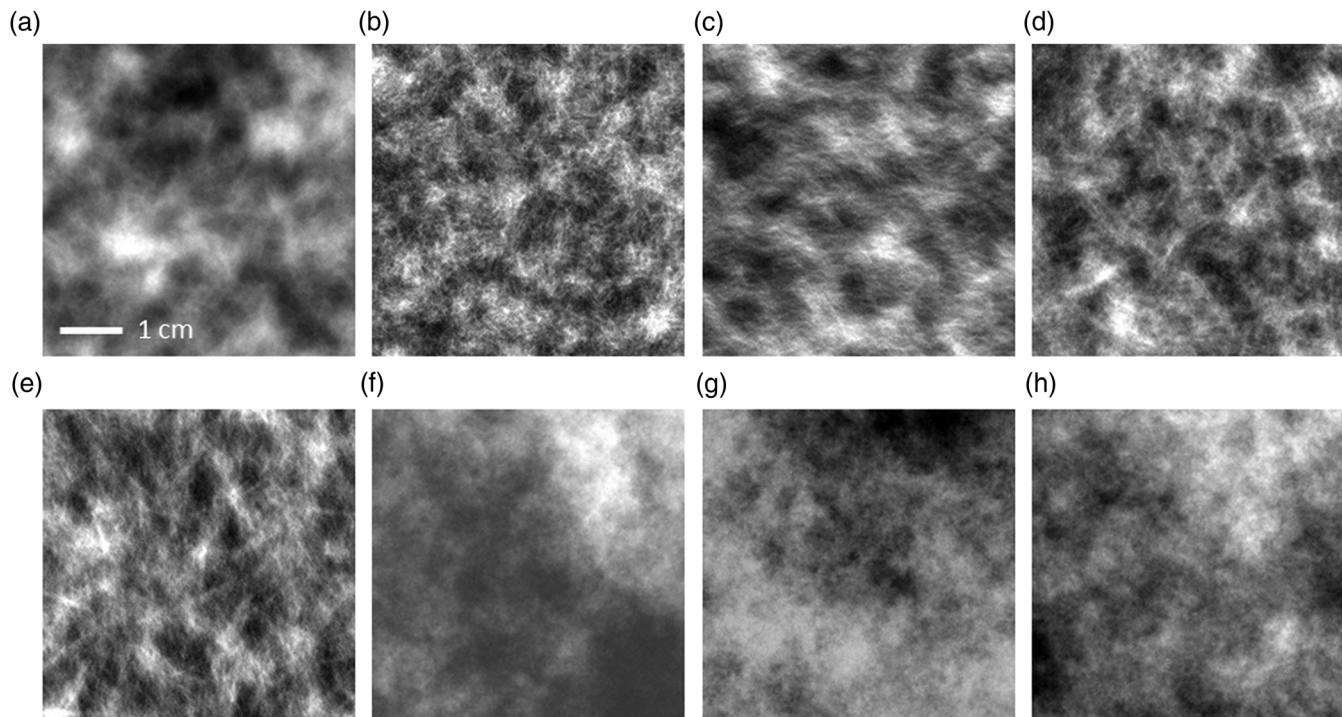


Fig. 2 CBLBs and BT backgrounds. Samples of CBLBs are shown (a–e) along with those generated using BT processes (f, g). A Gaussian texture generated at 100 μm sampling (h) is also shown for reference.

1024 × 2048 final image size. For each of the BT phantoms, a total of 50 independent backgrounds were generated for analysis. Examples of BT phantoms from these two processes are shown in Figs. 2(f) and 2(g).

2.4 UPenn Virtual Breast Phantom

A virtual breast phantom development project has been underway at the University of Pennsylvania for several years.^{21–23} This approach is based on simulating the major anatomical components of the breast, including skin, adipose tissue, fibroglandular tissue, and Cooper’s ligaments. The breast simulation is initially generated in an uncompressed state, to which a compression transform is applied using a finite element model.

We consider two versions of the UPenn phantom. One is the first published version of the phantom, tailored for projection mammography.²² In the second version, a new approach is used that involves simulating additional microstructure in the adipose compartments.²⁸ We refer to the original simulation procedure as the “coarse” phantom because of the strong texture of the projection images. The newer simulation is referred to as the “mixed” phantom because of the mixing of attenuation components in the adipose compartments. Mammograms are simulated from the 3-D phantom, assuming a 950-mL breast volume with 6.4 cm compressed breast thickness, a polyenergetic x-ray beam, and 70- μm detector pixels similar to the Hologic detector. The phantoms are generated at an image size of 3328 × 4096 pixels. For each of the UPenn phantoms, 50 independent backgrounds are generated for analysis. Example images from these two phantoms are given in Fig. 3, along with a power-law Gaussian texture sampled on a 70- μm pixel grid.

2.5 Image Power Spectra

This work makes extensive use of power spectra for both the sample images and the simulation procedures. Here, we briefly describe the windowed spectral estimation procedure we use. All image analysis (both power spectra and LFE) focuses on the ROIs within the image. Examples of the interior regions for the different phantoms are shown in Fig. 4. For the FFDM images, these ROIs are generated manually to indicate the interior of the breast, which is taken to be ~ 10 mm inside the skin line of the mammogram and avoiding pectoral muscle, ribs, skin-folds, etc. The average area of the ROIs, across all patients

and views, is 116.3 cm² and ranges from 15.9 to 308.4 cm². For the CBLB and BT phantoms generated on a 100- μm grid and an overall simulation size of 1024 × 2048, we used a central rectangle with an area of 112.3 cm². This ROI is also used for the Gaussian process sampled on the same scale. For the UPenn phantoms (and corresponding 70- μm Gaussian phantom), the ROI is also generated similar to the FFDM images. Since the UPenn phantoms all have the same breast contour with randomized interior structure, a single ROI (106 cm²) is appropriate for all of these phantoms.

The power spectrum is estimated from square image patches that are randomly selected so that the entire patch falls within the interior of the breast. Let $g_i[n, m]$, $n, m = 0, \dots, N - 1$, be the sample patch from the i ’th FFDM image or the i ’th image generated by one of the simulation procedures ($i = 1, \dots, I_{\text{Samp}}$). The patch size is fixed at 512 × 512 pixels (i.e., $N = 512$). For the FFDM images where the breast was too small to fit, the patches are excluded from the power-spectrum estimate. This constraint leaves 34 cases for use in each of the MLO views, and 33 cases for each of the CC views. The simulations all use 50 sample images. A windowed power spectrum is estimated for each ROI in a sample image as

$$\hat{S}_i[k, l] = |\text{FFT}(w[n, m](g_i[n, m] - \bar{g}[n, m]))|^2, \quad (1)$$

where $\bar{g}[n, m]$ is the average over all the patches and $w[n, m]$ is a spatial window that is centered in the middle of the patch and remains constant out to a radius of 10.0 mm before rolling off to zero with a cosine profile at a radius of 17.9 mm. The spatial frequency indices $[k, l]$ both range over $k, l = 0, \dots, N - 1$. The window function reduces spectral leakage in the power-spectrum estimate from periodic discontinuities at the edges of the patch.²⁹ To ensure that the power spectrum will have approximately the right amplitude, the window is normalized to have an average value of 1.0 over the patch.

We use radial averages to convert the 2-D power spectra in Eq. (1) to a one-dimensional function of radial frequency, consistent with previous works.^{4,6,30} This has typically been justified on the basis of approximate isotropy of power spectra for breast images, although other investigators have shown orientation dependence.³¹ Since our study is focused on the power spectrum at a more general level, we use radial averages and neglect any orientation dependence in our data. Let $f_{k,l}$ represent the radial frequency for the indices, then

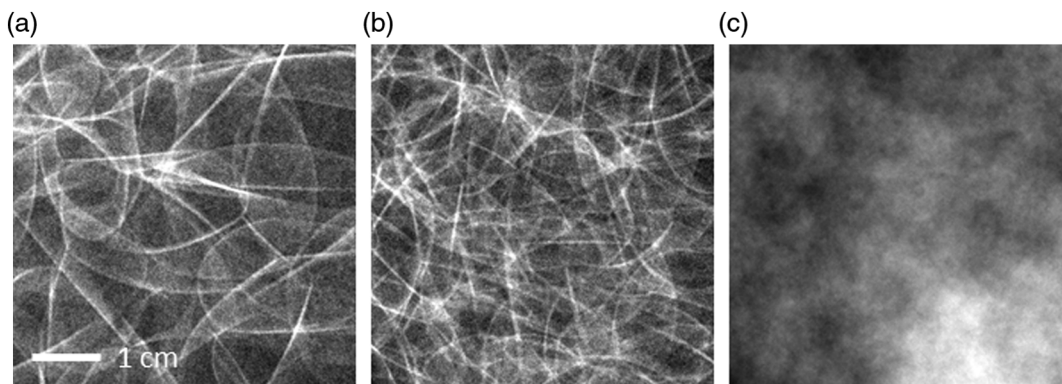


Fig. 3 Phantoms generated at 70 μm sampling. Samples from two versions of the UPenn virtual breast phantom are shown (a, b) along with a Gaussian texture (c) generated at 70- μm pixel size and shown for reference.

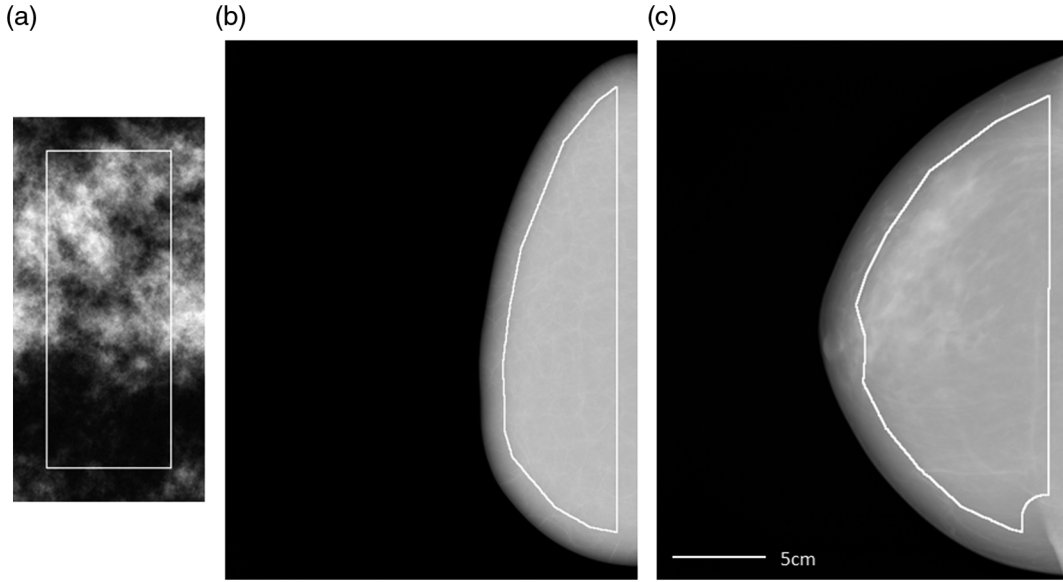


Fig. 4 ROIs used for analysis. Each image shows a phantom with the ROI indicated. (a) The CBLBs and BT processes used a rectangular ROI with an area of 112 cm². (b) The UPenn phantoms and (c) FFDM patient images used manually determined the ROIs. The UPenn phantom ROI has an area of 106 cm². The ROI in the FFDM image shown (LCC view) is roughly the average size of the area in the image set (111 cm²). In this image, a small skin fold in the lower right corner is explicitly avoided in the ROI.

$$f_{k,l} = \frac{\sqrt{k_w^2 + l_w^2}}{N\Delta},$$

where N is the patch size (512), Δ is the pixel size (0.1 or 0.07 mm), and k_w and l_w account for frequency “wrap-around” (i.e., $k_w = k$ for $0 \leq k < N/2$, and $k_w = k - N$ for $N/2 \leq k < N$, likewise for l_w).

Let q be the radial frequency index ($q = 0, \dots, N/2$) with associated radial frequency $\rho_q = q/N\Delta$. We define F_q as the set of all 2-D frequency index pairs in a band centered on ρ_q that has a width of one radial frequency sample (i.e., $F_q = \{[k, l] : -1/2N\Delta \leq f_{k,l} - \rho_q < 1/2N\Delta\}$); let N_q be the number of index pairs in the band, then the radial power spectrum is given as

$$\hat{S}_i[q] = \frac{1}{N_q} \sum_{[k,l] \in F_q} \hat{S}_i[k, l]. \quad (2)$$

The radial power spectrum for the process is generated by averaging over the image samples. In addition, we find it useful to work with the log-power spectrum at times. We generate the average log-spectrum by averaging over samples as well

$$L\hat{S}[q] = \frac{1}{I_{\text{Samp}}} \sum_{i=1}^{I_{\text{Samp}}} \log_{10}(\hat{S}_i[q]). \quad (3)$$

The sample standard deviation across i is used to quantify the variability of the log-power spectrum

$$\hat{\sigma}_{L\hat{S}[q]} = \left(\frac{1}{I_{\text{Samp}} - 1} \sum_{i=1}^{I_{\text{Samp}}} (\log_{10}(\hat{S}_i[q]) - L\hat{S}[q])^2 \right)^{1/2}. \quad (4)$$

2.6 Simulating Acquisition Noise

The above simulation procedures describe how the anatomical components of the phantom images are generated. However, a full simulation should also include the generation of acquisition noise. In the initial conference report of this study,²⁴ acquisition noise was modeled as white Gaussian noise. While this had the correct qualitative effect of flattening the power spectrum at high frequency, it is noticeably different than the actual power spectra of digital images. In addition, the different simulation procedures are normalized differently leading to large differences in the amplitude of the power spectra.

In this study we model the acquisition noise as another power-law process that is fitted to the power spectrum of the FFDM images over a frequency range from 0.1 cyc/mm to the Nyquist limit. The noise process has two free parameters: a power-law amplitude, A_{Nse} , and exponent, β_{Nse} . We also scale the simulation processes by a factor A_{Sc1} to have approximately the same amplitude as the FFDM images. These effects are achieved by fitting the power spectra of the simulation procedures to the power spectrum of the FFDM images. Let $\hat{S}_{\text{Bck}}[k, l]$ be the power spectrum estimated for one of the background simulation procedures. We fit the simulation power spectrum to the image spectrum using a three-parameter model

$$\hat{S}_{\text{Sim}}[q; \boldsymbol{\theta}] = A_{\text{Sc1}} \hat{S}_{\text{Bck}}[q] + A_{\text{Nse}} \rho_q^{\beta_{\text{Nse}}}, \quad (5)$$

where $\boldsymbol{\theta} = (A_{\text{Sc1}}, A_{\text{Nse}}, \beta_{\text{Nse}})$ is a vector comprising the three free parameters to be fit. Note that for $q = 0$, we replace ρ_0 with ρ_1 to avoid singularity. The parameters are fit by minimizing the weighted least squares error with the log-power spectrum of the FFDM images, $L\hat{S}_{\text{FFDM}}[q]$, according to

$$\text{Err}(\boldsymbol{\theta}) = \sum_{q=0}^{N/2} \frac{(L\hat{S}_{\text{FFDM}}[q] - \log_{10}(\hat{S}_{\text{Sim}}[q; \boldsymbol{\theta}]))^2}{\hat{\sigma}_{L\hat{S}[q]}^2}, \quad (6)$$

which is readily solved using standard optimization methods.

2.7 Laplacian Fractional Entropy

LFE is based on the relative entropy (K - L divergence) between the response histogram of a Gabor filter, normalized to be a probability distribution, and a best-fit Gaussian distribution over the same histogram bins. If the histogram represents an underlying Gaussian process, then it will be well fit by the Gaussian distribution, and the relative entropy will be negligible. Any sort of departure from a Gaussian distribution in the histogram will lead to nontrivial relative entropy. This relative entropy is normalized by the relative entropy of the best-fit Laplacian distribution with respect to this Gaussian. In this way, the Laplacian relative entropy serves as a yardstick for interpreting the Gaussian relative entropy. As a result, an LFE value of 100% means that the histogram exhibits as much departure from a Gaussian profile as the Laplacian distribution does. In principle, any non-Gaussian distribution could be used in place of the Laplacian. However, Laplacian distributions have played an important role in the analysis of natural scenes.³²⁻³⁴

The LFE was computed according to the procedure specified in a previous publication.¹⁴ Gabor filters that spanned 11 center frequencies from 0.125 to 4.0 cyc/mm were evaluated at six different orientations (0 deg, 30 deg, ..., 150 deg). The filters were sine-phase with a bandwidth of 1.4 octaves and an aspect ratio of 1. Filter responses were generated by convolving each filter with an image. All responses from 1 cm inside the boundary of the ROI were used to form response histograms from which the LFE measures at different frequencies were estimated. The histograms binned the central 99% of the responses, with an additional 1% bin for the remaining extremal values.

The computational scheme for determining LFE for a particular Gabor filter in a given image is shown in Fig. 5, resulting in an LFE value of 63%.

For the FFDM images, LFE estimates from a single filter are averaged across the six orientations and across the four views for each patient. The result is averaged over patients, to get a final estimate of LFE at each spatial frequency. The sample standard deviation over patients is used to quantify the variability of LFE. For the simulation processes, we average across orientations and samples, to get an estimate of mean LFE for the process. The standard error of the mean (across samples) is typically quite small (<4% of the mean on average).

3 Results and Discussion

The main results of this work are power spectra and LFE plots averaged over orientations and plotted as a function of spatial frequency.

3.1 Power Spectra

A comparison of radially averaged power spectra for simulation procedures and FFDM images are shown in Fig. 6. The simulation spectra have been fitted using the procedure described in Sec. 2.6. Each plot shows the results for a different class of simulation procedures. The fitted amplitudes of the acquisition power spectra ranged from 156.8 to 371.5, while the power-law exponents of the acquisition noise models ranged between -0.27 and -0.90 .

In most cases the fitting procedure results in good agreement between the fitted model and the FFDM data. The average relative error across all frequencies in all the simulation procedures is 6%. The largest divergence between fitted model and data occurs for the OpEx99 model of the CBLB [Fig. 6(a)] at low spatial frequencies (<0.3 cyc/mm) where relative error is over 100%, but these drop below 10% at higher frequencies. Fitted spectra from the binary processes [Fig. 6(b)] and the

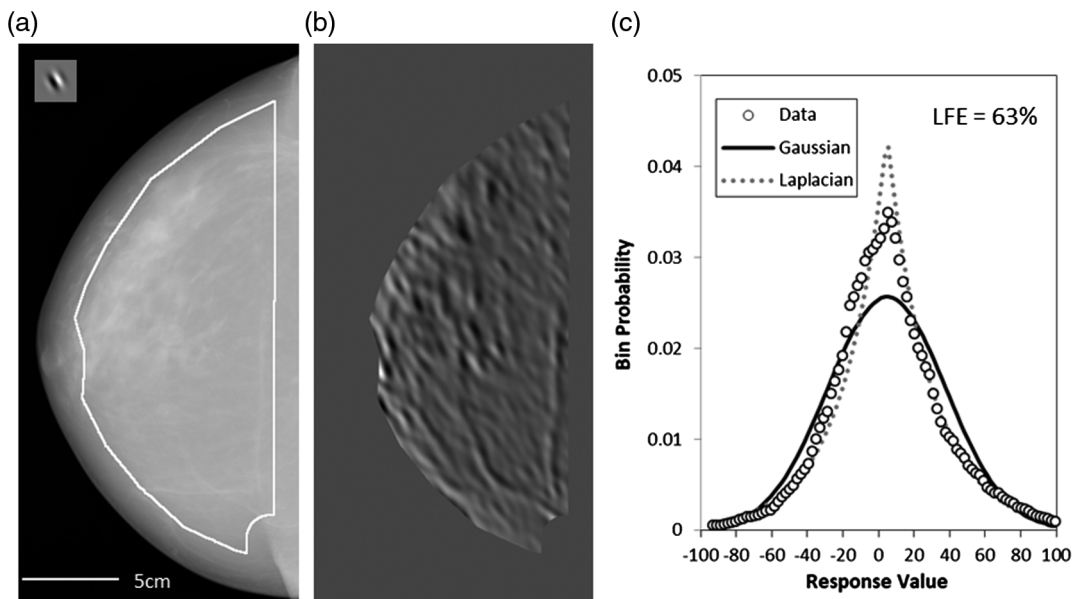


Fig. 5 Computation of LFE. The image (a) is convolved with a Gabor filter (upper left corner), and all filter responses in the ROI are retained (b). (c) The resulting histogram is normalized to a probability distribution by dividing by the total number of responses and is shown with fitted Gaussian and Laplacian distributions. The LFE measure is the relative entropy of the data to the Gaussian divided by the relative entropy of the Laplacian to the Gaussian, which is 63% in this case.

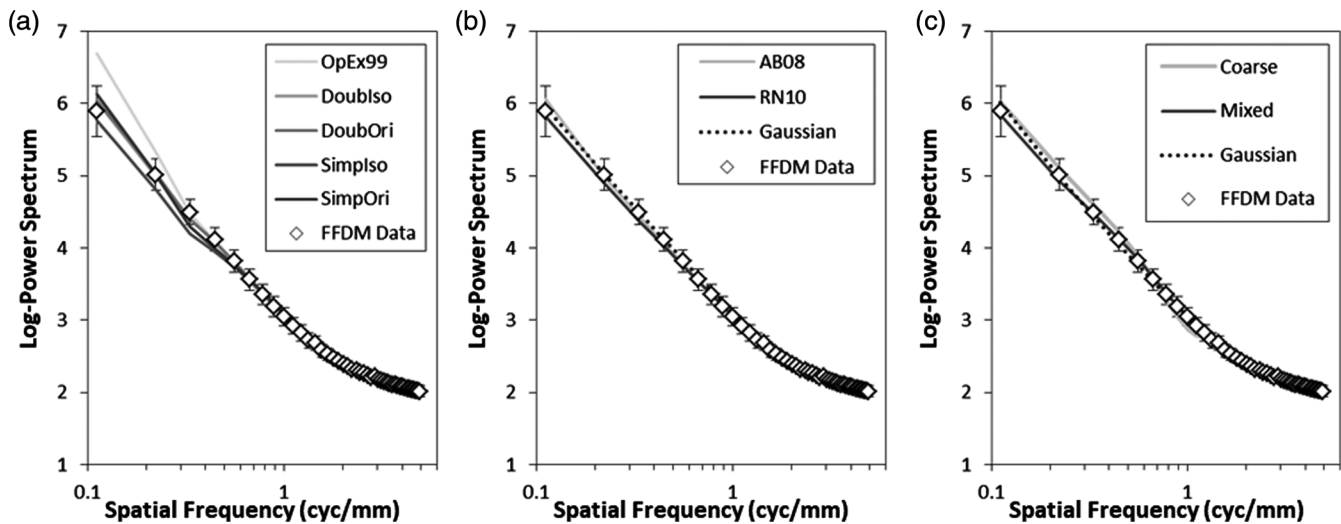


Fig. 6 Radial power spectra. Radial averages of the power spectrum are shown for each of the simulation approaches. Each of the phantom procedures has been scaled and has Gaussian noise added to match the observed FFDM power spectrum across a frequency range of 0.1 to 5 cyc/mm. The FFDM data are shown on each plot for reference with error bars of ± 1 sd for comparison.

UPenn breast phantom [Fig. 6(c)] agree with the sample estimates across the tested frequency range. The Gaussian processes used for reference [Figs. 6(b) and 6(c)] also agree well across the spectrum.

Measured from spatial frequencies between 0.1 and 1.0 cyc/mm, the power-law exponent of the FFDM power spectrum is $\beta = -3.01$, which is consistent with previously published results.⁶⁻⁸ With the exception of the OpEx99 process ($\beta = -3.81$), the relative errors in the power-law exponents are within 10% of the sample estimate. The Gaussian processes, generated with a power-law slope of 3.0, result in relative errors of <3%, which suggests that the spectral estimation procedure is not introducing a substantial bias in the power-law exponent.

Even with some departures at low frequencies by the CBLBs, the general finding is that the power spectra of the virtual phantoms are reasonably well fit by a power-law over the frequencies

in which anatomical variability dominates, even though the various phantoms have quite different visual appearances. This is a testament to the ubiquitous nature of the power-law spectrum. However, when the goal is to distinguish between various models, the generality of the power law may work against its usefulness. This serves as the motivation for investigating higher-order statistical properties.

3.2 Laplacian Fractional Entropy

LFE results for the various simulation procedures are shown in Fig. 7. Here again, the FFDM results for left CC view images are plotted (with ± 1 sd error bars across patients) for comparison with each class of simulation procedure. Other views have similar LFE values. The general profile of the LFE curve is similar to previous reports,¹⁴ which is not surprising, given that 26 of

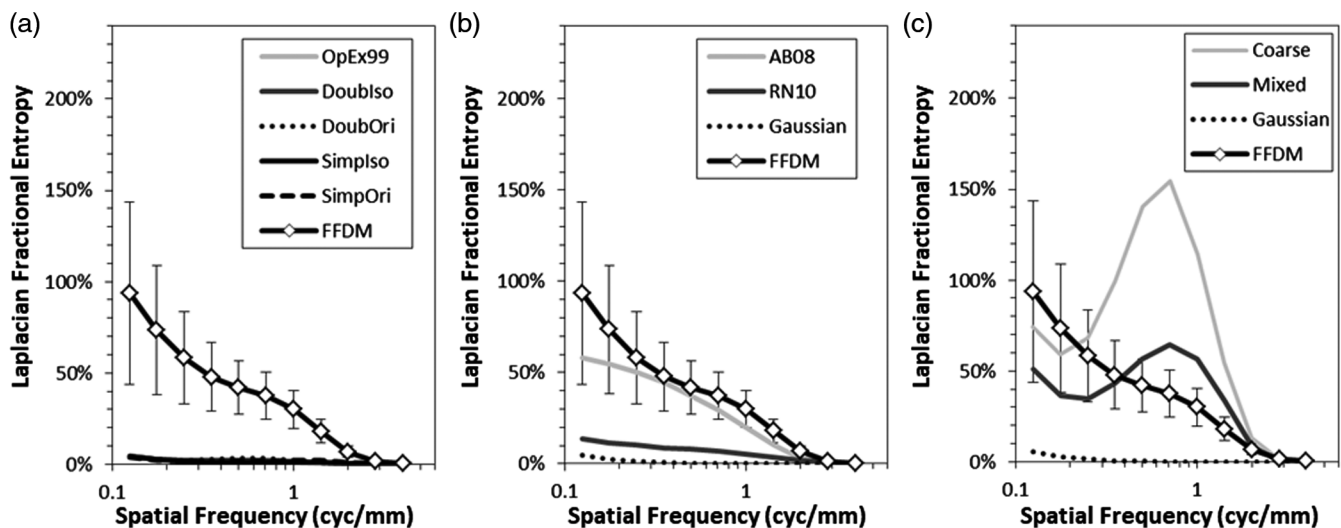


Fig. 7 LFE plots. Plots are shown for (a) the CBLBs, (b) BT backgrounds, and (c) two versions of the UPenn virtual breast phantom. The LFE data for the patient FFDM images are shown on each plot for reference, with error bars of ± 1 sd for comparison. In the binary processes plot (b), there is also a plot from the Gaussian process generated on a 100- μ m pixel size for reference and, on the UPenn virtual breast phantom plot, there is also a plot for the Gaussian process generated on a 70- μ m pixel size.

the 35 cases were used for those results. LFE appears to be dropping across the lower frequencies with a mild shoulder around 1 cyc/mm, where acquisition noise begins to dominate, driving the measure to zero.

Figure 7(a) shows LFE plots for the CBLBs, along with a matched Gaussian texture. These plots all have fairly small values across the frequency range (<5%). These are similar to the matched Gaussian process plotted in Fig. 7(b). Nonzero values of LFE for the Gaussian textures reflect departures from the Gaussian profile in the histogram due to the limited spatial extent of the ROI used to compute the histogram and the finite sample of histogram responses.

None of the CBLBs show any substantial non-Gaussian structure by the LFE measure. This is somewhat surprising given that the procedure for generating the phantoms is rigorously non-Gaussian.^{18,19} However, it appears that—seen through the lens of Gabor receptive fields—there is little higher-order structure in these backgrounds. Given that the number of background “blobs” in a simulation are generated using sums over Poisson random deviates, it may be that for the frequencies we are testing, Poisson variability is well approximated as Gaussian. It is also possible that substantial non-Gaussian structure in the images is not being captured in the Gabor-filter histograms used by the LFE measure. Further analysis is needed to resolve this issue.

Figure 7(b) shows LFE plots for the phantoms based on the BT processes, along with the matched (100 μm) Gaussian texture and the FFDM samples. The AB08 process results in higher LFE than the RN10 process, and this is mostly due to more regions of uniform intensity, which add to the response histograms near zero. This explains the increased LFE observed here compared to the earlier conference report.²⁴ Future implementations and optimizations of these processes should carefully consider the simulation parameters including sampling resolution and approaches to dealing with power-law processes near $f = 0$.

Figure 7(c) shows LFE plots for the UPenn phantoms along with the matched (70 μm) Gaussian texture, and the FFDM sample estimates. As expected, the Gaussian texture has very low LFE values similar to the 100 μm Gaussian texture plotted in Fig. 7(b). This serves as a check that the different pixel sizes do not produce any unexpected results. The coarse UPenn phantom shows a substantial peak in LFE at frequencies between 0.3 and 1.5 cyc/mm that is not present in the FFDM data. This peak is substantially modulated in the mixed UPenn phantom leading to a better agreement with the FFDM data over the entire spectrum.

The UPenn phantom results show how LFE can be used to guide the development of virtual breast phantom procedures. Both the coarse and the mixed phantoms can be made to closely match the power spectrum of breast images, which means that the power spectrum is not a useful way to identify which one is closer to the real images. In contrast, LFE shows a substantial difference between the two and suggests that mixing in the adipose compartments is moving the phantom in a more realistic direction. This is not entirely surprising given that there are physiological reasons to expect a distribution of attenuation values within a glandular compartment, but it is advantageous to have a quantitative measure that reflects this.

As a way to quantify how closely the phantom LFE matched the FFDM LFE at frequencies of interest, we compare differences in the average phantom LFE and FFDM LFE relative to the standard deviation in LFE across patients in the FFDM data. For frequencies between 0.125 and 1 cyc/mm, the average

error of the AB-08 phantom relative to the standard deviation of the FFDM LFE is 52%, and for the mixed UPenn phantom it is 125%. For the RN-10 phantom, the coarse UPenn phantom, all the CBLB phantoms, and the Gaussian processes, the errors are >200%.

The LFE results presented here are somewhat different than in our initial results.²⁴ This demonstrates the potential impact of varying the simulation procedures. For this reason, and others noted below, we hesitate to accept this as a reflection of the general validity of the virtual phantom procedures.

3.3 Study Limitations

Our study is fairly limited, and hence we caution against the overinterpretation of our results, and in particular the establishment of a “best” phantom on the basis of these results. The purpose of our study was to demonstrate that phantoms which have similar power-law spectra may exhibit considerable differences in higher-order statistics, and that LFE can be used to quantify those differences in a meaningful way. Comparison with the LFE derived from FFDM images from a limited set of patients scheduled for biopsy was made to give a general sense of how real FFDM images behave. This study was not designed to identify the most appropriate phantom. A more focused evaluation of phantom approaches should consider the population of patients from which sample images are used, and it would likely also be useful to allow for some training of parameters in the phantom procedure as well.

4 Conclusions

All of the virtual breast phantoms tested here have power spectra that resemble a power-law form at low spatial frequencies. This is generally consistent with the original findings of Burgess et al.⁷ and others^{5,8} for x-ray projection mammograms. At higher frequencies, near 1 cyc/mm and above, acquisition noise dominates and flattens the spectrum. With appropriate normalization, and the addition of acquisition noise with an appropriate power spectrum, the observed power spectra of all virtual breast phantoms could be made quite similar to the power spectrum of the mammograms. As a result, all of the phantoms may be considered consistent with the power-law model of breast anatomy. While this shows that the various phantoms have similar second-order statistical properties to clinical breast images, it also illustrates the limitations of the power spectrum as an endpoint for evaluating simulation models of the breast. The similarity of the phantoms at the level of the power spectrum along with clearly visible differences between the simulated images emphasizes the importance of higher-order structure in the appearance and realism of virtual breast phantoms.

LFE, which is relatively insensitive to the power spectrum of the phantoms, shows substantial variability between the different kinds of phantoms, and between the phantoms and the sample of FFDM images. This confirms the underlying motivation for this work, that the power-law spectrum should be considered a necessary condition for phantom realism, but not necessarily a sufficient condition. It also suggests a role for LFE in the optimization of phantoms realism beyond spectral similarity. This is exemplified in the UPenn virtual breast phantom results, where the use of an adipose/glandular mixing model within compartments results in closer agreement with the LFE of FFDM images, even though the power-law spectrum shows good agreement regardless of whether this component was used.

While our study was focused on the use of LFE as a measure of realism, one surprising incidental finding is noted. This is the relatively low LFE scores for all the CBLBs considered, which were difficult to distinguish (in LFE terms) from a Gaussian process. Further modeling and analysis is needed to understand how this finding comes about, and how the simulation procedure might be altered to make it more like breast images. For the present, we suggest some caution in relying on the CBLB to implement non-Gaussian structure in breast images.

As we have indicated, our results suggest the use of LFE for evaluating the realism of breast x-ray simulation approaches. But it may also be worth considering other approaches for evaluating higher-order structure. Presumably, there are many ways for physical processes to depart from a Gaussian distribution. So an array of features may be needed to adequately characterize these. One such possibility may be the use of co-occurrence statistics developed by Victor et al.,^{35,36} which are also invariant to power spectra and based on a different principle of statistical structure. While we believe that LFE has demonstrated useful properties for evaluating phantom realism, more work in this area is needed for a broader understanding of how higher-order structure is manifest in breast images.

Disclosures

The authors have nothing to disclose

Acknowledgments

Dr. Abbey and Dr. Boone were partially supported by the U.S. National Institutes of Health (R01-EB002138 and R01-EB025829). Dr. Abbey and Dr. Eckstein were partially supported by the U.S. National Institutes of Health (R01-EB018958 and R01-EB026427). Dr. Bakic and Dr. Maidment were partially supported by the U.S. National Institutes of Health (R01-CA154444) and the Burroughs Wellcome Fund (IRSA1016451). The content is solely the responsibility of the authors and does not necessarily represent the views of these funding agencies.

References

- R. F. Wagner and G. G. Brown, "Unified SNR analysis of medical imaging systems," *Phys. Med. Biol.* **30**, 489–518 (1985).
- C. E. Metz et al., "Toward consensus on quantitative assessment of medical imaging systems," *Med. Phys.* **22**, 1057–1061 (1995).
- K. Doi, "Diagnostic imaging over the last 50 years: research and development in medical imaging science and technology," *Phys. Med. Biol.* **51**, R5–R27 (2006).
- F. O. Bochud et al., "Importance of anatomical noise in mammography," *Proc. SPIE* **3036**, 74–80 (1997).
- J. J. Heine et al., "Multiresolution statistical analysis of high-resolution digital mammograms," *IEEE Trans. Med. Imaging* **16**, 503–515 (1997).
- J. J. Heine et al., "On the statistical nature of mammograms," *Med. Phys.* **26**, 2254–2265 (1999).
- A. E. Burgess, F. L. Jacobson, and P. F. Judy, "Human observer detection experiments with mammograms and power-law noise," *Med. Phys.* **28**, 419–437 (2001).
- F. O. Bochud et al., "Detectability of radiological images: the influence of anatomical noise," *Proc. SPIE* **2436**, 156–164 (1995).
- E. Engstrom, I. Reiser, and R. Nishikawa, "Comparison of power spectra for tomosynthesis projections and reconstructed images," *Med. Phys.* **36**, 1753–1758 (2009).
- 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).
- K. G. Metheany et al., "Characterizing anatomical variability in breast CT images," *Med. Phys.* **35**, 4685–4694 (2008).
- L. Chen et al., "Anatomical complexity in breast parenchyma and its implications for optimal breast imaging strategies," *Med. Phys.* **39**, 1435–1441 (2012).
- L. Chen, C. K. Abbey, and J. M. Boone, "Association between power law coefficients of the anatomical noise power spectrum and lesion detectability in breast imaging modalities," *Phys. Med. Biol.* **58**, 1663–1681 (2013).
- C. K. Abbey et al., "Non-Gaussian statistical properties of breast images," *Med. Phys.* **39**, 7121–7130 (2012).
- R. L. De Valois, D. G. Albrecht, and L. G. Thorell, "Spatial frequency selectivity of cells in macaque visual cortex," *Vision Res.* **22**, 545–559 (1982).
- D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *J. Opt. Soc. Am. A* **4**, 2379–2394 (1987).
- D. L. Ringach, M. J. Hawken, and R. Shapley, "Receptive field structure of neurons in monkey primary visual cortex revealed by stimulation with natural image sequences," *J. Vision* **2**, 2–24 (2002).
- F. Bochud, C. Abbey, and M. Eckstein, "Statistical texture synthesis of mammographic images with super-blob lumpy backgrounds," *Opt. Express* **4**, 33–42 (1999).
- C. Castella et al., "Mammographic texture synthesis: second-generation clustered lumpy backgrounds using a genetic algorithm," *Opt. Express* **16**, 7595–7607 (2008).
- C. K. Abbey and J. M. Boone, "An ideal observer for a model of x-ray imaging in breast parenchymal tissue," *Lect. Notes Comput. Sci.* **5116**, 393–400 (2008).
- P. R. Bakic et al., "Mammogram synthesis using a 3D simulation. I. Breast tissue model and image acquisition simulation," *Med. Phys.* **29**, 2131–2139 (2002).
- 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).
- D. D. Pokrajac, A. D. Maidment, and P. R. Bakic, "Optimized generation of high resolution breast anthropomorphic software phantoms," *Med. Phys.* **39**, 2290–2302 (2012).
- C. K. Abbey et al., "Non-Gaussian statistical properties of virtual breast phantoms," *Proc. SPIE* **9037**, 90370G (2014).
- J. P. Rolland and H. H. Barrett, "Effect of random background inhomogeneity on observer detection performance," *J. Opt. Soc. Am. A* **9**, 649–658 (1992).
- X. Gong et al., "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).
- M. J. Yaffe et al., "The myth of the 50-50 breast," *Med. Phys.* **36**, 5437–5443 (2009).
- P. R. Bakic et al., "Realistic simulation of breast tissue microstructure in software anthropomorphic phantoms," *Lect. Notes Comput. Sci.* **8539**, 348–355 (2014).
- J. S. Bendat and A. G. Piersol, *Random Data: Analysis and Measurement Procedures*, 2nd ed., Wiley, New York (1986).
- A. E. Burgess, "Mammographic structure: data preparation and spatial statistics analysis," *Proc. SPIE* **3661**, 642–653 (1999).
- I. Reiser, S. Lee, and R. Nishikawa, "On the orientation of mammographic structure," *Med. Phys.* **38**, 5303–5306 (2011).
- M. J. Wainwright, E. P. Simoncelli, and A. S. Willsky, "Random cascades on wavelet trees and their use in analyzing and modeling natural images," *Appl. Comput. Harmon. Anal.* **11**, 89–123 (2001).
- H. R. Sheikh, A. C. Bovik, and G. De Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," *IEEE Trans. Image Process.* **14**, 2117–2128 (2005).
- A. Srivastava et al., "On advances in statistical modeling of natural images," *J. Math. Imaging Vision* **18**, 17–33 (2003).
- J. D. Victor and M. M. Conte, "Local image statistics: maximum-entropy constructions and perceptual salience," *J. Opt. Soc. Am. A* **29**, 1313–1345 (2012).
- J. D. Victor et al., "A perceptual space of local image statistics," *Vision Res.* **117**, 117–135 (2015).

Craig K. Abbey is a researcher in the Department of Psychological and Brain Sciences at the University of California Santa Barbara. His training is in the field of applied mathematics and his research focuses

on assessment of medical imaging devices and image processing in terms of performance in diagnostic and quantitative tasks.

Predrag R. Bakic received his training in electrical engineering and medical physics. Currently, he is a research associate professor of radiology at the Hospital of the University of Pennsylvania. His research interests are computer simulation of human anatomy, virtual clinical trials, and optimization of breast tomosynthesis imaging.

David D. Pokrajac was a professor of computer sciences at Delaware State University (now with Boeing, Everett). He received his training in computer science and electrical engineering from Temple University. His research interests are in the field of spatial-temporal data mining.

Andrew D. A. Maidment received his training in medical biophysics from the University of Toronto, Canada. He is an associate professor of radiology at the Hospital of the University of Pennsylvania and chief of the physics section of the Department of Radiology. He is the author of more than 200 peer-reviewed papers. His research currently

focuses on the development of advanced imaging modalities to improve breast cancer detection.

Miguel P. Eckstein is a professor in the Department of Psychological and Brain Sciences at the University of California Santa Barbara. He received training in Cognitive Psychology at UCLA. His research uses a variety of tools including behavioral psychophysics, eye tracking, electro-encephalography (EEG), functional magnetic resonance imaging (fMRI) and computational modeling. These are applied to problems in medical imaging, computer vision systems, and interactions between robots/computer systems and humans.

John M. Boone is professor of radiology at the University of California Davis, and holds an appointment in the Department of Biomedical Engineering. He received his training in medical physics, and is board-certified by the American Board of Radiology in diagnostic radiological physics. His research interests focus on breast dosimetry and the development of breast computed tomography (bCT) for breast cancer screening and diagnostic evaluation.