Journal of Medical Imaging

MedicalImaging.SPIEDigitalLibrary.org

Location- and lesion-dependent estimation of mammographic background tissue complexity

Ali Avanaki
Kathryn Espig
Tom Kimpe

Location- and lesion-dependent estimation of mammographic background tissue complexity

Ali Avanaki,^{a,*} Kathryn Espig,^a and Tom Kimpe^b

^aBarco Healthcare, 9125 SW Gemini Drive, Suite 200, Beaverton, Oregon 97008, United States

^bBarco n. v., Beneluxpark 21, Kortrijk 8500, Belgium

Abstract. We specify a notion of perceived background tissue complexity (BTC) that varies with lesion shape, lesion size, and lesion location in the image. We propose four unsupervised BTC estimators based on: perceived pre and postlesion similarity of images, lesion border analysis (LBA; conspicuous lesion should be brighter than its surround), tissue anomaly detection, and local energy. The latter two are existing methods adapted for location- and lesion-dependent BTC estimation. For evaluation, we ask human observers to measure BTC (threshold visibility amplitude of a given lesion inserted) at specified locations in a mammogram. As expected, both human measured and computationally estimated BTC vary with lesion shape, size, and location. BTCs measured by different human observers are correlated ($\rho = 0.67$). BTC estimators are correlated to each other ($0.84 < \rho < 0.95$) and less so to human observers ($\rho \leq 0.81$). With change in lesion shape or size, LBA estimated BTC changes in the same direction as human measured BTC. Proposed estimators can be generalized to other modalities (e.g., breast tomosynthesis) and used as-is or customized to a specific human observer, to construct BTC-aware model observers with applications, such as optimization of contrast-enhanced medical imaging systems and creation of a diversified image dataset with characteristics of a desired population. © 2017 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: 10.1117/1.JMI.4.1.015501]

Keywords: human visual system properties; anthropomorphic numerical observer; virtual clinical trials; QUEST adaptive threshold seeking.

Paper 16174PR received Aug. 4, 2016; accepted for publication Dec. 13, 2016; published online Jan. 12, 2017.

1 Introduction

Validation of a medical imaging system is challenging due to the large number of system parameters that must be considered. Conventional methods involving clinical trials are limited by cost and duration, and in the instance of systems using ionizing radiation, the requirement for the repeated irradiation of volunteers. Alternatively, for preclinical validation and system optimization, virtual clinical trials (VCTs) may be used. A VCT consists of modeling human anatomy, image acquisition system,^{1–3} image processing and display, as well as image analysis and interpretation. We have developed anthropomorphic model observers^{4,5} that predict typical human observers better than commonly used model observers, which are designed based on ideal observers with some concessions for computational tractability.

In this paper (an earlier version of this work was presented at a conference⁶), we explore the estimation of the background tissue complexity (BTC). An automatic BTC estimator is necessary in tracking the performance of a human observer in reading images at various BTC levels. This is because the human observer's performance, unlike that of the existing model observers, varies considerably with BTC in a signal that is exactly known and a location that is exactly known (SKE/LKE) simulation paradigm.⁷ A model observer's capability of tracking the human observer performance with BTC is especially important when using VCT for design and/or optimization of contrast-enhanced (CE) breast imaging systems. In such systems, the uptake of the lesion relative to that of the normal fibroglandular tissue and the

dose of injected contrast agent affect the perceived BTC and thus lesion visibility.^{8–10} Moreover, the methods developed for BTC estimation may have applications in automatic measurement of breast density. Note that the concept of BTC to be studied herein is different from but related to the anatomical clutter (or noise); anatomical clutter is a tissue characteristic and does not directly concern lesion detection. Nevertheless, one can observe that lesion detection can become more difficult with higher anatomical clutter.

We believe that BTC as perceived by a human observer is a local property of the radiograph (i.e., varies with the location of the potential lesion). It also depends on the lesion characteristics (its shape and size) as explained below. These dependencies should be considered when designing a computational anthropomorphic BTC estimator.

1.1 Background Tissue Complexity is a Function of Location

The notion of BTC considered here relates to the richness of features in proximity of a target object (i.e., the lesion to be detected) that hinders the detection or search processes for a human observer. The proximity may be spatial or spatiotemporal depending on the type of visual stimulus (still image versus video) and viewing (e.g., video may be viewed as a set of frames shown as still images). We assume that a busy area far from the potential location of a lesion has no significant impact on detection, thus it should not affect BTC at that location. Evans et al.¹¹ suggested that an actual breast malignancy is a high-frequency

*Address all correspondence to: Ali Avanaki, E-mail: ali.avanaki@barco.com

Fig. 1 Copies of the same “lesion” are added to nonoverlapping tiles of a mammogram. Some lesions are harder to see than others (i.e., lesion visibility and thus BTC varies with location).

global signal that may be present even in the opposite and seemingly normal breast. Nevertheless, we believe that our locality assumption remains valid for the purposes of SKE/LKE simulations.

A visual example is provided in Fig. 1; a smooth round object, specifically one cycle of a two-dimensional (2-D) radial cosine pattern, is added at the centers of 29×29 nonoverlapping tiles of a mammogram. Care was taken to perform the addition in the luminance domain and to avoid clipping. It is clear that the visibility of the object is a function of its location in the background. We attribute lower visibility to higher perceived BTC, which we aim to estimate in this paper. If the insertion amplitudes were set using an ideal BTC estimator, all lesions in Fig. 1 would be equally conspicuous.

1.2 Background Tissue Complexity Depends on Lesion Shape and Size

The shape of the target object can also affect its detection and should be factored in BTC estimation. For example, a linear structure in the background may cover a circular lesion making it less conspicuous than an oval lesion with the same area and amplitude at the same location (Fig. 2).

Fig. 2 (a) A circular lesion and (b) an oval lesion, with the same area and amplitude, are superimposed at center of the same region of a mammogram. For demonstration purpose, the lesion insertion amplitude is larger than the required visibility thresholds. When the insertion amplitude is lowered, the circular lesion becomes harder to see perhaps because it blends in the linear structure of the background passing through that location. However, the oval lesion, which protrudes from the linear structure, remains visible at lower insertion amplitudes.

1.3 Related Research

In Ref. 12, we introduced a supervised method of BTC estimation for digital breast tomosynthesis (DBT) stacks. The differences between consecutive slices, measured in peak signal-to-noise ratio, were used as the input to the estimator. A Hotelling observer trained with perceived BTC values (one per DBT stack) for a subset of input was used as the estimator. The problems with this approach are as follows. (i) Supervised training of the estimator may not be practical or desirable in VCT scenarios. (ii) This method yields only one estimate of BTC per DBT stack. Hence, it is unusable for mammograms, as it cannot predict lesion visibility in different locations. Moreover, the result is insensitive to lesion shape and size.

Mainprize et al.¹³ proposed a local signal-to-noise ratio, d_{local} defined below, as a metric for potential masking of a lesion:

$$d_{\text{local}}^2 = \frac{(\int T^2(u, v) W^2(u, v) du dv)^2}{\int \text{NNPS}^2(u, v) T^2(u, v) W^2(u, v) du dv}, \quad (1)$$

where T is the modulation transfer function, W is the task function, NNPS is the normalized “noise” power spectrum (noise includes an anatomical component due to breast structure), and (u, v) are spatial frequencies. d_{local} is calculated for nonoverlapping regions of interest (ROIs) of the input mammogram. Direct calculation of the NNPS for each ROI is not possible. Therefore, a model of NNPS is calibrated using the noise measured for each ROI.¹⁴ Considering that the numerator of Eq. (1) is independent of the ROI, for a given mammogram and a given system, d_{local} is inversely proportional to ROI energy (i.e., local energy of the image). Note that this method is unsupervised and local (i.e., provides local complexity information) and is shown to perform worse than the supervised BTC estimator in Ref. 12. Recently, Alonzo-Proulx et al.¹⁵ and Mainprize et al.¹⁶ improved this method by using a calibration-free volumetric breast density estimator and by generalizing noise power spectrum estimation.

To locate anomalies in phantom CT images, Pezeshk et al.¹⁷ performed principal component analysis (PCA) over all overlapping ROIs across various scales and identified the anomaly ROIs as those far from mean ROI in the PCA coordinate system. They also showed that the phase-only-transform (PHOT), defined below, functions similarly:

$$\text{PHOT}(I) = F^{-1} \left[\frac{F(I)}{\|F(I)\|} \right], \quad (2)$$

where I is the input image, $\|\cdot\|$ denotes the absolute value (of complex number), and F and F^{-1} denote forward and inverse 2-D Fourier transform, respectively. The density of anomalies (considered potential lesions) calculated as such may be used as a local and unsupervised BTC estimate.

2 Methods

2.1 Estimate Background Tissue Complexity as Pre- and Postlesion Perceptual Similarity

In this method, to estimate BTC, we first superimpose the lesion at the location in question with a fixed amplitude. Next, we compare the postlesion and prelesion images using structural similarity metric (SSIM), a perceptual image similarity metric, defined below:¹⁸

Fig. 3 Derivation of immediate surround mask for a (a) binary lesion (the white disk) used by LBA estimator (Sec. 2.2), (b) lesion surround is white, and (c) immediate surround mask; brighter means more weight (thus more potential impact in determining BTC).

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}, \quad (3)$$

where x and y are the input signals (e.g., image ROI pre- and postlesion), μ and σ indicate the average and standard deviation, respectively, σ_{xy} is the signals' covariance, and C_1 and C_2 are the small positive constants keeping the denominator nonzero. We consider the perceptual similarity calculated as such as a predictor of BTC at the given location, for the given lesion. This is based on the premise that the greater the BTC, the less noticeable adding a lesion will be; thus, the more similar are the pre- and postlesion images. For better sensitivity of the estimate, only ROIs centered on the given location are compared. ROI size may be tuned for the desired estimation accuracy and sensitivity.

2.2 Background Tissue Complexity Estimator Based on Lesion Border Analysis

This method is based on the premise that BTC is correlated with the amplitude required to conspicuously superimpose the given lesion at a given location (i.e., the more complex the background, the higher the insertion amplitude needed for visibility). The superimposed lesion should be brighter than its immediate surround to be conspicuous. For a binary lesion, the surround is easily defined wherever there is no lesion (Fig. 3). Since far regions of background should not affect local BTC, we use a distance weighting function. Real lesions are not binary (Fig. 4). For such a lesion, we calculate the surround mask by inverting

the lesion and applying a distance weighting function, yielding an immediate surround mask. To estimate BTC at a given location, we multiply the background by the immediate surround mask centered at the location and use the maximum of the product as a predictor of BTC. The distance weighting function may be tuned for desired estimation performance.

2.3 Estimate Background Tissue Complexity by Local Anomaly Density

We adapt PHOT, described in Sec. 1.3 and in Ref. 17, to become sensitive to lesion shape and size by calculating the local average of anomalies (absolute value of PHOT output) with an averaging filter twice the lesion [in spatial size, to include the activities in lesion proximity, similar to lesion border analysis (LBA)] and normalized, as formulated below. This is based on the premise that detection of a larger lesion involves inspection of a larger neighborhood of the image and false lesions (i.e., background objects that may be mistaken with the lesion) are also larger (i.e., small-grain noise should be suppressed when dealing with detection of a larger lesion)

$$\text{BTC}_{\text{PHOT}}\{I, L\} = \text{PHOT}\{I\} * L_{\times 2}, \quad (4)$$

where $\text{BTC}_{\text{PHOT}}\{I, L\}$ is the PHOT-estimated BTC map (i.e., the value of each point in the map is set to the BTC estimate at that point) for image I in luminance domain and lesion L , PHOT is defined in Eq. (2), $*$ denotes the convolution, and $L_{\times 2}$ is the lesion expanded twice spatially and normalized (i.e., divided by sum of all pixels).

Fig. 4 Derivation of immediate surround mask for a (a) nonbinary lesion used by LBA estimator (Sec. 2.2), (b) surround mask, and (c) immediate surround mask; brighter means more weight (thus more potential impact in determining BTC).

2.4 Estimate Background Tissue Complexity by Local Energy

Inspired by Mainprize et al. (see Sec. 1.3 and Ref. 13), and similar to PHOT-based estimator derivation (Sec. 2.3), we adapt local energy to become sensitive to lesion shape and size by using twice the lesion (in spatial size) and normalized as the averaging filter, as formulated below. The reasons for selection of such an averaging filter are described in Sec. 2.3

$$\text{BTC}_{\text{Energy}}\{I, L\} = I^2 * L_{\times 2} - (I * L_{\times 2})^2, \quad (5)$$

where $\text{BTC}_{\text{Energy}}\{I, L\}$ is energy-estimated BTC map for image I in luminance domain lesion L , $*$ denotes the convolution, and $L_{\times 2}$ is the lesion expanded twice spatially and normalized (i.e., divided by sum of all pixels).

2.5 Human Measurement of Background Tissue Complexity

To measure BTC at a specific point, p , a human observer may adjust the insertion amplitude of the given lesion until it becomes visible. To find the insertion amplitude corresponding to threshold visibility, we use QUEST, an adaptive threshold seeking procedure.¹⁹ As compared to adjusting the amplitude manually, QUEST is more convenient for the user and yields a more accurate threshold as well as its confidence interval. We use a MATLAB[®] implementation of QUEST available from Ref. 20 with the default value of parameters and 41 trials per threshold measurement (40 is the typical number of trials used in QUEST example; we added one for reasons explained below).

The experiments are conducted on a Barco Uniti display (MDMC-12133) to ensure low noise and consistent presentation, provided by RapidFrame[™], and Color Per-Pixel-Uniformity[™]. In each trial, two panels are displayed; on one panel, a square mammogram region centered at p is shown, and on the other panel, the same region with the lesion superimposed to the center at the insertion amplitude being tested is shown. When p is too close to the mammogram margin, the rest of the square region is filled with a mirror of the mammogram along the nearest edge (using MATLAB's[®] padarray symmetric option) to preserve the observed texture continuity (see Fig. 5 as an example, where the top quarter of each ROI shown is filled by mirroring). The two panels are separated by one fifth of a panel width. The panels together with margins extend to about 15 visual degrees and are uniformly filled with average luminance of the region being displayed where there is no visual information for optimal

Fig. 5 Example of what is shown to the human observer, at an apparent size of about 15 deg, in one trial of a threshold measurement experiment. Observer's task is to pick the panel that has a lesion at the center (left in this example).

eye adaptation. That is because Barten²¹ noted that a surround luminance different from that of the target object adversely impacts effective contrast sensitivity. Lesion apparent size is about one fourth to half a degree which is the target object size for optimal visibility (spatial CSF remains flat at its peak at about 0.1 deg to 1 deg in typical viewing conditions²²).

In the first trial of each experiment, the lesion is shown with maximum possible amplitude to familiarize observer with the shape, size, and location (i.e., center of panel) of the lesion (Fig. 5). When adding the lesion, care was taken to avoid clipping (all pixel values remain between 0 and 1 modifying addition result), and that scaling (to affect insertion amplitude) and addition are performed in the luminance domain (not in pixel value). The task assigned to the human observer is to pick the panel with the lesion. Input choices are left, right, or "do not know." The order of the panels (left or right) is chosen randomly by the experiment program, which compares observer's input to the actual location of the lesion panel and based on this information (i.e., answered correct or incorrect) generates the next amplitude to be tested. A "do not know" input is assumed to be an incorrect response.

QUEST generates the probability distribution function (PDF) for the threshold being measured. In each experiment, we record the mean and standard deviation of the threshold PDF.

2.6 Evaluation

For a given mammogram, we generate BTC maps using pre- and postlesion perceptual similarity (P³S), LBA, PHOT-, and energy-based estimators as follows. For each BTC estimator and for each lesion under investigation, the mammogram pixel value was replaced with the estimated BTC there. We inspect the proposed methods by checking variations of their maps with location, lesion shape, and size, per the design criteria set out in Sec. 1. For the BTC maps to become comparable, we perform histogram equalization on the maps generated by each method. For example, for P³S, one histogram equalizing transform was calculated from its three BTC maps and then that transform was applied to the three P³S maps. Thus, the value at each point becomes proportional to the BTC rank within the maps generated by each method (approximately, because the number of bins used in histogram equalization is smaller than the number of possible BTC values), and the equalized BTC maps for different lesions can be correctly compared.

We compare the proposed methods against the human observers as follows. Since measuring BTC for a human is time consuming, generation of full BTC maps (for comparison against maps by proposed methods) using a human observer is not practical. Assuming that the proposed methods are good estimators, the BTC maps they generate are highly correlated with each other. Thus, we use the computationally generated maps (after histogram equalization described above, as the maps must be comparable for the following procedure) to pick a set of interesting points for BTC measurement by human as follows:

- The point with the highest sensitivity to shape (i.e., the maps for small circular and oval lesions differ most).
- The point with the highest sensitivity to size (i.e., the maps for small and large circular lesions differ most).
- The point where one BTC estimate most exceeds the maximum of the other three (e.g., $\arg \max$

$\{BTC_{PHOT} - \max\{BTC_{Energy}, BTC_{LBA}, BTC_{P3S}\}\}$, where both max operators are point-wise and across all pixels).

- iv. The point where the minimum of three BTC estimates most exceeds the fourth (e.g., $\arg \max \{\min\{BTC_{Energy}, BTC_{LBA}, BTC_{P3S}\} - BTC_{PHOT}\}$, where max and min operators are point-wise and across all pixels).

Since we proposed four methods, the human observer has to estimate BTC at a maximum (assuming no redundancy) of 16 points with the small circular lesion, at four points in category (i) with the oval lesion and at four points in category (ii) with the large circular lesion.

To check our assumption of good quality of the proposed BTC estimators, in addition to the interesting points above, we measure BTC on 10 randomly picked points and inspect the correlation of the measurements with the BTC estimates by the proposed methods at those points.

We gauge the sensitivity of the proposed estimators to change in lesion size and shape as follows. We asked a human observer to measure BTC at points where each estimator was most sensitive to change in lesion shape or size [i.e., category (i) and (ii) points defined above]. We inspect whether or not the measured BTC increases with the change in lesion size and do the same with the BTC estimates for the same point. If both measured and estimated BTC increase (or decrease) with the change in lesion size, we consider the estimator sensitive to lesion size (specifically sensitive to a change from small to large circular lesion). Sensitivity to lesion shape (from small circular to oval lesion with the same area) is derived similarly.

3 Results

We generated BTC maps using the computational estimators of Sec. 2 for the lesions and the mammogram region shown in Fig. 6. The lesions are Gaussians with 2σ of 7.5 (small round), 10.6 (also round, $2\times$ area), and vertical: 5.4 and horizontal: 10.6

Fig. 6 (a) Lesions used in the experiments, enlarged. Original sizes are 29×29 , 41×41 , and 21×41 pixels. (b) Mammogram region used in the experiments. Partial BTC maps shown in Fig. 7 correspond to the dashed portion.

(same area as the small round lesion, asymmetric) pixels are chosen to show the effect of lesion shape and size on BTC. The mammogram region shown is cropped from a 3262×2706 mediolateral oblique view digital mammogram available publicly²³ with a $94\text{-}\mu\text{m}$ pixel pitch and is an example image with a wide variety of perceived BTC in different locations. Partial BTC maps corresponding to the dashed part of the mammogram are shown in Fig. 7. We use Pearson's correlation coefficient²⁴ to explore the relationship of the proposed BTC estimates to each other and to the BTC measurements as follows. Correlation coefficients of full BTC maps for small circular lesion are reported in Table 1, and correlation coefficients between BTC maps generated for different lesions are given in Table 2. At different sets of points defined in Sec. 2.6, the correlation coefficients between measured (by human observers) and estimated BTC values (by proposed methods) for the small circular lesion are calculated and presented in Table 3. Correlation coefficients between the two human readers at various sets of points are also listed in Table 3. Two human observers were involved in the experiments. The human observers are experienced vision scientists (nonradiologists). The difference between estimated BTC for lesions of different shape or size, as well as the corresponding difference of BTC values measured by humans for category (i) and (ii) points defined in Sec. 2.6, are listed in Table 4. When the (human and estimator) differences are in the same direction (i.e., both positive or both negative) the corresponding p -values are listed as well. The p -values are calculated using the standard deviation of threshold (i.e., human measured BTC, about 0.07 for the values reported) provided by QUEST, on the premise that the BTC measurements are independent for different lesions (i.e., variance of difference is equal to sum of variances).

4 Discussion

The design criteria regarding BTC estimate dependence to location, lesion size, and shape are met by all of the proposed methods as described below. In Fig. 7, it may be observed that none of the partial BTC estimate maps shown is constant (i.e., they vary with location). BTC changes with location in all proposed methods satisfy our expectation (smaller BTC estimate in darker and lower activity areas in top right corner of the mammogram and the partial maps). Variations with lesion shape and size exist but are more subtle: the BTC maps in Fig. 7 slightly differ across columns corresponding to different lesions. The most prominent differences between BTC maps for different lesions are perhaps in the third row, which correspond to the energy-based BTC estimator. Moreover, the sensitivity to lesion size and shape may be observed in Table 2, where the correlation coefficients between the BTC maps generated by each method for different lesions are listed; though the correlations are strong, none of them is one. Based on Table 2, LBA estimator is the most sensitive to shape and size changes. This observation is further reinforced by human BTC measurements (Tables 3 and 4, to be discussed later in this section).

It may be observed that the BTC maps generated by the proposed methods are highly correlated. This was assumed in developing our evaluation method and can be verified by inspection of each column of Fig. 7 (partial BTC maps for the same lesion are generally similar) and from Table 1, where the lowest correlation of full BTC maps is still rather high (0.8433, between PHOT-based and P³S for small circular lesion).

BTC measured by a human observer as described in Sec. 2.5 varies from one observer to another. This may be observed in

Fig. 7 Rows from top: partial BTC maps generated by PHOT-based, P³S, energy-based, and LBA estimators. Columns from left: lesion used for BTC estimation: small circular, large circular, and oval lesion.

Table 3, “observer B” column: the correlation between BTC measurements by human observers A and B at 26 points is about 0.7. Therefore, a nonpersonalized BTC estimator, such as those proposed herein, is unlikely to produce estimates perfectly correlated to the measurements by a specific human observer.

Table 1 Pearson’s correlation coefficient between full BTC maps generated by the proposed BTC estimators.

BTC estimator	P ³ S	LBA	Energy-based
PHOT-based	0.8433	0.9485	0.9054
P ³ S	—	0.8941	0.8882
LBA	—	—	0.9222

Table 2 Pearson’s correlation coefficients between BTC maps generated for different lesions for different proposed BTC estimators.

BTC estimator	Oval and small circular	Small and large circular	Oval and large circular
PHOT-based	0.995	0.9899	0.9931
P ³ S	0.9685	0.9621	0.9656
LBA	0.9607	0.9287	0.9479
Energy-based	0.9742	0.9586	0.9707

Keeping that in mind, based on the first major row of Table 3, all proposed methods are good BTC estimators from observer A’s viewpoint, and the energy-based estimator is good from observer B’s viewpoint. It is interesting to note that for the special points defined in Sec. 2.6, corresponding to the second and third major rows in Table 3, the interobserver correlation is either significantly lower or higher, i.e., these points bring out agreement or disagreement in the observers, thus can help stratify BTC notions pertaining to the different varieties of human observers.

We noticed that BTC measurements from observer A were almost always lower (when higher, differed little) than those from observer B. That means observer A needed a lower lesion amplitude for visibility (Fig. 8), perhaps because of more familiarity with this type of study or a more acute vision in general. Our results suggest that the proposed BTC estimators are better predictors of the human observer that can correctly detect the lesion at lower amplitudes. In terms of sensitivity to lesion shape and size [i.e., correlation for category (i) and (ii) points listed in second major row of Table 3], LBA is the best estimator for observer A (as observed in Table 4, discussed later in this section, and Table 2) and PHOT-based estimator is best for observer B. PHOT-based estimator is perhaps the best overall choice for predicting both readers as it has the highest number of bold face correlation coefficients (i.e., row maximums) under its column.

The relatively large absolute value but negative correlation between human observers and P³S in the third major row of Table 3 is curious. Upon close inspection, we noted that points with high BTC estimates by P³S but low measured BTC by observers are the culprits. It may be observed that in such points (Fig. 9) there is a dark area close to the center that a human

Table 3 Pearson's correlation coefficients between estimated and human measured BTC values for small circular lesion at different sets of points defined in Sec. 2.6. Maximum correlations between humans and computational estimator for each set of points are shown in boldface. The weak correlations are shown in italics (i.e., zero-correlation hypothesis cannot be rejected at 0.05 significance level).

Point set		Observer B	PHOT-based	P ³ S	Energy-based	LBA
Random; 10 points	Observer A	0.665	0.8095	0.7906	0.7327	0.7336
	Observer B		<i>0.611</i>	<i>0.5801</i>	0.7098	<i>0.4767</i>
Category (i) and (ii), i.e., lesion sensitive; 8 points	Observer A	<i>0.425</i>	<i>0.0246</i>	<i>0.2334</i>	<i>0.0033</i>	0.5355
	Observer B		0.6564	<i>0.1464</i>	<i>0.5428</i>	<i>0.2894</i>
Category (iii) and (iv), i.e., method sensitive; 8 points	Observer A	0.9513	0.4399	−0.503	−0.2084	<i>0.305</i>
	Observer B		0.3981	−0.606	−0.1603	<i>0.369</i>
All 26 points	Observer A	0.7059	0.5522	<i>0.2345</i>	<i>0.2124</i>	0.5017
	Observer B		0.5158	<i>0.0979</i>	<i>0.2913</i>	0.4003

Table 4 Sensitivity of BTC estimates by proposed methods with respect to change in lesion shape and size, compared to the difference in BTC measured by observer A, in category (i) and (ii) points defined in Sec. 2.6. The *p*-values are listed when measured and estimated BTCs differences are in the same direction (i.e., both are positive, or both are negative, meaning measured and estimated BTC increase or decrease by change in lesion shape or size).

BTC estimator	Oval versus small circular lesion			Large versus small circular lesion		
	BTC measurement difference	BTC estimate difference	<i>p</i> -value, if sensitive	BTC measurement difference	BTC estimate difference	<i>p</i> -value, if sensitive
PHOT-based	−0.0484	0.1746		0.0088	0.2222	0.4461
P ³ S	−0.027	−0.3968	0.3406	0.0175	−0.4762	
LBA	0.0164	0.6825	0.4027	0.0028	0.7619	0.4856
Energy-based	−0.033	0.5397		0.05	0.5079	0.2239

Fig. 8 ROI at the center of which observer B measured a BTC much higher than observer A. We theorize that observer B needed the extra amplitude to distinguish the superimposed lesion (not shown here) from structures nearby (a small lesion-like structure the left of the center and a larger one immediately under the center).

observer can use to detect the lesion rather easily; even a low lesion amplitude can visibly affect such a dark area. P³S, however, considers the disturbance caused by adding the lesion in the whole ROI and not just the dark region. Moreover, the culprit points are on a high activity region but SSIM is formulated in such a way to discount the perceptual difference in a high activity area. Thus, lesion insertion in such areas will be less noticeable in terms of SSIM. Note that the energy-based method also correlates negatively with the human observers for the same points (cf. the third major row of Table 3 under “energy based”), perhaps for the same reason (i.e., nearby dark spot in a high activity region).

It is worth noting that all proposed BTC estimators outperform constant insertion amplitude (see Fig. 1). In Fig. 10, lesion insertion amplitude is modulated by P³S BTC estimates in the centers of 29 × 29 nonoverlapping tiles of a mammogram to produce the P³S counterpart of Fig. 1. It may be observed that conspicuity across the lesion grid is more uniform in Fig. 10 as compared to that in Fig. 1. That is in spite of the fact that P³S is not the best estimator in terms of correlation with human observers (Table 3), and that P³S correlates least with the PHOT-based estimator (Table 1), the highest correlating estimator with human observers.

Fig. 9 ROIs at the centers of which P^3S estimated a much higher BTC than what measured by human observers. We theorize this is due to a dark region on or in immediate proximity of the center that aid human observers in lesion detection but disturb estimation by P^3S . See the text for an explanation.

When a personalized BTC estimator is desired, one might combine different estimators to better match the measurements by a specific human observer. For example, a linear combination of the proposed estimators matching (in least squares sense) observer A measurements on 10 random points (same as the set mentioned in Table 3 major row 1) has an increased correlation of 0.8631 (from 0.8095, by PHOT-based estimator); similarly correlation with observer B can be increased to 0.7976 (from 0.7098, by energy-based estimator).

We gauge the sensitivity of the proposed estimators to change in lesion size and shape according to the method described in Sec. 2.6. The results are listed in Table 4 (p -value is given for lesion sensitive methods). As noted before (in Tables 2 and 3 analyses), LBA is sensitive to both lesion shape and size. Statistically stronger results may be reached if BTC for different lesions is measured at more points by human observers and/or if BTC can be measured more accurately (e.g., by lowering threshold standard deviation, using larger number of trials in each run of QUEST).

Estimator LBA is not far from an anthropomorphic model observer,¹² if the area within the immediate surround mask is brighter than BTC (i.e., maximum luminance with the immediate surround mask) by a certain margin and satisfies some other criteria (e.g., small gradient), it may be announced as a lesion. For a known lesion (SKE), this method may be used as the basis of an anthropomorphic search mechanism.

The procedure proposed in Sec. 2.5 is robust enough for a human observer to measure BTC with a complex lesion (e.g.,

spiculated breast lesion). BTC estimation with complex lesions is, however, beyond the scope of this research. That is because from human vision science, we know that humans detect a complex target by its parts and/or features (and not as a whole),²⁵ which are the basis for rendering a cognitively complex overall detection decision (e.g., vary with observer's experience and training). The good news is that BTC estimation for simple lesions is clinically relevant (e.g., small round lesion is a good approximation to a breast microcalcification) and can still be used in a limited way to predict detectability of a more complex lesion. For example, a breast microcalcification cluster is less detectable when all individual calcifications are less detectable due to high local BTC.

5 Conclusion

BTC estimated by proposed methods, as well as BTC measured by a human observer discussed in Sec. 2.5, are sensitive to lesion location, size, and shape. BTC estimators correlate with each other and correlate with human measured BTC as well. Therefore, all of the proposed methods, as-is or customized to a specific human observer, may be used to construct a BTC-aware model observer, with applications such as optimization of CE medical imaging systems and creation of a diversified image dataset matching a desired population.

None of the proposed BTC estimators correlates with human observers perfectly. This is reasonable since the human observers do not agree on their measurements either; they measure different BTC values for the same points of the same mammogram and for the same lesion. From the proposed methods, LBA sensitivity to changes in lesion shape and size is the closest to human observer and PHOT-based BTC estimates have highest correlation with measured BTC by humans. A combination of proposed BTC estimators (e.g., linear least squares) can better predict a specific human observer, though this requires training. Alternatively, free parameters of each of the proposed methods (i.e., weights of the three components of SSIM used in P^3S , size and shape of localization kernels in energy- and PHOT-based estimators, and size and shape of immediate surround mask in LBA) may be optimized for a specific human observer. Doing so also requires training.

To handle DBT stacks and browsing in time, the proposed methods may be generalized as follows. For LBA, a spatiotemporal immediate surround mask may be devised. For P^3S , a three-dimensional (3-D) generalization of SSIM should be used to estimate perceptual similarity of pre- and postlesion ROIs. For PHOT-based method, 3-D (inverse) Fourier transform and

Fig. 10 Lesions are added to nonoverlapping tiles of a mammogram with insertion amplitudes modulated by P^3S BTC estimates. Compare conspicuity as a function of location in this figure and in Fig. 1 (produced with constant lesion insertion amplitude).

a spatiotemporal lesion-dependent localization kernel may be used. The latter may be also used to generalize the energy-based BTC estimator to process DBT.

In the course of our experiments, we noted a set of points at which observers highly agreed on measured BTC, as well as a set of points at which observers highly disagreed on measured BTC. This may suggest that the decision on BTC (equivalently, visibility of given lesion in a given complex background) is not atomic and has to be broken apart to yet unknown subdecisions, to accurately model a specific human observer.

5.1 Limitations and Further Research

The fact that the proposed BTC estimators are highly correlated with each other means that they carry similar information. We did not intend this; rather by analyzing seemingly disjoint models of perceived BTC, we derived the proposed estimators. Further consideration of some new aspects of BTC may lead to BTC estimators carrying different information, therefore, correlating less with each other. Combination of such estimators may better predict a specific human observer's behavior in measuring BTC.

To show the importance of lesion shape in BTC measurement and estimation, we only considered small round and oval lesions. Though BTC estimation for complex lesions (e.g., spiculated breast lesion) is beyond the scope of this work (Sec. 4), a general indication of background complexity may be derived for such lesions by considering the BTC for their parts and/or features. An avenue of further research involves breaking complex lesions of interest to simple parts (e.g., a spiculated breast lesion = round central object + spikes of different eccentricities) and studying the effectiveness of computational BTC estimators in terms of correlation with human measured BTC for each part.

Our evaluation of BTC estimators may be enhanced by using more human observers with different experience levels (to explore the variation of estimators in predicting humans) measuring BTC in a larger set of images taken with different equipment (at different doses and employing different image processing techniques) from different tissue types (e.g., different BI-RADS).

For lesion insertion, both in proposed estimators and in measuring BTC by human observer, we used a simple additive superimposition. A more realistic lesion insertion model (e.g., Ref. 26) depends on the modality and perhaps even the specific medical imaging system being modeled and can generate artifacts that can affect estimated or measured BTC. Such artifacts may make the lesion more conspicuous in certain locations, thus lowering measured or estimated BTC. We theorize that our methods and conclusions remain valid by considering the realistically inserted lesion (i.e., including artifacts) as the new lesion since we rely on near (visibility) threshold phenomenon in BTC estimation and measurement.

To evaluate the performance of BTC estimators, first we make the estimates comparable using histogram equalization, inevitably with a number of bins smaller than the number of possible levels of estimated BTC. This is a potential limitation of the evaluation method since it can affect the correlation between the BTC estimates and/or human measured BTC.

To measure BTC in our reader study, we used ROIs centered on the lesion (or lesion location, for lesion absent alternative), assuming that far flung parts of image have no impact on lesion detection locally. While we believe this assumption is valid for our purpose, finding the minimum ROI size required for BTC

measurement can be beneficial for designing a more computationally efficient BTC estimator (the computational complexity of which is proportional to ROI size). To that aim, one should measure BTC at several locations for varying ROI sizes and find the lower threshold on ROI size at which measured BTC values remain highly correlated with those measured at the largest ROI size (this may be the whole image). For larger ROIs, location indicators may be required to keep the reader study LKE. The minimum ROI size calculated as such is a function of the image, the reader, and the lesion, which should be diversified to the extent that the BTC estimator is designed to operate. ROI size should also be considered in terms of estimation performance. For example, using a large ROI with P³S (Sec. 2.1) estimator decreases the sensitivity of BTC estimation since the perceived difference between lesion and no-lesion ROIs is averaged out over a larger area that are identical (since they only differ at lesion location) in both ROIs. Based on these considerations, the ROI size for each estimator may be optimized for both estimation performance and computational efficiency.

Our proposed method of BTC measurement by humans is slow (full measured BTC map is impractical and we resorted to measuring at a few points only) and not precise (measured BTC values have large standard deviation, causing large *p*-values in Table 4). We suspect simultaneous improvement in precision and speed of measurement may not be possible for the notion of BTC introduced herein, though each can be improved at the expense of the other. To speed up measurements, instead of adaptive threshold measurement by QUEST through several trials, the observer may adjust the insertion amplitude so that the lesion becomes "just-noticeable." The problem with this approach is that the threshold may change overtime (due to fatigue), thus may not be reproducible. Precision of BTC measurement may be improved by repeating our current method (based on QUEST) several times, averaging the results, and/or by increasing the number of trials per QUEST run.

Disclosures

The authors are employees of Barco Healthcare.

Acknowledgments

Ali Avanaki would like to thank Albert Xthona for fruitful discussions and Eddie Knippel for proofreading the paper.

References

1. D. Pokrajac, A. D. A. Maidment, and P. R. Bakic, "Optimized generation of high resolution breast anthropomorphic software phantoms," *Med. Phys.* **39**, 2290–2302 (2012).
2. M. A. Lago, A. D. A. Maidment, and P. R. Bakic, "Modelling of mammographic compression of anthropomorphic software breast phantom using FEBio," in *Proc. Int. Symp. on Computer Methods in Biomechanics and Biomedical Engineering (CMBBE)*, Salt Lake City, Utah (2013).
3. J. Kuo et al., "Dynamic reconstruction and rendering of 3D tomosynthesis images," *Proc. SPIE* **7961**, 796116 (2011).
4. A. N. Avanaki et al., "Development and evaluation of a 3D model observer with nonlinear spatiotemporal contrast sensitivity," *Proc. SPIE* **9037**, 90370X (2014).
5. A. N. Avanaki et al., "Integration of spatio-temporal contrast sensitivity with a multi-slice channelized Hotelling observer," *Proc. SPIE* **8673**, 86730H (2013).
6. A. R. Avanaki et al., "Location-and lesion-dependent estimation of background tissue complexity for anthropomorphic model observer," *Proc. SPIE* **9787**, 97870A (2016).

7. A. N. Avanaki et al., "It is hard to see a needle in a haystack: modeling contrast masking effect in a numerical observer," *Lect. Notes Comput. Sci.* **8539**, 723–730 (2014).
8. S. C. Chen et al., "Initial clinical experience with contrast-enhanced digital breast tomosynthesis," *Acad. Radiol.* **14**(2), 229–238 (2007).
9. M. L. Hill, J. G. Mainprize, and M. J. Yaffe, "An observer model for lesion detectability in contrast-enhanced digital mammography," in *Digital Mammography*, pp. 720–727, Springer, Berlin, Heidelberg (2010).
10. M. L. Hill et al., "Design and validation of a mathematical breast phantom for contrast-enhanced digital mammography," *Proc. SPIE* **7961**, 79615E (2011).
11. K. K. Evans et al., "A half-second glimpse often lets radiologists identify breast cancer cases even when viewing the mammogram of the opposite breast," *Proc. Natl. Acad. Sci. U. S. A.* **113**(37), 10292–10297 (2016).
12. A. R. Avanaki et al., "On anthropomorphic decision making in a model observer," *Proc. SPIE* **9416**, 941610 (2015).
13. J. G. Mainprize et al., "Towards a quantitative measure of radiographic masking by dense tissue in mammography," *Lect. Notes Comput. Sci.* **8539**, 181–186 (2014).
14. J. G. Mainprize and M. J. Yaffe, "Cascaded analysis of signal and noise propagation through a heterogeneous breast model," *Med. Phys.* **37**(10), 5243–5250 (2010).
15. O. Alonzo-Proulx et al., "Local detectability maps as a tool for predicting masking probability and mammographic performance," *Lect. Notes Comput. Sci.* **9699**, 219–225 (2016).
16. J. G. Mainprize et al., "Quantifying masking in clinical mammograms via local detectability of simulated lesions," *Med. Phys.* **43**(3), 1249–1258 (2016).
17. A. X. Pezeshk, L. Popescu, and B. Sahiner, "CT image quality evaluation for detection of signals with unknown location, size, contrast and shape using unsupervised methods," *Proc. SPIE* **9416**, 94160J (2015).
18. Z. Wang et al., "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.* **13**(4), 600–612 (2004).
19. A. B. Watson and D. G. Pelli, "QUEST: a Bayesian adaptive psychometric method," *Percept. Psychophys.* **33**(2), 113–120 (1983).
20. "Psychophysics Toolbox Version 3," <http://psychtoolbox.org> (January 2017).
21. P. G. Barten, "Formula for the contrast sensitivity of the human eye," *Proc. SPIE* **5294**, 231–238 (2003).
22. P. G. Barten, *Contrast Sensitivity of the Human Eye and its Effects on Image Quality*, Vol. **72**, SPIE Press, Bellingham, Washington (1999).
23. 40 F, "Bilateral digital mammography images of the breasts, mediolateral oblique view," https://commons.wikimedia.org/wiki/File:40F_MLO_DMMG.png (January 2017).
24. L. F. Smith, Z. S. Gratz, and S. G. Bousquet, *The Art and Practice of Statistics*, Cengage Learning, Belmont, California (2008).
25. D. G. Pelli, B. Farell, and D. C. Moore, "The remarkable inefficiency of word recognition," *Nature* **423**, 752–756 (2003).
26. M. S. Vaz, Q. Besnehard, and C. Marchessoux, "3D lesion insertion in digital breast tomosynthesis images," *Proc. SPIE* **7961**, 79615Z (2011).

Ali Avanaki holds a PhD in electrical and computer engineering from University of British Columbia, Vancouver, Canada, and MASc and BASc in electrical engineering from Sharif University of Technology, Tehran, Iran. His research interests include modeling perception of medical images using human visual system properties.

Kathryn Espig graduated with a master of science in Computer Science from the Georgia Institute of Technology. She is a software engineer and project manager in the Technology and Innovation Group at Barco Healthcare. She has been involved with graphics software for over 20 years.

Tom Kimpe received his PhD in electrical engineering from University of Ghent and a Master in Business Administration from Vlerick Management School. His areas of research and interest include (medical) display technology, image and signal processing, image quality modelling, network technology and medical regulatory aspects. He is Barco Healthcare's vice president of technology and innovation.