Original paper

Determination of contrast-detail curves in mammography image quality assessment by a parametric model observer

T. Kretz*, M. Anton, T. Schaeffter, C. Elster

Physikalisch-Technische Bundesanstalt, Braunschweig and Berlin, Germany

ABSTRACT

A novel approach is proposed for the determination of contrast-detail curves in mammography image quality assessment. The approach is compared with current practice using virtual mammography.

A binary parametric model observer is applied to images of the CDMAM phantom. The observer accounts for the simple disc shaped objects in the phantom and is applied separately to each cell of the phantom. For each of these applications, the area under the ROC curve (AUC) of the model observer is determined. The different AUCs, calculated from different applications of the parametric model observer, are then combined to a single contrast-detail curve quantifying the ability of the observer to detect details in the images. Virtual mammography is developed as a tool to simulate X-ray images of single CDMAM cells and to quantitatively assess the approach in comparison with current practice.

It is shown that the proposed approach can lead to similar contrast-detail curves as current practice. The precision of the estimated contrast-detail curves is increased, i.e. using 5 images yields about the same precision for the proposed approach as 16 images when applying current practice.

We conclude that contrast-detail curves in mammography image quality assessment can also be determined through the AUC of a binary parametric model observer. Since the proposed approach has higher precision than current practice, it is a promising candidate for contrast-detail analysis in mammography image quality assessment.

1. Introduction

According to health statistics provided by the statistical office of the European Union, breast cancer is the most frequent type of cancer among women in Europe [1]. Early detection through cancer screening reduces breast cancer mortality rates significantly [2]; currently, mammography using soft X-ray radiation is applied for this purpose [3]. Although increasing the radiation dose applied in mammography allows better discrimination between tumors and healthy tissue, it also risks causing damage (e.g. [4]). This becomes even more important for screening programs where the participants are assumed to be predominantly healthy. Therefore, reliable image quality assessment is important in order to minimize the applied dose while maintaining sufficient diagnostic power.

In mammography image quality assessment one distinguishes between technical image quality and clinical image quality (e.g. [5]). The former applies to the raw images, while the latter is based on clinical images presented to radiologists. The images presented to clinicians are derived from the raw images by employing nonlinear processing

algorithms to enhance the low contrast between breast tissue and pathological findings [6]. While CDMAM phantoms may not be suitable to detect an impact of these processing algorithms (cf. [7]), several studies point towards their importance for the detection of microcalcifications in clinical images (e.g. [8,9]).

According to the European Reference Organization for Quality Assured Breast Screening and Diagnostic Services (EUREF) Guideline [2], a task-based approach is currently applied for technical image quality assessment in mammography. More precisely, a contrast-detail curve [10,11] is determined from images of the contrast-detail mammography (CDMAM) phantom shown in Fig. 1. The contrast-detail curve characterizes the ability to detect small details in the image. The CDMAM phantom consists of 205 cells, each containing two gold discs. One of the two gold discs is in the center of each cell, and the other is located at one of the four corners of the cell. Diameter and thickness of the discs vary across cells, and the contrast-detail curve displays the minimum thickness required for a given diameter that will allow the employed (model) observer to correctly locate the position of the second disc in a cell with a specified probability [10]. The potential
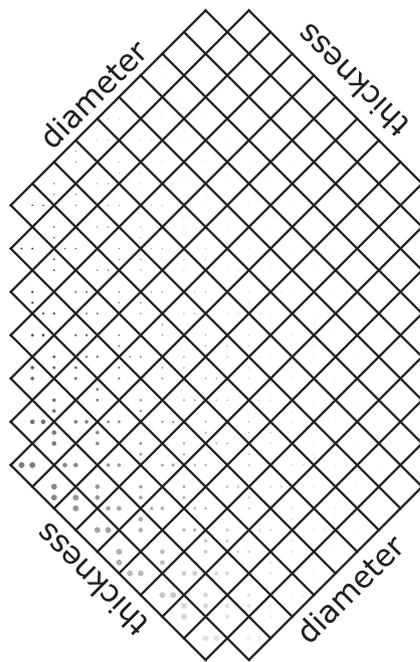
Fig. 1. Schematic picture of a commonly used contrast-detail phantom for image quality assessment in mammography (CDMAM). The diameter and thickness of the two gold discs in each cell vary between [0.06–2.00] mm and [0.03–2.00] μm, respectively. Variation of thickness is indicated as color intensity in the sketch. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

value of quality control using the CDMAM phantom has been explored in [12], where it was demonstrated that the analysis of CDMAM images can be related to the task of detecting calcification clusters.

In order to conform with current requirements on mammography quality, the resulting contrast-detail curve needs to be below a specified limiting curve [2]. To ensure a reliable conformity assessment with high reproducibility, the precision of the estimated contrast-detail curves needs to be sufficiently high. The EUREF Guideline procedure recommends to use 16 images. However, doubts have been raised whether this number of images is large enough for reliable decisions to be taken [13]. Raising the number of images may, on the other hand, not be practical.

Task-based image quality has been proposed and applied for quality assessment in medical imaging for a long time, cf. [14–17]. In task-based image quality assessment image quality is measured in terms of the performance of a so-called observer when carrying out a specific task such as the detection of a lesion. The observer is either a human (e.g. a radiologist) or a mathematical model observer. Task-based image quality assessment using mathematical model observers can be applied to processed and unprocessed images. A popular figure of merit for image quality in binary classification tasks (e.g. presence or absence of a lesion) is the area under the ROC curve (AUC) which characterizes a binary classifier [16]. Mathematical model observers are usually used which provide an automatic and objective assessment of image quality. These model observers are typically nonparametric; examples comprise the non-prewhitening observer or the channelized Hotelling observer [18]. Bouwman et al. demonstrate that the non-prewhitening model observer [19] as well as the channelized Hotelling observer [20] can predict the human observer performance, when including aspects of the human visual system. Training and testing nonparametric model observers require large sets of images ($\gtrsim 100$), which is impractical for daily routine in mammography image quality assessment.

Recently, the parametric model observer MSOpi (Modified Simple Observer for pooled images) has been proposed [21]. This observer applies to simple geometries such as the CDMAM phantom which combines simple disc shaped objects. Since this observer utilizes knowledge about the geometry, significantly fewer images are needed to reliably estimate its AUC than for established nonparametric model observers.

The goal of this paper is to explore the potential of the parametric model observer MSOpi for its application in mammography image quality assessment. Furthermore, use of this observer can provide a link between current practice in mammography quality control and task-based image quality assessment. In order to apply the MSOpi for mammography image quality assessment along the lines of the EUREF Guideline procedure, an approach is developed which yields a contrast-detail curve for this model observer based on its AUC.

A framework for virtual mammography is built to simulate X-ray images of single cells of the CDMAM phantom. Virtual mammography allows the precision associated with a contrast-detail curve to be assessed, cf. [13]. The development of virtual mammography follows state-of-the-art simulation methods for ray casting [22,23], complemented by a phenomenological model of image degradations, such as noise, blurring and scattering.

Virtual mammography is used to compare the conventional determination of contrast-detail curves in mammography image quality assessment with the developed alternative based on the MSOpi observer. We investigate how close contrast-detail curves based on the MSOpi observer are related to those of the current EUREF Guideline procedure, and we explore the potential benefit of the MSOpi observer for an improved precision in the estimation of contrast-detail curves.

The paper is structured as follows. Section 2 recalls the currently recommended analysis of CDMAM phantom data for mammography image quality assessment, presents the proposed alternative analysis, and describes the implementation of virtual mammography. A comparison of the two approaches based on virtual measurements of the CDMAM phantom is then presented in Section 3 and discussed in Section 4. Finally, some conclusions are drawn.

2. Methods

2.1. Determination of contrast-detail curves

2.1.1. EUREF Guideline approach

16 images of the CDMAM phantom are acquired and analyzed. For each single image and each cell, a decision is taken as to which of the four corners contains the gold disc. This decision can principally be taken either by a human observer or by a data-driven rule. We restrict ourselves to a data-driven rule which, according to the EUREF Guideline procedure, selects that corner for which the average signal is minimum [2,10]. The data-driven rule is then applied a second time, with the corner containing the gold disc being replaced with the center disc. In this way, 32 trials of correctly localizing a gold disc out of 4 candidates are carried out when conducting 16 images, and the obtained frequency of correct localizations is taken as an estimate of the probability of correct localization.

For given disc area, the probability of a correct localization increases with increasing thickness of the disc. Similarly, the probability increases with increasing disc area for given thickness of the disc.

For each cell, the relative frequency of correct decisions is determined across the 32 trials. A logistic regression is then carried out for each diameter by fitting a psychometric curve to the observed frequencies of correctly localized discs in dependence on their thickness.

Fig. 2 shows examples of psychometric curves for typical data in mammography image quality assessment. For very small thickness, the decision approaches pure guessing and the probability tends to 25%. On the other hand, when the thickness becomes very large, the decision probability approaches 100%. Each psychometric curve refers to a fixed disc area; the larger the disc area, the more the corresponding psychometric curve is shifted to the left.

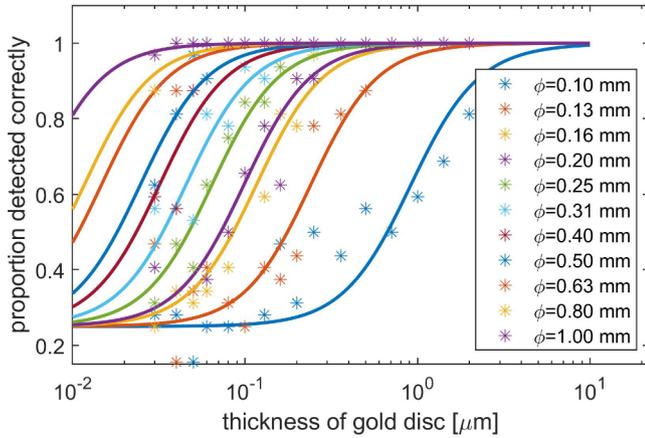The psychometric curves are parameterized as follows:

Fig. 2. EUREF Guideline procedure: examples of logistic regressions (Eq. (1)) in dependence on the thickness of the gold discs for various disc diameters ϕ and typical CDMAM data using 16 images.

$$p(t) = \frac{0.75}{1 + \exp[-\xi(C(t) - C_T)]} + 0.25, \qquad (1)$$

where $p(t)$ denotes the probability of correctly locating a gold disc of thickness t, and $C(t)$ denotes the contrast of a gold thickness (gold attenuation coefficient $\mu = 0.047 \mu m^{-1}$ at $31 keV$) of thickness t with

$$C(t) = \ln(1 - e^{-\mu t}).$$

$\xi > 0$ and $C_T < 0$ are free parameters [2,10,24] which are estimated using a nonlinear least square fit procedure, where ξ is constrained to be constant across diameters. Hence, all curves are fitted simultaneously, with individual values of C_T for each curve and a common value of ξ.

For each diameter, the critical thickness is determined as the thickness at which the fitted psychometric curve takes the value 0.625, which marks the center of the interval ranging from pure guessing (i.e., 0.25) to certainty (i.e., 1). The resulting function of critical thickness in dependence on the diameter is the contrast-detail curve. Fig. 3 shows the contrast-detail curve resulting from the single logistic regressions from Fig. 2, together with the limit curve [2]. In order to meet the quality assurance criteria, the contrast-detail curve needs to be below this limit curve.

The precision of the fitted psychometric curves and thus of the derived contrast-detail curve depends on the frequency of correct localization of the gold disc. These frequencies follow a binomial
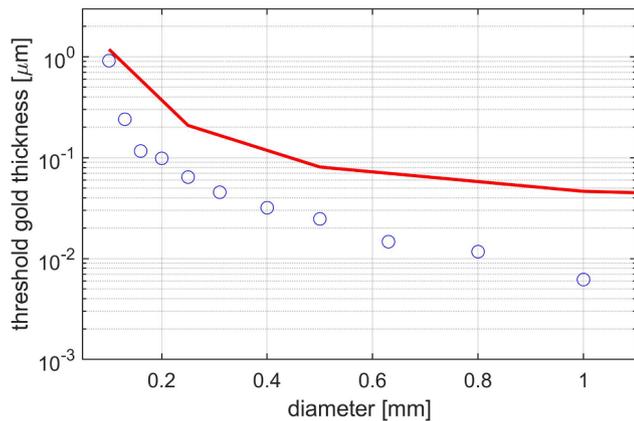

Fig. 3. EUREF Guideline procedure: example of a contrast-detail curve, based on the fit results of Fig. 2, showing the minimum disc thickness for each diameter that is needed to correctly locate the disc position with at least 62.5% probability. The solid curve shows the acceptable limit curve, see [2]. In order to meet the quality assurance criteria, the contrast-detail curve needs to be below the acceptable limit curve.

distribution, and the precision of the logistic regression therefore depends primarily on the number of repeated images.

2.1.2. Proposed alternative

An alternative way of defining and estimating a contrast-detail curve is proposed. The approach is based on the parametric binary model observer MSOpi [21]. Binary model observers are binary classifiers whose performance is characterized through the receiver operating characteristic (ROC) curve [16]. A single figure of merit derived from the ROC curve is the area under the ROC curve (AUC). For a perfect classifier, AUC equals 1, while for pure guessing AUC = 0.5 results. Current task-based image quality assessment utilizes AUC which can be viewed as an established metric.

The MSOpi model observer is a parametric observer that determines the presence or absence of a structure that has a constant signal height in a selected region of an image which otherwise consists of background noise only. The MSOpi model observer is directly applicable separately to each cell of the CDMAM phantom in that it classifies a selected corner as containing, or not containing, a disc. Hence, the signal region is defined as the second disc in the corner, and the background region as the remaining part of the cell with the exception of the central gold disc. When applying the observer to real data, the area corresponding to a gold disc can be identified through cross-correlation, utilizing the known structure of the phantom [25]. For our virtual mammography, the positions of the gold discs are known in advance and thus no prior segmentation has been applied.

Assuming that the position and size of the signal region is known exactly and the noise in the data is independent Gaussian noise with constant variance, the following analytic expression for the AUC of the MSOpi model observer has been derived in [21],

$$AUC = \Phi\left(\frac{1}{\sqrt{2}} \frac{|\theta - \mu|}{\sigma/\sqrt{m}}\right), \qquad (2)$$

where μ, θ denote the scalar mean pixel value of the background and the signal region, respectively, and m stands for the number of pixels in the signal region. For both regions, the variance of the noise is assumed to equal σ^2, and Φ is the standard normal distribution function.

The mean intensities observed for the central gold disc and the gold disc located in one of the corners can be expected to be the same. Hence, expression (2) remains equally valid if we define the signal region as the union of the central gold disc and the gold disc in the corner, which has been done in the following. Note that for each cell of the CDMAM phantom a different AUC of the MSOpi observer results in general, i.e. the parameters entering (2) vary across cells.

The AUC of the observer depends on the number m of pixels in the signal region, and it can simply be tuned by taking only a fraction, f say, of all pixels in the signal region. We will make use of this possibility to reach a contrast-detail curve that is close to that obtained by the current EUREF Guideline procedure. Thus, henceforth expression (2) is written as

$$AUC = \Phi\left(\frac{1}{\sqrt{2}} \frac{|\theta - \mu|}{\sigma/\sqrt{fn_1}}\right), \qquad (3)$$

where $m = fn_1$ denotes the actual number of pixels taken from the signal region for classification, and n_1 denotes the number of all pixels in the signal region. Note that for the estimation of θ, all pixels in the signal region are taken. Expression (3) can be extended to non-integer values of $m = fn_1$ in a natural way, and we will therefore employ expression (3) also for such cases.

The analytic expression (3) allows the AUC to be determined without actually applying the observer to a test set of images. Instead, the AUC can be estimated once estimates for the three parameters θ, μ, σ are available, and these parameters could in principle already be estimated from a single image of the CDMAM phantom.

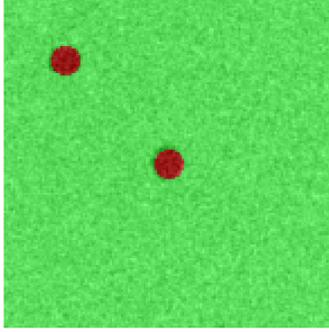To apply the MSOpi for the analysis of CDMAM cells, we utilize the

Fig. 4. Example image of a single CDMAM cell after grid removal. The red regions shade the two disc signals and are used to estimate θ and σ_1. The green region highlights the background and is used to estimate μ and σ_0. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

following estimate of AUC [21],

$$\widehat{AUC} = \Phi\left(\frac{1}{\sqrt{2}}\frac{|\hat{\theta} - \hat{\mu}|}{\hat{\sigma}/\sqrt{fn_1}}\right),\tag{4}$$

where $\hat{\mu}$ and $\hat{\theta}$ denote the pixel averages of background and (complete) signal region, respectively, and $\hat{\sigma}$ is a pooled estimate of the standard deviation of the noise,

$$\hat{\sigma} = \sqrt{\frac{1}{n_0 + n_1 - 2}((n_0 - 1)\sigma_0^2 + (n_1 - 1)\sigma_1^2)},$$

where σ_0 and σ_1 are the standard deviations observed in the noise and the signal region, respectively, and n_0 and n_1 denote the corresponding number of pixels. Fig. 4 shows the image of a single CDMAM cell containing two discs: one in the center and one in the upper left corner. The reddish colored pixels are used to estimate θ, n_1, and σ_1, while the greenish region marks the background and is taken to estimate μ, n_0 and σ_0.

In order to improve the precision of the estimation, more than a single image can be used for estimating μ, θ and σ. In that case the signal regions of the different images are merged into a single signal image region, and the same is done for the noise regions of the different images. After merging, the estimation of μ, θ and σ is done as described above.

The fraction f in (3) is chosen as

$$f = f(d) = \gamma\frac{d^2}{n_1 d_{min}^2},\tag{5}$$

with d_{min} denoting the minimum diameter of all discs, which is $0.1mm$ in our case (0.1 mm is also the minimum diameter in the EUREF Guideline procedure). The variable d denotes the known diameter of the disc in the current cell of the CDMAM phantom. The factor γ is a free factor. We chose $\gamma = 5/8$ such that the resulting contrast-detail curves are close to those of the EUREF Guideline approach, see Fig. 11. The choice of $f = f(d)$ in (5) ensures that the number of signal pixels taken for the observer scales with d^2, i.e., that it is proportional to the overall number of signal pixels and that the AUC remains unaffected if several images are merged, i.e. n_1 is enlarged. Hence, by inserting (5) into (4), we get

$$\widehat{AUC} = \Phi\left(\frac{1}{\sqrt{2}}\frac{|\hat{\theta} - \hat{\mu}|}{\hat{\sigma}}\frac{\sqrt{\gamma}d}{d_{min}}\right).\tag{6}$$

The remaining steps in the determination of a contrast-detail curve are similar to those in the EUREF Guideline approach described above. Specifically, the logistic regression carried out in the EUREF Guideline approach for the probability of correct localization is replaced with a logistic regression of the AUC in dependence on the thickness of the
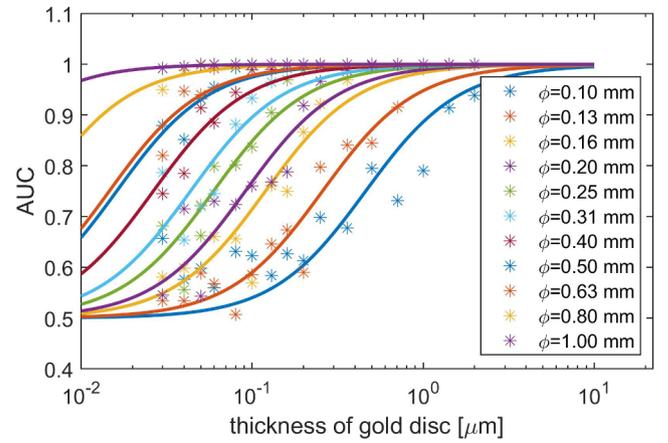

Fig. 5. Proposed approach: example of the logistic regressions (Eq. (4)) for the AUC of the MSOpi observer in dependence on the thickness of the gold discs for different diameters ϕ from typical mammography data using 5 images.

gold discs, and from the corresponding results a contrast-detail curve is then derived. The cells are grouped according to their diameter and for each diameter a psychometric curve is fitted to the estimated AUC in dependence on the thickness of the gold disc. Fig. 5 illustrates these logistic regressions. Since the AUC varies between 1/2 and 1, the following function is taken

$$AUC(t) = \frac{0.5}{1 + \exp[-\xi(C(t) - C_T)]} + 0.5,\tag{7}$$

where $AUC(t)$ denotes the AUC for gold disc thickness t.

The parameters ξ, $C(t)$ and C_T have the same meaning as those in (1), and they are estimated by nonlinear least squares in the same way as for the EUREF Guideline procedure. The curve in (7) tends to 0.5 for very small thickness, corresponding to an AUC of pure guessing, and approaches 1 for large thickness. For each diameter, the critical thickness is determined as the thickness for which the fitted logistic curve takes the AUC value of 0.75, which marks the center between pure guessing and certainty for a binary task. The resulting function of critical thickness in dependence on the diameter then defines the contrast-detail curve. Fig. 6 shows the contrast-detail curve resulting from the single logistic regressions from Fig. 5. Note that the choice of the AUC threshold for the proposed approach between pure guessing (i.e. AUC = 0.5) and certainty (i.e. AUC = 1) has been made in analogy to the choice made in the EUREF Guideline procedure.
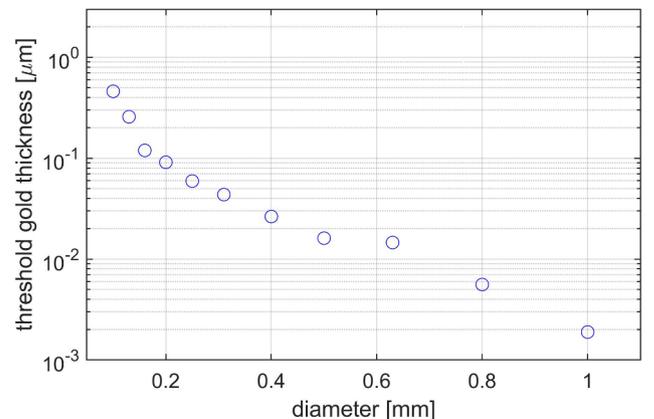

Fig. 6. Proposed approach: example of contrast-detail curve, based on the results of Fig. 5.
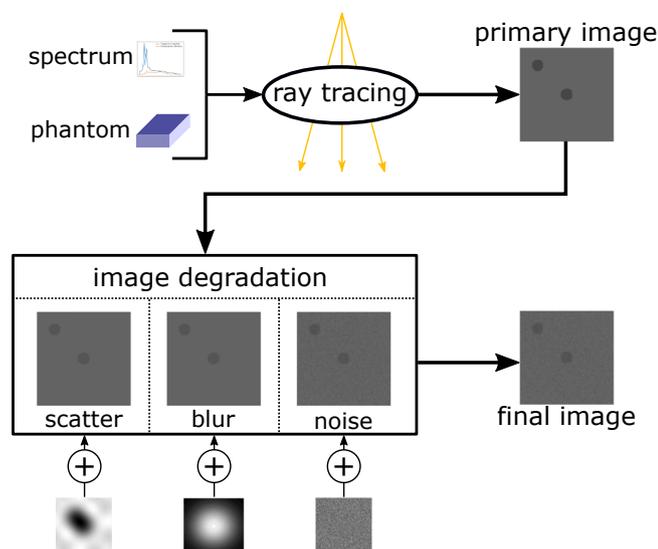
Fig. 7. Schematic overview of the simulation workflow. The primary image is determined by a ray-casting procedure and subsequently degraded by adding scattering, blurring, and noise to the image.

2.2. Virtual mammography

Virtual mammography enables the performance of a model observer to be evaluated on huge datasets. Simulated mammograms can achieve a remarkable similarity to real images, cf. [22,26,27].

Recently, virtual mammography was used to quantify the uncertainty associated with the contrast-detail curve as determined by the EUREF Guideline procedure [13]. Following this strategy, we developed a tool to simulate images of the CDMAM phantom. Fig. 7 shows a schematic overview of the image formation workflow used to simulate images of single CDMAM cells. Each cell consists of the corresponding grid boundary, the background and the two discs. A primary image is computed and subsequently degraded by adding scattering, blurring, and noise.

More specifically, the primary image is determined by a backward ray-casting procedure: for every detector pixel, the energy dependent attenuation of a given voxel phantom is determined using Siddon's algorithm [28]. The energy distribution of the X-rays is determined by an initial spectrum produced according to [29] considering a combination of a tungsten anode and a rhodium filter (). The focal spot is considered as a single spot. We computed the image to have a pixel-pitch of 0.06 mm/px, yielding a total of $\approx 85,000$ single rays to simulate.

At the degradation stage, we computed scattering, blurring and noise, and added them to the primary image. In film/screen imaging scattering is the main effect that reduces the contrast of mammographic images [30]. In digital imaging scattering affects the contrast-to-noise ratio as it introduces quantum noise. In our approach, we computed scattering using a convolution technique: we extracted a scattering image (I_S) by a convolution of the primary image (I_P) with a scatter-point spread function [31,32]. More precisely, we multiplied the Fourier transformation of the primary image with the Fourier transform of an ideal low pass filter with $D_0 = 2$ pixels. The scattering image is then obtained via inverse Fourier transformation of this product.

The amount of scattering can be controlled by using a parameter called the scatter to primary ratio (SPR) [33,34], which ensures that the pixel values of the scattering image are comparable to those of the primary image. SPR was chosen according to [33] for an equivalent breast thickness of 6 cm. The scattering degraded image is calculated by the sum of I_S and I_P

$$I = I_P + I_S \tag{8}$$

Note that the consideration of scattering is phenomenological and

not derived from a Monte-Carlo simulation of scattering for the described setting. A Gaussian filter with a standard deviation of 0.7 pixel was applied to model blurring.

Ravaglia et al. [35] discuss noise properties of different imaging devices, showing that noise consideration depends strongly on the kind of the detector used. Here, we applied Gaussian white noise in combination with a Gaussian filter technique designed to simulate spatially correlated noise. Specifically, white Gaussian noise was taken with standard deviation $\sigma = 10.15$ and subsequently spatially filtered by an isotropic Gaussian filter with standard deviation 0.55 pixel. The resulting standard deviation across images of the noise produced in this way was $\sigma_0 = 5.7$, which matched the standard deviation of the pixel values of a homogeneous area in a real image. The noise was then added to the signal. Note that σ and hence σ_0 are given in arbitrary units and their choice was made such that resulting images closely resemble real images.

Three datasets of 500 CDMAM images each were created for different signal-to-noise ratios (SNRs); here, SNR is taken as

$$SNR = \frac{|\theta - \mu|}{\sigma},$$

where θ and μ are average pixel values of regions with and without a signal. Note that the SNR varies across the cells of the CDMAM phantom for a given diameter, since $|\theta - \mu|$ changes with the thickness, while σ stays the same. The difference $|\theta - \mu|$ between the signal and the background was ≈ 2 for the gold disc with lowest thickness ($t = 0.03\,\mu m$) and ≈ 80 for the highest thickness ($t = 2.00\,\mu m$).

SNR of the three datasets was controlled through the choice of the standard deviation σ of the noise component, with $\sigma_l = \frac{1}{2}\sigma_0$, $\sigma_0 = 5.7$ and $\sigma_h = \frac{3}{2}\sigma_0$. The larger the standard deviation of the noise, the smaller the SNR. The choice of σ_l and σ_h was made such that significant differences in the resulting contrast-detail curves could be observed. For each simulation, the position of the phantom is slightly changed.

The remaining simulation parameters kV, SPR were chosen as follows: kV = 31 kV, SPR = 0.18 (cf. [33]) for a tungsten target and rhodium filter combination. For σ_0, the resulting images are visually similar with images measured by the German reference center in Münster. These images are acquired by a Hologic® Selenia® Dimensions® with a voltage of 31 kV and a X-ray tube current of 120 mA.

Fig. 8 shows the autocorrelation and the normalized noise power spectrum of the simulated images (left) and the real images (right). The autocorrelation indicates that the correlation length is about 1 pixel for the simulated images as well as the real images. However, the simulated images appear to have a higher correlation between adjacent pixels.

Also the noise power spectra of our simulated images match the noise characteristics of the real image. Thus, the phenomenological treatment of noise in the simulation chain achieves realistic noise characteristics in terms of the noise power spectrum and the correlation length.

Note, that the results presented here show a qualitative similarity between the noise characteristics of the simulated and the real images. The noise considered in the simulation falls in the range of that of a real detector and has no unusual noise features.

3. Results

The three datasets obtained for $\sigma = \sigma_l$, $\sigma = \sigma_0$ and $\sigma = \sigma_h$, respectively, with 500 virtual CDMAM images each were taken to explore the proposed approach for image quality assessment in mammography, and to compare its performance with that of the EUREF Guideline procedure. The EUREF Guideline procedure consists of two steps: First the position of the phantom and the image are aligned. After that an observer is applied to localize the second gold disc in the cell. Since for the simulation the positions are known in advance, alignment of the position of the phantom and image has not been done.
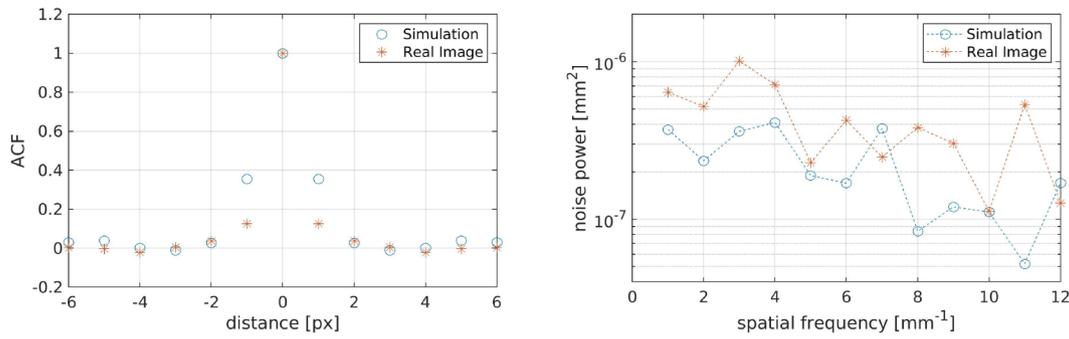
Fig. 8. Left: noise auto correlation for simulated images (blue) and the real images (red). Right: normalized noise power for the simulated images (blue) and the real images (red). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Figs. 2 and 3 show example results of the EUREF Guideline procedure based on 16 images randomly drawn from dataset 2 ($\sigma = \sigma_0$). The results illustrate the logistic regression (Fig. 2) and the resulting contrast-detail curve (Fig. 3). Figs. 4 and 5 show corresponding results for the proposed alternative approach based on 5 images from dataset 2.

To achieve a better reproducibility the EUREF Guideline proposes a smoothing of the data used for the logistic curve fitting. This smoothing can be done for the proposed method in the same way by applying the smoothing kernel from [11] to the estimates of AUC across the CDMAM phantom. Both procedures were repeated by applying the smoothing method prior to carrying out the logistic regression. The results show the expected behavior, i.e. for both methods the logistic curves become more smooth. Thus, the precision of the contrast-detail curves is increased. However, this increase in precision applies similarly to both methods, and the observed differences in precision between both approaches remain the same. In the following, results without prior smoothing are reported.

Application of the two approaches was then repeated 5000 times, each time drawing different images from dataset 2. In a few cases (<10) outlying results were produced by both approaches due to an unsuccessful termination of the nonlinear fit. Table 1 shows quantitative results for the threshold gold thickness for the different methods after removing the outliers where the fit failed. Fig. 9 shows the mean of the contrast-detail curves for the two approaches and illustrates their difference. The mean contrast-detail curves of both approaches are similar, albeit not identical, and hence a small bias between the two approaches remains. As can be seen, the two methods mainly deviate for small disc diameters ($d < 0.25$ mm). Note that for the two smallest diameters the method difference equals roughly the precision of the methods; for the other diameters the method difference is significantly smaller than the precision of the methods.

Table 1

Mean threshold gold thickness for the EUREF Guideline approach from 16 images and for the proposed approach from both 5 and 16 images. Uncertainties indicate the intervals containing (pointwise) 95% of all contrast-detail curves.

| Diameter [mm]	Mean threshold gold thickness [µm]		
	EUREF 16 images	MSOpi 5 images	MSOpi 16 images
0.10	0.78 ± 0.21	0.56 ± 0.15	0.568 ± 0.089
0.13	0.201 ± 0.038	0.248 ± 0.042	0.248 ± 0.025
0.16	0.116 ± 0.022	0.120 ± 0.023	0.120 ± 0.014
0.20	0.094 ± 0.018	0.095 ± 0.019	0.094 ± 0.010
0.25	0.057 ± 0.011	0.054 ± 0.011	0.054 ± 0.006
0.31	0.044 ± 0.008	0.040 ± 0.009	0.040 ± 0.005
0.40	0.030 ± 0.007	0.028 ± 0.007	0.028 ± 0.004
0.50	0.022 ± 0.006	0.020 ± 0.005	0.020 ± 0.003
0.63	0.014 ± 0.005	0.013 ± 0.004	0.013 ± 0.002
0.80	0.009 ± 0.005	0.007 ± 0.003	0.007 ± 0.002
1.00	0.005 ± 0.005	0.004 ± 0.003	0.004 ± 0.001
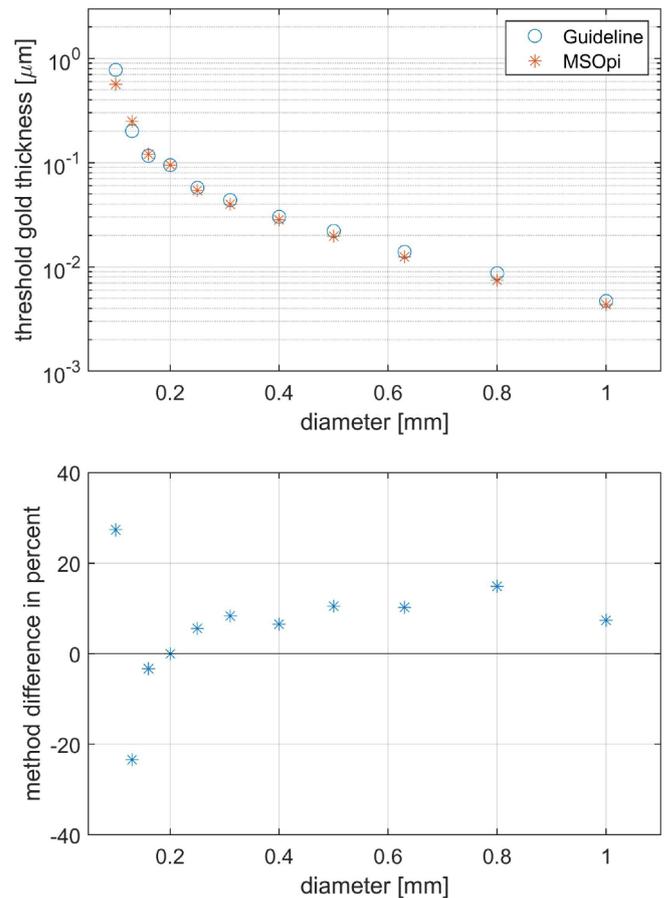
Fig. 9. Top: mean of the contrast-detail curves, obtained from repeatedly applying either the EUREF Guideline approach to 16 images from virtual mammography in each case (blue) and the proposed approach using 5 images (red). Bottom: difference of the two curves relative to the EUREF Guideline curve presented above. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 10 illustrates the precision in the estimation of the contrast-detail curves. The results have been determined for the three different datasets reflecting different signal-to-noise ratios ($\sigma = \sigma_l$, $\sigma = \sigma_0$ and $\sigma = \sigma_h$). The precision was determined from 500 repetitions of estimating contrast-detail curves, and the error bars in Fig. 10 indicate intervals containing (pointwise) 95% of all contrast-detail curves under repeated sampling of 16 images (EUREF Guideline procedure) and 5 images (proposed approach), respectively. Fig. 10 also shows the mean of the estimated contrast-detail curves which is shifted downwards for larger SNRs, indicating an improvement in image quality as expected. While for the two lower SNRs the contrast-detail
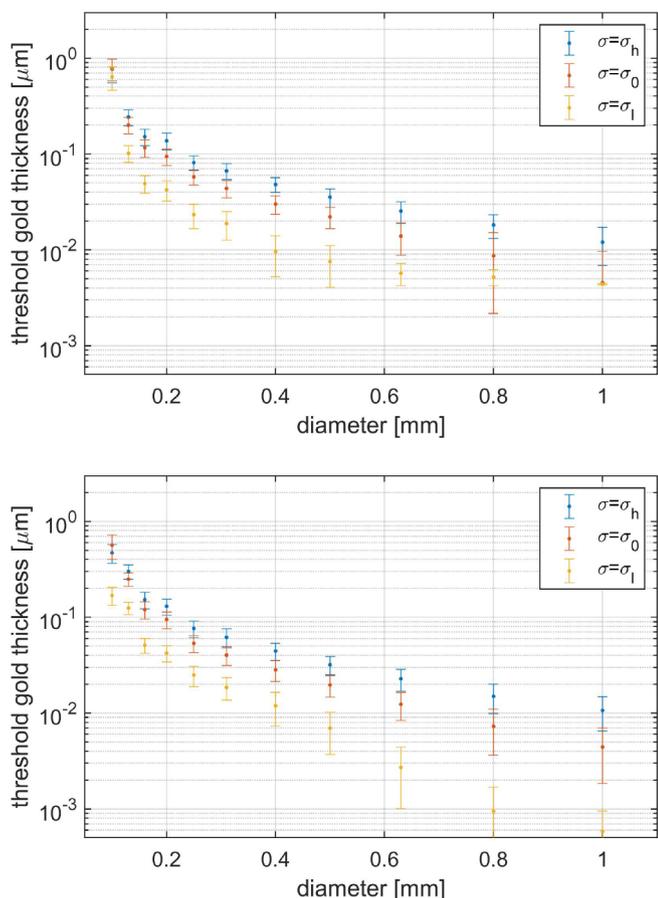
Fig. 10. Top: mean and intervals containing (pointwise) 95% of the contrast-detail curves, obtained from repeatedly applying the EUREF Guideline approach to 16 images from virtual mammography in each case, for different noise levels. Bottom: same results for the proposed alternative procedure, using 5 images.
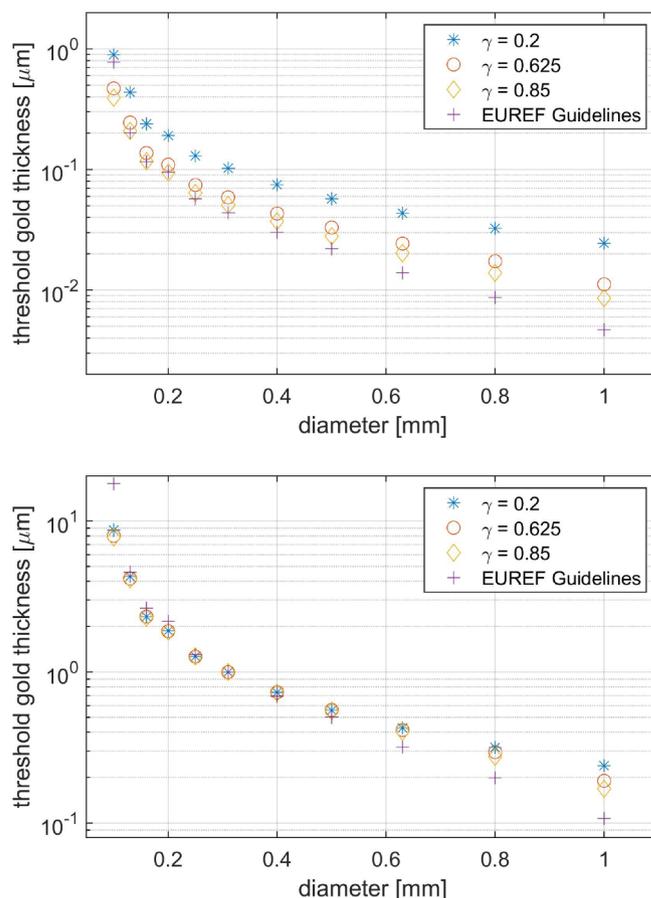
Fig. 11. Top: Variation of the contrast-detail curves as obtained by applying the parametric model observer for different choices of the free scaling factor γ. Bottom: Normalized contrast-detail curves with respect to the corresponding threshold thickness at $d = 0.25$ mm.

curves obtained by the EUREF Guideline procedure and the proposed approach are very similar, larger deviations can be observed for high SNR at 0.1 mm diameter and diameters beyond 0.5 mm. The fact that the contrast-detail curves of both procedures for the high SNR lie clearly below the curves obtained for lower SNR for most of the diameters may raise some doubts on the particular results of the EUREF Guideline procedure at diameters 0.1 mm and 1.0 mm. Note that for the proposed approach the contrast-detail curve at high SNR stays well beyond the curves for lower SNRs for all diameters, which appears to be reasonable.

Some of the above results are also given in Table 1, which in addition includes the outcome of the proposed approach when instead of 5 images 16 images are taken as recommended in the EUREF Guideline procedure.

The proposed approach provides the choice of a free parameter γ, see (6). Fig. 11 demonstrates how the choice of γ affects the resulting mean contrast-detail curve. γ directly scales the argument of the AUC function in (3) and changing γ hence shifts the resulting logistic curves (e.g. in Fig. 5) either to the right or to the left, which results in the change of the resulting contrast-detail curve indicated in Fig. 11 (top). However, varying γ only shifts the entire contrast-detail curve with respect to the threshold gold thickness. Fig. 11 (bottom) shows the normalized contrast-detail curve. The normalization was achieved by dividing a calculated contrast-detail curve by its threshold gold thickness at $d = 0.25$ mm. The figure indicates that different values of γ do not affect the shape of the contrast-detail curves.

We chose the factor γ such that the obtained difference between the

mean contrast-detail curve as obtained by the EUREF Guideline method and our approach is small for the considered scenario.

The choice of the critical AUC is another parameter that influences the resulting contrast-detail curves and could be used to further decrease the bias between the contrast-detail curves of the two approaches. However, no such fine-tuning has been carried out.

4. Discussion

Implementing virtual mammography allows huge image datasets to be produced that, in turn, can be used for the development of analysis methods. Knowledge about the ground truth allows the analysis methods to be quantitatively assessed. Following state-of-the-art virtual mammography [13,22] for primary image simulation in addition with a phenomenological image degradation model, we produced several virtual datasets of images of the CDMAM phantom with different SNRs. These datasets were used to assess a proposal for an alternative determination of a contrast-detail curve in mammography quality assessment and to compare it with the currently applied EUREF Guideline procedure.

The alternative determination of a contrast-detail curve is based on AUC, which has been established for image quality assessment in medical imaging [16]. One advantage of our proposal is thus to link the (proposed) contrast-detail curve to measures currently used in image quality assessment. However, the AUC is based on the evaluation of a technical phantom which actually poses a simpler problem than detecting non-spiculated lesions in a clinical study. Another advantage is that the precision of the estimated contrast-detail curve can be

significantly higher than that of the contrast-detail curve currently used in mammography image quality assessment.

The ultimate goal in current mammography image quality assessment is to ensure that the image quality is good enough for a reliable diagnosis. This goal can be seen as being achieved when the true contrast-detail curve falls below a limit curve (cf. Fig. 3). Due to the uncertainty in the estimation of the contrast-detail curve, there is an uncertainty in such a decision, particularly when the difference between the estimated contrast-detail curve and the limit curve is small. The smaller the uncertainty in the estimation of a contrast-detail curve, the more reliably a decision can be taken. Since the current EUREF Guideline procedure produces rather large uncertainties (cf. Figs. 7 and 8), decisions can become unreliable in cases in which the contrast-detail curve is close to the limit curve, and taking only 16 images may be deemed too few, cf. [13]. However, for practical reasons, the number of images ought to be kept small. In [13] it is also stated that the signal obtained for a diameter of $d = 0.1$ mm shows the highest uncertainty and thus it is crucial that this uncertainty is kept reasonably small. The proposed alternative procedure estimates contrast-detail curves with a significantly higher precision which is thus promising from a practical point of view.

The precision of the proposed approach depends on the precision in the estimation of the noise and the mean pixel values of signal and background region. Due to the large number of pixels within each cell, the precision in this estimation can be high even for a single image, which is the reason for the improvement in the precision of estimated contrast-detail curves. For the smallest diameter $d = 0.1$ mm the estimates are derived on 2–4 pixels only and the uncertainty is the highest here for both methods. As shown in Table 1 the MSOpi calculates a lower threshold gold thickness than the EUREF Guideline procedure. Additionally, the fluctuation under repeated sampling is lower for the proposed approach than for the EUREF Guideline procedure, allowing for a more reliable conformity assessment for the smallest diameter ($d = 0.1$ mm). Increasing the number of images further decreases the fluctuation of the estimates.

The proposed approach yields contrast-detail curves whose mean is similar, albeit not identical, to that obtained by the EUREF Guideline method.

The employed parametric model observer has a free parameter γ in the definition of the fraction $f(d)$ in (5) of points used to define the signal region. We demonstrated that the choice of γ allows to shift the entire contrast-detail curve. Thus, we scaled our method such that the magnitude of the obtained contrast-detail curve matches that of the EUREF Guideline procedure. However, γ does not influence the shape of the curve and it is encouraging, that both methods yield curves of similar shape. For our setting we selected γ to minimize the difference between the contrast-detail curves obtained by the automatic readout procedure from the EUREF Guideline and the outcome of our method. In Eq. (6) it is shown, that with the definition of $f = f(d)$ in (5) the AUC depends on the diameter d but is independent of the total number of pixels in the signal region. Thus, neither γ nor the AUC needs to be changed when taking another pixel pitch into account. However, different types of detectors typically have different noise characteristics. While a choice of γ could be appropriate for varying noise strength it needs to be adapted if the noise correlation changes. For a different noise power spectrum the value of γ has to be adapted so that the resulting contrast-detail curves match those of the EUREF Guideline procedure.

However, perfect agreement cannot be expected and applying the proposed approach would basically require to define a new limit curve in order to reach an equivalent assessment. This could be done by again utilizing virtual experiments, and by determining through them the bias between the contrast-detail curves of the two approaches. Since the current EUREF Guideline procedure aims to predict human performance, another way to calibrate the proposed approach (i.e. to determine its limit curve) would be to correlate its results with those of a

human observer study, similarly to the investigations carried out in [13,36]. Note that the bias between the contrast-detail curves of the two approaches observed for the three different datasets of virtual mammograms is essentially the same. Nevertheless, since the proposed approach is based on current measures of image quality [16,37,17], one might want to apply it even if it is not in perfect equivalence with the present EUREF Guideline procedure. One advantage of doing so is that the results of the proposed approach are directly traceable to established measures in image quality.

The results of this paper mainly provide a proof-of-principle and further investigations are required prior to applying the approach in daily routine. For example, the performance of the suggested method should be compared for mammography systems that are known to have varying quality. Furthermore, the approach needs to be tested on real data. Application to real data requires the alignment of the discs and maybe also introduces colored noise. Nevertheless, by replacing the observer in the automatic reading procedure of the current EUREF Guideline with the MSOpi, some simple modifications are sufficient to implement the proposed approach. The potential of the proposed approach to produce contrast-detail curves with significantly higher precision makes it a promising method.

5. Conclusions

Image quality in mammography is currently assessed in terms of contrast-detail curves derived from several images of the CDMAM phantom. The derivation of the contrast-detail curves is based on the probabilities of correctly locating gold discs of varying diameter and thickness. We have proposed a different way of arriving at contrast-detail curves from these images using measures which have been established in task-based image quality assessment. These measures are evaluated in combination with a parametric mathematical observer that takes advantage of the simple disc shaped objects in the CDMAM phantom. Virtual mammography software has been developed to evaluate and compare both approaches. One result is that the estimation of contrast-detail curves is more precise for the proposed approach than the currently applied one, and that it needs significantly fewer images in order to arrive at reliable decisions. Prior to its application in daily routine the proposed approach still needs further testing, e.g. on real data from systems with varying quality, or a comparison of its results to those of human observers. This requires in particular automatic localization of the discs using a method such as that applied in the current EUREF Guideline procedure.

Nevertheless, the encouraging results of the approach achieved for virtual mammography suggest that such an endeavor could be worthwhile, and that the approach altogether can prove useful for automatic mammography image quality assessment.

References

[1] Eurostat, 1 in 4 deaths caused by cancer in the EU28, Eurostat, the statistical office of the European Union; 2014.
[2] Perry N, Puthaar E. European guidelines for quality assurance in breast cancer screening and diagnosis. Eur Commun 2006.
[3] ICRU. ICRU Report No. 82: mammography – assessment of image quality. J. ICRU 9 (2).https://doi.org/10.1093/jicru/ndp029.
[4] Brenner DJ, Doll R, Goodhead DT, Hall EJ, Land CE, Little JB, et al. Cancer risks attributable to low doses of ionizing radiation: assessing what we really know. Proc

Natl Acad Sci 2003;100(24):13761–6.

[5] Gennaro G, Avramova-Cholakova S, Azzalini A, Chapel ML, Chevalier M, Ciraj O, et al. Quality Controls in Digital Mammography protocol of the EFOMP Mammo Working group. Phys Med 2018;48:55–64.

[6] Rojas LJ, Fausto AMF, Mol AW, Velasco FG, Abreu POS, Henriques G, et al. Optimization of the exposure parameters in digital mammography using contrast-detail metrics. Phys Med 2017;42:13–8.

[7] Zanca F, Bosmans H, Jacobs J, Michielsen K, Sisini F, Nens J, et al. Contrast-detail comparison between unprocessed and processed CDMAM images. Proc.SPIE. 2009. vol. 7258. pp. 7258–7258–10.

[8] Zanca F, Jacobs J, Van Ongeval C, Claus F, Celis V, Geniets C, et al. Evaluation of clinical image processing algorithms used in digital mammography. Med Phys 2009;36(3):765–75.

[9] Warren LM, Given-Wilson RM, Wallis MG, Cooke J, Halling-Brown MD, Mackenzie A, et al. The effect of image processing on the detection of cancers in digital mammography. Am J Roentgenol 2014;203(2):387–93.

[10] Karssemeijer N, Thijssen MAO. Determination of contrast-detail curves of mammography systems by automated image analysis. Digit Mammogr 1996;96:155–60.

[11] Young KC, Cook JJH, Oduko JM, Bosmans H. Comparison of software and human observers in reading images of the CDMAM test object to assess digital mammography systems. Med. Imaging 2006 Phys. Med. Imaging. International Society for Optics and Photonics, vol. 6142. 2006. pp. 614206–614206–13.

[12] Mackenzie A, Warren LM, Wallis MG, Given-Wilson RM, Cooke J, Dance DR, et al. The relationship between cancer detection in mammography and image quality measurements. Phys Med 2016;32(4):568–74.

[13] Mackenzie A, Eales TD, Dunn HL, Yip Braidley M, Dance DR, Young KC. Simulation of images of CDMAM phantom and the estimation of measurement uncertainties of threshold gold thickness. Phys Med 2017;39:137–46. https://doi.org/10.1016/j.ejmp.2017.06.019.

[14] ICRU. ICRU Report No. 54: medical imaging – the assessment of image quality. J ICRU 6 (2).https://doi.org/10.1093/jicru/ndl021.

[15] Barrett HH. Objective assessment of image quality: effects of quantum noise and object variability. JOSA A 1990;7(7):1266–78.

[16] Barrett HH, Myers KJ. Foundations of Image Science. In: Barrett Harrison H, Myers Kyle J, editors. Found. Image Sci. Wiley-VCH; 2003. p. 1584.

[17] Barrett HH, Myers KJ, Hoeschen C, Kupinski MA, Little MP. Task-based measures of image quality and their relation to radiation dose and patient risk. Phys Med Biol 2015;60(2):R1–75.

[18] Barrett HH, Yao J, Rolland JP, Myers KJ. Model observers for assessment of image quality. Proc Natl Acad Sci 1993;90(21):9758–65.

[19] Bouwman RW, van Engen RE, Broeders MJM, den Heeten GJ, Dance DR, Young KC, et al. Can the non-pre-whitening model observer, including aspects of the human visual system, predict human observer performance in mammography? Phys Med 2016;32(12):1559–69.

[20] Bouwman RW, Goffi M, van Engen RE, Broeders MJM, Dance DR, Young KC, et al. Can the channelized Hotelling observer including aspects of the human visual system predict human observer performance in mammography? Phys Med 2017;33:95–105.

[21] Anton M, Khanin A, Kretz T, Reginatto M, Elster C. A simple parametric model observer for quality assurance in computer tomography. Phys Med Biol 2018;63(7):16. 75011.

[22] Elangovan P, Warren LM, Mackenzie A, Rashidnasab A, Diaz O, Dance DR, et al. Development and validation of a modelling framework for simulating 2D-mammography and breast tomosynthesis images. Phys Med Biol 2014;59(15):4275–93.

[23] Bliznakova K, Sechopoulos I, Buliev I, Pallikarakis N. BreastSimulator: a software platform for breast x-ray imaging research. J Biomed Graph Comput 2012;2(1):1.

[24] Veldkamp WJH, Thijssen MAO, Karssemeijer N. The value of scatter removal by a grid in full field digital mammography. Med Phys 2003;30(7):1712–8.

[25] van Engen RE, Bosmans H, Dance DR, Heid P, Lazzari B, Marshall N, et al. Digital mammography update. European protocol for the quality control of the physical and technical aspects of mammography screening. S1, part 1: acceptance and constancy testing. Eur Guid Qual Assur Breast Cancer Screen Diagnosis 2013:1–54.

[26] Gong X, Glick SJ, Liu B, Vedula AA, Thacker S. A computer simulation study comparing lesion detection accuracy with digital mammography, breast tomosynthesis, and cone-beam CT breast imaging. Med Phys 2006;33(4):1041–52. https://doi.org/10.1118/1.2174127.

[27] Bellon C, Jaenisch G-R. aRTist–analytical RT inspection simulation tool, in. Proc DIR 2007:25–7.

[28] Siddon RL. Fast calculation of the exact radiological path for a three-dimensional CT array. Med Phys 1985;12(2):252–5. https://doi.org/10.1118/1.595715.

[29] Boone JM, Fewell TR, Jennings RJ. Molybdenum, rhodium, and tungsten anode spectral models using interpolating polynomials with application to mammography. Med Phys 1997;24(12):1863–74.

[30] Ducote JL, Molloi S. Scatter correction in digital mammography based on image deconvolution. Phys Med Biol 2010;55(5):1295–309.

[31] Ersahin A, Molloi S, Qian Y-J. A digital filtration technique for scatter-glare correction based on thickness estimation. IEEE Trans Med Imaging 1995;14(3):587–95.

[32] Love LA, Kruger RA. Scatter estimation for a digital radiographic system using convolution filtering. Med Phys 1987;14(2):178–85.

[33] Boone JM, Lindfors KK, Cooper VN, Seibert JA. Scatter/primary in mammography: comprehensive results. Med Phys 2000;27(10):2408–16.

[34] Dance DR, Day GJ. The computation of scatter in mammography by Monte Carlo methods. Phys Med Biol 1984;29(3):237–47.

[35] Ravaglia V, Bouwman R, Young K, van Engen R, Lazzari B. Noise analysis of full field digital mammography systems. Med. Imaging 2009 Phys. Med. Imaging. International Society for Optics and Photonics, vol. 7258. 2009. pp. 72581B–72581B–11.

[36] Bouwman RW, Mackenzie A, van Engen R, Broeders MJM, Young KC, Dance DR, et al. Toward image quality assessment in mammography using model observers: detection of a calcification-like object. Med. Phys. 2017;44(11):5726–39.

[37] Pepe MS. The statistical evaluation of medical tests for classification and prediction. Medicine 2003.