

Journal of Medical Imaging

MedicalImaging.SPIEDigitalLibrary.org

Region of interest based Hotelling observer for computed tomography with comparison to alternative methods

Adrian A. Sanchez
Emil Y. Sidky
Xiaochuan Pan

Region of interest based Hotelling observer for computed tomography with comparison to alternative methods

Adrian A. Sanchez,^{a,*} Emil Y. Sidky,^a and Xiaochuan Pan^{a,b}

^aThe University of Chicago, Department of Radiology, 5841 South Maryland Avenue, Chicago, Illinois 60615, United States

^bThe University of Chicago, Department of Radiation and Cellular Oncology, 5758 South Maryland Avenue, Chicago, Illinois 60615, United States

Abstract. We compare several approaches to estimation of Hotelling observer (HO) performance in x-ray computed tomography (CT). We consider the case where the signal of interest is small so that the reconstructed image can be restricted to a small region of interest (ROI) surrounding the signal. This reduces the dimensionality of the image covariance matrix so that direct computation of HO metrics within the ROI is feasible. We propose that this approach is directly applicable to systems optimization in CT; however, many alternative approaches exist, which make computation of HO performance tractable through a range of approximations, assumptions, or estimation strategies. Here, we compare several of these methods, including the use of Laguerre-Gauss channels, discrete Fourier domain computation of the HO (which assumes noise stationarity), and two approaches to HO estimation through samples of noisy images. Since our method computes HO performance exactly within an ROI, this allows us to investigate the validity of the assumptions inherent in various common approaches to HO estimation, such as the stationarity assumption in the case of the discrete Fourier transform domain method.

© The Authors. Published by SPIE under a Creative Commons Attribution 3.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.JMI.1.3.031010](https://doi.org/10.1117/1.JMI.1.3.031010)]

Keywords: image reconstruction; objective assessment; image quality; Hotelling observer.

Paper 14036SSPRRR received Mar. 31, 2014; accepted for publication Nov. 7, 2014; published online Dec. 8, 2014.

1 Introduction

Objective assessment of image quality through task-specific metrics has a long history in medical imaging and is regarded by many as ultimately being the most meaningful approach to medical image evaluation.^{1–4} However, the application of task-based assessment to x-ray computed tomography (CT) is recent relative to its application to planar imaging modalities and nuclear medicine. One reason for this delay is that metrics based on the Hotelling observer (HO),³ such as those considered in this work, involve the image covariance matrix, and in CT, this matrix is often extremely large (well over 10^9 elements), poorly conditioned, and possesses few, if any, simplifying structural properties. In order to address the challenge of large dimensionality, efficient channels have been proposed^{5–7} which essentially constitute a transformation of the image into a new basis, where the number of basis functions is substantially less than the number of image pixels. Another common means of circumventing the dimensionality problem is to assume noise stationarity, so that HO metrics can be obtained with relative computational efficiency through discrete Fourier transform (DFT) operations (see Sec. 2). Meanwhile, in order to address the somewhat unpredictable structure of the image covariance, various estimation strategies have been proposed which rely on samples of noisy images in order to construct an estimate for the image covariance when an analytic formulation of image covariance is impossible or infeasible.^{8,9}

Building on previous work,¹⁰ we propose a formalism, which, in certain cases, may be more appropriate than the

use of efficient channels or estimation techniques, while still addressing the issues of dimensionality and structural complexity of the image covariance matrix. Our approach differs from alternative methods in that the resulting metrics are nonstochastic, and we impose no restrictions on the structure of the signal or on the nature of the noise correlations. Instead, similar to Refs. 11–13, we only assume that the relevant noise correlations and signal extent can be restricted to a region of interest (ROI) in the image. In order to investigate the impact and validity of various assumptions that simplify the computation of HO metrics, we herein compare the results of parameter optimization using our proposed method to two alternatives: a set of efficient [Laguerre-Gauss (LG)] channels and a DFT domain approach. Each of these approaches corresponds to a different assumption regarding the signal, the image noise, or both. We then also investigate two statistical estimation strategies for HO metrics and compare these results with our proposed method.

The specific context in which our formalism is intended to be applied is detection or classification tasks where the object of interest is small (on the order of several pixels), and the reconstruction algorithm is a direct, linear algorithm, such as filtered backprojection (FBP). In this case, we hypothesize that most of the relevant information for performing the detection or classification task is likely to be contained in pixels within a small ROI surrounding the signal. For direct linear algorithms, an analytic form of the image covariance matrix for a small ROI can be constructed and its elements can be stored directly in computer memory so long as the ROI contains at most several thousand pixels. While we consider only two-dimensional (2-D) reconstruction in this work, the extension to three-dimensional (3-D) reconstruction is straightforward so long as the corresponding 3-D ROI contains only several thousand voxels.

*Address all correspondence to: Adrian A. Sanchez, E-mail: adrianalberto@uchicago.edu

Since the metrics evaluated in this work are task-specific, we restrict ourselves to two relevant tasks in a specific CT application, namely dedicated breast CT. The tasks considered are microcalcification detection and the Rayleigh discrimination task, which measures the ability of an observer to distinguish between a single object and two smaller objects. The specific metric we investigate is the HO efficiency, which is a summary scalar metric describing the preservation of relevant information from the projection data measurements to the reconstructed image. The mathematical construction of this metric is described in Sec. 2, along with a description of the two imaging tasks considered. Section 3 demonstrates the use of the proposed ROI HO method for the optimization of reconstruction filter width in the FBP algorithm. Section 3 also contains results for the application of a channelized HO (CHO), a DFT-domain HO, and two estimation approaches to the same parameter optimization. Finally, a discussion and brief conclusion are included in Sec. 4.

2 Methods

2.1 Task Modeling

Two tasks are considered in this work. The first is a Rayleigh resolution task, wherein an observer must classify an image as either an image of a single 1.4 mm bar convolved with a Gaussian with a full width at half maximum (FWHM) of 0.53 mm or two distinct Gaussians of FWHM = 0.53 mm, separated by a 0.8 mm trough. These two signals were modeled on a discrete grid with pixels an order of magnitude smaller than the detector pixels backprojected to the center of the field of view (FOV). The two signals are shown in Fig. 1.

The second task is a signal detection task, where we model a microcalcification as a Gaussian with FWHM of 80 μm . Here the two classes into which an observer classifies the image are a signal-present class and a signal-absent class. For both tasks, we operate under the signal-known-exactly-but-variable paradigm,³ where three signal locations are considered: the center of the FOV, the position (2 cm, 2 cm), and the position (4 cm, 4 cm); however, results among the three signal locations are not substantially different.

2.2 Breast CT Simulation

Since the tasks considered in this work involve small signals, all of the relevant geometric and scanning parameters used in the breast CT simulations are based on Ref. 14, the purpose of which is to characterize spatial resolution properties of a dedicated breast CT scanner. In a separate submitted work,¹⁵ we fully explore parameter ranges for this system and compare

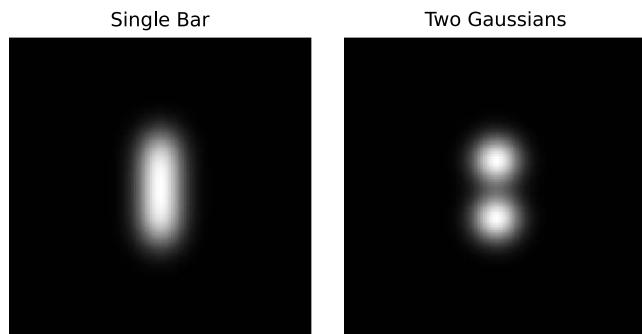


Fig. 1 The two signals simulated for the Rayleigh discrimination task.

with other work in optimization of breast CT systems. However, we here investigate only a few parameter settings and instead focus on HO performance estimation methods. To summarize, we simulate a 40 cm flat panel detector, rebinned to a sampling of 1024 detector bins. The source-to-isocenter distance is 45.8 cm, and the source-to-detector distance is 87.8 cm. The FOV is restricted to 16.59 cm, and this is sampled on a 512 \times 512 image pixel grid, for an image pixel size of 0.324 mm. Meanwhile, the detector element size, backprojected into the center of the FOV, is \sim 0.2 mm.

The simulated x-ray spectrum corresponds to an 80 kVp setting with added Be and Al filtration of 0.8 and 2.5 mm, respectively. Methods from Refs. 16–18 were used in simulation of the x-ray spectrum. Finite detector bin size is modeled by subsampling each detector bin at 16 evenly spaced locations. The peak attenuation value for both types of signals considered is that of calcium. The attenuation of the background medium is the mean attenuation of a breast composed of 50% adipose tissue and 50% glandular tissue, as determined in Ref. 19. The numerical phantom diameter used is 14 cm, and the total photon fluence at isocenter necessary to achieve the same dose as two-view mammography for this diameter ($\approx 2 \times 10^8 \text{ mm}^{-2}$) is obtained from Ref. 20. We restrict the dose in our simulation to the corresponding dose from mammography since breast CT is being investigated as an alternative screening modality. In order to determine the 2-D fluence necessary for our fan-beam simulations, we assume a slice thickness of 1 mm. The total number of incident photons in our fan-beam simulations is then $\bar{N}_0 = 4.17 \times 10^{10}$. For the range of parameters investigated here, 50 projection views were sufficient to ensure that angular undersampling artifacts remained farther from the signal of interest than the typical correlation length of the image noise. For this reason, the HO optimization was insensitive to these artifacts. We, therefore, used only 50 projection views in order to further minimize the computation burden of the optimization. The total fluence is divided equally among these 50 projection views. The reconstruction algorithm used is the FBP algorithm with a Hanning filter. The optimization of the Hanning filter width is the system optimization task demonstrated in this work.

2.3 Model Observer Method

For each of the approaches to system optimization considered, we employ the HO (Ref. 3) and its associated metric, the HO signal-to-noise ratio (SNR). The HO quantifies the degree to which an ideal observer with full knowledge of the relevant distributions can classify an image as belonging to one of two classes (e.g., signal-absent or signal-present), using only linear operations. Let us first consider the performance of the HO in the line-integral data domain (i.e., the projection data after applying the negative logarithm). Hereafter, bold font indicates a random variable, and bars over variables indicate population averages. Given a data vector \mathbf{g} , the HO will classify an image based on the outcome of a scalar value t , computed as

$$t = \mathbf{w}_g^T \mathbf{g}. \quad (1)$$

If we take $\Delta \bar{\mathbf{g}}$ to be the mean difference between data belonging to class 1 and data belonging to class 2, then the Hotelling template, \mathbf{w}_g , is defined as

$$\mathbf{w}_g = K_g^{-1} \Delta \bar{\mathbf{g}}, \quad (2)$$

where K_g is the data covariance matrix, whose (i, j) 'th component is given by $\text{Cov}(\mathbf{g}_i, \mathbf{g}_j)$. We take K_g to be diagonal, with diagonal elements given by²¹

$$(K_g)_{i,i} = (e^{\bar{g}_i} + 1)/\bar{N}_0, \quad (3)$$

where \bar{N}_0 is the average total number of incident photons in a blank scan. Further, throughout this work, we consider the data noise to be Gaussian. Since the matrix K_g is clearly signal dependent, in all further discussions, we consider K_g to be the average covariance matrix between the two signal classes. Under the assumption of equal class prevalence, the resulting HO is still the optimal linear observer, but may not be equivalent to the ideal observer.³ We then have that the HO SNR in the data domain (SNR_g) is given by

$$\text{SNR}_g^2 = \mathbf{w}_g^T \Delta \bar{\mathbf{g}}. \quad (4)$$

In order to evaluate the optimal reconstruction filter width for the task considered, we consider classification in the reconstructed image domain rather than in the projection data domain. In other words, the reconstructed image in vector form is given by $\mathbf{y} = \mathbf{A}\mathbf{g}$. The matrix \mathbf{A} denotes the FBP reconstruction operator. Since this operator is linear and discrete-to-discrete, its action is equivalent to left-multiplication with a matrix. We then denote this matrix as \mathbf{A} . The HO SNR in the image domain (SNR_y) is then given by

$$\text{SNR}_y^2 = \mathbf{w}_y^T \Delta \bar{\mathbf{y}}, \quad (5)$$

where $K_y \mathbf{w}_y = \Delta \bar{\mathbf{y}}$, $K_y = \mathbf{A}K_g\mathbf{A}^T$, \mathbf{w}_y is the Hotelling template in the image domain, and $\Delta \bar{\mathbf{y}}$ is the mean difference in the reconstructed image between the two hypotheses. Finally, we define the HO efficiency as $\varepsilon = (\text{SNR}_y/\text{SNR}_g)^2$.

We consider ε a more useful parameter for algorithm optimization since for many CT applications, one would not operate at the threshold for observer performance within which SNR is a meaningful metric. Instead, ε is a useful metric for algorithm and system evaluation that remains independent of the difficulty of a given task, while still retaining the exactness of computation

inherent in HO performance calculation through the exact covariance matrix K_y . A further motivation for selection of the efficiency metric over more conventional metrics, such as SNR or area under the ROC curve (AUC),²² is that we are interested in performing system component optimization rather than full system evaluation. A single value of SNR or AUC carries little information regarding how much one can hope to improve performance by optimizing a single component. Rather, we would prefer a metric like efficiency, which relates the quality of the input to the system component to the quality of the component's output. Here, the component that we optimize is a parameter of the reconstruction algorithm, so that the efficiency values are a reflection of how well the reconstruction preserves information relevant to the classification task.

In order to allow for direct manipulation of the image covariance matrix K_y , we restrict the reconstruction to an ROI image. As stated previously, 50 projection views are simulated. This angular sampling is sufficient for accurate reconstruction in the immediate vicinity of the signal; however, undersampling artifacts are visible elsewhere in the full image. We, therefore, restrict the ROI size based on the radius from the signal at which these artifacts begin to appear. Specifically, we consider the signal energy in the reconstructed image within an annulus of radius r and width $2\Delta r$, given by

$$E_y(r) = \int_0^{2\pi} d\theta \int_{r-\Delta r}^{r+\Delta r} [\Delta \bar{\mathbf{y}}(r', \theta)]^2 r' dr'. \quad (6)$$

The integrand of Eq. (6) is shown in Fig. 2(a) for the microcalcification signal and a single Hanning filter width. In order to select the ROI size, we approximate the above integral at evenly spaced values of r and set the ROI radius to the value of r that minimizes $E_y(r)$. An example of the function $E_y(r)$ is shown in Fig. 2(b), corresponding to the image shown in Fig. 2(a).

For the ROI sizes investigated here, ranging from 4×4 to 40×40 , the reconstruction matrix \mathbf{A} ranges in size from 16×4100 to 1600×4100 . Meanwhile, the covariance matrix K_y ranges from 16×16 to 1600×1600 . The method proposed remains numerically feasible for ROI sizes up to roughly 100×100 . Hereafter, we refer to this approach as ROI-HO.

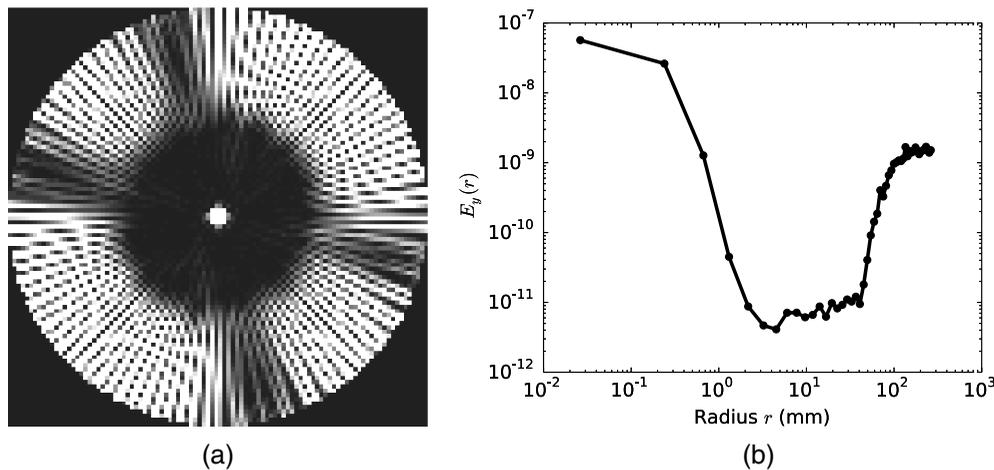


Fig. 2 (a) The integrand of Eq. (6) for the microcalcification detection task. (b) The reconstructed microcalcification signal energy as a function of radius from the signal, $E_y(r)$. In this case, the resulting region of interest (ROI) would have dimension of 14×14 pixels. The results shown here correspond to 50 projection views.

2.4 Approximation Strategies

One can show (see, for example, Ref. 3, Sec. 7.4.4) that if one assumes that the matrix K_y is circulant, the equation $K_y w_y = \Delta \bar{y}$ can readily be solved by application of the DFT operator. The assumption that K_y is Toeplitz corresponds to an assumption of noise stationarity, and the more restrictive assumption of a circulant K_y further implies cyclic correlations at the boundaries of the image. In order to test the validity of these assumptions, and, hence, of the DFT approach in determining HO performance for the present task, we extract a single row of the matrix K_y , corresponding to a pixel near the center of the signal location, and base a circulant approximation of K_y on this single row. We then repeat the optimization of filter width for microcalcification detection and Rayleigh discrimination under this approximation.

An alternative approach suggested by Gallas and Barrett^{3,5} is the use of efficient LG channels for estimation of HO performance. These channels are discretizations of circularly symmetric basis functions, formed as the product of an exponential function and the Laguerre polynomials. They form an orthonormal basis for all circularly symmetric square-integrable functions in \mathbb{R}^2 , and hence, they are an appropriate choice of channels for the task of detecting a circularly symmetric microcalcification. We, therefore, also apply 50 LG channels to the optimization of filter width for the microcalcification detection task. Rationale for this number of channels, as well as for the scale factor that determines the width of the Gaussian envelope of the LG channels is given in Sec. 3.

2.5 Estimation Strategies

Yet another approach to the estimation of HO metrics is to train and then test a linear classifier on samples of noisy images from each of the classes being considered.^{23,24} Using this approach, a set of training images from each class is used to compute a sample covariance matrix, along with sample mean images corresponding to each class. These are used to compute a template estimate, which is then applied to a series of testing images. The outcomes of the test statistic for each testing

image are then recorded and a Mann-Whitney U statistic is computed on the test statistic values, yielding an estimate of HO AUC,²² which can be related to HO SNR through the equation

$$\text{SNR} = 2 \operatorname{erf}^{-1}(2\text{AUC} - 1), \quad (7)$$

where erf^{-1} denotes the inverse error function.²⁵

For this approach, the authors of Ref. 5 suggest that, as a rule of thumb, 10 to 100 images are required for each row of the covariance matrix. For the case of a 30×30 ROI, this corresponds to anywhere between ~ 9000 to 90,000 images. While this is infeasible for a parameter optimization study in general, we demonstrate the outcome of this approach for image training and testing sets ranging from 500 to 3000 images in order to demonstrate the general trend of results obtained using this method. Due to the large number of independent noisy data sets required, determination of HO performance for each reconstruction filter width is performed using the same sets of noisy data. Two approaches, commonly termed the hold-out method and resubstitution are applied. In the first case, the training and testing phase use separate, independent sets of noisy data. In the second, the HO is tested on the same images with which it was trained. These two approaches yield estimates with negative and positive biases, respectively.^{23,24}

A more feasible approach based on sample images can be taken when prior knowledge of the mean images under each hypothesis is exploited in order to reduce the bias and variance of the estimator of HO performance, thereby reducing the necessary number of sample images. The construction of an unbiased estimator of HO SNR in this case is detailed in Ref. 9. We demonstrate the application of this approach using 300 and 700 noisy image samples (far fewer than the number of samples required for training and testing) and compare to the previous methods. Two scenarios are considered: (1) the same 300 or 700 noisy data realizations are used to reconstruct noisy images at each filter width considered and (2) a separate set of 300 or 700 data realizations is used to create noisy images at each filter width.

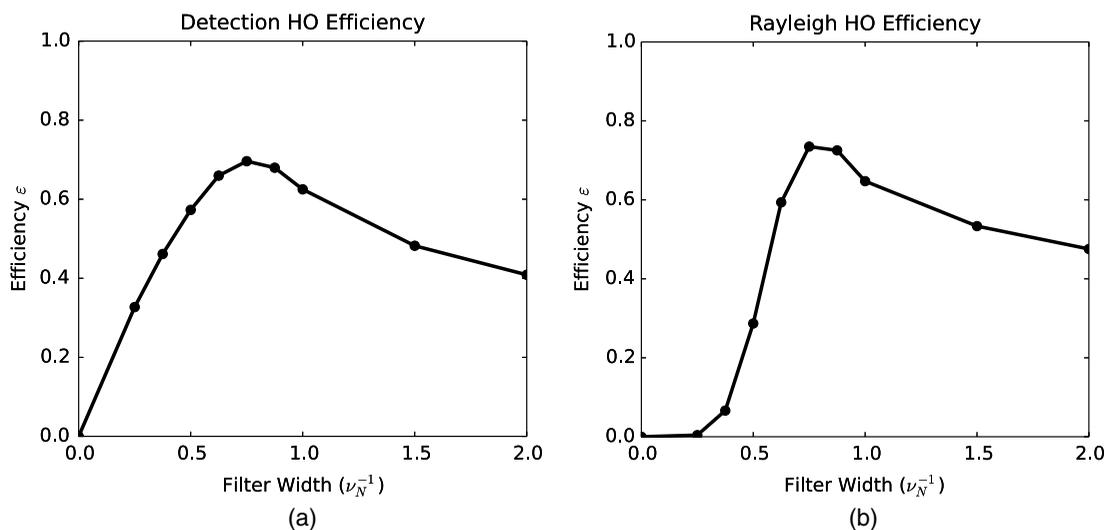


Fig. 3 Efficiency values are shown for various views and Hanning window widths (relative to the Nyquist frequency on the detector) for both the detection task (a) and the Rayleigh task (b). The Nyquist frequency in this case is $\nu_N = (1/2\Delta u) \approx 1.3 \text{ mm}^{-1}$. While the efficiency values for moderate filtering to no filtering were seen to have a dependence on ROI size, the same trend pictured here was seen for ROI sizes up to $\sim 1.5 \text{ cm}$ diameter.

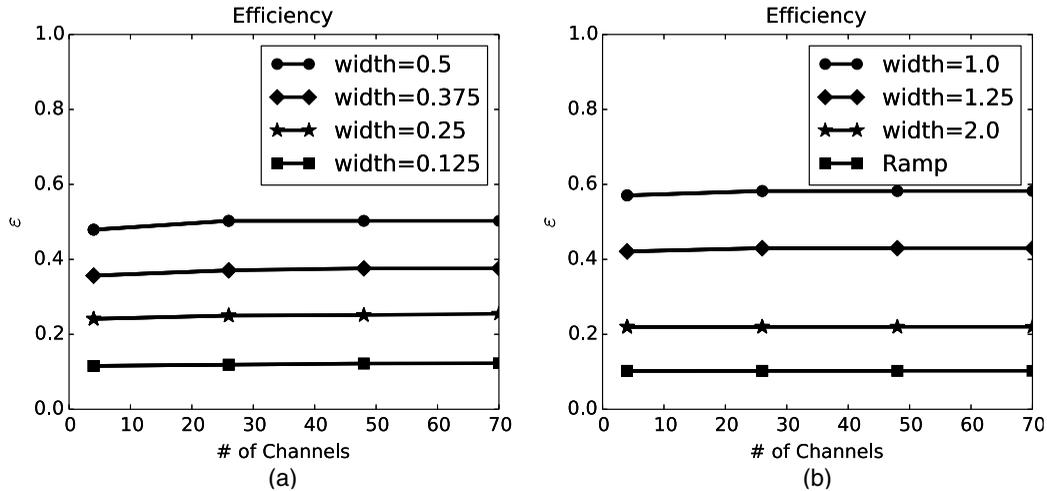


Fig. 4 Dependency of channelized Hotelling observer (CHO) efficiency on number of channels for a range of reconstruction filter widths from 0.125 to 0.5 (a), and 1.0 to the (unapodized) ramp filter (b), in units of the detector Nyquist frequency. For most filter widths, the CHO efficiency stabilizes with 50 channels; however, the performance estimates for this task and system are not sensitive to the number of channels used, in general. Subsequent results shown are for 50 channels.

3 Results

The results of applying the ROI-HO for microcalcification detection and Rayleigh discrimination are shown in Fig. 3 for a range of Hanning filter widths. The ROI-HO is noticeably sensitive to the reconstruction filter width, showing a clear maximum in performance for Hanning windows in the range of $0.75\nu_N$ to $0.825\nu_N$ for each task. However, this result should not be interpreted as giving a universally optimal filter width, but rather as a demonstration of the sensitivity of HO efficiency to relevant reconstruction algorithm parameters.

Next, we consider the use of channels for estimating HO performance in microcalcification detection. As with the ROI-HO, we restrict the CHO to an ROI, however, due to the Gaussian envelope that modulates the LG channels, this only has an effect for the smallest ROIs used. For the CHO using LG channels, results correspond solely to the microcalcification task, since the

Rayleigh task involves a signal that is not radially symmetric. Figures 4 and 5 demonstrate the dependence of CHO efficiency on the number of channels and the scale factor that modulates the width of the Gaussian envelope of the LG functions. For the majority of filter widths considered, the CHO efficiency estimate is completely stable above ~ 50 channels, while the optimum scale factor of the channels (full width at half maximum of the Gaussian) is roughly seven times the width of the microcalcification diameter. The remainder of the results presented corresponds to these CHO parameters. In our case, however, the dependence of the CHO performance estimates upon each of these parameters is weak, so that fewer channels or a slightly different scale factor could likely produce comparable results.

Figure 6 compares three approaches to optimization of the reconstruction filter width parameter, namely the ROI-HO approach proposed in this work, the CHO with 50 LG channels,

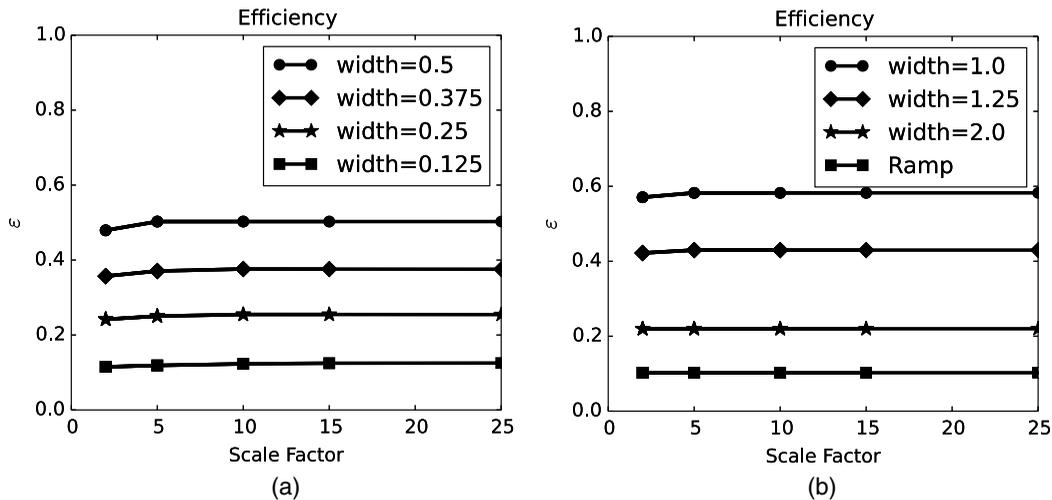


Fig. 5 CHO efficiency using 50 channels and 50 projection views for a range of Gaussian envelope widths in the Laguerre-Gauss channels. The x axis is normalized to the $80 \mu\text{m}$ microcalcification diameter. A scale factor of roughly seven times the microcalcification width was determined to be optimal. As in Fig. 4, (a) and (b) correspond to narrower and wider filters, respectively.

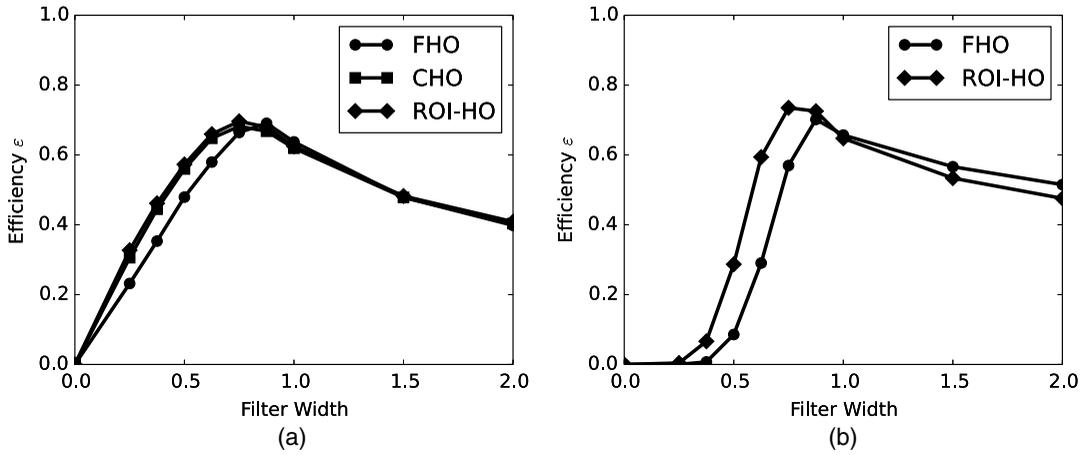


Fig. 6 A comparison of optimization of reconstruction filter width for 50 projection views. Shown are results obtained with the proposed ROI-HO, an approximate HO computed using the discrete Fourier transform (labeled FHO), and the CHO. The curves in (a) correspond to the detection task, while (b) corresponds to the Rayleigh task.

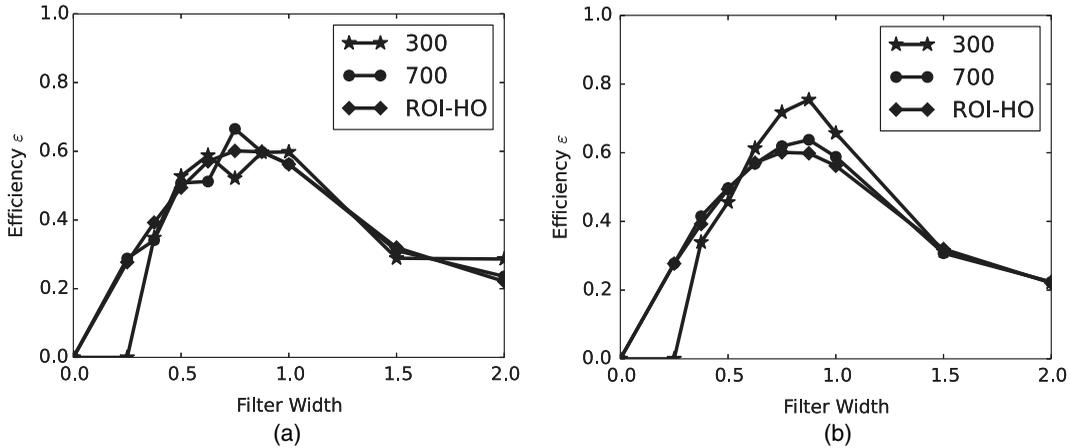


Fig. 7 Results of estimating the HO efficiency from sample images using the method proposed in Ref. 9 with 300 and 700 noisy sample images. (a) An independent set of images is used for each filter width. (b) The same noisy data realizations are used for each filter width. Error bars corresponding to two standard deviations of the estimates for 700 sample images are shown in Fig. 8.

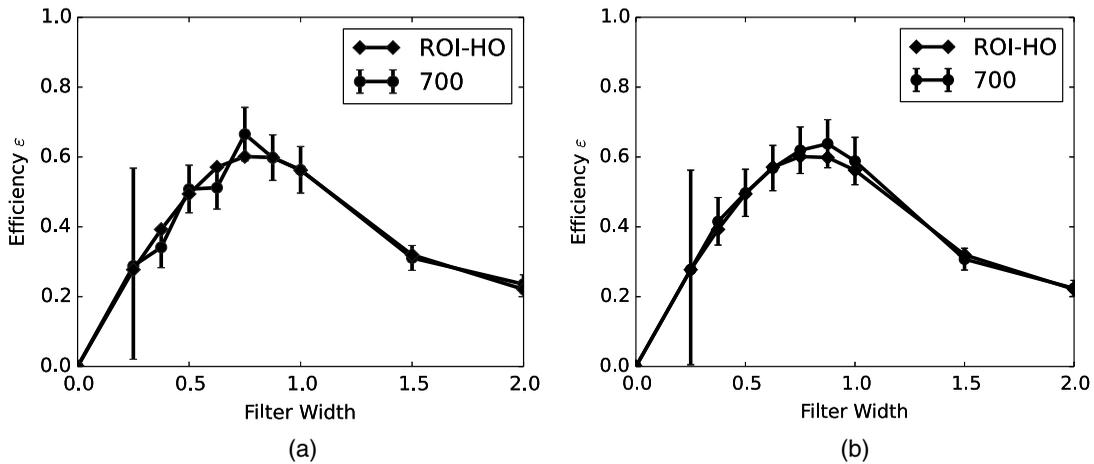


Fig. 8 Results from Fig 7 for 700 noisy sample images are shown with error bars corresponding to two standard deviations derived from jackknifing. The error bars illustrate statistical variation, but do not account for inherent estimator bias, which can be seen by comparing with the analytically computed ROI-HO curves. It is possible that other sources of uncertainty not investigated here, aside from statistical variations, could influence these results. As in Fig. 7, independent sets of images were used in (a), while the same set of images was used for each point in (b).

and the DFT-domain approach (labeled FHO). Since each of the methods utilizes the same ROI images, the ROI-HO that computes the exact HO performance should be taken as the ground truth for the sake of evaluating the alternative approaches. Clearly, for microcalcification detection, the CHO is an excellent approximation of the true HO. Further, the CHO allows for greater computation efficiency, since, in general, the ROIs used contained >50 pixels. However, the efficiency of this CHO comes with a loss of generality, as it is not applicable to general signals lacking radial symmetry, as in the Rayleigh discrimination task.

Meanwhile, the DFT-domain HO was capable of reproducing the general trend of the HO results; however, the quantitative reliability of this method varied widely depending on the

specific parameter setting used. For example, while wider filter widths allowed for accurate computation of detection task performance with the DFT-domain approach, the estimates of Rayleigh task performance using this method could vary from the HO by as much as 40%, as seen in the right side of Fig. 6. While this may not impact the optimization of filter width substantially, if absolute HO performance with a fixed filter width of, say, $0.625\lambda_N$ were of interest, the DFT-domain HO could be misleading.

Figure 7 demonstrates the use of noisy sample images to estimate HO efficiency when the mean image under each hypothesis is known. In our case, since we consider only linear image reconstruction algorithms, the mean image is produced simply by reconstructing an image without noise. The left and

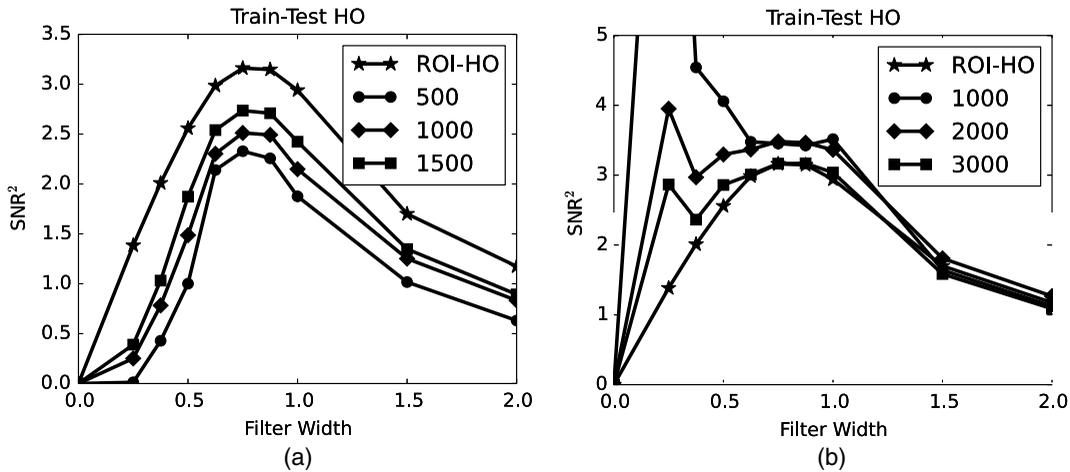


Fig. 9 (a) HO SNR² estimates from training and testing performed using the hold-out approach for 500, 1000, and 1500 training images, with an equal number of testing images. The prevalence of images from each class is also equal. (b) HO SNR² estimates resulting from training and testing performed using resubstitution for 1000, 2000, and 3000 total images. The bias and variance of the estimates is worst for narrow filter widths, where the size of the ROI used is largest. Variance of the estimates is illustrated in Fig. 10 through 95% confidence intervals derived from bootstrapping.

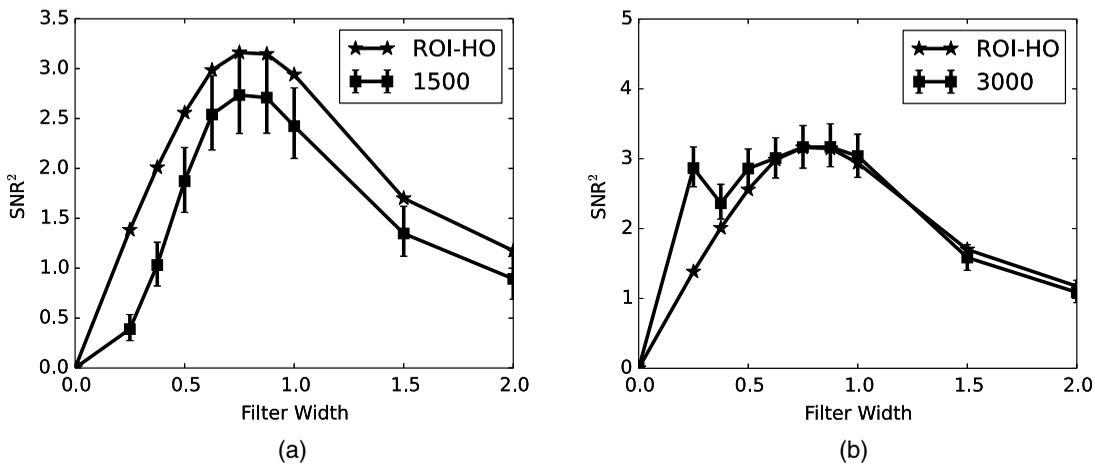


Fig. 10 Shown here are the results from Fig. 9 for the largest number of sample images with error bars denoting 95% confidence intervals derived from 1000 bootstrap samples. These errors derive from the variance of the training and testing estimator, while inherent bias in the estimator contributes additional error, especially for small filter widths where the number of ROI pixels is large. As in Fig. 8, the error bars here only convey statistical uncertainties, and bias can be inferred by comparison with the analytically computed ROI-HO curves. The results in (a) correspond to the hold-out approach, while the results in (b) correspond to resubstitution.

right plots correspond, respectively, to the use of independent noisy data sets for each filter width and the reuse of the same simulated data for each filter width. The variability seen in the left-hand figure provides intuition as to the variance of the efficiency estimates, while the figure on the right illustrates that, although subject to the same variability, reusing the same data realizations correlates the estimates for different filter widths, potentially allowing for more reliable rank ordering of parameter settings. In other words, the estimated curves on the right could undergo vertical translation due to the variance of the estimator, but are more likely to preserve their shape than the curves shown on the left.

Figure 9 illustrates the use of linear classifier training and testing in order to estimate HO performance. In this case, prior knowledge of the mean images is not exploited, and sample estimates of the images are computed instead. The left-hand plot corresponds to the hold-out approach, where independent image sets are used for the training and testing phases, while the right-hand plot is generated using resubstitution. As seen in the figure, these two approaches introduce negative and positive biases in the estimates, respectively. Note that thousands of images are required in order to construct quantitatively meaningful estimates. However, the general trend of the results can be seen for comparatively few samples. This suggests that for certain tasks, such as rank-ordering of only a few options for algorithm implementation, the training and testing approach could be useful if one lacks an adequate model of the image covariance or class means.

4 Conclusions

We have demonstrated that for classification tasks involving small signals in CT, the HO performance computed within an ROI constitutes an efficient and objectively meaningful approach to reconstruction algorithm and system optimization. Specifically, we have employed the HO to optimize reconstruction algorithm filter width for two tasks: microcalcification detection in dedicated breast CT and Rayleigh discrimination.

We performed a comparison of this approach with several alternatives for objective assessment. In broad terms, these alternatives fall into two categories: methods based on computation of statistical estimates and methods based on approximations regarding the signal and/or image noise. While our proposed methodology assumes that information beyond a small ROI is irrelevant to the classification task, the CHO imposes assumptions of rotational symmetry on the signal and Hotelling template, and the Fourier-domain approach assumes stationary noise. Strictly speaking, none of these assumptions are valid in CT for the tasks considered here; however, each approach possesses strengths and weaknesses. The Fourier-domain HO is the most computationally efficient method for large ROIs, but is less accurate than the ROI-HO or CHO approach. Meanwhile, the CHO with LG channels is limited to cases when the signal of interest and Hotelling template possess radial symmetry.

The statistical estimation approaches investigated possess an attractive degree of flexibility, in that they do not rely on any assumptions regarding the image covariance matrix or, in the case of the training/testing approach, the signals themselves. However, this flexibility comes at the expense of computational efficiency, as this family of approaches requires excessively large numbers of noisy samples, making any extensive system or

algorithm optimization prohibitively time-consuming. Finally, while we have illustrated some of the trade-offs, benefits, and shortcomings of several methods for objective assessment in CT, ultimately future work will be necessary to ensure that the proposed methodology yields metrics that correlate with improved performance of humans for the given tasks.

Acknowledgments

The authors would like to thank Brandon Gallas and Adam Wunderlich for helpful discussions regarding this work. This work was supported in part by NIH R01 Grant Nos. EB000225, CA158446, CA182264, EB018102 and NIH F31 Grant No. EB018742. The contents of this article are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health.

References

1. R. F. Wagner and D. G. Brown, "Unified SNR analysis of medical imaging systems," *Phys. Med. Biol.* **30**(6), 489 (1985).
2. A. E. Burgess et al., "Efficiency of human visual signal discrimination," *Science* **214**(4516), 93–94 (1981).
3. H. H. Barrett and K. J. Myers, *Foundations of Image Science*, John Wiley & Sons, Inc., Hoboken, New Jersey (2004).
4. C. E. Metz, "ROC methodology in radiologic imaging," *Invest. Radiol.* **21**(9), 720733 (1986).
5. B. D. Gallas and H. H. Barrett, "Validating the use of channels to estimate the ideal linear observer," *JOSA A* **20**(9), 1725–1738 (2003).
6. S. Park, J. M. Witten, and K. J. Myers, "Singular vectors of a linear imaging system as efficient channels for the Bayesian ideal observer," *IEEE Trans. Med. Imaging* **28**, 657–668 (2009).
7. J. Witten, S. Park, and K. Myers, "Partial least squares: a method to estimate efficient channels for the ideal observers," *IEEE Trans. Med. Imaging* **29**, 1050–1058 (2010).
8. H.-P. Chan et al., "Classifier design for computer-aided diagnosis: effects of finite sample size on the mean performance of classical and neural network classifiers," *Med. Phys.* **26**(12), 2654–2668 (1999).
9. A. Wunderlich and F. Noo, "New theoretical results on channelized Hotelling observer performance estimation with known difference of class means," *IEEE Trans. Nucl. Sci.* **60**, 182–193 (2013).
10. A. A. Sanchez, E. Y. Sidky, and X. Pan, "Task-based optimization of image reconstruction in breast CT," *Proc. SPIE* **9037**, 90370J (2014).
11. M. P. Eckstein et al., "Automated computer evaluation and optimization of image compression of x-ray coronary angiograms for signal known exactly detection tasks," *Opt. Express* **11**(5), 460–475 (2003).
12. A. H. Baydush et al., "Computer aided detection of masses in mammography using subregion Hotelling observers," *Med. Phys.* **30**(7), 1781–1787 (2003).
13. C. Chou and M. A. Anastasio, "Noise texture and signal detectability in propagation-based x-ray phase-contrast tomography," *Med. Phys.* **37**(1), 270–281 (2010).
14. A. L. C. Kwan et al., "Evaluation of the spatial resolution characteristics of a cone-beam breast CT scanner," *Med. Phys.* **34**, 275–281 (2007).
15. A. A. Sanchez, E. Y. Sidky, and X. Pan, "Task-based optimization of dedicated breast CT via Hotelling observer metrics," *Med. Phys.* **41**(10), 101917 (2014).
16. J. M. Boone and J. A. Seibert, "An accurate method for computer-generating tungsten anode x-ray spectra from 30 to 140 kV," *Med. Phys.* **24**(11), 1661–1670 (1997).
17. J. M. Boone, T. R. Fewell, and R. J. Jennings, "Molybdenum, rhodium, and tungsten anode spectral models using interpolating polynomials with application to mammography," *Med. Phys.* **24**(12), 1863–1874 (1997).
18. J. M. Boone, "Spectral modeling and compilation of quantum fluence in radiography and mammography," *Proc. SPIE* **3336**, 592–601 (1998).
19. P. C. Johns and M. J. Yaffe, "X-ray characterisation of normal and neoplastic breast tissues," *Phys. Med. Biol.* **32**(6), 675 (1987).
20. J. M. Boone et al., "Technique factors and their relationship to radiation dose in pendant geometry breast CT," *Med. Phys.* **32**, 3767 (2005).

21. H. H. Barrett and W. Swindell, *Radiological Imaging: The Theory of Image Formation, Detection, and Processing*, Academic Press, San Diego, California (1996).
22. J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (ROC) curve," *Radiology* **143**, 29–36 (1982).
23. R. F. Wagner et al., "Multivariate Gaussian pattern classification: effects of finite sample size and the addition of correlated or noisy features on summary measures of goodness," *Lec. Notes Comput. Sci.* **687**, 507–524 (1993).
24. K. Fukunaga and R. Hayes, "Effects of sample size in classifier design," *IEEE Trans. Pattern Anal. Mach. Intell.* **11**, 873–885 (1989).
25. J. Beutel et al., *Handbook of Medical Imaging. Vol. 1. Physics and Psychophysics*, SPIE Press, Bellingham, Washington (2000).

Adrian A. Sanchez received his BS degree in physics from the University of Tennessee, Knoxville, in 2010, and is currently a PhD candidate at The University of Chicago in the Committee on Medical Physics, graduate program in medical physics with an anticipated graduation in 2015. His primary research interests are currently

tomographic image reconstruction and objective image quality assessment.

Emil Y. Sidky received his BS degree in physics and mathematics from the University of Wisconsin, Madison, in 1987 and his PhD degree in physics from The University of Chicago in 1993. He held academic positions in physics at the University of Copenhagen and Kansas State University. He joined the University of Chicago in 2001, where he is currently a research associate professor. His current interests are in CT image reconstruction and large-scale optimization.

Xiaochuan Pan received his BS degree from Beijing University, Beijing, China, in 1982, his MS degree from the Institute of Physics, Chinese Academy of Sciences, Beijing, in 1986, and the master's and PhD degrees from The University of Chicago in 1988 and 1991, respectively, all in physics. He is currently a professor with the Department of Radiology, The University of Chicago. His research interests include physics, algorithms, and applications of tomographic imaging.