To be published in Journal of the Optical Society of America A:

Title:	Channelized Hotelling observers for the assessment of volumetric imaging data sets
Authors:	Ljiljana Platisa, Bart Goossens, Ewout Vansteenkiste, Subok Park, Brandon Gallas, Aldo Badano, and Wilfried Philips
Accepted:	4 March 2011
Posted:	4 March 2011
Doc. ID:	125530

Channelized Hotelling observers for the assessment of volumetric imaging data sets

Ljiljana Platiša,[1,*] Bart Goossens,[1] Ewout Vansteenkiste,[1] Subok Park,[2]

Brandon D. Gallas,[2] Aldo Badano,[2] and Wilfried Philips[1]

[1] *TELIN-IPI-IBBT, Ghent University, St-Pietersnieuwstraat 41, B-9000 Ghent, Belgium*

[2] *CDRH\FDA, 10903 New Hampshire Ave, Silver Spring, MD 20993, USA*

[*] *Corresponding author: Ljiljana.Platisa@Telin.UGent.be*

Current clinical practice is rapidly moving in the direction of volumetric imaging. For two-dimensional (2D) images, task-based medical image quality is often assessed using numerical model observers. For 3D images, however, these models have been little explored so far. In this work, first, two novel designs of a multi-slice channelized Hotelling observer (CHO) are proposed for the task of detecting 3D signals in 3D images. The novel designs are then compared and evaluated in a simulation study with five different CHO designs: a single-slice model, three multi-slice models and a volumetric model. Four different random background statistics are considered, both Gaussian (non-correlated and correlated Gaussian noise) and non-Gaussian (lumpy and clustered lumpy backgrounds). Overall, the results show that the volumetric model outperforms the others, while the disparity between the models decreases for greater complexity of the detection task. Among the multi-slice models, the second proposed CHO could most closely approach the volumetric model whereas the first new CHO seems to be least affected by the number of training samples.

OCIS codes: 110.2960, 110.2970, 110.3000, 110.4155, 330.1880, 330.5510.

1. Introduction

Today, medical imaging is an essential part of clinical practice. The primary goal of medical imaging is to assist physicians in the diagnostic process. Given the seriousness of a diagnostic error, reliable and valid *image quality* assessment is of fundamental importance in optimization and evaluation of medical imaging systems.

In its most general sense, image quality is often characterized as a measurement of the image degradation. To that end, a number of "task-independent" metrics have been defined to evaluate a great range of factors which may affect the quality of a medical image: noise [1, 2], contrast resolution [1], and spatial resolution [3], to mention just a few. However, medical images are inherently *task-specific* rather than task-independent. In this respect, image quality for medical applications shall be defined in terms of how well, given the images, the specific diagnostic task can be performed by a physician [4, 5].

In that manner, the task-based image quality assessment is determined by the task of interest, the image data, the observer to perform the task and the measure of observer performance [6]. The diagnostic task in medical imaging is generally either estimation, quantifying one or more parameters of interest using the given image data; classification, deciding to what class an image belongs; or hybrid estimation-classification, when estimation and classification are combined [7]. In this work, we focus on one particular classification task called *signal detection*, where the image is classified as signal-absent (normal clinical case) or signal-present (abnormal clinical case). Tumor detection in PET scans, bone metastasis detection in bone SPECT scans, and mass detection in breast tomosynthesis are some common examples of relevant clinical tasks.

Until recently, medical images and thus the detection tasks were limited to single-slice or two-dimensional (2D) views, *i.e.*, detection of planar signals in 2D images, often projections or reconstructed 2D images. In recent years, the advent of volumetric image acquisition and visualization (PET/SPECT, MRI, breast tomosynthesis, CT) has profoundly shifted the paradigm towards the detection of lesions (signals) using multi-slice reconstructed image data [8–10]. Following these trends, assessing and optimizing image quality for volumetric

image analysis is one of the major challenges in medical imaging today.

The most obvious and currently still the most widely used task-based assessment of medical image quality is a reader study. In reader studies, the observers, often true medical experts, read a set of test images, synthetic or real clinical ones, and make a diagnostic decision for those images. To their disadvantage, reader studies are often time consuming and expensive. As an alternative, mathematical model observers may be used [11,12]. In general, two major types of model observers can be identified [7]: *ideal observers* which set an upper bound on the signal-detection performance of any observer [12–15], and *anthropomorphic observers* which are designed to mimic human observer mechanisms and performance in a given detection task [16, 17]. Commonly, two figures of merit are used to quantify observer performance in a binary classification task [6, 18–20]: the area under the receiver operating characteristic curve (AUC) and the task signal-to-noise ratio (SNR).

In signal detection theory [21], the observer which has a full knowledge of the statistical information of the image data is known as the Bayesian ideal observer (IO). The IO is optimal among all observers, either human or model, in the sense that it maximizes the diagnostic accuracy as measured by the AUC. Consequently, for design and optimization of data acquisition hardware, detection performance of the IO is preferred over any other observer. In practice, however, it is often difficult, if not impossible, to derive or estimate the IO performance. This is due to high dimension and great complexity of the image statistics that are unknown and poorly estimated for real clinical data sets. The IO is tractable only for simple stylized settings, like when the data is Gaussian, in which case the IO is linear.

It needs no debate that the clinical detection task is a complex mechanism to model, already in 2D, let alone in 3D, and thus simplifications are inevitable. This concerns both the

observer model and the image data. One practical alternative to the ideal observer is the ideal *linear* observer known as the Hotelling observer (HO). The HO is optimal among all linear observers in that it maximizes the SNR [7]. Additionally, when the image data are Gaussian distributed, the HO is equal to the IO. Another simplification for the observer models is the so-called *channelized Hotelling observer* (CHO) proposed by Myers and Barrett [22]. In essence, the CHO is an HO constrained to a number of channels. Originally, the channels were inspired by the properties of human visual system (HVS) related to examination of the data through frequency selective channels. An important advantage of the channelized models over the non-channelized ones is dimensionality reduction of the problem, which has been discussed by Barrett *et al.* [23].

Depending on the properties of the channels relative to the image statistics in the task, the CHO can be used either to approximate the ideal observer (*efficient* channels) or to track humans (*anthropomorphic* channels). For example, in 2D images, Gallas and Barrett [15] found Laguerre-Gauss (LG) channels to be efficient in detection tasks using various lumpy backgrounds and rotationally symmetric signals. Not limited to types of backgrounds and signals are the singular-value-decomposition channels used by Park *et al.* [24, 25] which only require the system to be linear and the system's response functions to be known. Most recently, Witten *et al.* [26] investigated channels chosen by the partial least squares (PLS) method, which identifies channels based on the image and truth data covariance. Regarding anthropomorphic channels, their most common feature is that they have low or no response to low-frequency data, such as Gabor filters used in the study of Eckstein *et al.* [16] or the difference-of-Gaussian (DOG) and square channels which Abbey *et al.* [17] used in their experiments.

As an attempt to incorporate higher order statistics to the CHO and better approximate the IO, Park *et al.* [24,25,27,28] developed what they call a channelized ideal observer (CIO). The authors demonstrate that the CIO using LG channels [27, 28], or more generally using system singular vectors [24, 25] or PLS channels [26], could well approximate the IO even for non-Gaussian images.

On the side of image data, the most simplified approach of task-based image quality assessment restricts the task of interest to detecting whether a known object (signal) is present at one specified location in a known background, the so-called binary signal-known-exactly and background-known-exactly (SKE/BKE) detection task [22, 29, 30]. More complicated and more clinically relevant, are the paradigms of background-known-statistically (BKS) [13, 15, 17, 28, 31–38] and signal-known-statistically (SKS) [38–40] which incorporate background and signal variability, respectively. For the scope of this work, we focus on SKE/BKS tasks.

In recent publications, several authors proposed different approaches for manipulating the 3D image data during the process of signal detection. The most direct way to migrate the model observer for 2D detection task to the 3D detection task is to use a conventional 2D (planar) CHO and apply it on a single image slice only, the slice where the signal is centered. We refer to this approach as *single-slice* CHO (ssCHO). It has been used by Liang *et al.* [41], for example, to estimate observer performance in stack-mode reading of volumetric images. As the authors pointed out, the limitation of the ssCHO is that model observers which are designed for use in pure 2D detection tasks do not incorporate information about signal contrast in the z-direction nor the spatial correlation of the background and signal in the adjacent slices.

A similar motivation underlies the analysis of Kim *et al.* [30] who compared the behavior of 2D and 3D (volumetric) implementations of the numerical observers for simulated whole-body PET oncology imaging. Their results indicate that there is a significant increase in SNR or detectability of volumetric numerical observers relative to planar observers. Similarly, Lartizien *et al.* [35] used 3D implementations of model observers with 3D channels to compare different acquisition protocols in whole-body PET imaging, and found these to be a useful tool for their task of interest. We call a 3D implementation of the CHO a *volumetric* CHO (vCHO).

Chen *et al.* [34] proposed a more sophisticated two-layer model which combines 2D CHOs followed by an HO. The model which they called a multi-slice CHO-HO was used to process simulated multi-slice multi-view images similar to SPECT myocardial perfusion scans. First, the image slices of each of the three orthogonal views (coronal, sagittal and axial) were channelized and the 2D CHO was computed for each slice and each view, giving arrays of the decision variables. Then, an HO was applied on these decision variable arrays to obtain a single scalar detection score for the 3D image, known in statistical hypothesis testing as the test statistic. This approach was guided by the assumption that, for multi-slice images, human observers make their detection decision in a two-stage process. The first stage assessing each slice separately and the second stage, integrating these slice assessments to yield the final classification decision. Later, Gifford *et al.* [38] tested two different processes for modeling the observer capacity for integrating the information from multiple slices in the image stack. One process describes an observer that is able to integrate the slice information by computing the sum of the decision variables for each slice. There, this sum represents the final test statistic for the image stack. The other process supposes that the observer is unable

to do any integration, and instead the image test statistic is assigned to the maximum of the decision variables across slices. Further on, we use the term *multi-slice* CHO (msCHO) to refer to any approach which treats the 3D image as a conglomerate of multiple slices rather than just a single volume.

Most recently, Young *et al.* [36] used 2D projections of 3D breast tomosynthesis data to approximate the performance of ideal linear observer. Unlike the conventional CHO that would use a single 2D projection only, they built a CHO model that uses concatenated channelized angular projections. By doing so, Young *et al.* were able to incorporate correlations between multi-projections. Again, their preliminary results indicate that the observer using multiple projections outperforms the single-slice observer in their considered range of image acquisition parameters.

The purpose of this paper is to provide a well-founded overview of the properties of different model observer designs for 3D images which can then serve as a basis for building the anthropomorphic models. Our investigations assume a sequential approach to modeling human performance: (1) begin with the concept of the ideal observer; (2) compare performance predictions to human performance results on actual classification tasks, (3) modify the model to better predict human performance. Barrett and Myers [7] refer to this framework for building a human-like model as *modified-ideal-observer* approach. Thus, in the future investigations, during step 2 of the process, we will include human data in the analysis and compare those to the models, looking for the model which performs closest to humans. There, the comparative model analysis from the present work will assist us in better understanding of how incorporating different approaches to multi-slice image treatment can limit optimal detection performance, which will give intuition for explaining sub-optimal performance of

the humans. Finally, in step 3 of the process, these findings will be used to guide the design modifications of the selected model observer such that it can better predict detection performance of the human observers.

In particular, we consider a single-slice CHO [15, 22], three designs of a multi-slice CHO, and a volumetric CHO model [30, 35]. The three multi-slice designs include the model proposed by Chen *et al.* [34] only restricted to a single view (either coronal, sagittal or axial), and two novel multi-slice CHO models introduced in this paper: one guided by the assumptions from the work of Chen *et al.* [34] and one inspired by the recent work of Young *et al.* [36]. To better understand the properties of the considered models in the practical cases where only limited clinical images are available, we conduct a series of multiple-reader multiple-case (MRMC) experiments (relying on the training and testing paradigm) and use different statistical measures to characterize the models. To account for potential influence of the background structure, we analyze all five CHO models within four different setups, all SKE/BKS: white Gaussian noise (WNB), correlated Gaussian noise (CNB), lumpy backgrounds (LB) [31] and clustered lumpy backgrounds (CLB) [41, 42]. Especially, for the two Gaussian data setups we also estimate the IO strategy. We remark though that the IO is mainly used as a point of reference in evaluating the range of disparity among the CHO models which we compare. In addition, to provide guidelines for future research, we explore and discuss the major considerations involved in selecting and using these models in specific applications or approaches.

Overall, our results show that the volumetric model outperforms the others in all four setups. The multi-slice observers are the next best, and the single-slice model expectedly achieves the lowest detection scores. At the same time, the disparity between the models

is most notable for high complexity detection tasks (e.g. detecting a Gaussian signal in a correlated Gaussian noise background when their parameters are very similar) and it gets less pronounced as the complexity of the task drops (e.g. detecting a Gaussian signal in a white noise background). The concept of task complexity is detailed in Section 5.A.

The paper is organized as follows. Section 2 describes the models of image objects we use in the study and provides the essential background information about the model observers. In Section 3 we describe the aforementioned five CHO model designs investigated in this work and in Section 4 we explain the setup of our experimental study. The results are presented and discussed in Section 5. Finally, Section 6 draws some conclusions from this work.

2. Mathematical background

We are interested in a binary classification task determined by two hypotheses: signal is absent (H_1) or signal is present (H_2). An observer decides which of these two is true for a given image.

Let an image under consideration be denoted by $\mathbf{g}$, a vector whose entries $\mathrm{g}_m, m = 1, ..., M$, are the intensity of image pixels in 2D data or image voxels in 3D data, and M is the number of elements (pixels or voxels) in the image. An observer is defined by its discriminant function which maps an image $\mathbf{g}$ to its test statistic, $t = t(\mathbf{g})$. The decision is made by comparing the test statistic to a certain threshold, t_0. When t is greater than t_0, the signal is considered detected, hence H_2 holds, and the image is classified as signal-present. Otherwise, H_1 is satisfied and the image is classified as signal-absent.

In the remaining of this section, we will introduce the image models considered in our study, briefly outline the fundamentals of the ideal observer and review the mathematical

framework for the linear observer models.

2.A. Object models

Since our work is aimed at investigating fundamental aspects of observer models for multi-slice images, we use three-dimensional images with known statistical properties and different levels of complexity. This provides a controllable test environment and allows for automated generation of a large number of random realizations which increases the statistical significance of the results.

Let us denote $\mathbf{s}$ the signal to be detected, $\mathbf{b}$ the noiseless image background and $\mathbf{n}$ the measurement of noise in the image. Then the data under the two hypotheses are given by

$$H_1 : \mathbf{g} = \mathbf{b} + \mathbf{n}, \tag{1}$$

$$H_2 : \mathbf{g} = \mathbf{b} + \mathbf{n} + \mathbf{s}. \tag{2}$$

In our case, four different models are considered for $\mathbf{b}$ while the model of $\mathbf{s}$ is kept the same for all four background models. The noise is white Gaussian noise with mean 0 and standard deviation 1, $n \sim N(0, 1)$. The amount of noise is small so that it does not disturb the statistical properties of the background. The models we use for background and signal simulations are described in the remaining of this subsection, and their parameters are summarized in Table 1.

2.A.1. Image backgrounds

The criteria for choosing the background models are twofold: on one side, given that the purpose of the model observers is assessment of medical images, we aim at image data models which may be of clinical relevance, and on the other side, as a point of reference for

comparing the CHO models, we are interested in estimating the ideal observer strategy for the selected data. In most cases, these two criteria exclude each other: the IO performance is very difficult to estimate for clinical images, because their statistics are understandably complex and often unknown.

In order to keep the analysis general, we select four different categories of background images to be used in the study: white Gaussian noise (WNB), correlated Gaussian noise, or colored noise backgrounds (CNB), lumpy backgrounds (LB) [31] and clustered lumpy backgrounds (CLB) [41, 42]. Example background images with added noise are shown in Fig. 1. These correspond to signal-absent images in the study.

The first two models are Gaussian so that the IO strategy is readily calculable for these problems [7]. We will use these IO calculations to evaluate the non-ideal model observers (variants of CHO) against the theoretical upper bounds of the performance, to be explained in Section 2.B. In contrast, the LB and CLB models are used as representatives of non-Gaussian data. The two-dimensional CLB have been shown by Bochud *et al.* [42] to have a close visual appearance to real mammographic backgrounds. Recently, Castella *et al.* [43] used a genetic algorithm to optimize the CLB generation and achieve even more realistic mammographic texture synthesis. At the same time, both the LB and CLB models are statistically well described which allows automated generation of large ensembles of images required for the observer experiments. Due to their complexity, the IO strategy for LB and CLB are not included in the present analysis. We remark here that for the non-Gaussian LB data Kupinski *et al.* [14] and Park *et al.* [13] have been able to estimate the IO and the CIO, respectively, using Markov-chain Monte Carlo (MCMC) techniques.

The simplest background considered in the paper is the fixed background, $\mathbf{b}_{WNB} = 0$.

Since the statistics of these backgrounds are determined by the added measurement white noise, $\mathbf{n}$, (see Eq. (1,2)) we refer to them as white noise backgrounds, WNB. Next in order of background complexity is CNB data, generated by convolving white noise following the distribution $N(0,1)$ with a 3D Gaussian kernel characterized by σ_b. The correlated Gaussian random backgrounds are sometimes also referred to as *lumpy backgrounds*, not to be confused with the LB as we use them in this study, which are non-Gaussian. We describe these next.

As defined by Rolland and Barrett [31], a lumpy background, $\mathbf{b}_{LB}$, is produced by placing a random number of lumps, $l(\mathbf{r})$, at random locations in the image, $\mathbf{r}_k$. In our simulation, $\mathbf{b}_{LB}$ is extracted from a larger field of view (FOV), in order to avoid a boundary problem in generating the LB images. In particular, the size of $\mathbf{f}_{LB}$ is $M_{FOV} = 128^3$ voxels and the size of $\mathbf{b}_{LB}$ is $M = 64^3$ voxels (see also Table 1). Formally, the LB images can be described as

$$\mathbf{f}_{LB}(\mathbf{r}) = \sum_{k=1}^{K} l(\mathbf{r} - \mathbf{r}_k), \tag{3}$$

where $\mathbf{r}$ is a 3D vector and K is the number of lumps selected using a Poisson probability distribution with mean $\overline{K}$. For the LB images, the values of lump locations, $\mathbf{r}_k$, are selected using a uniform probability distribution over the support of the FOV, $\mathbf{f}_{LB}$. The set of K lump locations may be referred to as a "lump map" of the image. We choose the lumps to be 3D Gaussian signals of magnitude a_b and with the spread parameter σ_b,

$$l(\mathbf{r}) = a_b \exp\left(\frac{-|\mathbf{r}|^2}{2{\sigma_b}^2}\right). \tag{4}$$

Finally, the most complex background we treat in this paper is the CLB. The original concept of the two-dimensional CLB was introduced by Bochud *et al.* [42]. In [41], the 2D concept is extended to 3D with the assumption that the projection of a 3D CLB yields a

2D CLB with related characteristics in terms of the parameters of cluster and lump size and density.

As with the LB, to prevent potential boundary effects, the clustered lumpy background $\mathbf{b}_{CLB}$ of size $M = 64^3$ voxels is extracted from a larger FOV, $\mathbf{f}_{CLB}$ of size $M_{FOV} = 128^3$ voxels. The $\mathbf{f}_{CLB}$ is created in a two step process. The first step is similar to the process with the LB, only now we shall refer to the "lump map" as the "cluster map" and use $\mathbf{r}_k$ to denote cluster, rather than lump, location. In the next step, each cluster position, $\mathbf{r}_k$, is used as the spatial origin for placing a random number, N_k, of lumps. These N_k lumps are randomly positioned in the kth cluster at locations $\mathbf{r}_{kn}$. Thus,

$$\mathbf{f}_{CLB}(\mathbf{r}) = \sum_{k=1}^{K} \sum_{n=1}^{N_k} l(\mathbf{r} - \mathbf{r}_k - \mathbf{r}_{kn}), \tag{5}$$

where K stands for the number of clusters in the field of view $\mathbf{f}_{CLB}$. Again similar to the LB, both K and N_k are selected using a Poisson probability distribution with mean values $\overline{K}$ and $\overline{N_k}$, respectively. The location of the kth cluster, $\mathbf{r}_k$, is selected using a uniform probability distribution over the support of the $\mathbf{f}_{CLB}$. To create CLB images, anisotropic 3D exponential blobs are used with characteristic lengths L_x, L_y and L_z in x, y and z directions, respectively. The details can be found in [41].

2.A.2. Spherically symmetric signal (SKE)

Signal-present images are created by adding the signal $\mathbf{s}$ to a background image $\mathbf{b}$. In particular, we use a spherically symmetric Gaussian blob created in 3D Cartesian space and centered in the image volume. Similar to the lump in LB backgrounds, the signal is defined by Eq. (4) only now we use a_s to represent signal magnitude and σ_s to denote signal spread parameter. The central slice from a sample signal volume is depicted in Fig. 2(a) and the

radial profile of the signal used in the study is given in Fig. 2(b).

Parameters of both the backgrounds and the signals used in the paper are listed in Table 1.

2.B. Observer models

According to the signal detection theory [21], the observer is completely characterized by its discriminant function which assigns a scalar test statistic to each image object, $t = t(\mathbf{g})$. In the following, we introduce the ideal and the channelized mathematical model observers, and define their discriminant functions.

2.B.1. The ideal observer

The Bayesian ideal observer (IO) is defined as one that has full knowledge of the problem in terms of the conditional probability density functions of image data, $\mathbf{g}$, under each hypothesis, $\mathrm{pr}(\mathbf{g}|H_i)$, $i = \{1, 2\}$. Hence, the test statistic of the ideal observer is defined as the likelihood ratio [21],

$$\Lambda(\mathbf{g}) = \frac{\mathrm{pr}(\mathbf{g}|H_2)}{\mathrm{pr}(\mathbf{g}|H_1)}. \tag{6}$$

Clearly, calculation of the likelihood ratio, or more conveniently the log-likelihood ratio $\lambda(\mathbf{g}) = \ln \Lambda(\mathbf{g})$, requires knowledge of the probability density functions that make up the Eq. (6). In practical applications, these are often complicated or even unknown. Here, though, the IO can be derived for the WNB and CNB image models. The analytical expressions for calculating the SNR and AUC of the IO for these two image categories are given later in this section.

On the other hand, for greater complexity of the image data statistics, even in cases of simulated data such as LB or CLB from our study, and especially in cases of real clinical

data, analytical formulas for calculating the theoretical upper bounds for the observer performance cannot be derived. Rather, computation of the likelihood ratio in those cases requires specialized procedures to be developed. As we mention earlier, in current literature this has been done for LB and CLB backgrounds using MCMC techniques [13, 14, 25].

2.B.2. Channelized observers

In case of real clinical images, it is often complicated or impossible to know the probabilities required to calculate λ. Primarily, this is caused by random variations in both anatomical background (bones, veins, organs) and the signal (size, shape and location of the lesion) which are not all well understood to date and thus accurate models of those are not yet available. To circumvent this problem, a linear approximation of the ideal observer has been defined, where linearity refers to the discriminant function

$$t(\mathbf{g}) = \sum_{m=1}^{M} w_m g_m, \tag{7}$$

where M is the number of elements in the image $\mathbf{g}$. The weights, w_m, $m = \{1, ..., M\}$, form an image $\mathbf{w}$ called the *template* of the observer. Thus, the discriminant function may be written as a scalar product

$$t(\mathbf{g}) = \mathbf{w}^t \mathbf{g}. \tag{8}$$

Within the framework of linear discriminant analysis, the optimal linear discriminant is defined as the one which maximizes the SNR. In this context, the ideal *linear* observer is known as the Hotelling observer (HO) [7]. The template of the HO is defined as

$$\mathbf{w}_{HO} = {\mathbf{K}_\mathbf{g}}^{-1} \Delta\overline{\mathbf{g}}, \tag{9}$$

where $\Delta\overline{\mathbf{g}} = \langle \mathbf{g}|H_2\rangle - \langle \mathbf{g}|H_1\rangle$ and $\langle\cdot\rangle$ denotes ensemble average. The average of the ensemble covariance matrices of the signal-absent and signal-present data is denoted $\mathbf{K_g}$. It is defined as follows:

$$\mathbf{K_g} = \frac{1}{2}(\mathbf{K_{g,1}} + \mathbf{K_{g,2}}), \tag{10}$$

with $\mathbf{K_{g},}_i = \langle(\mathbf{g} - \overline{\mathbf{g}_i})(\mathbf{g} - \overline{\mathbf{g}_i})^t|H_i\rangle$, $i = \{1, 2\}$, and $\overline{\mathbf{g}_i} = \langle \mathbf{g}|H_i\rangle$. When the images are Gaussian random vectors, the HO equals the IO. However, to their disadvantage, both the IO and HO encounter the difficulty of high-dimensionality computations [23]. The main difficulty in computing the HO stems from the inversion of a large covariance matrix, $\mathbf{K_g}$, which is used in Eq. (9) to estimate the observer template, $\mathbf{w}_{HO}$.

The template is often estimated from the data for which the ground truth is known a priori (trainer data). We refer to this as the *training phase*. Next, in the *testing phase*, the estimated observer template is used to classify the data for which the ground truth is unknown (tester data).

To overcome the dimensionality problem of the HO model, another variant of the linear observer named the *channelized* Hotelling observer was defined [22]. The CHO may be seen as a specialization of the HO model which makes use of the frequency selective channels to model the HVS while reducing the dimensionality of the problem. The channels can be seen as M-dimensional images, $\mathbf{u}_p$, $p = \{1, ..., P\}$, where P is the number of channels. In contrast to the HO where all image data is used to build the template $\mathbf{w}_{HO}$, the CHO model only makes use of the channel outputs,

$$\mathbf{v} = \mathbf{U}^t\mathbf{g}, \tag{11}$$

where $\mathbf{U}$ denotes the channel matrix, $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_P]$. We note that processing the

images through channels has greatly reduced the dimensionality of the problem, $P \ll M$. To illustrate this reduction, we refer to the parameter values used in the study: the size of the image is $M = 64 \times 64 \times 64$ voxels while the number of channels P is of the order of 10. Note that real clinical images are usually larger while the number of channels usually remains of the order of 10.

If we denote the ensemble covariance matrix of the channelized data as $\mathbf{K_V}$, the template of a CHO model is

$$\mathbf{w}_{CHO} = {\mathbf{K_V}}^{-1} \Delta\overline{\mathbf{v}}, \tag{12}$$

where $\mathbf{K_V} = \mathbf{U}^t \mathbf{K_g} \mathbf{U}$ and $\Delta\overline{\mathbf{v}} = \mathbf{U}^t \Delta\overline{\mathbf{g}}$. Finally, the test statistic is calculated as a linear combination of all channel responses, $t_{CHO}(\mathbf{v}) = {\mathbf{w}_{CHO}}^t \mathbf{v}$.

In selecting the channels for the present study, we refer to the primary objective of comparing optimal performance of the observer models rather than their ability to mimic humans. Additionally, the following assumptions apply to the study: there is no preferred orientation in the correlation structure of the background and the signal is spherically symmetric in a known location. Accordingly, we chose LG channels centered on the location of the signal. The details about the specific use of the channels in different CHO designs are given in the next section.

The LG functions are a product of Laguerre polynomials and Gaussian functions, and defined by

$$u_p(r) = \frac{\sqrt{2}}{a_u} \exp\left(\frac{-\pi r^2}{{a_u}^2}\right) \; L_p\left(\frac{2\pi r^2}{{a_u}^2}\right), \tag{13}$$

where $r \in \Re^2$, a_u is the spread parameter of the LG channel, and L_p denotes Laguerre

polynomials defined by

$$L_p(x) = \sum_{k=0}^{p} (-1)^p \binom{p}{k} \frac{x^k}{k!}. \tag{14}$$

The weight of the polynomials is concentrated within a Gaussian envelope with spread σ_u, where $a_u^2 = 2\pi\sigma_u^2$.

The procedure used for selecting the LG channel parameters is described in Section 4.B and the corresponding results are summarized in Table 2. In Fig. 3 we show the first 5 LG channels used in the study for CNB and CLB images when $\sigma_s = 8$.

2.C. Performance measures

In objective image quality assessment [20], AUC and SNR are often used to quantify the performance of the model observers. In the first step, we use the test statistics, t, for signal-present and signal-absent images and apply the Wilcoxon-Mann-Whitney statistic to estimate the AUC of a CHO model. Then, AUC is used to calculate the SNR. The relationship between SNR and AUC can be expressed as

$$\mathrm{SNR}_{AUC} = 2\ \mathrm{erf}^{-1}(2\mathrm{AUC} - 1), \tag{15}$$

where erf$(\cdot)$ represents the error function. In the literature [7], the SNR_{AUC} is also called the detectability index, d_A, and it is commonly used for performance comparison in the domain of observer studies.

For two out of four image categories considered in the study, we also estimate the IO strategy and use it as a reference to evaluate performances of the CHO models. In this study, the IO performance will be calculated for WNB and CNB images while for LB and CLB these calculations shall not be included.

In the case of the IO, we first calculate the SNR and then use it in Eq. (15) to get the AUC. When the signal is assumed exactly known, the SNR of the IO equals

$$\mathrm{SNR}_\lambda = (\mathbf{s}^t \mathbf{K}^{-1} \mathbf{s})^{1/2}. \tag{16}$$

Here, $\mathbf{K}$ stands for the covariance matrix of the background: $K_{WNB} = {\sigma_{WNB}}^2 I$ in case of WNB, or $K_{CNB}(r_i, r_j) = {a_b}^2 (\pi {\sigma_b}^2)^{3/2} \exp\left(-(r_j - r_i)^2/(4{\sigma_b}^2)\right)$ in case of 3D CNB with the Gaussian kernel determined by Eq. (4).

Observer performance experiments are often limited in size, especially when real data is used. In these cases, it is important to determine the errors in the estimated AUC or SNR. The source of the errors is twofold: variation in test case difficulty (case variability) and variation in estimating the reader performance (reader variability) [44]. In the terminology of linear model observers, a reader is determined by the template of the model, and we estimate the template. So for our study, we generate multiple estimates of the template and assess them within a fully-crossed multiple-reader multiple-case (MRMC) study design which assumes that every reader reads every case. One non-parametric estimate of the variance of AUC in such MRMC study design is the one-shot method defined by Gallas [45]. The one-shot algorithm gives the estimate of AUC averaged over the readers, $\widehat{AUC}$, together with the estimate of its variance, $\widehat{V}_{\widehat{AUC}}$. We use these two metrics to calculate the error bars when measuring the performance of the CHO models.

In the course of comparing the CHO models, we also make use of the metric named statistical efficiency. Commonly, the relative efficiency η of the current observer characterized by SNR_{curr} relative to the reference observer characterized by SNR_{ref} is defined as follows

$$\eta = \frac{{\mathrm{SNR}_{curr}}^2}{{\mathrm{SNR}_{ref}}^2}. \tag{17}$$

The efficiency metric is used within the study to investigate several different parameters of the model observer designs (see Section 4.C).

3. Methods

We compare three different designs of the channelized Hotelling observer model: (1) a single-slice model (ssCHO), (2) three variants of a multi-slice model (msCHO), one existing and two novel ones, and (3) a volumetric model (vCHO). The models are defined in this section along with the corresponding notation.

To avoid confusion, in this work we will use 2D-CHO and 3D-CHO, respectively, to denote the CHO for 2D and 3D images in general, without implying any specific model design.

3.A. Single-slice CHO (ssCHO)

Let N denote the number of slices in the image and Q the number of voxels in each slice so that the number of elements in the image $\mathbf{g}$ is $M = Q \times N$. In our study, there are $N = 64$ image slices and $Q = 64^2 = 4096$ voxels per slice, thus the number of voxels in the image is $M = Q \times N = 64^3 = 262144$. Lastly, the column vector of voxel intensities of the nth slice in the stack is denoted $\mathbf{g}_{(n)}$, $n = 1, ..., N$.

We use the name single-slice CHO to refer to the conventional 2D-CHO [15] when it is run on a single slice in the volume, the central slice of the signal; this is shown in Fig. 4(a). In our study, for example, the signal is centered in the central slice of the image, $\mathbf{g}_{(N/2)} = \mathbf{g}_{(32)}$. Thus, for the purpose of ssCHO, we limit our experiments to this particular slice in the image, $\mathbf{g}_{ssCHO} = \mathbf{g}_{(N/2)}$. Consequently, in view of Eq. (11), the size of both the channels, $\mathbf{u}_p, p = 1, ..., P$, and the concerned image data, $\mathbf{g}_{ssCHO}$, is equal to the size of image slice,

Q. The resulting vector of the channelized image data is equal to the vector of channelized data of the selected image slice, $\mathbf{v}_{ssCHO} = \mathbf{U}^t\mathbf{g}_{(N/2)}$. Once the channel responses are known, the template of the ssCHO model, $\mathbf{w}_{ssCHO}$, is estimated using Eq. (12).

Since the ssCHO design is only using limited image information to perform the detection task, it is expected and proven [30, 34, 47] to perform not as high as the model designs described next, and for the scope of this work it is used merely as a reference method.

3.B. Volumetric CHO (vCHO)

As known from the literature [7], the definition of the CHO model is not limited by the dimensionality of the problem as long as the related calculations are manageable. Therefore, a straightforward approach in solving a 3D detection task could remain in the scope of Eq. (11), just as it was in case of ssCHO.

In contrast to the ssCHO where the conventional CHO is applied only on a single slice in the image, $\mathbf{g}_{(N/2)}$ of size Q, the vCHO makes use of the complete image volume, $\mathbf{g} = [\mathbf{g}_{(1)} ... \mathbf{g}_{(N)}]$ of the size $M = Q \times N$ as we depict in Fig. 4(b). Similarly, instead of planar channels of the size Q used in ssCHO design, we now use volumetric channels $\mathbf{u}_p$, $p = 1, 2, ..., P$, of the size $M = Q \times N$. In this way, the vCHO becomes "aware" of the contrast and correlation between the adjacent image slices which was not the case with the ssCHO.

Specifically, given the fact that the signal in our study is smooth and spherically symmetric, we select to use 3D LG functions which are isotropic in all three dimensions (see Fig. 3). The 3D LG channels are created in 3D Cartesian space ($r \in \Re^3$), they are the same size as the image and they are centered on the location of the 3D signal.

The same as with the ssCHO, the channelized data $\mathbf{v}_{vCHO}$ is used in Eq. (12) to estimate

the template of the model $\mathbf{w}_{vCHO}$. It is important to note that the size of the vector of the channelized data $\mathbf{v}_{vCHO}$ is the same as those of $\mathbf{v}_{ssCHO}$, that is $1 \times P$.

3.C. Multi-slice CHO (msCHO)

Three different designs of multi-slice CHO are considered in this work: type a (msCHO_a), type b (msCHO_b) and type c (msCHO_c). To define these, the volumetric image $\mathbf{g}$ is referred to as an array of slices $\mathbf{g}_{(1)}, ..., \mathbf{g}_{(N)}$, where N is the number of slices in the image.

Unlike the ssCHO which exploits information of a single slice only, the multi-slice model design makes use of multiple slices in the image stack. Similar to ssCHO, yet unlike vCHO, the multi-slice observer makes use of 2D rather than 3D channels to filter the image prior to estimating the linear discriminant (see Fig. 5).

While the present work is not focused on modeling human observer performance, the design of msCHO model is partly inspired by the postulates about how humans actually view the volumetric image data sets while using the stack-mode presentation. For example, we may think of a radiologist who is inspecting a multi-slice CT image of the chest. We follow a simplifying assumption of Chen *et al.* [34] that humans interpret the multi-slice image in a two stage process. First, they pre-process the image in planar views (xy-plane), slice after slice, and buffer the scores obtained for each slice. Next, these scores are processed in the z-direction to result in the stack test statistic, t, which is used to make the classification decision: normal or abnormal case. Further on, we refer to these two phases as *pre-processing stage* and *integration stage*, respectively. The array of intermediate scores assigned to the slices in the pre-processing stage is hereafter called *planar scores*.

3.C.1. Pre-processing stage

For all three multi-slice designs, the slice data is first processed with a set of 2D-LG channels, as illustrated in Fig. 5. Here, each slice in the image, $\mathbf{g}_{(n)}$, $n = 1, ..., N$, is channelized by the set of P two-dimensional channels $\mathbf{u}_p$, $p = 1, ..., P$, to get the channel outputs $\mathbf{v}_{(n)} = [v_{1(n)}, ..., v_{P(n)}]$, where $v_{p(n)} = \mathbf{u}_p{}^t \mathbf{g}_{(n)}$. This resembles ssCHO design only now the channels are applied on each slice in the stack rather than on a single slice only. For simplicity, for each N slices in the stack, we use exactly the same set of 2D channels. This approach is practical but may be not the most efficient. In line with Eq. (11), the channelized data of the nth slice is $\mathbf{v}_{(n)} = \mathbf{U}^t \mathbf{g}_{(n)}$, where $n = 1, ..., N$ and $\mathbf{U}$ is the channel matrix. The matrix of the channel outputs for all slices in the image is denoted $\mathbf{v}_{msCHO} = [\mathbf{v}_{(1)}, \mathbf{v}_{(2)}, ..., \mathbf{v}_{(N)}]$.

The models differ in how they use the *channelized slice data*, $\mathbf{v}_{msCHO}$. In general, two approaches have been taken in handling the $\mathbf{v}_{msCHO}$; these are illustrated in Fig. 6. For one approach, applied for msCHO$_c$, $\mathbf{v}_{msCHO}$ is seen as output of the stage, or planar scores (see Fig. 6(c)). The other approach, applied for msCHO$_a$ and msCHO$_b$, extends the pre-processing stage to calculate a test statistic for each slice, $t_{(n)}$, $n = 1, ..., N$. In view of model design, this corresponds to a 2D-CHO which is run on each slice in the stack to build an array of test statistics for all slices denoted $\mathbf{t}_{planar} = [t_{(1)}, t_{(2)}, ..., t_{(N)}]$ (see Fig. 6(a),(b)). Here, $\mathbf{t}_{planar}$ is considered the vector of planar scores and it is used as input to the following stage, the integration stage. The details of the three variants of the model are discussed next.

3.C.2. msCHO, type **a** (msCHO$_a$)

This model design is illustrated in Fig. 6(a) and it corresponds to the work of Chen *et al.* [34] and Gifford *et al.* [38]. The channelized slice data obtained in the early pre-processing stage

is used to estimate 2D-CHO templates at each slice position in the stack. Let us denote the template matrix for slice position n as

$$\mathbf{w}_{(n)} = \mathbf{K}_{\mathbf{v}_{(n)}}{}^{-1} \Delta\overline{\mathbf{v}_{(n)}}, \quad n = 1, ..., N. \tag{18}$$

This type of msCHO model uses equally positioned slices of the trainer image stacks to estimate the 2D-CHO template which is to be applied on the tester slices at exactly the same position. For example, to build a template for the first slice in the tester stack ($n = 1$) we use only the first slices of the trainer images. As such, there is a total of N different templates, $\mathbf{w}_{(n)}$.

Next, the templates are used to calculate the test statistic for each slice in the planar view. The output data may be summarized in a vector of planar CHO metrics, $\mathbf{t}_{planar} = [t_{(1)}, t_{(2)}, ..., t_{(N)}]$, where

$$t_{(n)} = \mathbf{w}_{(n)}{}^{t} \, \mathbf{v}_{(n)}, \quad n = 1, ..., N. \tag{19}$$

In the final step, the integration phase, $\mathbf{t}_{planar}$ is used by the one-dimensional HO to calculate the final scalar statistic of the msCHO model; namely

$$t(\mathbf{t}_{planar}) = (\mathbf{K}_{planar}^{-1} \Delta\overline{\mathbf{t}_{planar}})^{t} \, \mathbf{t}_{planar} = \mathbf{w}_{HO_a}{}^{t} \, \mathbf{t}_{planar}. \tag{20}$$

It is important to remark that the HO template is also estimated using the trainer data, just as the 2D-CHO templates are. To do this, the 2D-CHO templates from Eq. (18) are applied to the trainer images in order to estimate vectors of trainer slice test statistics which are then used to estimate the HO template, $\mathbf{w}_{HO_a} = \mathbf{K}_{planar}^{-1} \Delta\overline{\mathbf{t}_{planar}}$.

3.C.3. msCHO, type **b** (msCHO$_b$)

In contrast to the msCHO$_a$ where a different template was used for each of the consecutive slices, we propose the first new model design using one 2D-CHO template over multiple adjacent slices in the image. The model is illustrated in Fig. 6(b). Assuming the 3D signal is centered on the image volume, the template is estimated using only the central slice, $\mathbf{g}_{(N/2)}$, from the trainer image stacks.

The number of consecutive slices to be processed with the same signal template depends on the inter- and intra-slice thickness as well as on the signal properties, especially the signal spread, and the background variability. Ideally, when the slice thickness is small and there are few or no missing slices between slices, the background variability is not too high and the signal characteristics are not changing significantly across slices, a single template could be applied on every slice in the stack, independent of the slice position within the stack [46]. In view of Eq. (19), we shall call this template $\mathbf{w}_{planar}$, $\mathbf{w}_{(n)} = \mathbf{w}_{planar}, n = 1, ..., N$. For simplification, in this study we assume that the aforementioned conditions are approximately satisfied and use a single template for each slice in the image stack. The exact number of slices used in our study is discussed at the end of the section.

If, on the other hand, the signal would be spread over fewer slices and there would be pronounced disturbances in its isotropy, together with the greater variability of the background, it might be not correct to apply the same template on all slices in the stack. Rather, a separate template should be estimated for each subset of "similar" adjacent slices of the testing stacks. Eventually, for the greatest variability of the data, a separate template should be estimated for each slice position in the stack, hence msCHO$_b$ would converge to msCHO$_a$.

The same as with type a of the msCHO model, in the integration phase of the model msCHO$_b$, the vector of slice test statistics is used by the HO with the template $\mathbf{w}_{HO_b}$ to infer the image test statistics, t.

3.C.4. msCHO, type **c** (msCHO$_c$)

Inspired by the work of Young *et al.* [36], we propose an alternative multi-slice CHO approach and the second novel CHO design in this study. Here, the channelized slice data $\mathbf{v}_{msCHO}$ are fed directly to a HO to integrate into a final observer score for the image, as depicted in Fig. 6(c). This approach is most similar to vCHO in that the correlation between slices are directly incorporated in the model though the channels and the exact algorithm are different from those of the vCHO. In the scenario of type c of the msCHO, the test statistic of the model is

$$t(\mathbf{v}_{msCHO}) = (\mathbf{K}_{msCHO}^{-1}\ \Delta\overline{\mathbf{v}_{msCHO}})^t\ \mathbf{v}_{msCHO} = \mathbf{w}_{HO_c}{}^t\ \mathbf{v}_{msCHO}. \tag{21}$$

We notice that the size of the covariance matrix $\mathbf{K}_{msCHO}$ is determined by the number of slices in the image, N, and by the number of channels, P. As mentioned earlier, P is usually of the order of 10 while N may well exceed this range. This suggests potential difficulties in estimating the template $\mathbf{w}_{HO_c}$ in Eq. (21) caused by the large dimensionality of the covariance matrix, similar as in [23], especially when the available trainer data set is limited in size. For example, when $N = 64$ and $P = 10$, the number of elements in $\mathbf{K}_{msCHO}$ is $(N \times P)^2 = 409600$.

3.C.5. Region of interest (ROI)

As we have defined them so far, the multi-slice CHO models can use all slices in the image stack. However, Wells *et al.* [47] found that, for the task of detection of small lesions in thoracic Ga-67 SPECT data, the benefit of a multi-slice display comes primarily from the two slices immediately adjacent to the central slice. The authors used a 1 cm diameter sphere to model the signal, where each voxel width was 0.317 cm. We shall henceforth refer to this subset of significant adjacent slices the region of interest (ROI), where the number of slices in the ROI is denoted R.

The preferred size of the ROI is influenced by the imaging technology (slice thickness and separation), as well as by the statistical properties of image data, including smoothness and symmetry of the signal, the range of its spread over slices, and the variability of the background content. All considered, the value of R shall be chosen to fit the properties of the given data. In case of the human observer study from [47], for example, it was shown that increasing the ROI (in their case $R > 3$) brings less significant improvement in the observer's performance.

In our experiments, each of the three msCHO designs illustrated in Fig. 6 are applied on the ROI of size R for which the channelized slice data is depicted in Fig. 5. The value of R is varied among the values of 3, 5 and 11 adjacent slices centered around the slice with the peak signal intensity.

4. Experimental setup

4.A. Sample images

For the experiment setup, the testbed of image ensembles is comprised of four categories: WNB, CNB, LB and CLB, as described in Section 2.A. Detailed parameters of all background images are summarized in Table 1. The total number of synthesized backgrounds is 22000 for each WNB and CNB categories, and 14000 for each LB and CLB categories, where the simulation time is significantly longer. We aim at a data set which is large enough to ensure statistical significance of the results while the computational time and computer power required for both image generation and observer calculations remain within reasonable limits. In each category, half of the set is used as signal-absent images and the remaining half is used to create signal-present images by inserting a 3D Gaussian signal in the center of the background volume (see Fig. 2). Given the parameters of the background images and aiming at non-trivial detection tasks, the spread of the 3D Gaussian signal is assigned $\sigma_s = 8$ throughout the study. In addition, for CNB data we also consider $\sigma_{s2} = 5$ and $\sigma_{s3} = 3$. For each image category and each considered σ_s, the peak intensity of the signal, a_s, is varied in the range of four different values, selected to approximately fit the criterion of AUC covering the range from 0.6 to 0.9 in equal steps. Due to different parameters of the backgrounds, a_s values differ across four categories, as specified in Table 1.

The image data is used as follows. For WNB and CNB categories, 10000 pairs (hereafter called *trainer pairs*) of signal-present and signal absent images are used as training data. For LB and CLB categories, the number of trainer pairs is 6000. In all categories, 1000 image pairs (hereafter called *tester pairs*) are used as test data. Tester data are kept independent

from the trainer data.

4.B. Study design

We test the performance of five CHO designs: ssCHO, msCHO$_a$, msCHO$_b$, msCHO$_c$, and vCHO, for four image categories: WNB, CNB, LB and CLB. Initially, we run a set of experiments to select the parameters of LG channels. Next, the performance and variance of the CHO models are evaluated in MRMC studies. For CNB data, we investigate the influence of signal parameters σ_s and a_s. This will allow the influence of the signal size to be assessed. In addition, for WNB and CNB images, the IO performance is estimated using Eq. (16) and Eq. (15). Finally, for the three multi-slice observers, we investigate the influence of ROI size on the model observer performance.

For all considered model observers, the observer templates are estimated using the trainer data. For a given CHO, all template parameters (the covariance matrix $\mathbf{K}$, the mean channelized signal $\Delta\overline{\mathbf{v}}$, the mean planar test statistics $\Delta\overline{\mathbf{t}_{planar}}$) are estimated using the exact same pairs of signal-absent and signal-present trainer 3D images. In the testing phase, the observer templates are used in estimating the test statistics for each of the tester 3D images. There is no overlap between the trainer and the tester image sets.

As defined in the previous section, 2D channels are required for both ssCHO and msCHO experiments while 3D channels are used by vCHO only. To that end, we explore two basic types of the CHO models: ssCHO to select parameters of the 2D channels and vCHO to select parameters of the 3D channels. Given that the sampled 3D LG channels as used in the study are not exactly orthonormal, we considered also the orthonormalized version of the 3D LG channels. In line with the work of Gallas and Barrett [15], each model is investigated

for several values of the channel spread parameter: for $\sigma_{s1} = 8$, $a_u = \{7, 12, 18, 24, 32\}$; for $\sigma_{s2} = 5$, $a_u = \{4, 7, 12, 15, 21\}$; and for $\sigma_{s3} = 3$, $a_u = \{3, 5, 7, 9, 12\}$. For each spread parameter, the number of LG channels is varied in the range of $P = 1, ..., 30$. The experiments are conducted with $N_{tr} = 2000$ trainer pairs and $N_{ts} = 1000$ tester pairs, and for the second largest among four considered values of signal magnitude a_s given in Table 1. Further in the study, these selected channel parameters are used. Within the same image category, a unique set of 2D LG channels is used for both the ssCHO and msCHO, while the 3D LG channels are used for the vCHO. The exact same set of 2D LG channels are used for all three types of the msCHO and for all slices in the image stack.

The MRMC studies are characterized by the following parameters: the number of trainer image pairs (N_{tr}), the number of tester image pairs (N_{ts}) and the number of readers (N_{rd}).The exact values of these parameters are given in Table 3. A range of different values of N_{tr}, while N_{rd} and N_{ts} are kept fixed, will allow the influence of the size of trainer data set to be evaluated. Each of the specified MRMC configurations is repeated for every signal spread value, σ_s, and related range of four signal magnitudes, a_s, all as specified in Table 1.

4.C. Figures of merit

The metrics used in the study include: AUC and the estimate of its variance, SNR, and model efficiency η; these are all defined in Section 3. For each MRMC configuration, we first estimate AUC and then use it in Eq. (15) to calculate the SNR. To evaluate the variability associated with the results, we use the one-shot variance analysis [45]. Eventually, in analyzing the influence of particular parameters of the CHO designs on their performances, we focus on CNB category of the data and use Eq. (17) to estimate efficiency, η, of the observers.

Three different types of the observer model efficiency are considered: efficiency of the CHO relative to the IO, η_{CHO}, efficiency of the CHO trained with fewer trainer pairs relative to its performance for the largest considered number of trainer pairs, $\eta_{N_{tr}}$, and efficiency of the ssCHO model relative to the vCHO model, $\eta_{ss,v}$. In view of Eq. (17), the actual SNR values used in place of SNR_{curr} and SNR_{ref} for each different type of the efficiency are specified in Table 4.

5. Results and discussion

To facilitate the interpretation of the results, we first refer to the IO performance in 2D versus 3D detection tasks and analyze the influence of image parameters and level of task complexity on the performance gap between 2D and 3D observer. We then proceed to elaborate on the selection of the channel parameters used in the study and continue to present a detailed comparative analysis of the five CHO models described in Section 3. Finally, we point to the major differences among these models and think about their potential applications in the future.

5.A. Complexity of the detection task: 2D versus 3D

Before we get into the analysis of the CHO model performances, it is worthwhile looking at the performance of the ideal observer for the 2D (2D-IO) versus the 3D problem (3D-IO). In Fig. 7, we show these results for the two image categories in the study for which the data is Gaussian: WNB (top graph) and CNB (bottom graph). As stated in Section 2.B.1, when the image data are Gaussian the ideal linear observer, the HO, equals the IO. We first calculate the SNR of the IO using Eq. (16) and then use this to calculate the AUC of the

IO by inverting Eq. (15).

Looking at Fig. 7, we notice that the 3D-IO outperforms the 2D-IO for both the WNB and CNB. Such results confirm our intuition about the gain in the observer performance from using the information from more than a single slice in the detection process. This is in line with the fact that 3D observer, unlike the 2D one, exploits also the information about signal contrast in z-direction and about spatial correlation structure between the slices of the data which yields more accurate estimates of the signal $\mathbf{s}$ and the covariance matrix $\mathbf{K}$ in Eq. (16).

Moreover, the difference between 2D-IO and 3D-IO performance is much more significant in the case of WNB compared to the CNB images. This may be explained by different levels of difficulty of the detection tasks in the two categories of image data. Namely, going from 2D to 3D adds more information on the signal which results in 3D-IO outperforming the 2D-IO. However, when there is correlation in the backgrounds, such as in CNB, 3D also adds more complexity to the background which makes the detection task more difficult and diminishes the positive impact of the additional signal information. Having together the benefit that comes from extra signal information and the detriment that comes from increased background complexity, the performance difference between 2D and 3D is narrower when the backgrounds have correlation, that is when the complexity of the detection task is higher.

Clearly, similar performance trends are expected to be observed among ssCHO and vCHO model designs, both in terms of the 2D versus 3D approach and uniform versus inhomogeneous image contents (backgrounds).

Last, we note that the difficulty of the detection task, either 2D or 3D, depends not only on the correlation of the background data but also on other parameters of image objects.

For example, in our study the signal is of Gaussian shape with the spread $\sigma_s = 8$ for both WNB and CNB while the spread of the CNB Gaussian kernel is determined by $\sigma_b = 8$. Given that the size and shape of the signal are the same as those of the filter kernel for the noise, the CNB detection task may be described as difficult. In contrast, the task is relatively easy for the WNB. When, for example, we would change the signal size to be smaller such as $\sigma_s = 5$ the difficulty of the WNB task would increase while the difficulty of the CNB task would decrease. Indeed, by looking at the parameters of CNB data in Table 1, we observe that the decrease in the signal spreads, $\sigma_{s1} > \sigma_{s2} > \sigma_{s3}$, is followed by the decrease in the signal amplitudes, $s_{s1} > s_{s2} > s_{s3}$, while the background structure is fixed, $\sigma_b = \sigma_{s1}$. This decreasing trend in the level of the signal, while preserving the value of the AUC, confirms the decrease in the difficulty of the detection task.

5.B. Exploring channel parameters

On the way to evaluate the CHO models, we first run a series of experiments for each of the four image categories aiming to select the parameters of 2D and 3D LG channels such that they capture as much information as possible for the purpose of signal detection. The results of this investigation for $\sigma_s = 8$ are depicted in Fig. 8. Here, the graphs in the left and right column depict results for ssCHO and vCHO, respectively, while the rows represent the image categories: WNB, CNB, LB and CLB, from top to bottom. Each curve in a graph corresponds to a different a_u. The solid lines labeled "ideal observer" show AUC performance of the IO calculated using Eq. (16) and Eq. (15): 2D-IO for 2D image data and 3D-IO for 3D image data. Further in the text, IO is used to refer to 3D-IO unless otherwise indicated. The number of trainer images used in these experiments is $N_{tr} = 2000$ which allows a meaningful

estimate of the involved data covariance matrices. For each image category, the selected number of 2D LG channels is denoted P_{2D} and the selected number of 3D LG channels is denoted P_{3D}.

We observe in all plots that the curves nicely converge to an asymptote as the number of channels increase. Looking in more detail, for both ssCHO and vCHO, the curves for narrower channels reach to the higher range of performance with fewer channels but then approach the asymptote more slowly. For wider channels, on the other hand, performance improves more gradually as the number of channels increase but does not have the long approach to the asymptote. Further on, we notice that in most categories ssCHO converges faster than vCHO reaching to the asymptote with a fewer channels hence there $P_{3D} > P_{2D}$. We found these results independent of the orthonormality of the 3D LG channels. In line with the task complexity discussion at the beginning of this section, the distance between the CHO asymptote and the IO score is more pronounced in the case of ssCHO compared to vCHO where the linear model nearly approaches the IO.

Aiming at the best and stable performance of the CHO with a reasonable number of channels and given the plots in Fig. 8, we select the channel parameters which are used further in the study. For all four image categories and related signal size, the selected parameters of the LG channels are listed in Table 2, these particular values of a_u, P_{2D} and P_{3D} are used in the remaining of the study. With respect to the category, the narrowest and fewest channels are used in case of WNB while wider and more of those are used for other image categories. Again, the tendency conforms with the difficulty of the detection tasks. Thus, for example, $a_u = 12$, $P_{2D} = 3$, $P_{3D} = 4$ for WNB while for more complex CNB these values increase to $a_u = 24$, $P_{2D} = 9$, $P_{3D} = 12$.

5.C. Comparing CHO performances

The performance results for the five CHO model designs are summarized in Fig. 9 for all four categories of the backgrounds and their image parameters as defined in Table 1. The signal size is the same in all images, $\sigma_s = 8$. The results correspond to the study design of $N_{tr} = 2000$ trainer pairs and $N_{ts} = 1000$ tester pairs, and $N_{Rd} = 5$ readers (templates) for WNB and CNB backgrounds or $N_{rd} = 3$ readers for LB and CLB data, all in accordance with Table 3. The details about the size of image data sets and MRMC study configurations can be found in Section 4. For the msCHO models the size of ROI is $R = 11$ with approximately 65% of the signal energy included in the decision process. Here, the energy of the signal is calculated as $E(\mathbf{s}) = \sum_{m=1}^{M} s_m^2$, $m = 1, ..., M$ where $\mathbf{s}$ is defined by Eq. (4) and M is the number of voxels in the signal image. The size of ROI is selected such that the covariance matrix of msCHO$_c$ is not unbearable, all other observers are less demanding. In each study experiment, the AUC is averaged over the total number of readers. The error bars are ± 2 standard deviations estimated by the one-shot method [45]. For the purpose of this analysis, and in view of the remarks from Section 4.A concerning the selection of the signal magnitudes, we shall avoid directly comparing the absolute values of AUC for different categories. Instead, we look at the absolute AUC values only within the same category, and use different categories to compare relative trends amongst the CHO variants with regards to image data properties.

In all four data categories, vCHO clearly outperforms the other models. Among multi-slice designs which are ranked next, msCHO$_c$ which infers the classification decision directly from the channelized slice data, $\mathbf{v}_{msCHO}$, outperforms the other two which use $\mathbf{v}_{msCHO}$ to build the slice test statistics prior to estimating the final image statistic. On the lower side,

expectedly, is the ssCHO design. For all five models, the error bars slightly decrease as the magnitude of the signal grows.

Across four image categories, the most striking difference between the model performances is observed for WNB images where ssCHO performs significantly worse than the other four models. As explained earlier, the reason for this remarkable benefit of using information from multiple slices in the process of signal detection stems from the low complexity of the detection task. Even more, given the uniform structure of the white noise background and the relatively "large" spread of the signal used in our study ($\sigma_s = 8$), the detection task gets relatively "easy" as the observer gets access to all three-dimensions of the image. The least amount of disagreement between the model performances is observed for CLB images which use the most complex backgrounds in the study.

In further analysis and discussion, we focus on CNB data and explore the influence of specific parameters: signal size, signal magnitude and size of trainer data set.

The results of MRMC studies for CNB images when the signal size is $\sigma_{s2} = 5$ and $\sigma_{s3} = 3$ are presented, respectively, in the top and bottom graphs of Fig. 10. For msCHO, the size of ROI is the same as in Fig. 9, $R = 11$. The approximate percent of signal energy included in the decision process is now 88% for σ_{s2} and 99% for σ_{s3}. Overall, in Fig. 10 we observe similar tendencies in CHO model performances as those in Fig. 9. Only now the absolute difference between performances of the different models is more pronounced.

Let us first look at the ssCHO versus vCHO. We refer to the results for CNB with σ_{s1} from Fig. 9 and those for CNB with σ_{s2} and σ_{s3} from Fig. 10. For example, let us examine the experiment setups when the AUC of vCHO is in the range of 0.9 (the second largest a_s for a given σ_s). By comparing these, we observe that the absolute differences in performance

of vCHO and ssCHO for a given σ_{si}, denoted $\mathrm{D}|\sigma_{si}$, $i = 1, 2, 3$, are ordered as follows: $(\mathrm{D}|\sigma_{s1} \approx 0.07) < (\mathrm{D}|\sigma_{s2} \approx 0.2) < (\mathrm{D}|\sigma_{s3} \approx 0.22)$. Earlier in this section, we established that the complexity of the detection task in three CNB image setups, each with the kernel size of $\sigma_b = 8$, is highest for $\sigma_{s1} = 8$, lower for $\sigma_{s2} = 5$ and lowest for $\sigma_{s3} = 3$.Here again, the ordering of performance differences nicely agrees with the earlier discussion that the benefit of vCHO over ssCHO is most significant when the task complexity is low ($\mathrm{D}|\sigma_{s3} \approx 0.22$), and it gets smaller for higher complexity tasks ($\mathrm{D}|\sigma_{s1} \approx 0.07$). Similar trends appear with respect to the difference between the ssCHO and msCHO. There also, the difference in performance is largest when the task is of lowest complexity (σ_{s3}).

Another interesting aspect to these results is the influence of ROI size, the number of slices used with the msCHO models. Even for σ_{s3} when 99% of the signal energy is included in the ROI, there is a difference between the msCHO models and vCHO. This may indicate that the msCHO still has insufficient information on background statistics. The extent of the vCHO is not limited to the ROI size. It is possible that the msCHO performance can be increased by choosing more slices but still fewer than the whole volume. On the other hand, especially with msCHO_a and msCHO_b, involving more slices that have little or no signal in them might only add unnecessary noise. Of course, the specific choice of ROI size should represent the best compromise between the aforementioned considerations. Moreover, it would depend on the type of image data and its background statistics, and for msCHO_c, on the number of training images available to adequately estimate the covariance matrix. Eventually, we remind that in our experiments the same 2D LG channels are used for each slice of a given image stack which may not be optimal. The influence of the ROI size will be discussed later in the section, yet detailed analysis in this respect requires future research.

We continue with comparing CHO performances to the IO over a range of CNB image parameters σ_s and a_s. By doing this, we aim at evaluating the range of disparity among different CHO models. It is not the explicit focus of this study to select a CHO model which approximates the IO; ; the IO model is used as a point of reference. The statistical efficiency of the five CHO model observers relative to the IO, η_{CHO}, is calculated using Eq. (17) with the corresponding SNR values from Table 4. The results are summarized in Table 5.

In general, the definition of SNR from Eq. (16) suggests a linear increase of SNR with the increasing signal magnitude, a_s. Thus, for each σ_{s1}, σ_{s2} and σ_{s3}, we expect the efficiency η_{CHO} as defined in Eq. (17), to be constant with respect to a_s. Indeed, for CNB setups with $\sigma_{s2} = 5$ or $\sigma_{s3} = 3$, and given the results in Table 5, the efficiencies of the CHO models relative to the IO are approximately constant with the considered values of a_s. However, the efficiencies observed with $\sigma_{s1} = 8$ do not meet the expectations. Even, with very low $a_s = 0.25$, it happens that $\eta_{vCHO} > 100\%$ which, in theory, is not really possible. Such unstable behavior of the efficiency η_{vCHO} in the case of σ_{s1} could be attributed to the effect of training the CHO models (see later discussion of $\eta_{N_{tr}|a_s}$ and Table 6).

In comparing the η_{CHO} across three values of σ_s, we notice that the benefit of vCHO over the other models is more significant for smaller σ_s that is for lower complexity of the signal-to-background parameters, or lower complexity of the detection tasks. This is confirmed with the calculations of the efficiency of ssCHO relative to the vCHO, $\eta_{ss,v}$ aimed to illustrate the difference in observer efficiency caused by the restricted amount of information used by the ssCHO compared to the vCHO model design. These are also included in Table 5. The value of $\eta_{ss,v}$ varies significantly from approximately 60% for $\sigma_{s1} = 8$ to approximately 17% for $\sigma_{s2} = 5$ or 14% for $\sigma_{s3} = 3$.

Knowing that the limited size of image ensembles is often encountered with sets of real clinical data, we aim to evaluate the influence of the number of trainer pairs, N_{tr}, on the efficiency of the CHO models. To do that, we calculate CHO efficiencies relative to the scores obtained with the largest considered trainer data set. Eventually, the greater the efficiency of the model for a smaller value of N_{tr}, the less the CHO depends on the number of available trainer images and the better it suits the experiments with a limited number of images. Specifically, we use $\eta_{N_{tr}}$ to investigate the influence of the size of trainer data set in a twofold manner: with reference to the signal spread parameter, $\eta_{N_{tr}|a_s}$ - for all ssCHO, msCHO and vCHO models; and with reference to the size of ROI, $\eta_{N_{tr}|R}$ - for the three msCHO models.

The results of $\eta_{N_{tr}|a_s}$ calculations for all five CHO models and different levels of the signal a_s are given in Table 6. Here, all experiment parameters correspond to the results in Fig. 9. The rows labeled "$\text{SNR}_{N_{tr}=5000}$" give the SNR values for $N_{tr} = 5000$. These are included to indicate the absolute range of the observer performance for different signal levels, a_s. We notice that the efficiency $\eta_{N_{tr}|a_s}$ for ssCHO and vCHO are greater than those of the msCHO models. This difference is more noticeable for the lower levels of the signal ($a_s = 0.25$) and it gets less significant for higher signal levels ($a_s = 1$). Also, for each observer model, the values of $\eta_{N_{tr}|a_s}$ significantly increase with the increase of N_{tr} for lower signal levels, and this variability is greatly reduced for higher signal levels. Hence, the influence of the size of trainer data set is less significant when the observer performance is higher. Given the parameter values in our study, the CHO models are most sensitive to the size of trainer data set when $a_s = 0.25$ where $SNR_{N_{tr}=5000}$ is below 1, and they are least sensitive to the value of N_{tr} when $a_s = 1$ where $SNR_{N_{tr}=5000}$ is in the range of 2 or greater. This is in line with

the conclusions from Fukunaga and Hayes [48] who discussed the effect of finite sample size on training a classifier showing that the bias is a function of the performance level.

For the msCHO models and CNB images with $\sigma_{s1} = 8$ and $a_s = 0.75$, we vary the size of ROI among 3, 5 and 11 adjacent slices and for each of them we calculate $\eta_{N_{tr}|R}$. Additionally, msCHO$_a$ and msCHO$_b$ models are applied on all slices in the image, $R = 64$. These results are presented in Table 7 where columns denote the size of ROI and rows indicate the number of trainers. In case of msCHO$_c$, the covariance matrix of channelized slice data, $\mathbf{K}_{msCHO}$ in Eq. (21) is of the greatest dimension, $(R \times P)^2$ compared to R^2 of the other two models. When the number of slices in ROI increase to $R = 64$ and given that $P_{2D} = 9$, the size of our data set ($N_{trMAX} = 5000$) is insufficient to properly estimate $\mathbf{K}_{msCHO}$. Thus, for msCHO$_c$ the analysis is restricted to the lower three values of R.

In Table 7, we first observe the row labeled "SNR$_{N_{tr}=5000}$" where the SNR values for $N_{tr} = 5000$ are presented. For $R = 3$ and $R = 5$, SNR$_{N_{tr}=5000} = 1.49$ for either msCHO$_a$ or msCHO$_b$. Given these scores, and reading from Table 6 that for ssCHO SNR$_{N_{tr}=5000} = 1.48$ in the same test environment ($a_s = 0.75$), we conclude that for type a and type b of multi-slice observers the benefit from first 3 or 5 adjacent slices is minor. With msCHO$_c$, the contribution of the first few slices around the signal is slightly greater yet notably less compared to those of $R = 11$. All in all, from the results presented in Table 7, it is clear that the major contribution in msCHO performance comes from the next few slices, mainly from the ROI of 11 consecutive slices centered around the central slice of the 3D signal. Further growing the ROI might be considered to fine tune R for a given data. To that end, we note that msCHO$_c$ is able to reach SNR$_{N_{tr}=5000} = 1.75$ already with $R = 11$ while the other two models need all $R = 64$ slices to approach this level of the performance.

Eventually, we evaluate the overall influence of the number of trainer pairs on the model performances. As noted before, for all three msCHO designs the efficiency degrades as the size of ROI grows. However, this is more pronounced for fewer trainer pairs and it gradually disappears as N_{tr} grows. Looking back at Table 6 and together with Table 7, msCHO$_a$ and msCHO$_b$ are less sensitive to N_{tr} than msCHO$_c$. Even further, the msCHO$_b$ compared to msCHO$_a$ appears slightly more robust to the changes of ROI size especially when N_{tr} is in the lower range. To illustrate this, when $a_s = 0.75$ and $R = 11$, the msCHO$_b$ achieves $\eta_{N_{tr}} > 70\%$ with $N_{tr} = 50$ but then progresses to $\eta_{N_{tr}} > 90\%$ already with $N_{tr} = 200$. The msCHO$_a$ is a few percent lower while msCHO$_c$ is able to reach $\eta_{N_{tr}} > 90\%$ only with $N_{tr} = 1000$ trainer image pairs, which is in line with the earlier remarks about dimensionality restrictions of the latter model. The least affected by the limited number of trainer images are vCHO and ssCHO models, reaching over 80% of efficiency with as few as $N_{tr} = 50$.

5.D. Some practical considerations

In conclusion of this section, we think about potential applications of volumetric versus multi-slice versus single-slice observer designs in the actual 3D signal detection tasks.

Based on the results of our study, vCHO approaches the IO scores most closely. Therefore, it comes forward as a preferred model for optimization of the system to maximize detection of the 3D signal. In contrast to vCHO, ssCHO performs worst among all five CHO models in terms of actual performance metrics. Still, it follows the trends of the other models, and it is the simplest and fastest to apply. Consequently, it might be considered for preliminary experiments in 3D detection tasks, especially when the initial parameter space is large and shall be downsized prior to further in-depth analysis.

Another important aspect to consider when selecting the preferred CHO design is characteristics of the signal. Throughout this study, the signal is a spherically symmetric 3D Gaussian function. In practice, however, this would most often not be the case. Certainly, as the signal gets more asymmetric the choice of LG channels as we use them in the study might not be adequate and alternative channels shall be considered. Nonetheless, when this asymmetry is in the z-direction, perhaps even due to the increased slice thickness, it might be desirable to reconsider not only the channel selection but also the preferred model design for a given application. It may well be that the vCHO design which seems to be the most efficient design in the case of a spherically symmetric 3D signal compares differently to the msCHO designs when the signal characteristics are changed. Encouraged by the results from [46] where the detection of 2D signal in 3D backgrounds is investigated, we favor further investigating msCHO models when the signal is spread over a very limited number of slices only or its symmetry in z-direction is noticeably distorted.

Last but not least, given the possible applications of the model observers from this study, we are driven to think about the CHO model designs from the perspective of mimicking humans. While the anthropomorphic models as such are outside the scope of this work, we refer to some of their basic considerations to stimulate the discussion. As proposed by Myers and Barrett [22], the property of frequency selective channels which are known to exist in the HVS is used to model the process of signal detection in the two-dimensional environment. This mechanism certainly extends to three-dimensional problems. For video imaging applications, for example, it has been modeled with a three-dimensional filter bank which is separable in spatial and temporal frequency components [49]. However, current literature does not tell us how exactly the HVS is channelizing the data when viewing it in stack

browsing mode where the speed of browsing is not predetermined and the forward-backward looping is allowed. Conveniently, the stack-mode viewing scenario itself resembles the technique of msCHO signal detection. Henceforth, it might be worthwhile to further explore the msCHO model designs to better understand their relation to the human performance.

Even more challenging is the design of anthropomorphic models which operate on real clinical images. Inevitably, there is a number of factors to be considered here, ranging from the anatomical properties of the signal as well as of the background, through the parameters of the underlying imaging technology (inter- and intra-slice reconstructed thickness), the speed of browsing through the stack, and the limitations of the medium of image presentation such as the temporal effects in slow medical displays. Again, considerations about robustness of the model designs to the number of trainer images may play an important role in applications dealing with real clinical images where a limited number of samples are available. In addition, an important aspect of modeling human observers is the issue of channel selection. Undoubtedly, in-depth further investigations are necessary before the preferred design of the anthropomorphic 3D model can be proposed.

6. Conclusions

This work was set to investigate potential CHO model observer designs for the task of signal detection in a three-dimensional problem: three previously used in the literature (ssCHO, msCHO_a, vCHO) and two novel designs (msCHO_b, msCHO_c). In that sense, we have presented the theoretical background for the selected models and conducted an experimental comparative analysis of those for a range of statistically different images. Where applicable, the models were compared to the ideal observer known to set the theoretical boundary for

the signal detection performance.

When the signal is known and spherically symmetric, our results indicate that the CHO using volumetric channels is outperforming the other model designs. Even more, when the data statistics are Gaussian, the vCHO closely approaches the scores of the IO. Next ranked are the multi-slice observers, where the novel proposed msCHO$_c$ performed better than the msCHO$_a$ and msCHO$_b$, while the latter two were within the same range of performance. Expectedly, on the low end of the detection performance scale was the ssCHO. Importantly, the disparity between the models gets less pronounced as the complexity of the image content grows, *i.e.*, the difficulty of the detection task drops.

Further on, we found that the major benefit of multi-slice versus single-slice observer comes from a number of adjacent slices centered around the signal referred to as ROI, rather than all slices in the stack. This agrees with the conclusions from [47]. The exact size of the ROI is subject to the properties of a particular data set (slice thickness, signal spread, background statistics, etc.) and shall be determined on a case-by-case basis. Among msCHO designs, the new msCHO$_b$ seems to be least affected by the number of training samples, assuming the size of ROI is appropriately selected. Due to its design, in particular the relatively large size of the covariance matrix, the msCHO$_c$ model is most sensitive to the size of training ensemble and thus most susceptible to the dimensionality problem. For future work, one important aspect to study is the size of ROI in relation to the thickness of image slices but also in relation to the spatial spread of the signal over slices (signal parameters in the z-direction).

Finally, our work explored and discussed some basic aspects of the potential use of the different CHO designs considered. Most fundamentally, the high level of vCHO performance makes it a good candidate for what is called an efficient model, a model which can approx-

imate the ideal linear observer performance. Along the assumptions about human visual system which motivated the design of three msCHO models, this concept of CHO appears as a candidate for anthropomorphic model design. In order to test this candidacy, it is essential to acquire human observer performance data. These shall be used for analysis of the CHO performance with human-like channels and modifying the models to better predict human performance. This investigation is in progress and shall be reported in a separate publication.

Acknowledgments

The authors would like to thank Aleksandra Pižurica for critically reading the manuscript. We also thank the reviewers of the manuscript for their thoughtful comments and suggestions that improved the clarity of the manuscript.

This work is financially supported by the IBBT in the context of project MEdical Virtual Imaging Chain (MEVIC). MEVIC is an IBBT project in cooperation with the following companies and organizations: Barco, Hologic and Philips. IBBT is an independent multidisciplinary research institute founded by the Flemish government to stimulate ICT innovation.

References

1. American Association of Physicists in Medicine, "Specification and acceptance testing of computed tomography scanners," Tech. Rep. 39, New York, NY; American Association of Physicists in Medicine (1993).
2. National Electrical Manufacturers Association, *NEMA NU 2-2007 Performance Measurements of Positron Emission Tomographs* (ROSSLYN, The National Electrical Manufacturers Association, 2007).

3. M. A. Lodge, A. Rahmim, and R. L. Wahl, "A practical, automated quality assurance method for measuring spatial resolution in pet," J. Nucl. Med. **50**, 1307–1314 (2009).

4. P. F. Judy, R. G. Swensson, and M. Szulc, "Lesion detection and signal-to-noise ratio in CT images," Med. Phys. **8**, 13-23 (1981).

5. K. J. Myers, and H. H. Barrett, and M. C. Borgstrom, and E. B. Cargill, and A. V. Clough, and R. D. Fiete, and T. D. Milster, and D. D. Patton, and R. G. Paxman, and G. W. Seeley, and W. E. Smith, and M. O. Stempski, "A systematic approach to the design of diagnostic systems for nuclear medicine," in Information Processing in Medical Imaging: Proceedings of the Ninth Conference (Bacharach, S. L., Ed.), Martinus Nijhoff, Dordrecht, The Netherlands, 431444, (1986).

6. H. H. Barrett, "Objective assessment of image quality: effects of quantum noise and object variability," J. Opt. Soc. Am. A **7**, 1266–1278 (1990).

7. H. H. Barrett and K. J. Myers, *Foundations of Image Science* (John Wiley and Sons, New York, 2004).

8. B. I. Reiner, E. L. Siegel, F. J. Hooper, S. Pomerantz, A. Dahlke, and D. Rallis, "Radiologists productivity in the interpretation of CT scans: A 216 comparison of PACS with conventional film," Am. J. Roentgenol. **176**, 861-864 (2001).

9. A. Rahmim A, and H. Zaidi, "PET versus SPECT: strengths, limitations and challenges," Nucl. Med. Commun. **29**, 193–207 (2008).

10. I. Andersson, D. M. Ikeda, S. Zackrisson, M. Ruschin, T. Svahn, P. Timberg, and A. Tingberg, "Breast tomosynthesis and digital mammography: a comparison of breast cancer visibility and birads classification in a population of cancers with subtle mammo-

graphic findings," Eur. Radiol. **18**, 2817-25 (2008).

11. H. H. Barrett, J. Yao, J. P. Rolland, and K. J. Myers, "Model observers for assessment of image quality," Proc. Natl. Acad. Sci. USA **90**, 9758-9765 (1993).
12. H. H. Barrett, J. L. Denny, R. F. Wagner, and K. J. Myers, "Objective assessment of image quality. II. Fisher information, Fourier crosstalk, and figures of merit for task performance," J. Opt. Soc. Am. A **12**, 834–852 (1995).
13. S. Park, M. A. Kupinski, E. Clarkson, and H. H. Barrett, "Ideal-observer performance under signal and background uncertainty," in *Information Processing in Medical Imaging*, Vol. 2732 in Lecture Notes in Computer Science, C. J. Taylor and J. A. Noble, eds. (Springer-Verlag, 2003) pp. 342353.
14. M. A. Kupinski, J. W. Hoppin, E. Clarkson, and H. H. Barrett, "Ideal-observer computation in medical imaging with use of Markov-chain Monte Carlo techniques," J. Opt. Soc. Am. A **20**, 430–438 (2003).
15. B. D. Gallas and H. H. Barrett, "Validating the use of channels to estimate the ideal linear observer," J. Opt. Soc. Am. A **20**, 1725–1738 (2003).
16. M. P. Eckstein, C. K. Abbey, and J. S. Whiting, "Human vs. Model Observers in Anatomic Backgrounds," Proc. SPIE vol. 3340, p. 16-26 (1998).
17. C. K. Abbey and H. H. Barrett, "Human- and model-observer performance in ramp-spectrum noise: effects of regularization and object variability," J. Opt. Soc. Am. A **18**, 473–488 (2001).
18. J. A. Swets, and R. M. Pickett, "Evaluation of Diagnostic Systems: Methods from Signal Detection Theory," (Academic Press, New York, 1982).

19. C. E. Metz, "Quantification of failure to demonstrate statistical significance. The usefulness of confidence intervals," Invest. Radiol. **28**, 59–63 (1993).

20. H. H. Barrett, C. K. Abbey, and E. Clarkson, "Objective assessment of image quality. III. ROC metrics, ideal observers, and likelihood-generating functions," J. Opt. Soc. Am. A **15**, 1520–1535 (1998).

21. D. M. Green and J. A. Swets, "Signal Detection Theory and Psychophysics" (John Wiley & Sons, New York, 1966) [reprint (Krieger, New York, 1974)].

22. K. J. Myers and H. H. Barrett, "Addition of a channel mechanism to the ideal-observer model," J. Opt. Soc. Am. A **4**, 2447–2457 (1987).

23. H. H. Barrett, K. J. Myers, B. D. Gallas, E. Clarkson, and H. Zhang, "Megalopinakophobia: its symptoms and cures," Proc. SPIE , **4320**, p. 299–307 (2001).

24. 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. Img., **28**, 657–668, (2009).

25. S. Park and E. Clarkson, "Efficient estimation of ideal-observer performance in classification tasks involving high-dimensional complex backgrounds," J. Opt. Soc. Am. A **26**, 59–71 (2009).

26. J. M. Witten, S. Park, and K. J. Myers, "Partial least squares: a method to compute efficient channels for the Bayesian ideal observers," IEEE Trans. Med. Img., in press (2010).

27. S. Park, E. Clarkson, H. H. Barrett, M. A. Kupinski, and K. J. Myers, "Performance of a channelized-ideal observer using Laguerre-Gauss channels for detecting a Gaussian

signal at a known location in different lumpy backgrounds," Proc. SPIE vol. 6146, p. 61460P (2006).

28. S. Park, H. H. Barrett, E. Clarkson, M. A. Kupinski, and K. J. Myers, "Channelized-ideal observer using Laguerre-Gauss channels in detection tasks involving non-Gaussian distributed lumpy backgrounds and a Gaussian signal," J. Opt. Soc. Am. A **24**, B136–B150 (2007).

29. K. Myers, H. Barrett, M. Borgstrom, D. Patton, and G. Seeley, "Effect of noise correlation on detectability of disk signals in medical imaging," J. Opt. Soc. Am. A **2**, 1752–1759 (1985).

30. J. S. Kim, P. Kinahan, C. Lartizien, C. Comtat, and T. Lewellen, "A comparison of planar versus volumetric numerical observers for detection task performance in whole-body PET imaging," IEEE Trans. Nucl. Sci. **51**, 34–40 (2004).

31. J. P. Rolland and H. H. Barrett, "Effect of random background inhomogeneity on observer detection performance," J. Opt. Soc. Am. A **9**, 649–658 (1992).

32. A. E. Burgess, F. L. Jacobson, and P. F. Judy, "Human observer detection experiments with mammograms and power-law noise," Medical Physics **28**, 419–437 (2001).

33. S. Park, B. D. Gallas, A. Badano, N. A. Petrick, and K. J. Myers, "Efficiency of the human observer for detecting a Gaussian signal at a known location in non-Gaussian distributed lumpy backgrounds," J. Opt. Soc. Am. A **24**, 911–921 (2007).

34. M. Chen, J. Bowsher, A. Baydush, K. Gilland, D. DeLong, and R. Jaszczak, "Using the hotelling observer on multi-slice and multi-view simulated SPECT myocardial images," Conf. Rec. IEEE Nucl. Sci. Symp. **4**, 2258–2262 (2001).

35. C. Lartizien, P. E. Kinahan, and C. Comtat, "Volumetric model and human observer comparisons of tumor detection for whole-body positron emission tomography," Acad. Radiol. **11**, 637–648 (2004).

36. S. Young, S. Park, S. K. Anderson, A. Badano, K. J. Myers, and P. Bakic, "Estimating breast tomosynthesis performance in detection tasks with variable-background phantoms," Proc. SPIE vol. 7258, p. 72580 (2009).

37. S. Park, A. Badano, B. Gallas, and K. Myers, "Incorporating human contrast sensitivity in model observers for detection tasks," IEEE Trans. Med. Img. **28**, 339–347 (2009).

38. H. Gifford, M. King, P. Pretorius, and R. Wells, "A comparison of human and model observers in multislice LROC studies," IEEE Trans. Med. Img. **24**, 160–169 (2005).

39. S. Park, E. Clarkson, M. A. Kupinski, and H. H. Barrett, "Efficiency of the human observer detecting random signals in random backgrounds," J. Opt. Soc. Am. A **22**, 3–16 (2005).

40. C. Castella, M. P. Eckstein, C. K. Abbey, K. Kinkel, F. R. Verdun, R. S. Saunders, E. Samei, and F. O. Bochud, "Mass detection on mammograms: influence of signal shape uncertainty on human and model observers," J. Opt. Soc. Am. A **26**, 425–436 (2009).

41. H. Liang, S. Park, B. D. Gallas, K. J. Myers, and A. Badano, "Image browsing in slow medical liquid crystal displays.," Acad. Radiol. **15**, 370–382, (2008).

42. F. O. Bochud, C. K. Abbey, and M. P. Eckstein, "Statistical texture synthesis of mammographic images with clustered lumpy backgrounds," Opt. Express **4**, 33–43 (1999).

43. C. Castella, K. Kinkel, F. Descombes, M. P. Eckstein, P.-E. Sottas, F. R. Verdun,

and F. O. Bochud, "Mammographic texture synthesis: second-generationclustered lumpy backgrounds using agenetic algorithm," Opt. Express **16**, 7595–7607 (2008).

44. E. Clarkson, M. A. Kupinski, and H. H. Barrett, "A probabilistic development of the MRMC method," Acad. Radiol. **13**, 1410-1421 (2006).

45. B. D. Gallas, "One-shot estimate of MRMC variance: AUC," Acad. Radiol. **13**, 353–62 (2006).

46. Lj. Platisa, B. Goossens, E. Vansteenkiste, A. Badano, and W. Philips, "Channelized Hotelling observers for the detection of 2D signals in 3D simulated images," in Proceedings of IEEE Conference on Image Processing (ICIP, 2009), 1781 – 1784

47. R. Wells, M. King, H. Gifford, and P. Pretorius, "Single-slice versus multi-slice display for human-observer lesion-detection studies," IEEE Trans. Nucl. Sci. **47**, 1037–1044 (2000).

48. K. Fukunaga, and R. R. Hayes, "Effects of sample size in classifier design," IEEE Trans. Pattern Analysis and Machine Intelligence, **11**, 873–885 (1989).

49. C. J. van den Branden Lambrecht, "A working spatio-temporal model of the human visual system for image restoration and quality assessment applications," in Proceedings of IEEE Conference on Acoustics, Speech, and Signal Processing (ICASSP, 1996), **4**, 2291–2294

List of Figures

4 (a) Single-slice CHO. The model is constrained to use only the information of one slice in the volume, the slice in which the signal is centered, $\mathbf{g}_{(N/2)}$. The $\mathbf{g}_{(N/2)}$ from all images are first channelized using a set of 2D LG channels, $\mathbf{u}_p$, $p = 1, ..., P$ where P is the total number of channels. The vector of channel outputs $\mathbf{v}$ of the size P is than processed by the template $\mathbf{w}_{CHO}$ to estimate the test score of the ssCHO, t. (b) Volumetric CHO. The main difference from ssCHO model is that vCHO exploits not only a single slice from the volume but the image volume as a whole, $\mathbf{g} = [\mathbf{g}_{(1)}...\mathbf{g}_{(N)}]$. Here, the channels match the dimension of the image volume and they are 3D LG functions in a 3D Cartesian space. In any other aspect, the vCHO model is the same as ssCHO. 57

5 Multi-slice CHO. Processing slice data with 2D-LG channels. The multi-slice image $\mathbf{g}$ is represented as an array of slices $\mathbf{g}_{(1)}, ..., \mathbf{g}_{(N)}$, where N is the number of slices in the image. Each slice in the array $\mathbf{g}_{(n)}$ is channelized by the same set of P two-dimensional channels $\mathbf{u}_{(p)}$, $p = 1, ..., P$, to get the channel outputs $\mathbf{v}_{(n)} = [v_{1(n)}, ..., v_{P(n)}]$, where $v_{p(n)} = \mathbf{u}_{(p)}{}^t\mathbf{g}_{(n)}$. The matrix of the channel outputs for all slices in the image is denoted $\mathbf{v}_{msCHO} = [\mathbf{v}_{(1)}, \mathbf{v}_{(2)}, ..., \mathbf{v}_{(N)}]$. The same procedure applies on both signal-present and signal-absent images. The concept of ROI is explained in Section 3.C.5. 58

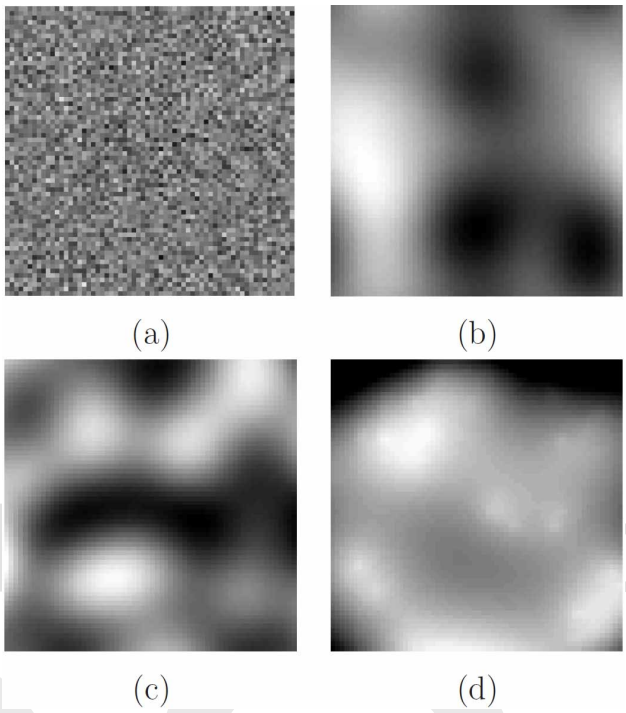

Fig. 1. Four image categories: (a) white noise background (WNB), (b) correlated Gaiussian noise background (CNB), (c) Gaussian lumpy background (LB), and (d) clustered lumpy background (CLB). In each case, a randomly selected slice from the image volume is presented. Detailed parameters of the background images are given in Table 1.

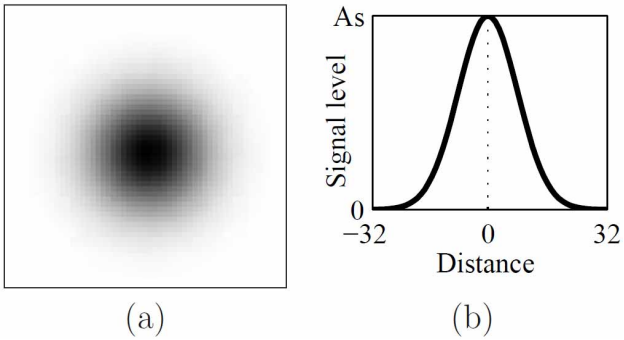

Fig. 2. Sample signal image. (a) central slice of the signal volume, size of the slice is 64×64 voxels, (b) contrast profile in the central slice of a simulated 3D Gaussian signal.

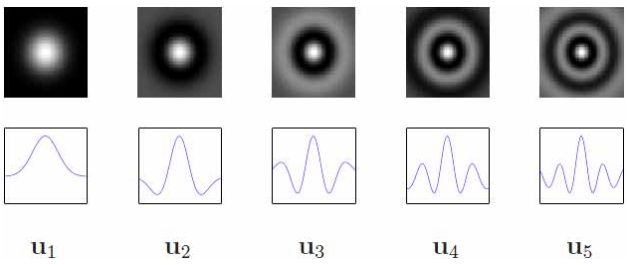

Fig. 3. The first five Laguerre-Gauss channels with the spread parameter $a_u = 24$. Top: The images illustrate 2D channels or central slices of 3D channels. Bottom: Plots of the LG functions. For 3D channels, these plots are the same in planar view (xy-plane) as in z-direction.

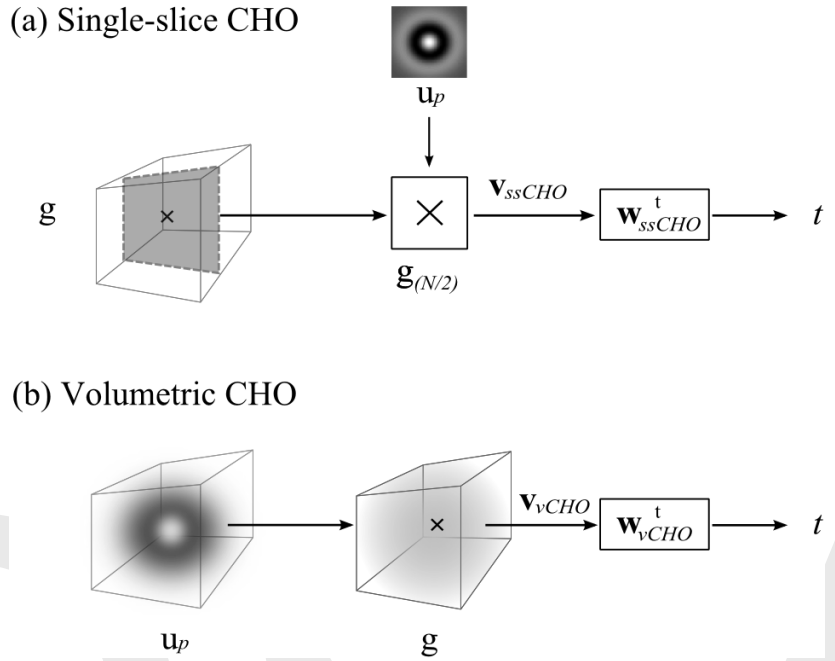

Fig. 4. (a) Single-slice CHO. The model is constrained to use only the information of one slice in the volume, the slice in which the signal is centered, $\mathbf{g}_{(N/2)}$. The $\mathbf{g}_{(N/2)}$ from all images are first channelized using a set of 2D LG channels, $\mathbf{u}_p$, $p = 1, ..., P$ where P is the total number of channels. The vector of channel outputs $\mathbf{v}$ of the size P is than processed by the template $\mathbf{w}_{CHO}$ to estimate the test score of the ssCHO, t. (b) Volumetric CHO. The main difference from ssCHO model is that vCHO exploits not only a single slice from the volume but the image volume as a whole, $\mathbf{g} = [\mathbf{g}_{(1)}...\mathbf{g}_{(N)}]$. Here, the channels match the dimension of the image volume and they are 3D LG functions in a 3D Cartesian space. In any other aspect, the vCHO model is the same as ssCHO.

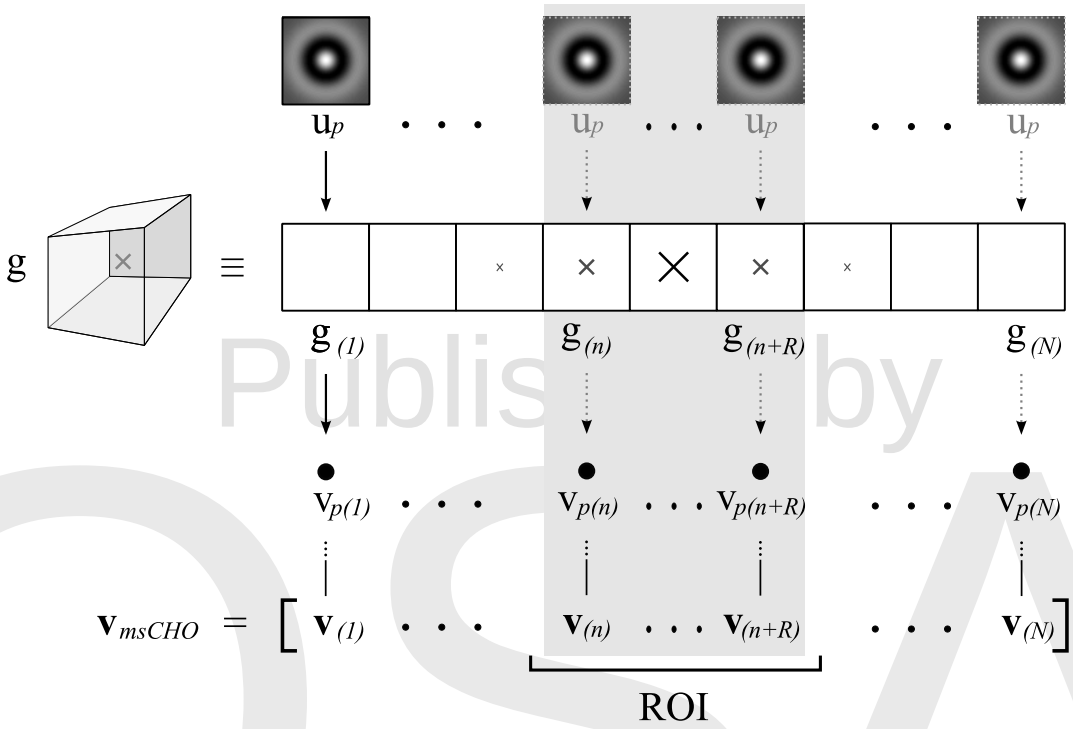

Fig. 5. Multi-slice CHO. Processing slice data with 2D-LG channels. The multi-slice image $\mathbf{g}$ is represented as an array of slices $\mathbf{g}_{(1)}, ..., \mathbf{g}_{(N)}$, where N is the number of slices in the image. Each slice in the array $\mathbf{g}_{(n)}$ is channelized by the same set of P two-dimensional channels $\mathbf{u}_{(p)}$, $p = 1, ..., P$, to get the channel outputs $\mathbf{v}_{(n)} = [v_{1(n)}, ..., v_{P(n)}]$, where $v_{p(n)} = \mathbf{u}_{(p)}{}^t \mathbf{g}_{(n)}$. The matrix of the channel outputs for all slices in the image is denoted $\mathbf{v}_{msCHO} = [\mathbf{v}_{(1)}, \mathbf{v}_{(2)}, ..., \mathbf{v}_{(N)}]$. The same procedure applies on both signal-present and signal-absent images. The concept of ROI is explained in Section 3.C.5.

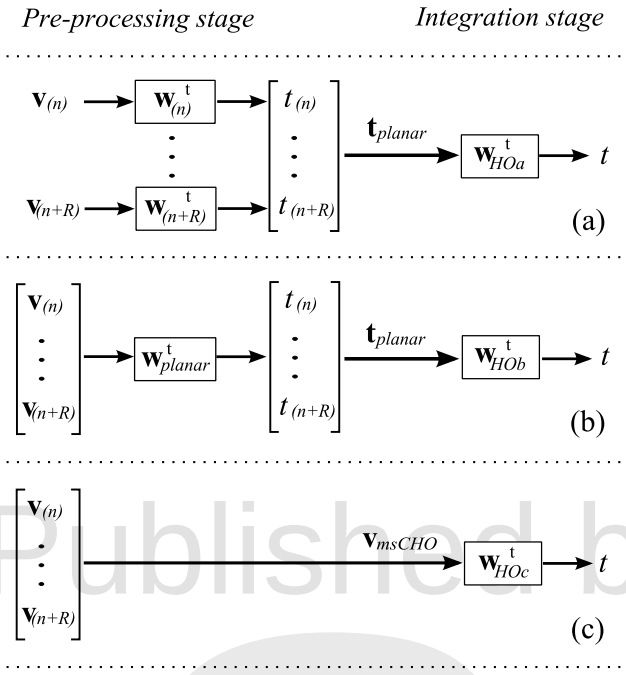

Fig. 6. Three different designs of the multi-slice CHO: (a) msCHO$_a$, (b) msCHO$_b$, (c) msCHO$_c$. Each observer is applied on the region-of-interest, ROI, consisting of R consecutive slices where $R \leq N$ and $R = N$ corresponds to the whole image sequence; for details about ROI see Section 3.C.5. First, in the early pre-processing stage, the channelized slice data $\mathbf{v}_{(n)}, ..., \mathbf{v}_{(n+R)}$ is obtained, as illustrated in Fig. 5. In two out of three msCHO designs, (a) and (b), this channelized data is used to calculate the vector of test statistics, $\mathbf{t}_{planar} = [t_{(n)}, ..., t_{(n+R)}]$, using different templates in Eq. (8): (a) a separate 2D template $\mathbf{w}_{(n)}, ..., \mathbf{w}_{(n+R)}$ is used for each slice, (b) one 2D template $\mathbf{w}_{planar}$ is used for all slices in the ROI. Next, either $\mathbf{t}_{planar}$, in case of (a) and (b), or directly the channelized slice data, in case of (c), is used as input to the integration stage. There, all three types of msCHO use 1D-HO to estimate the final test statistic, t.

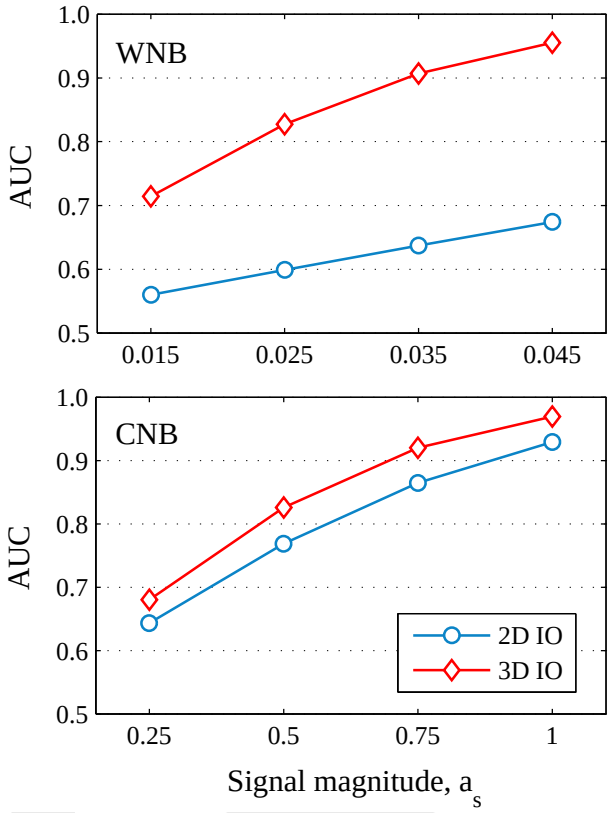

Fig. 7. Ideal observer performances. Top: Category of white noise images, WNB. Bottom: Category of colored noise images, CNB. The two curves in each graph correspond to a two-dimensional (2D IO) and a three-dimensional problem (3D IO). The 2D images are of size $M = 64^2$ with a 2D Gaussian signal inserted in the center of the image, while the 3D images are of size $M = 64^3$ with 3D spherically symmetric Gaussian signal inserted in the center of the volume. For both 2D and 3D Gaussian signal, the value of signal spread parameter is $\sigma_{s1} = 8$. Further details about image parameters are given in Table 1. The AUC values are obtained using Eq. (16) to calculate SNR and then Eq. (15) to calculate the AUC of the IO.

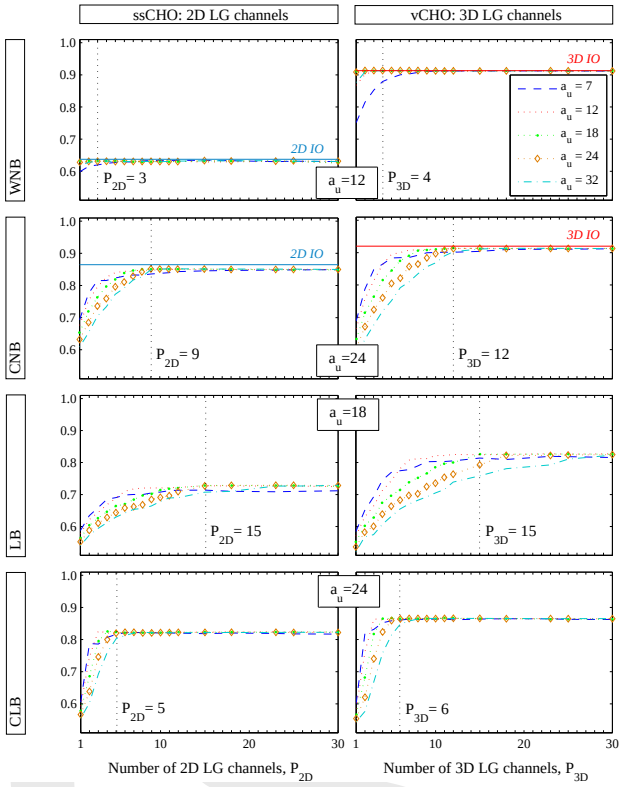

Fig. 8. Plots of estimated AUC as a function of a number of channels, $P = 1, ..., 30$. Right: AUC for a single-slice CHO design (ssCHO) using 2D LG channels applied on the central slice of the image stack. Left: AUC for a volumetric CHO design (vCHO) using 3D LG channels. For both model designs, a set of different spread parameters is considered, $a_u = \{7, 12, 18, 24, 32\}$. Top to bottom: The results for WNB ($a_s = 0.035$), CNB ($a_{s1} = 0.75$), LB ($a_s = 12$), and CLB ($a_s = 12$). The plots are obtained for $N_{tr} = 2000$ trainer image pairs and $N_{ts} = 1000$ test image pairs. Selected channel parameters are listed in Table 2: channel spread parameter a_u, and number of channels P_{2D} for ssCHO, and P_{3D} for vCHO.

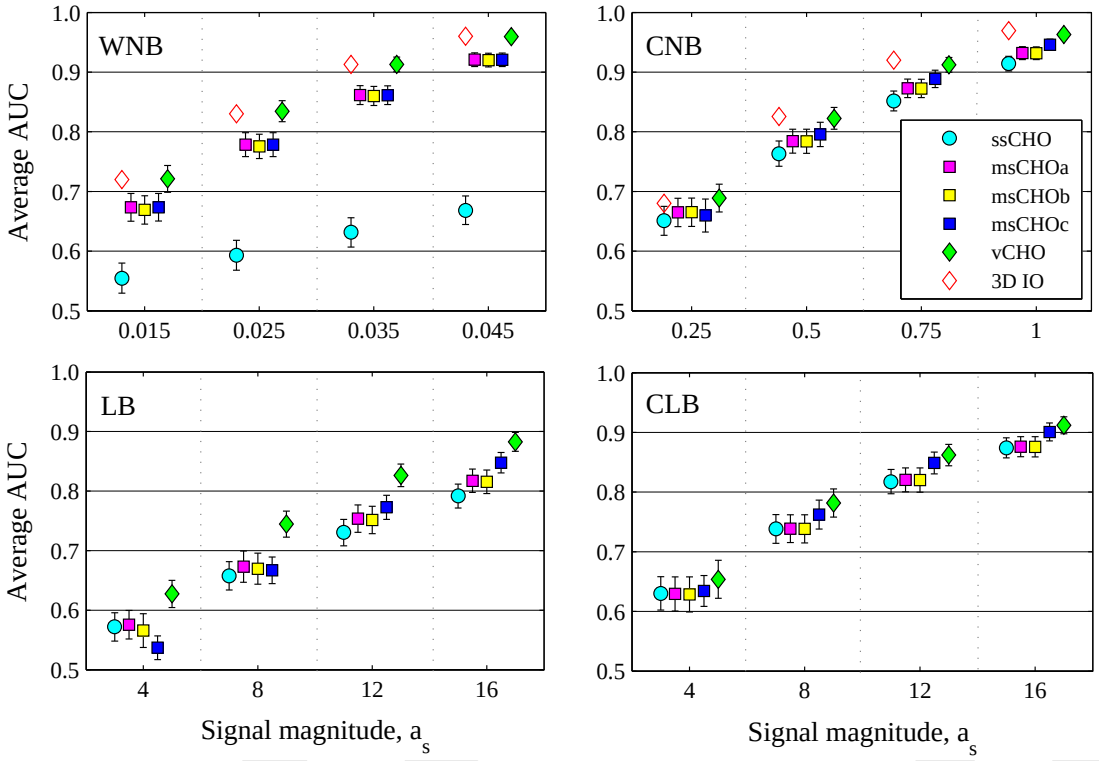

Fig. 9. Average AUC for the five model observer designs: ssCHO run on the central slice in the image; msCHO$_a$, msCHO$_b$ and msCHO$_c$ each applied on the region-of-interest comprised of $R = 11$ adjacent slices centered on the central signal slice; vCHO applied on the whole image volume. Each graph corresponds to one of the four background categories (left to right, top to bottom): WNB, CNB, LB and CLB. The value of signal spread parameter is $\sigma_{s1} = 8$ and the related signal magnitudes a_s correspond to those defined in Table 1. Number of trainer image pairs per reader $N_{tr} = 2000$, and number of tester pairs $N_{ts} = 1000$. Number of readers N_{rd} corresponds to the applicable study configurations from Table 3. Error bars are ± 2 standard deviations estimated by the one-shot method.

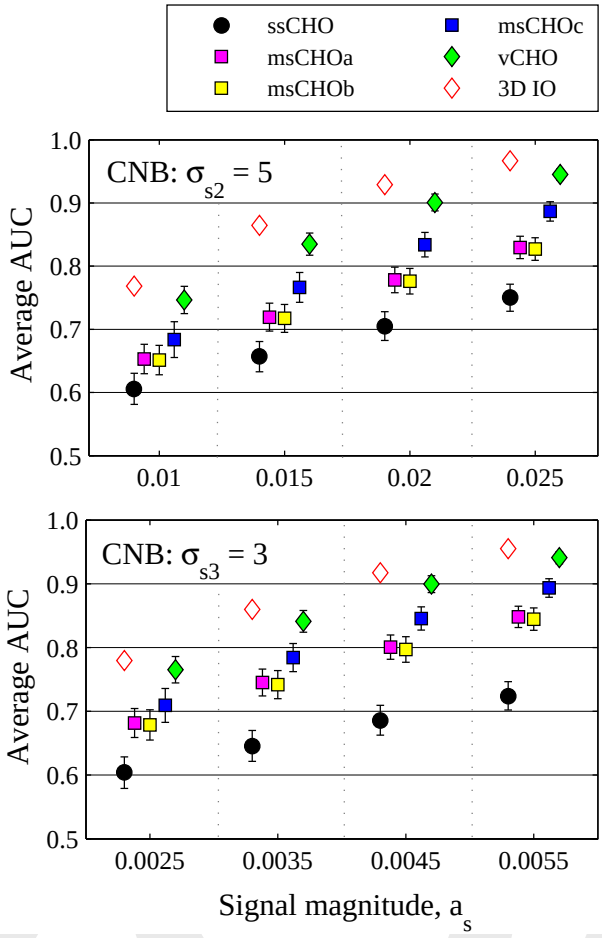

Fig. 10. For CNB image category, average AUC of the five CHO model designs and the ideal 3D IO when the value of signal spread parameter is (top) $\sigma_{s2} = 3$ and (bottom) $\sigma_{s3} = 5$. The related signal magnitudes a_s correspond to those defined in Table 1. The three msCHO models are applied on the region-of-interest comprised of $R = 11$ adjacent slices centered on the central signal slice. Number of trainer image pairs per reader $N_{tr} = 2000$, and number of tester pairs $N_{ts} = 1000$. Number of readers $N_{rd} = 5$. Error bars are ± 2 standard deviations estimated by the one-shot method.

Table 1. Signal and background parameters

Background category	Background 3D image	Gaussian 3D signal
White noise	$\sim N(0,1)$	$\sigma_s = 8$,
(WN)	$M = 64^3$	$a_s = \{0.015, 0.025, 0.035, 0.045\}$
Coloured noise	$\sigma_b = 8$,	$\sigma_{s1} = 8$,
(CN)	$M = 64^3$	$a_{s1} = \{0.25, 0.5, 0.75, 1\}$
		$\sigma_{s2} = 5$,
		$a_{s2} = \{0.01, 0.015, 0.02, 0.025\}$
		$\sigma_{s3} = 3$,
		$a_{s3} = \{0.0025, 0.0035, 0.0045, 0.0055\}$
Lumpy background	$\sigma_b = 8$, $a_b = 255$,	$\sigma_s = 8$,
(LB)	$M = 64^3$	$a_s = \{4, 8, 12, 16\}$
	$M_{FOV} = 128^3$, $\overline{K} = 800$	
Clustered lumpy	$L_x = 3$, $L_y = 2$, $L_z = 3$,	$\sigma_s = 8$,
background	$a_b = 255$, $M = 64^3$,	$a_s = \{4, 8, 12, 16\}$
(CLB)	$M_{FOV} = 128^3$,	
	$\overline{K} = 80$, $\overline{N} = 20$	

The following notation applies: M - number of voxels in the image; M_{FOV} - number of voxels in the FOV (LB, CLB); σ_s - spread parameter of the 3D Gaussian signal; a_s - magnitude of the 3D Gaussian signal; σ_b - standard deviation of the 3D Gaussian kernel (CNB); or spread parameter of the 3D Gaussian lump (LB); a_b - peak intensity level in the background image; $\overline{K}$ - mean number of lumps in the FOV (LB, CLB); L_x, L_y and L_z - characteristic lengths of asymmetrical lumps in x, y and z directions, respectively (CLB).

Table 2. Parameters of the Laguerre-Gauss (LG) channels

Background category	Signal size	a_u	P_{2D}	P_{3D}
WNB	$\sigma_s = 8$	12	3	4
CNB	$\sigma_{s1} = 8$	24	9	12
	$\sigma_{s2} = 5$	21	11	12
	$\sigma_{s3} = 3$	12	12	12
LB	$\sigma_s = 8$	18	15	15
CLB	$\sigma_s = 8$	24	5	6

For each image category and their related signal size, parameters of the LG channels are determined: the size of the channels, a_u, the number of 2D LG channels, P_{2D}, and the number of 3D LG channels, P_{3D}. The parameters of 2D and 3D LG channels are selected in the experiments with ssCHO and vCHO models, respectively. The models are investigated in the space of five families of LG channels defined by the value of the channel spread parameter, $a_u = \{7, 12, 18, 24, 32\}$. For each family, the number of LG channels is varied in the range of $P = 1, ..., 30$. The experiments are conducted with $N_{tr} = 2000$ trainer pairs and $N_{ts} = 1000$ tester pairs, and for the second largest among four considered values of signal magnitude a_s given in Table 1. The results of these experiments are illustrated in Fig. 3.

Table 3. MRMC study configurations

Background category	Number of trainer image pairs (N_{tr})	Number of readers (N_{rd})
WNB, CNB	$N_{tr} = \{50, 100, 200, 500, 1000, 2000\}$	$N_{rd} = 5$
	$N_{tr} = \{5000\}$	$N_{rd} = 2$
LB, CLB	$N_{tr} = \{50, 100, 200, 500, 1000\}$	$N_{rd} = 5$
	$N_{tr} = \{2000\}$	$N_{rd} = 3$

The total number of each of WNB and CNB images is 11000 image pairs, and the total number of each of LB and CLB images is 7000 image pairs. For all study configurations, the number of tester image pairs is fixed to $N_{ts} = 1000$. No overlap exists between the trainer images and the tester images.

Table 4. Terms of Eq. (17) for three different types of model observer efficiency, η

Type of efficiency	SNR_{curr}	SNR_{ref}
η_{CHO}	SNR of a given CHO model	SNR of the IO
$\eta_{N_{tr}}$	SNR of the CHO trained with N_{tr} image pairs, $N_{tr} < 5000$ (see Table 3)	SNR of the CHO trained with maximum considered number of trainer pairs, $N_{tr} = 5000$
$\eta_{ss,v}$	SNR of the ssCHO model	SNR of the vCHO model

Table 5. Efficiency of CHO models applied on CNB images with different spread of the signal: efficiency of CHO model relative to the IO performance, η_{CHO}; efficiency of ssCHO relative to the vCHO performance, $\eta_{ss,v}$

		ssCHO	msCHO$_a$	msCHO$_b$	msCHO$_c$	vCHO	
σ_s	a_s	η_{CHO} [%]	η_{CHO} [%]	η_{CHO} [%]	η_{CHO} [%]	η_{CHO} [%]	$\eta_{ss,v}$ [%]
8	0.25	69	85	86	91	>100	62
	0.5	59	71	71	82	98	60
	0.75	55	66	66	77	93	59
	1	53	63	63	75	91	59
5	0.01	13	28	27	35	82	16
	0.015	13	27	27	36	78	17
	0.02	14	27	26	37	77	18
	0.025	14	26	26	37	76	18
3	0.0025	12	36	35	46	88	13
	0.0035	12	36	36	47	86	14
	0.0045	12	36	35	47	85	14
	0.0055	12	36	35	47	85	15

Three different values of signal spread parameter are considered: $\sigma_{s1} = 8$, $\sigma_{s2} = 5$ and $\sigma_{s3} = 3$. For each σ_s the exact same backgrounds are used and their lump spread parameter is $\sigma_b = 8$. For multi-slice CHO models (msCHO), the efficiency for the ROI size of $R = 11$ are given. The values of η_{CHO} and $\eta_{ss,v}$ are calculated using the formula in Eq. (17) and as explained in Section 4.C. The calculations are done for MRMC configuration with the number of trainer image pairs $N_{tr} = 5000$.

Table 6. Efficiency of five CHO models for different levels of the signal a_s while the number of trainer images increase: $\eta_{N_{tr|a_s}}$

a_s	0.25	0.5	0.75	1	0.25	0.5	0.75	1	0.25	0.5	0.75	1
	$\eta_{N_{tr\|a_s}}$ [%] for ssCHO				$\eta_{N_{tr\|a_s}}$ [%] for vCHO							
$\text{SNR}_{N_{tr}=5000}$	0.55	1.02	1.48	1.94	0.70	1.31	1.92	2.53				
$N_{tr} = 50$	41	74	84	88	55	82	87	88				
100	56	81	89	92	61	84	90	93				
200	69	91	95	97	73	92	96	97				
500	89	96	98	99	90	97	98	99				
1000	96	99	99	100	94	99	99	100				
2000	99	100	100	100	99	100	100	100				
	$\eta_{N_{tr\|a_s}}$ [%] for msCHO_a				$\eta_{N_{tr\|a_s}}$ [%] for msCHO_b				$\eta_{N_{tr\|a_s}}$ [%] for msCHO_c			
$\text{SNR}_{N_{tr}=5000}$	0.61	1.12	1.62	2.11	0.61	1.12	1.62	2.11	0.63	1.20	1.75	2.30
$N_{tr} = 50$	13	51	69	75	23	58	72	78	0	0	3	5
100	34	71	81	85	37	70	82	87	11	26	36	42
200	44	83	91	93	54	86	93	95	16	43	60	68
500	78	94	96	97	81	94	97	98	41	70	80	85
1000	88	97	98	99	89	97	98	99	61	85	92	94
2000	97	99	99	100	97	99	99	100	85	95	97	98

For CNB images, the efficiency of CHO models: ssCHO, msCHO_a, msCHO_b, msCHO_c, and vCHO, trained with fewer image pairs relative to their performance for the largest considered number of trainer images, $\eta_{N_{tr|a_s}}$, are calculated using the formula in Eq. (17) and as explained in Section 4.C. For three msCHO models, the efficiency for the ROI size of $R = 11$ are given.

Table 7. Efficiency of msCHO models for different size of ROI while the number of trainer images increase: $\eta_{N_{tr|R}}$

R	3	5	11	64	3	5	11	64	3	5	11	64
	$\eta_{N_{tr\|R}}$ [%] for msCHO$_a$				$\eta_{N_{tr\|R}}$ [%] for msCHO$_b$				$\eta_{N_{tr\|R}}$ [%] for msCHO$_c$			
SNR$_{N_{tr}=5000}$	1.49	1.49	1.62	1.73	1.49	1.49	1.62	1.73	1.60	1.65	1.75	-
$N_{tr} = 50$	84	83	69	19	82	81	72	25	55	36	3	-
100	88	87	81	49	87	86	82	52	77	68	36	-
200	95	94	91	68	95	94	93	73	88	80	60	-
500	97	97	96	88	97	97	97	87	94	89	80	-
1000	99	99	98	94	99	99	98	94	98	96	92	-
2000	100	100	99	98	100	100	99	98	99	99	97	-

For CNB images, the efficiency of msCHO models: msCHO$_a$, msCHO$_b$, and msCHO$_c$, trained with fewer image pairs relative to their performance for the largest considered number of trainer images, $\eta_{N_{tr|R}}$, are calculated using the formula in Eq. (17) and as explained in Section 4.C. In particular, the efficiency for the signal magnitude of $a_s = 0.75$, each for four different ROI size, $R = \{3, 5, 11, 64\}$, are presented. Here, $R = 64$ implies that the CHO is applied to all slices in the image.