Journal of Medical Imaging

MedicalImaging.SPIEDigitalLibrary.org

Channelized Hotelling observer correlation with human observers for low-contrast detection in liver CT images

Alexandre Ba
Craig K. Abbey
Damien Racine
Anaïs Viry
Francis R. Verdun
Sabine Schmidt
François O. Bochud

SPIE.

Alexandre Ba, Craig K. Abbey, Damien Racine, Anaïs Viry, Francis R. Verdun, Sabine Schmidt, François O. Bochud, "Channelized Hotelling observer correlation with human observers for low-contrast detection in liver CT images," *J. Med. Imag.* **6**(2), 025501 (2019), doi: 10.1117/1.JMI.6.2.025501.

Channelized Hotelling observer correlation with human observers for low-contrast detection in liver CT images

Alexandre Ba,[a] Craig K. Abbey,[b] Damien Racine,[a] Anaïs Viry,[a] Francis R. Verdun,[a] Sabine Schmidt,[c] and François O. Bochud[a,*]
[a]Lausanne University Hospital and Lausanne University, Institute of Radiation Physics, Lausanne, Switzerland
[b]UC Santa Barbara, Department of Psychological and Brain Sciences, California, United States
[c]Lausanne University Hospital and Lausanne University, Department of Radiology, Lausanne, Switzerland

Abstract. Task-based image quality procedures in CT that substitute a human observer with a model observer usually use single-slice images with uniform backgrounds from homogeneous phantoms. However, anatomical structures and inhomogeneities in organs generate noise that can affect the detection performance of human observers. The purpose of this work was to assess the impact of background type, uniform or liver, and the viewing modality, single- or multislice, on the detection performance of human and model observers. We collected abdominal CT scans from patients and homogeneous phantom scans in which we digitally inserted low-contrast signals that mimicked a liver lesion. We ran a rating experiment with the two background conditions with three signal sizes and three human observers presenting images in two reading modalities: single- and multislice. In addition, channelized Hotelling observers (CHO) for single- and multislice detection were implemented and evaluated according to their degree of correlation with the human observer performance. For human observers, there was a small but significant improvement in performance with multislice compared to the single-slice viewing mode. Our data did not reveal a significant difference between uniform and anatomical backgrounds. Model observers demonstrated a good correlation with human observers for both viewing modalities. Human observers have very similar performances in both multi- and single-slice viewing mode. It is therefore preferable to use single-slice CHO as this model is computationally more tractable than multislice CHO. However, using images from a homogeneous phantom can result in overestimating image quality as CHO performance tends to be higher in uniform than anatomical backgrounds, while human observers have similar detection performances. [DOI: 10.1117/1.JMI.6.2.025501]

Keywords: channelized Hotelling observers; image quality; observer performance; computed tomography; liver.

Paper 18227RR received Oct. 7, 2018; accepted for publication Apr. 15, 2019; published online May 20, 2019.

1 Introduction

Image quality plays a central role in clinical CT. A relevant way to assess image quality could be by measuring the diagnostic accuracy of an observer performing a task of interest. At the same time, there is an advantage in terms of reducing time/cost of reader studies in substituting a human observer with a numerical model. This has led to the development of many model observers that can predict human observer detection performance in medical images. In recent years, the channelized Hotelling model observer (CHO) has been increasingly used[1–4] and during the last decade, numerous studies have documented a good correlation with human observers in CT dose optimization or assessment of iterative reconstruction algorithms.[5–10]

Many studies for low-contrast detection with model observers use images with a uniform background from homogeneous phantoms.[7,11] However, anatomical structures and inhomogeneities in organs generate noise that can affect the detection performance of human observers.[4,12] In addition, a recent study of model observers showed that anatomical-like texture in CT images affects low-contrast lesion detectability.[13] Image quality assurance procedures in CT that use uniform background images and CHO would therefore greatly benefit from a comparison of performances with clinical images of patients containing anatomical textures and features. This could be achieved by assessing the correlation of CHO detection performance compared with human observers in both uniform and anatomical images.

CT produces a volumetric representation of an object that radiologists usually explore slice-by-slice in a multislice viewing mode. However, low-contrast detection studies with CHO have only focused on single-slice images. Strategies to model multislice reading with CHO (msCHO) have been proposed[14,15] and recent studies[16,17] found a good correlation with human observers in CT images of a homogeneous phantom with a uniform background. While it is closer to clinical practice to model a multislice reading task with msCHO, single-slice CHO (ssCHO) is more tractable[17] and has already been extensively validated. In addition, ssCHO and human observer detection performance correlates well in a multislice reading task in CT images with a uniform background.[17] However, for some two-dimensional (2-D) imaging modalities, e.g., in mammography,[12] human observer performance is affected by anatomical noise, and it is therefore worthwhile to investigate if this is valid for a three-dimensional imaging modality such as CT. This can be

*Address all correspondence to François O. Bochud, E-mail: Francois.Bochud@chuv.ch

2329-4302/2019/$25.00 © 2019 SPIE

achieved by comparing ssCHO and msCHO with human performance for anatomical images. Ultimately, if ssCHO and uniform background images are to be used for quality assurance procedures, it could be worthwhile to investigate how it correlates with msCHO in anatomical images.

The purpose of this study is to answer the following questions: Does detection in anatomical backgrounds lead to different performances than in a uniform background? What is the impact of the reading modality (single- versus multislice) on detection performance in anatomical backgrounds? What is the level of correlation for detection performance between CHO and human observers in anatomical backgrounds?

To answer these questions, we collected both abdominal CT scans from patients and homogeneous phantom scans in which we digitally inserted low-contrast signals mimicking a liver lesion. We ran a rating experiment with the two background conditions with three signal sizes and three human observers in two reading modalities: single- and multislice. We estimated the receiver operating characteristic (ROC) for all observers and computed the area under the ROC curve as a performance metric. In addition, a CHO for single- and multislice detection was implemented and evaluated according to its degree of correlation with the human observer performance.

2 Materials and Methods

2.1 *CT Data Acquisition*

Two categories of CT images were used in this study: liver CT images from patients with an anatomical background and CT images of an abdominal phantom with a uniform background (QRM, Moerendorf) (Fig. 1). The first category involved 16 CT exams of contrast-enhanced abdominal cases without pathologies that were collected in our hospital database [Fig. 1(b)]. The images were acquired with a local protocol for abdominal exploration using a multidetector CT (Discovery HD 750, GE Healthcare) at a pitch of 1.375, 0.6 s rotation time, 40 mm collimation width, 120 kVp tube voltage, and a noise index of 25. The images were reconstructed using statistical iterative algorithm (ASIR) at a level strength of 50%, image thickness was 2.5 mm, and slice interval was 2 mm. The second image category contained 13 repeated acquisitions of an abdominal phantom with additional shell mimicking medium to large patient attenuation (100 cm perimeter).

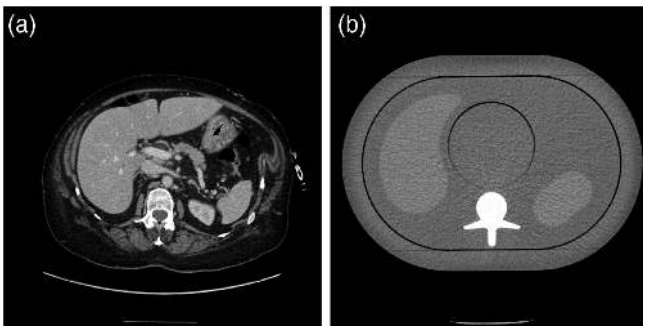

Fig. 1 Sample slices of the images used in this study. (a) Contrast-enhanced abdominal CT acquisitions collected in our hospital database. (b) CT image of the QRM abdominal phantom with additional shell.

2.2 *Image Texture Characterization*

The uniform and anatomical textures were characterized according to their noise power spectrum (NPS) using a homemade Igor Pro-based software (WaveMetrics Inc., Portland). For each clinical exam, 50 ROIs of 70×70 mm were manually selected within the respective liver structure. No overlap between ROIs was allowed. The 2-D NPS was estimated by calculating the average squared amplitude of the finite Fourier transform of the ROIs. To facilitate comparison, the 2-D NPS was radially averaged and resampled using linear interpolation.

2.3 *Image Sets for Reader Studies*

We used the same collected images to generate the signal-present and the signal-absent dataset. The signal-present images were hybrid CT images generated by inserting 6, 8, and 10 mm low-contrast spherical signals, mimicking hypodense focal liver lesions for both anatomical and uniform image classes. The signal profile in all directions was fitted to real liver lesion profiles[18,19] [Fig. 2(a)]. The method used an alpha blending technique, which replaces anatomical structures in the volume of interest by another obtained by blending a uniform region and the signal. For anatomical images, lesions were manually inserted into the regions of interest within the liver. We avoided superimposing the lesions onto main anatomical structures such as veins, arteries, or known focal liver lesions. For uniform images, the lesions were randomly inserted in the liver-like part of the phantom and distributed within eight nonoverlapping specific locations.

All the images contained 512×512 pixels (one pixel = [0.58 to 0.98] mm, depending on patient size). In signal-present and signal-absent datasets, two subsets were generated with single and multiple slices. The multislice sets contained 10 consecutive slices. The signal was present in 3, 4, or 5 consecutive slices, for 6, 8, and 10 mm signal size, respectively, centered in the middle slice. The single-slice image category was extracted from the middle slice image of the multiple slice image subsets. Each dataset contained 100 anatomical and 100 uniform independent sets of images with a signal prevalence of 1:2. For the human observer study, four contrasted fiducial cues were inserted around the signal location and were presented [Fig. 2(b)]. For the model observer study, 78×78-pixel regions of interest around signal locations, without fiducial cues, were cropped from the human observer study dataset. For multislice image sets, only slices containing the signal were used for model observer computation (Fig. 3).

The signal amplitude was set in a prestudy to reach a targeted human performance defined as an area under the ROC curve (AUC) equal to 0.95 with the uniform multislice dataset and the largest signal (10 mm). The same signal amplitude was then used in all the other conditions investigated in this study.

In total, we investigated 12 different image categories: two background types (uniform and anatomical), two reading modes (single- and multislice), and three signal sizes (6, 8, and 10 mm).

2.4 *Human Reader Study*

The participants were one woman and two men and had approximately equal experience with detection study as they were involved in previous psychophysical studies with similar phantom and images. We used a Python-based (Python Software Foundation, Beaverton) homemade software for presentation

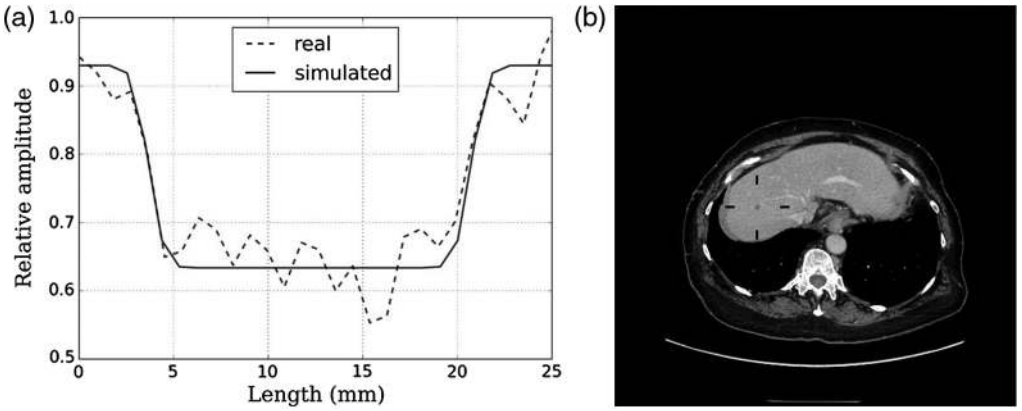

Fig. 2 (a) Synthetic signal profile (solid line) matching with real lesion profile (dashed line) in the slice. (b) Simulated lesion inserted in abdominal images (signal contrast is enhanced for visualization purpose). For human reader studies, four contrasted fiducial cues were inserted around the signal location.

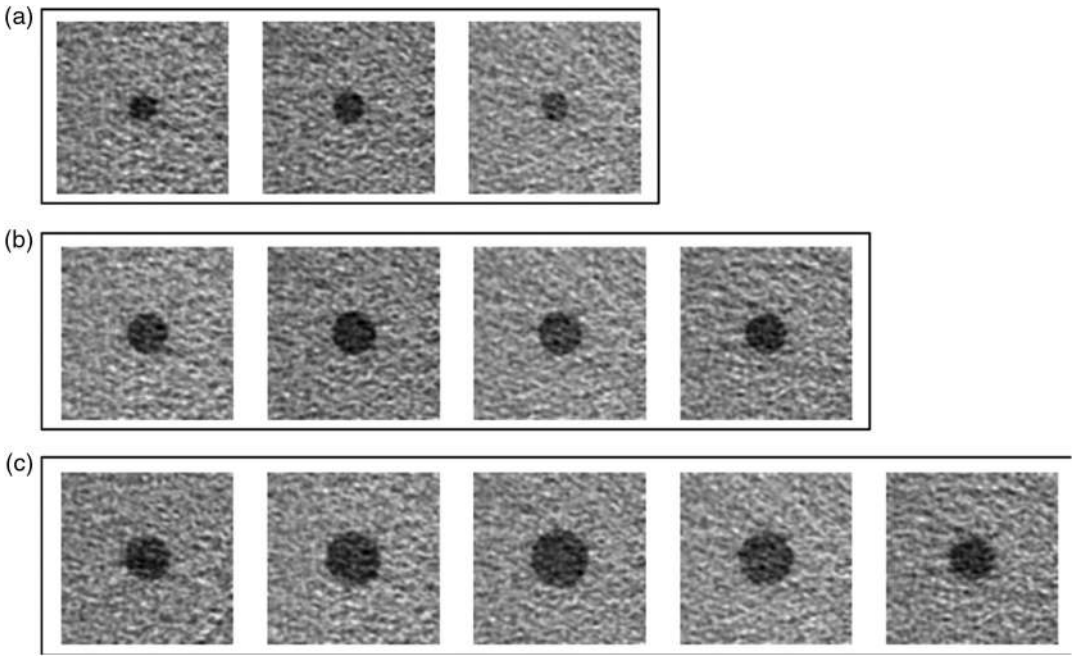

Fig. 3 Ensemble average over 120 image samples of signal present images used in model observer study for multislice sets in uniform background. The ROI with signal present contains three slices for (a) 6 mm, four slices for (b) 8 mm, and five slices for (c) 10 mm signal size. Single-slice sets were generated from multislice ensemble central slice: slice in second position for 6 mm signal size, slice in second position for 8 mm signal size, and slice in third position for 10 mm signal size.

and evaluation of medical images in observer performance studies. The readers were instructed to inspect the images that could contain a single signal in a location known exactly (SKE) and quantify their confidence about the presence of the stimulus on a nine-level ordinal rating scale: from "signal is definitely present" (1) to "signal is definitely absent" (9) with the cutoff in the middle of the scale. With the multiple slice condition, the readers were allowed to scroll through the 10-slice stack at any speed or direction.

A white square on the top left of the image indicated the middle slice in multislice reading mode. No time limitation to read the cases was imposed and no zoom or pan was allowed. Figure 4 shows screenshots of the user interface of the software used for the reader studies. Note that each image was accompanied with the rating scale.

This study consisted of two reading sessions per image background type, reading mode, and signal size. Each reader was first trained with demo images of 20 trials. The demo images were randomly chosen in the testing set and the prevalence of signal present images was 1:2. For training and testing, each image sample was presented once. A feedback was given after each answer in order to allow the readers to learn about their strategy. The training session was followed by a testing session with respectively 100 single-slice uniform, 100 single-slice anatomical, 100 multislice uniform, and 100 multislice anatomical images. For each reading session, the reader was informed about the signal prevalence (which was always 1:2).

The images were displayed with a zoom factor of two on a diagnostic monitor (EIZO Radiforce MX210) calibrated to the DICOM grayscale display function and TG18 standards.

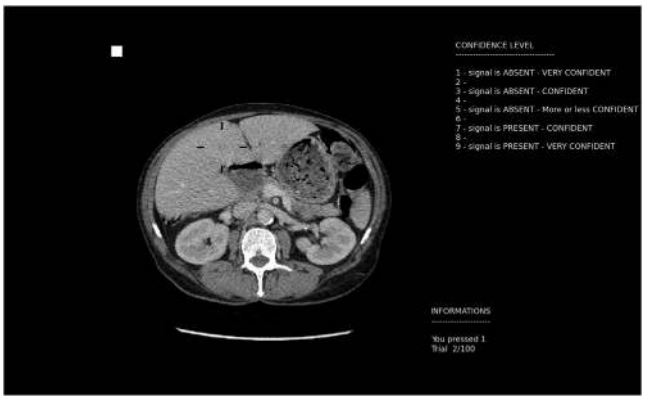

Fig. 4 Screenshot of the software user interface used for the reader studies. The figure shows the central slice of an anatomical image set with cues indicating the possible signal location.

2.5 *Model Observers for Detection in Single- and Multislice Datasets*

Linear model observers compute a scalar quantity called a decision variable from a linear combination of the image's parameters. The decision variable λ of a given model with template $\mathbf{w}$ to an image $\mathbf{g}$ is given as

$$\lambda = \mathbf{w}^t\mathbf{g}.$$

In this study, we chose the channelized Hotelling observer (CHO).[1,20,21] This observer decorrelates image noise from image samples using the ensemble covariance matrices of the signal-present and signal-absent image classes to compute the decision variable. In practice, CHO template, $\mathbf{w}_{\mathrm{CHO}}$ is obtained by first processing each image $\mathbf{g}$ by a set of 2-D channels $\mathbf{u}_j$, with $j = 1, \ldots, J$, which reduce the dimensionality of the image to the number of channels $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$ with $\mathbf{V} = \mathbf{U}^t\mathbf{g}$ and $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_n]$. Then, the channels' outputs are linearly combined using the dot product between the inverse of the covariance matrix $\mathbf{K}_v$ and an estimation of the mean difference signal in the channel space following $\mathbf{w}_{\mathrm{CHO}} = \mathbf{K}_v^{-1}[\langle \mathbf{v}_{\mathrm{signal-present}} \rangle - \langle \mathbf{v}_{\mathrm{signal-absent}} \rangle]$.

For the multislice dataset, we used the decision variable from the multiple-slice image with CHO proposed by Platisa et al.[14] and Chen et al.[15] In summary, we first computed the decision variable from the CHO template $\mathbf{w}_{\mathrm{CHO}}$ to each slice composing the multislice image to provide N decision variables $\lambda_{ms} = [\lambda_1, \ldots, \lambda_n]$ with $\lambda_n = \mathbf{w}_{\mathrm{CHO}}^T\mathbf{v}_n$ These N decision variables were then integrated using a dot product with a one-dimensional Hotelling observer (HO) following $\lambda = \mathbf{w}_{\mathrm{HO}}^T\boldsymbol{\lambda}_{\mathrm{ms}}$. The HO template $\mathbf{w}_{\mathrm{HO}}$ is computed from the ensemble covariance matrices of the signal-present and signal-absent decision variables ensemble and an estimation of the mean difference of signal-present and signal-absent decision variable classes following $\mathbf{w}_{\mathrm{HO}} = \mathbf{K}^{-1}[\langle \lambda_{\mathrm{signal-present}} \rangle - \langle \lambda_{\mathrm{signal-absent}} \rangle]$.

In this paper, the CHO for signal detection in a single-slice image is called an ssCHO and the CHO for signal detection in a multislice image is called a multislice CHO (msCHO).

We chose four types of different anthropomorphical[22] sets of channels, all of which have been successfully used to predict human observer detection performance: dense-difference-of-Gaussian channels (D-DOG), sparse-difference-of-Gaussian (S-DOG), square (SQR), and Gabor (GBR). In this study, D-DOG, S-DOG, and SQR are used as described by Abbey and Barrett.[23] GBR channels functions are defined as

$$V(x,y) = \exp\left[-4\ln(2)\frac{(x^2+y^2)}{w_s^2}\right] \times \cos[2\pi f(x\cos\theta + y\sin\theta) + \beta],$$

where f is the spatial frequency, θ is the orientation, w_s is the bandwidth, and β is the phase. The Gabor channels' parameters proposed in this study used five orientations, seven frequencies, and one phase resulting in 35 channels.

The D-DOG channels' radial spatial frequency profile functions are defined as

$$C_j(\rho) = \exp\left[-\frac{1}{2}\left(\frac{\rho}{Q\sigma_j}\right)^2\right] - \exp\left[-\frac{1}{2}\left(\frac{\rho}{\sigma_j}\right)^2\right],$$

where $\sigma_j = \sigma_0\alpha^j$ is the channel standard deviation of the j'th channel, and σ_0 is the initial standard deviation, $\sigma_0 = 0.005$ pixels^{-1}, $\alpha = 1.4$, $Q = 1.66$. We used 10 channels.

The S-DOG channels' radial spatial frequency profile functions are defined by the same function as D-DOG with different parameters, where $\sigma_j = \sigma_0\alpha^j$, $\sigma_0 = 0.015$ pixels^{-1}, $\alpha = 2$, $Q = 2$. We used three channels.

The SQR channels' radial spatial frequency profile functions are defined as

$$C_j(\rho) = \begin{cases} 0 & \text{for } \rho \le \rho_0\alpha^{j-1} \\ 1 & \text{for } \rho_0\alpha^{j-1} < \rho \le \rho_0\alpha^j \\ 0 & \text{for } \rho > \rho_0\alpha^j \end{cases}.$$

$\rho_0 = 0.015$, $\alpha = 2$. We used four channels.

2.6 *Figure of Merit and Reader Performance Analysis*

The figure of merit for the human observer experiment was the area under the ROC curve (AUC) obtained with the multiple-reader, multiple-case (MRMC) paradigm.[24]

Reader performance analysis was performed using the software package OR-DBM MRMC 2.51, written by Kevin M. Schartz et al. and freely available at Ref. 25. This program is based on the methods initially proposed by Dorfman et al.[26] and Obuchowski and Rockette[27] and later unified and improved by Hillis and colleagues.[28–30] Six analyses were performed, one for each background-size combination. The three readers were considered as fixed effects and the cases were considered as random effects.

The figure of merit for model observer experiments was the detectability index d'. The model observer performance was estimated using the resubstitution method with corrected bias from Ref. 31. The advantage of this method is to provide an unbiased direct estimation of d' and the associated 95% confidence interval.

To enable the comparison with model observers, AUC from the human observer study was converted to a detectability index from a rating experiment using

$$d_A = \sqrt{2}\Phi^{-1}(\mathrm{AUC}).$$

Both d_A and d' are equivalent if the decision variable is normally distributed under each hypothesis with equal variance.[32]

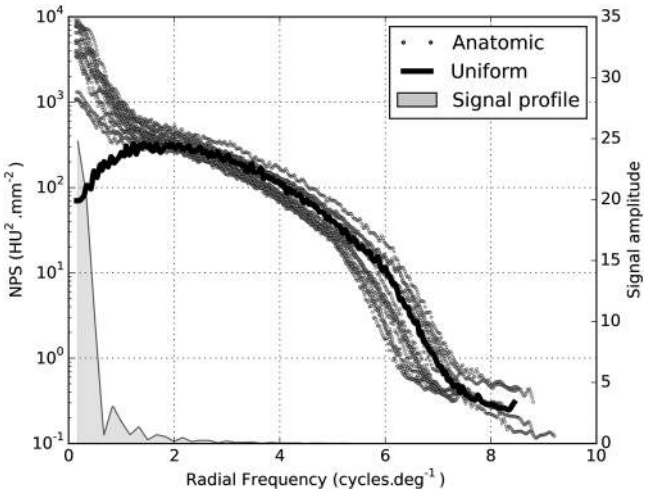

Fig. 5 Plot of NPS (left axis) with respect to the radial frequency profile in cycle per degrees for the 16 cases collected in the anatomical image category (dot markers) and for one image realization of the phantom in the uniform image category (bold solid line). Frequency range was calculated for a screen observer distance of 50 cm. The shaded area represents the 6-mm signal frequency profile (right axis).

3 Results

3.1 *Textures Characterization*

Figure 5 presents the measured NPS averaged over all orientations and displayed versus their radial spatial frequency for anatomical and uniform background image sets. Both NPS presented a relatively close agreement from approximately two to eight cycles deg^{-1}. For the lower frequency range, the uniform image NPS decreases and the anatomical image NPS increases. The small-size signal profile is plotted in the same figure. Its contribution is essentially below two cycles deg^{-1}, where the two NPS differ.

3.2 *Comparison of Detection Performance between Single-Slice and Multislice Reading Modality, and Uniform and Anatomical Backgrounds for Human Observers*

Figure 6 presents the human observers d' for single- and multislice viewing modalities. As expected, d' increased with signal size. Detectability in the multislice reading modality was consistently higher than or equal to the single-slice reading modality, but no individual comparison revealed a statistically significant difference.

Figure 7 presents the performance of human observers for uniform and anatomical backgrounds. As expected, d increased with signal size. Detectability in uniform backgrounds was in all but one case consistently higher than in anatomical backgrounds, but no individual comparison demonstrated a statistically significant difference.

3.3 *Multiway Statistical Analysis*

The MRMC variance analysis shows a small but significant effect of the viewing condition when pooling all the data together (across signal size and anatomical background type) with the average AUC going from 0.787 to 0.820 ($p < 0.007$). This effect was no more significant when the results are stratified based on background type ($p = 0.197$ and $p = 0.282$ for uniform and anatomical backgrounds, respectively).

The same analysis could not be performed by pooling all the data together in order to compare the effect of the backgrounds, because the anatomic background cases are different from the uniform cases. If we just ignore this and assume that performance is independent in multislice, single-slice, uniform, and anatomic images, the calculation reveals that the difference in performance between anatomic and uniform background is not significant.

3.4 *Correlation between Human Observers and CHO*

Figure 8 shows the human and CHO detectability index. Human and model observers d' show a relatively linear relationship for all investigated channel types. The least performing model observer was the square channel CHO. As expected, all model observers overestimated human observers. We also calculated the Pearson rank correlation between human and model observers. All models investigated demonstrated a high rank correlation (rho = 1, p-value = 0) except for CHO with SQR channels in an anatomical background multislice viewing modality (rho = 0.5, p-value = 0.667).

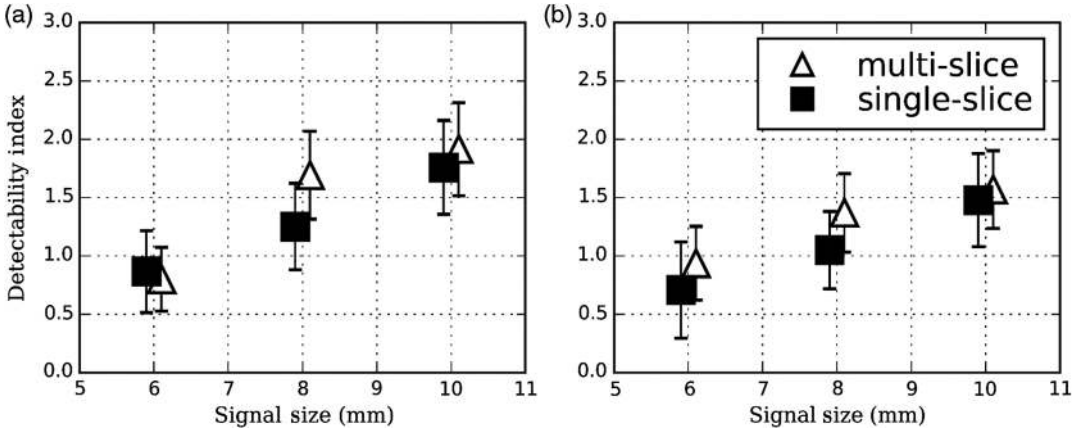

Fig. 6 Plot of d' for human observers comparing single- and multislice viewing mode for (a) uniform and (b) anatomical backgrounds. Error bars represent 95% interval confidence.

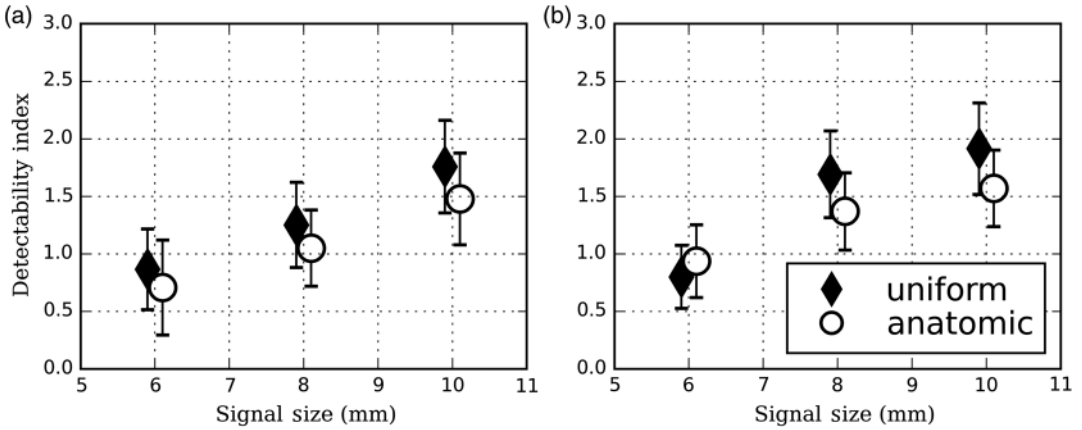

Fig. 7 Plot of d' for human observers comparing uniform and anatomic background for (a) single- and (b) multislice viewing mode. Error bars represent 95% interval confidence.

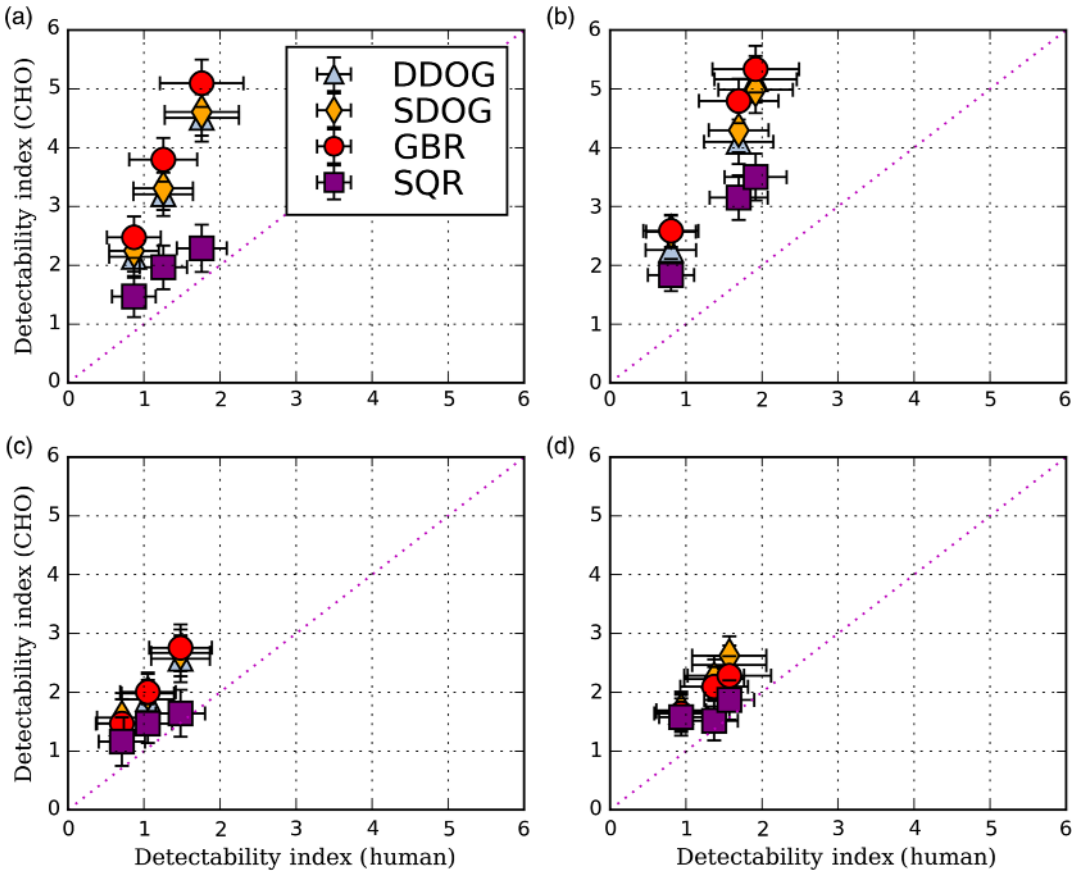

Fig. 8 Plot of d' for human observers versus d' for model observers in (a) single-slice reading modality with uniform background, (b) multislice reading modality with uniform background, (c) single-slice reading modality with anatomical background, and (d) multislice reading modality with anatomical background. Error bars represent 95% interval confidence.

4 Discussion

Figure 6 shows that human performance in a multislice reading modality is consistently higher than or equal to a single-slice viewing mode thus suggesting that humans take significant advantage of additional slices to detect low-contrast signals. This is statistically significant when the data are pooled altogether. So it appears that humans derive a small benefit from multislice data (about a 0.03 increase of AUC on average going from single-slice to multislice).

The results of our study suggest that detection in anatomical backgrounds lead to performances similar to those in uniform background. Although Fig. 7 shows a small tendency toward better performance for uniform than for anatomical backgrounds; our data could not show that this was significant.

The performances of human observers in both backgrounds are correlated, which suggests that human observers can at least partly process anatomical fluctuations present in liver CT images. At first, this seems surprising as the NPS of a liver texture is different from a uniform background, especially in the low-frequency domain of the signal. However, the low-frequency domain accounts for large structures in the liver parenchyma such as arteries or blood vessels. Moreover, the liver parenchyma texture is locally uniform as described by the NPS, and these structures appear as a hypersignal in contrast-enhanced abdominal CT and are contrasted relative to the background. Low-contrast targets can therefore be easily distinguished from such structures. This could explain why the anatomical part of the liver image noise has a weak effect on the detection of low-contrast signals.

The level of correlation for detection performance between CHO and human observers is high as suggested by Fig. 8. This result stands for any kind of reading modality and background type. However, according to the type of channels, the correlation might be different as illustrated by the different slopes in Fig. 8. CHO with D-DOG and S-DOG channels present a similar correlation with human observer as the radial spatial frequency profile functions for these types of channels is identical. Therefore, the number of channels, the channels' bandwidth, and the initial channels' spread parameter have marginal effects on DOG channels response for this kind of task. CHO with GBR channels are similar to CHOs with D-DOG or S-DOG channels. The use of GBR channels in our case was probably a bit of overkill since our signals are circularly symmetric and therefore do not require a large number of free parameters (orientation and phase) and channels. Finally, SQR channels had the poorest agreement among the channels investigated, thus demonstrating the lowest correlation with a human observer for any condition. This result was expected and shows that SQR channels should not be preferred to D-DOG, S-DOG, or GBR channels for this kind of task.[4,23]

The level of correlation between human and CHOs is consistent among the types of reading modality and therefore in line with previous results,[17] suggesting that modeling a single-slice viewing mode for low-contrast detection is sufficient for volumetric imaging assessment.

There were a number of limitations in this study. The first was the number of independent conditions investigated. Although we chose clinically representative dose levels and reconstruction algorithms, these parameters can influence image quality and we did not evaluate that aspects in this work. Another limitation was the small number of human observers. However, the good coherence between their performances gives us confidence in our results. Finally, we chose the NPS to quantify the statistical properties of the background because this quantity is widely accepted in the medical imaging community. However, the NPS is one of many features that could describe an image's texture; a thorough investigation of the differences between the textures of a uniform and an anatomical image could include other descriptors.[33]

5 Conclusion

For the assessment of image quality in low-contrast liver CT images, human observers have very similar performances in both multi- and single-slice viewing mode. As a consequence, it is preferable to use ssCHO as this model is computationally more tractable than mmCHO and has already been extensively validated by the medical imaging community.

However, using images from a homogeneous phantom can result in overestimating image quality as measured by CHOs, as their performance tends to be higher in uniform than anatomical backgrounds, while human observers have similar performances.

Disclosures

No conflicts of interest, financial or otherwise, are declared by the authors.

Acknowledgments

This work was supported by Grant SNF 320030_156032/1 and was based partly on scientific content previously reported in SPIE proceedings.

References

1. H. H. Barrett et al., "Model observers for assessment of image quality," *Proc. Natl. Acad. Sci. U. S. A.* **90**(21), 9758–9765 (1993).
2. H. C. Gifford et al., "A comparison of human and model observers in multislice LROC studies," *IEEE Trans. Med. Imaging* **24**(2), 160–169 (2005).
3. J. G. Brankov, "Evaluation of the channelized Hotelling observer with an internal-noise model in a train-test paradigm for cardiac SPECT defect detection," *Phys. Med. Biol.* **58**(20), 7159–7182 (2013).
4. C. Castella et al., "Mass detection in breast tomosynthesis and digital mammography: a model observer study," *Proc. SPIE* **7263**, 72630O (2009).
5. H.-W. Tseng et al., "Assessing image quality and dose reduction of a new x-ray computed tomography iterative reconstruction algorithm using model observers," *Med. Phys.* **41**(7), 071910 (2014).
6. J. Fan et al., "Evaluation of low contrast detectability performance using two-alternative forced choice method on computed tomography dose reduction algorithms," *Proc. SPIE* **8318**, 83181F (2012).
7. L. Yu et al., "Prediction of human observer performance in a 2-alternative forced choice low-contrast detection task using channelized Hotelling observer: impact of radiation dose and reconstruction algorithms," *Med. Phys.* **40**(4), 041908 (2013).
8. S. Leng et al., "Correlation between model observer and human observer performance in CT imaging when lesion location is uncertain," *Med. Phys.* **40**(8), 081908 (2013).
9. D. Racine et al., "Objective assessment of low contrast detectability in computed tomography with channelized Hotelling observer," *Phys. Med.* **32**(1), 76–83 (2016).
10. I. Hernandez-Giron et al., "Low contrast detectability performance of model observers based on CT phantom images: kVp influence," *Phys. Med.* **31**(7), 798–807 (2015).
11. I. Hernandez-Giron et al., "Automated assessment of low contrast sensitivity for CT systems using a model observer," *Med. Phys.* **38**(S1), S25–S35 (2011).
12. F. O. Bochud et al., "Estimation of the noisy component of anatomical backgrounds," *Med. Phys.* **26**(7), 1365–1370 (1999).
13. J. Solomon et al., "Comparison of low-contrast detectability between two CT reconstruction algorithms using voxel-based 3D printed textured phantoms: voxel-based 3D printed textured phantoms," *Med. Phys.* **43**(12), 6497–6506 (2016).
14. L. Platisa et al., "Channelized Hotelling observers for the assessment of volumetric imaging data sets," *J. Opt. Soc. Am. A* **28**(6), 1145–1163 (2011).
15. M. Chen et al., "Using the Hotelling observer on multislice and multiview simulated SPECT myocardial images," *IEEE Trans. Nucl. Sci.* **49**(3), 661–667 (2002).
16. A. Ba et al., "Anthropomorphic model observer performance in three-dimensional detection task for low-contrast computed tomography," *J. Med. Imaging* **3**(1), 011009 (2015).

17. L. Yu et al., "Correlation between a 2D channelized Hotelling observer and human observers in a low-contrast detection task with multislice reading in CT," *Med. Phys.* **44**(8), 3990–3999 (2017).
18. H. Shin et al., "Insertion of virtual pulmonary nodules in CT data of the chest: development of a software tool," *Eur. Radiol.* **16**(11), 2567–2574 (2006).
19. J. Solomon and E. Samei, "A generic framework to simulate realistic lung, liver and renal pathologies in CT imaging," *Phys. Med. Biol.* **59**(21), 6637–6657 (2014).
20. B. D. Gallas and H. H. Barrett, "Validating the use of channels to estimate the ideal linear observer," *J. Opt. Soc. Am. A* **20**(9), 1725–1738 (2003).
21. K. J. Myers and H. H. Barrett, "Addition of a channel mechanism to the ideal-observer model," *J. Opt. Soc. Am. A* **4**(12), 2447–2457 (1987).
22. X. He and S. Park, "Model observers in medical imaging research," *Theranostics* **3**(10), 774–786 (2013).
23. 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**(3), 473–488 (2001).
24. J. A. Swets et al., "Assessment of diagnostic technologies," *Science* **205**(4408), 753–759 (1979).
25. The University of Iowa, "Dpt of Radiology, Medical Image Perception Laboratory, Software OR-DBM-MRMC," 2018, http://perception.radiology.uiowa.edu (25 April 2019).
26. D. D. Dorfman, K. S. Berbaum, and C. E. Metz, "Receiver operating characteristic rating analysis: generalization to the population of readers and patients with the jackknife method," *Invest. Radiol.* **27**(9), 723–731 (1992).
27. N. A. Obuchowski and H. E. Rockette, "Hypothesis testing of diagnostic accuracy for multiple readers and multiple tests an anova approach with dependent observations," *Commun. Stat. Simul. Comput.* **24**(2), 285–308 (1995).
28. S. L. Hillis et al., "A comparison of the Dorfman–Berbaum–Metz and Obuchowski–Rockette methods for receiver operating characteristic (ROC) data," *Stat. Med.* **24**(10), 1579–1607 (2005).
29. S. L. Hillis, "A comparison of denominator degrees of freedom methods for multiple observer ROC analysis," *Stat. Med.* **26**(3), 596–619 (2007).
30. S. L. Hillis, K. S. Berbaum, and C. E. Metz, "Recent developments in the Dorfman–Berbaum–Metz procedure for multireader ROC study analysis," *Acad. Radiol.* **15**(5), 647–661 (2008).
31. A. Wunderlich et al., "Exact confidence intervals for channelized hotelling observer performance in image quality studies," *IEEE Trans. Med. Imaging* **34**(2), 453–464 (2015).
32. H. H. Barrett, *Foundations of Image Science*, Wiley-Interscience, Hoboken, New Jersey (2004).
33. R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst. Man Cybern.* **SMC-3**(6), 610–621 (1973).

Biographies of the authors are not available.