Published in final edited form as:

Proc SPIE Int Soc Opt Eng. 2016 February 27; 9783: . doi:10.1117/12.2216962.

Predicting detection performance with model observers: Fourier domain or spatial domain?

Baiyu Chen, Lifeng Yu, Shuai Leng, James Kofler, Christopher Favazza, Thomas Vrieze, and Cynthia McCollough

Department of Radiology, Mayo Clinic, Rochester, MN 55905

Abstract

The use of Fourier domain model observer is challenged by iterative reconstruction (IR), because IR algorithms are nonlinear and IR images have noise texture different from that of FBP. A modified Fourier domain model observer, which incorporates nonlinear noise and resolution properties, has been proposed for IR and needs to be validated with human detection performance. On the other hand, the spatial domain model observer is theoretically applicable to IR, but more computationally intensive than the Fourier domain method. The purpose of this study is to compare the modified Fourier domain model observer to the spatial domain model observer with both FBP and IR images, using human detection performance as the gold standard. A phantom with inserts of various low contrast levels and sizes was repeatedly scanned 100 times on a third-generation, dual-source CT scanner at 5 dose levels and reconstructed using FBP and IR algorithms. The human detection performance of the inserts was measured via a 2-alternative-forced-choice (2AFC) test. In addition, two model observer performances were calculated, including a Fourier domain non-prewhitening model observer and a spatial domain channelized Hotelling observer. The performance of these two model observers was compared in terms of how well they correlated with human observer performance. Our results demonstrated that the spatial domain model observer correlated well with human observers across various dose levels, object contrast levels, and object sizes. The Fourier domain observer correlated well with human observers using FBP images, but overestimated the detection performance using IR images.

Keywords

Computed tomography (CT); iterative reconstruction (IR); model observer; non-prewhitening; detectability index (d'); channelized Hotelling observer (CHO)

I. INTRODUCTION

Mathematical model observers have been used as an alternative to human observers to assess the performance of detection tasks. Previous studies have shown that Fourier domain model observers correlates well with human performance for detection task using filtered backprojection (FBP) images [1]. However, the Fourier domain model observer is challenged by iterative reconstruction (IR), because IR algorithms are nonlinear and IR images have noise texture different from that of FBP. Recently, a modified Fourier domain model observer with nonlinear noise and resolution properties incorporated has been proposed to assess IR algorithms [2, 3], which needs to be validated with human

performance. On the other hand, the spatial domain model observers are theoretically applicable to IR, but more computationally intensive than Fourier domain model observers. Therefore, the purpose of this study is to compare the modified Fourier domain model observer to the spatial domain model observer under both FBP and IR images, using human detection performance as the gold standard.

II. METHODS

A cylindrical phantom (Helical CT phantom, CIRS, Norfolk, VA) was scanned on a third-generation, dual-source CT scanner (SOMATOM Force, Siemens Healthcare, Forchheim, Germany) at 5 dose levels (120 kV and 25, 50, 100, 200, and 400 effective mAs, which corresponded to CTDI_{vol} of 1.67, 3.35, 6.70, 13.39, and 26.79 mGy), respectively. For each dose level, scans were repeated 100 times. Images were reconstructed using 1 mm thickness, 0.5 mm increment, 180 mm field of view, and 2 reconstruction algorithms (Weighted FBP [4] and ADMIRE, the state-of-art Siemens IR algorithm). The phantom has inserts of 3 contrast levels (−7, −14, and −21 HU) and 7 sizes (with diameters of 2.4, 3.2, 4.0, 4.8, 6.3, 9.5 and 10.0 mm), as shown in Figure 1(a). The 10 mm insert is cylindrical, which was used for the task transfer function (TTF) measurement, which will be described below. The rest of the inserts are spherical, which were used for the human observer study and the calculation of spatial domain model observer described below. The phantom also has a uniform section (Figure 1(b)), which was used for the NPS measurement described below.

The human observer performance in detecting the low-contrast spherical objects was measured via a 2-alternative-forced-choice (2AFC) test. The two alternatives consisted of an signal-present region of interest (ROI) and an signal-absent ROI: The signal-present ROI was cropped from the slice through the center of the spherical object and the signal-absent ROI was cropped from the uniform region next to the object. The two alternatives were presented side-by-side to the observers (four board certified medical physicists) in a randomized and blinded fashion, with the observers instructed to choose the ROI that they thought contained the signal. A template of the low-contrast object (noise-free signal, which was approximated from the average of 100 repeats acquired at 400 effective mAs) was shown to the observers together with the two alternatives, such that the 2AFC was conducted under a signal known exactly scenario. Each 2AFC test contained 100 trials, which corresponded to the 100 repeated scans of the phantom. The percent correct of the choices in the 100 trials was calculated as an estimate of the area under the receiver operating characteristic (ROC) curve, A_z .

The Fourier domain model observer was calculated using a non-prewhitening matched filter model with eye filter and internal noise (NPWEi). This particular model was chosen because it has been shown to correlate well with human detection performance using FBP images [1, 5]. In this study, as an attempt to apply NPWEi to IR images as well, the NPWEi models image resolution using tasktransfer function (TTF) instead of modulation transfer function (MTF), as detailed below. The NPWEi was calculated as

$$d'_{NPWEi}^2 = \frac{\left[\int_0^{2\pi} d\theta \int_0^{Nyquist} |W(u)|^2 \cdot TTF^2(u) \cdot E^2(u) \cdot u \, du \right]^2}{\int_0^{2\pi} d\theta \int_0^{Nyquist} \left(|W(u)|^2 \cdot TTF^2(u) \cdot E^4(u) \cdot NPS(u) + N_i \right) u \, du},$$

where d' is the detectability index, TTF is the task transfer function, NPS is the noise power spectrum, W is the task function, E is the eye filter, N_i is the internal noise, and u and θ are the radial and angular spatial frequencies, respectively [1, 6]. The TTF models the resolution, which is an extension of the conventional MTF to accommodate the dependency of resolution on object contrast with nonlinear reconstruction algorithms [7–9]. The TTF was calculated from the edge spread function, which was measured from the phantom images using the circular edges around the 10 mm insert for each contrast level separately. NPS models the noise, which was measured from the phantom images using ROIs with 64×64 pixels in the uniform regions. Details about the TTF and NPS calculation can be found in a previous study [8]. The task function mathematically modeled the detection task as the Fourier transform of the noise-free insert. E is the eye filter that models the observer visual response to various frequencies. The eye filter was calculated using the formula $E(f) = f \cdot e^{-cf}$, where c is a constant depending on the viewing distance. This study assumed a viewing distance of 50 cm ($c=2.18$). N_i is the internal noise that models the human inefficiency caused by cognitive inconsistency. The internal noise was modeled as 60% of

the NPS[6]. A_z values were estimated from d' as $A_z = \frac{1}{2} \left(1 + \operatorname{erf} \left(\frac{d'}{2} \right) \right)$ under the assumptions that the decision variables for signal-absent and signal-present images follow normal distribution and have equal variance [6].

The spatial domain model observer was calculated using a channelized Hotelling observer (CHO) with Gabor filters and internal noise. This particular model was chosen because previous studies have shown good correlation between this model and human detection performance using both FBP and IR images [10, 11]. The CHO was calculated in terms of

the decision variable λ , as $\lambda_{CHO} = \omega_{CHO}^t \cdot g_c = \sum_{m=1}^M \omega_{CHOm} g_{cm}$, where ω_{CHO} is the template, g_c is the channel output of the images, and M is the number of channels [10, 11]. The template was calculated as $\omega_{CHO} = S_c^{-1} [\bar{g}_{sc} - \bar{g}_{bc}]$, where $\bar{g}_{sc}$ and $\bar{g}_{bc}$ are the signal-present and signal-absent channel output means, respectively, and S_c is the intraclass channel scatter matrix. The channel filters used in this study are Gabor filters [10]. In this study 4 channel bands ([1/8, 1/4], [1/16, 1/8], [1/32, 1/16], and [1/64, 1/32]), 5 orientations (0, $2\pi/5$, $4\pi/5$, $6\pi/5$, and $8\pi/5$), and 2 phases were used, leading to a total of 40 channels. Such selection of channels ensured both accuracy and precision of the λ calculation, as demonstrated in a previous study [12]. CHO used internal noise to model the intra-variation of human observer decisions. In this study, we added the internal noise to λ_{CHO} , and modeled it as a normally distributed random variable with a zero mean and a standard deviation proportional to the standard deviation of the decision variable in signal-absent images. The ratio between the standard deviation of internal noise and the standard deviation of the decision variable was

determined through a calibration procedure using the images of 4 mm, -14 HU insert and the human observer performance. By comparing the λ_{CHO} values for the two alternatives in the 2AFC test, the alternative with a higher λ_{CHO} value was considered to be the signal-present alternative. The decision was compared to the truth for all 100 trials, and the percent correct was calculated as A_z .

III. RESULTS

The A_z values calculated from the model observers and the A_z values measured from the human observers are compared in Figure 2. With FBP images, both NPWEi and CHO correlated well with the performance of human observers across various dose levels, object sizes, and object contrasts, except for small object (3.2 mm). With IR images, CHO maintained good correlation (mostly within the errorbar of human observers), but NPWEi overestimated the detection performance in general (CHO in general have lower RMSE).

The A_z values of the model observers were also compared against the A_z values of the human observers, as shown in Figure 3. If the model observers agree well with the performance of human observers, the data should align along $y=x$ (the dashed line). For FBP, both Fourier domain and spatial domain model observers align along the dashed line, except for small object. For IR, spatial domain model observer is closer to the dashed line than Fourier domain observer. The Fourier domain observer is mostly above the line, which again shows that it overestimated the detection performance. This is also shown in the lower R^2 values of Fourier domain observer.

IV. DISCUSSION

This study demonstrated that, while Fourier domain and spatial domain model observers both correlated well with human observers for FBP images, the Fourier domain method tends to overestimate the detection performance of IR images, even though the contrast-dependent spatial resolution of IR has been taken into account by TTF. This is possibly because 1) FBP images and IR images have different noise correlations, and 2) the non-prewhitening model observer is suboptimal in correlated noise [13]. In particular, when the internal noise is trained to account for the human inefficiency caused by the noise correlation of FBP, the same form of internal noise may not accurately account for the human inefficiency caused by the noise correlation of IR images.

The overestimation of detection performance needs to be taken into consideration if Fourier domain model observer is used to assess IR images, especially if it is used to estimate the dose reduction potential of IR.

V. CONCLUSION

The spatial domain model observer correlated well with human observers across various dose levels, object contrast levels, and object sizes for both FBP and IR algorithms. The Fourier domain observer correlated well with human observers using FBP images, but overestimated the detection performance using IR images.

Acknowledgments

Research reported in this publication was supported by the National Institute of Biomedical Imaging and Bioengineering of the National Institutes of Health under award number U01EB017185 and R01EB017095. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

1. Gang GJ, Lee J, Stayman JW, et al. Analysis of Fourier-domain task-based detectability index in tomosynthesis and cone-beam CT in relation to human observer performance. *Medical Physics*. 2011; 38(4):1754–1768. [PubMed: 21626910]
2. Chen B, Ramirez Giraldo JC, Solomon J, et al. Evaluating iterative reconstruction performance in computed tomography. *Medical physics*. 2014; 41(12):121913. [PubMed: 25471973]
3. Solomon J, Wilson J, Samei E. Characteristic image quality of a third generation dual-source MDCT scanner: Noise, resolution, and detectability. *Medical Physics*. 2015; 42(8):4941. [PubMed: 26233220]
4. Stierstorfer K, Rauscher A, Boese J, et al. Weighted FBP—a simple approximate 3D FBP algorithm for multislice spiral CT with good dose usage for arbitrary pitch. *Physics in medicine and biology*. 2004; 49(11):2209. [PubMed: 15248573]
5. Richard S, Siewerdsen JH. Comparison of model and human observer performance for detection and discrimination tasks using dual-energy x-ray images. *Medical Physics*. 2008; 35(11):5043–53. [PubMed: 19070238]
6. Burgess AE. Visual perception studies and observer models in medical imaging. *Semin Nucl Med*. 2011; 41(6):419–36. [PubMed: 21978445]
7. Richard S, Husarik DB, Yadava G, et al. Towards task-based assessment of CT performance: system and object MTF across different reconstruction algorithms. *Medical physics*. 2012; 39(7):4115–22. [PubMed: 22830744]
8. Chen B, Christianson O, Wilson JM, et al. Assessment of volumetric noise and resolution performance for linear and nonlinear CT reconstruction methods. *Medical Physics*. 2014; 41(7):071909. [PubMed: 24989387]
9. Yu L, Vrieze TJ, Leng S, et al. Technical Note: Measuring contrast- and noise-dependent spatial resolution of an iterative reconstruction method in CT using ensemble averaging. *Medical physics*. 2015; 42(5):2261–2267. [PubMed: 25979020]
10. Yu L, Leng S, Chen L, 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. *Medical Physics*. 2013; 40(4):041908. [PubMed: 23556902]
11. Leng S, Yu L, Zhang Y, et al. Correlation between model observer and human observer performance in CT imaging when lesion location is uncertain. *Medical Physics*. 2013; 40(8):081908. [PubMed: 23927322]
12. Ma, C.; Yu, L.; Chen, B., et al. Impact of number of repeated scans on model observer performance for a low-contrast detection task in CT. Orlando, FL: 2015.
13. Eckstein, M.; Abbey, C.; Buchod, F. A practical guide to model observers for visual detection in synthetic and natural noisy images. SPIE Press; 2009. p. 10

Figure 1.

(a) The CIRS helical CT phantom with the slice across the center of the spherical inserts shown. (b) The CIRS helical CT phantom with the slice across the uniform region shown. Both images were averaged from 100 repeated scans to reduce image noise for illustration purpose.

Figure 2.

The A_z values calculated from Fourier-domain model observer, spatial domain model observer, and human observers for FBP and IR images.

Figure 3.

Comparison between the A_z values calculated from model observers and the A_z values measured from human observers for (a) FBP images and (b) IR images.