Published in final edited form as:

Proc SPIE Int Soc Opt Eng. 2015 February 21; 9416: . doi:10.1117/12.2082836.

Impact of Number of Repeated Scans on Model Observer Performance for a Low-contrast Detection Task in CT

Chi Ma, Lifeng Yu, Baiyu Chen, Thomas Vrieze, Shuai Leng, and Cynthia McCollough

Department of Radiology, Mayo Clinic, Rochester, MN

Abstract

Channelized Hotelling observer (CHO) has been validated against human observers for detection/classification tasks in clinical CT and shows encouraging correlations. However, the goodness of correlations depends on the number of repeated scans used in CHO to estimate the template and covariance matrices. The purpose of this study is to investigate how the number of repeated scans affects the CHO performance in predicting human observers. A phantom containing 21 low-contrast objects (3 contrast levels and 7 sizes) was scanned on a 128-slice CT scanner at three dose levels. Each scan was repeated 100 times. Images were reconstructed using a filtered-backprojection kernel and a commercial iterative reconstruction method. For each dose level and reconstruction setting, the low-contrast detectability, quantified with the area under receiver operating characteristic curve (Az), was calculated using a previously validated CHO. To determine the dependency of CHO performance on the number of repeated scans, the Az value was calculated for each object and dose/reconstruction setting using all 100 repeated scans. The Az values were also calculated using randomly selected subsets of the scans (from 10 to 90 scans with an increment of 10 scans). Using the Az from the 100 scans as the reference, the accuracy of Az from a smaller number of scans was determined. The minimum necessary number of scans was subsequently derived. For the studied signal-known-exactly detection task, results demonstrated that, the minimal number of scans required to accurately predict human observer performance depends on dose level, object size and contrast level, and channel filters.

Keywords

Computed tomography (CT); Task-based image quality assessment; Model observer; Channelized Hotelling observer (CHO); Radiation dose reduction

1. Purpose

Task-based image quality metrics using model observers were proposed to assess image quality in clinical CT. In our previous works, we demonstrated that the performance of a channelized Hotelling observer (CHO) is highly correlated with human observer performance in several phantom-based detection/classification tasks [1-3]. For each of these studies, scans at each condition were repeated a large number of times to obtain a reasonable estimate of the template and covariance matrices for precise prediction. The number of repeated scans was chosen as a result of balance between accuracy and labor intensity.

Theoretically, the variance and bias of the CHO predicted figure of merit (FOM) can be reduced to be arbitrarily small by increasing the number of repeated scans[4]. For computer simulation, this can be readily achieved by simply generating a large number of simulated images, given sufficient computing time and power. In clinical CT image quality assessment, however, due to labor and machine usage limitations on scan acquisitions, it becomes more important to quantify the error associated with limited but reasonable number of repeated scans. To determine this associated error, we applied a validated model observer to predict the low-contrast detectability for both a filtered-backprojection (FBP) algorithm and an iterative reconstruction (IR) method on a signal-known exactly detection task for various contrast levels, object sizes, and dose levels. Using the predicted performance from 100 scans as the reference, the number of repeated scans that is required to yield accurate prediction with CHO, which is important information when designing experiments to perform task-based image quality assessment using model observers in clinical CT.

2. Methods

2.1 Experimental setup

A cylindrical phantom (Helical CT Phantom, CIRS Inc.) was scanned on a dual-source 128-slice CT scanner (Definition Flash, Siemens Healthcare). Only one of the two x-ray sources was used in the scan. The phantom had a diameter of 18 cm and a length of 4 cm, which contains three groups of low-contrast objects at different contrast levels (nominal values: -5, -10, and -20 HU below the liver equivalent background). Each group has 7 disks with diameters of 10, 9.5, 6.3, 4.8, 4, 3.2, and 2.4 mm. A cross section of the low-contrast module is shown in Figure 1. The scanning parameters are as follows: 120 kV, 64x0.6 mm detector collimation with the z-flying focal spot, 0.5 second rotation time, and helical pitch 0.8. Three mA settings were used: 240, 120, and 60 effective mAs (mAs/pitch), corresponding to a volume CTDI (CTDI_{vol}) of 16, 8, and 4 mGy CTDI_{vol}, respectively. The CARE Dose4D was off. Each dose level was scanned 100 times to generate the ensemble images. Images were reconstructed at 1 mm slice thickness using 2 different configurations, B40 and I40-3. The B40 is an FBP kernel. The I40-3 is an IR kernel (SAFIRE, Sinogram Affirmed Iterative Reconstruction, Siemens Healthcare) with a strength setting of 3 (on a scale of 1-5, with 1 leads to the least amount of noise reduction and 5 the most). Figure 2 shows example images reconstructed by FBP and IR methods acquired from one of the repeated scans at three dose levels.

2.2 model Observer Study

The data set was analyzed using a CHO model observer [5]. The test variables in CHO can be expressed as $\lambda = \omega_{\text{CHO}}^t g_c$ [6], where g_c is the channel output of the image, and ω_{CHO}

is the template, which can be obtained by $\omega_{\text{CHO}} = S_c^{-1} [\bar{g}_{sc} - \bar{g}_{bc}]$, where

$S_c = \frac{1}{2} [K_{sc} + K_{bc}]$ is the intraclass channel scatter matrix and $\bar{g}_{sc} = U^T \bar{g}_s$ and $\bar{g}_{bc} = U^T \bar{g}_b$ are the mean of channel output for images with and without signal. K_{sc} and K_{bc} are related to the image covariance matrices of K_s and K_b by $K_{sc} = U^T K_s U$, $K_{bc} = U^T K_b U$; and U is the matrix representation of the channel profiles. In this study, we considered both

Gabor and Laguerre-Gauss (LG) channel filters in CHO [7-9]. Only the results from LG will be shown below. Internal noise was added to the test variables. Area under the ROC curve (Az) was calculated based upon the test variables at all 100 realizations and the ground truth of signal presence and absence, at each dose level and object size using a non-parametric approach.

2.3. Determining the minimum number of scans

The covariance matrices and template in CHO were obtained using 100 repeated scans, which was chosen as a result of balance between accuracy and labor intensity. The Az predicted by the CHO from the 100 repeated scans will inherently have a positive bias, because previous theoretical analysis showed that a positive bias in Az exists for a resubstitution method in which the CHO was trained and tested on the same finite

samples[10]. The bias is proportional to $\frac{1}{N_t}$, where N_t is the number of training samples.

Although the absolute value of such a bias is difficult to estimate in realistic CT scans where no ground truth can be established, our previous studies showed that the model observer calculated using 100 repeated scans can achieve excellent agreement with human observers. Therefore, in this study, we will use the Az value from all 100 scans as the reference to determine the number of scans that is minimally required to yield accurate prediction of the human observer performance. We calculated the Az value for each of the 21 objects and 6 dose/reconstruction settings by using a subset of the scans, from 10 to 90. The subsets of the scans were randomly selected from the 100-image pool for 50 independent realizations. Mean Az values and standard deviations were calculated for each subset of the scans from 10 to 90. The minimally required number of repeated scans was determined to meet the following criteria: At and beyond this number of repeated scans, the difference of Az compared to the reference is within 2%.

3. Results

Figure 3 shows Az values measured by CHO for all low-contrast objects at all three mAs levels, with 100 repeated scans. At high dose levels, e.g. 240 mAs, the larger lesions (10 mm and 9.5 mm) at all three contrast levels were almost perfectly identified. Az values decreased with decreasing radiation dose, contrast and object size, and became close to 0.5 at 60 mAs for smaller objects (3.2 mm and 2.4 mm) with 7 HU contrast, indicating it was almost a random guess at this dose level.

Figure 4 presents the Az value changes with the number of repeated scans used in the CHO calculation. Results from the 4 mm contrast object and the FBP reconstruction were used to demonstrate the performance differences at three different lesion contrasts. Using the Az from all 100 scans as the reference, it can be observed the accuracy of the low-contrast detectability increases with the increased number of scans. It can also be observed from the error-bar that the Az values tend to be stabilized when number of scans approaches 100. The minimal number of repeated scans required to achieve less than 2% error in Az for three object sizes (6.3, 4.0, and 3.2 mm) and for both FBP and IR is listed in Table 1. It appeared that more scans are required to yield adequate accuracy for smaller objects at lower contrast.

With the decrease of the radiation dose, the minimally required number of scans also increased. A total of 60 repeated scans are sufficient for accurate prediction of detectability even for the object with a 3.2 mm diameter.

4. Conclusions/New and Breakthrough Work

We performed an experimental study to investigate the impact of the number of repeated scans on the performance of model observers; the task is to predict human observer performance for a lesion-known-exactly phantom-based detection task. We determined the minimum number of scans required to accurately predict the performance of human observer for both FBP and IR methods at different dose levels. This study provides important insights into the experimental design of task-based image quality assessments in clinical CT using model observers. Theoretical analysis of the relation between variance/bias of CHO and the number of repeated scans are underway.

Acknowledgements

This paper has not been published elsewhere. The authors thank Dr. Matthew Kupinski for his helpful discussions on model observers. The project was supported by NIH Grant Number R01 EB017095.

Reference

- [1]. 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]
- [2]. 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]
- [3]. Zhang Y, Leng S, Yu L, et al. Correlation between human and model observer performance for discrimination task in CT. *Physics in medicine and biology*. 2014; 59(13):3389–404. [PubMed: 24875060]
- [4]. Barrett, HH.; Myers, KJ. *Foundations of image science*. Wiley-Interscience; 2004. p. 971-973.
- [5]. Beutel, J. *Handbook of medical imaging: Physics and psychophysics*. Spie Press; 2000.
- [6]. Barrett HH, Yao J, Rolland JP, et al. Model observers for assessment of image quality. *Proc Natl Acad Sci U S A*. 1993; 90(21):9758–65. [PubMed: 8234311]
- [7]. Eckstein M, Bartroff J, Abbey C, et al. Automated computer evaluation and optimization of image compression of x-ray coronary angiograms for signal known exactly detection tasks. *Optics Express*. 2003; 11(5):460–475. [PubMed: 19461753]
- [8]. Zhang Y, Abbey CK, Eckstein MP. Adaptive detection mechanisms in globally statistically nonstationary-oriented noise. *J Opt Soc Am A Opt Image Sci Vis*. 2006; 23(7):1549–58. [PubMed: 16783416]
- [9]. Chawla AS, Samei E, Saunders R, et al. Effect of dose reduction on the detection of mammographic lesions: a mathematical observer model analysis. *Medical physics*. 2007; 34(8): 3385–98. [PubMed: 17879801]
- [10]. Wagner RF, Chan H-P, Sahiner B, et al. Finite-sample effects and resampling plans: applications to linear classifiers in computer-aided diagnosis. 3034:467–477.

Figure 1.

A phantom with three groups of low-contrast inserts (21, 14, and 7 HU), each with 7 different sizes with diameters of 10, 9.5, 6.3, 4.8, 4, 3.2, and 2.4 mm.

Figure 2.

Sample CT images of low-contrast inserts (21, 14, and 7 HU) at different dose levels with FBP and IR reconstructions.

Figure 3.

Az value measured by CHO for each low-contrast object at three different mAs levels, with both FBP and IR reconstruction.

Figure 4. Az value changes with the number of repeated scans that were used in the CHO calculation, for the 4 mm contrast object and the FBP reconstruction.

Table 1

Minimal number of repeated scans required to be within 2% accuracy using the Az from all 100 scans as the reference.

			6.3 mm			4.0 mm			3.2 mm			
FBP	240mAs		120mAs	60mAs	240mAs		120mAs	60mAs	240mAs		120mAs	60mAs
	21HU	10	10	10	21HU	10	50	50	21HU	50	50	60
	14HU	30	40	40	14HU	50	50	50	14HU	50	50	60
	7HU	40	50	50	7HU	50	50	60	7HU	50	50	60
IR	240mAs		120mAs	60mAs	240mAs		120mAs	60mAs	240mAs		120mAs	60mAs
	21HU	10	10	10	21HU	10	50	50	21HU	50	50	50
	14HU	20	40	50	14HU	50	60	50	14HU	50	50	50
	7HU	50	50	60	7HU	50	50	60	7HU	50	50	60