Published in final edited form as:
Proc SPIE Int Soc Opt Eng. 2018 February ; 10577: . doi:10.1117/12.2294955.

Correlation between model observers in uniform background and human observers in patient liver background for a low-contrast detection task in CT

Hao Gong, **Lifeng Yu**, **Shuai Leng**, **Samantha Dilger**, **Wei Zhou**, **Liqiang Ren**, and **Cynthia H McCollough**[*]
Department of Radiology, Mayo Clinic, Rochester, 55901, MN

Abstract

Channelized Hotelling observer (CHO) has demonstrated strong correlation with human observer (HO) in both single-slice viewing mode and multi-slice viewing mode in low-contrast detection tasks with uniform background. However, it remains unknown if the simplest single-slice CHO in uniform background can be used to predict human observer performance in more realistic tasks that involve patient anatomical background and multi-slice viewing mode. In this study, we aim to investigate the correlation between CHO in a uniform water background and human observer performance at a multi-slice viewing mode on patient liver background for a low-contrast lesion detection task. The human observer study was performed on CT images from 7 abdominal CT exams. A noise insertion tool was employed to synthesize CT scans at two additional dose levels. A validated lesion insertion tool was used to numerically insert metastatic liver lesions of various sizes and contrasts into both phantom and patient images. We selected 12 conditions out of 72 possible experimental conditions to evaluate the correlation at various radiation doses, lesion sizes, lesion contrasts and reconstruction algorithms. CHO with both single and multi-slice viewing modes were strongly correlated with HO. The corresponding Pearson's correlation coefficient was 0.982 (with 95% confidence interval (CI) [0.936, 0.995]) and 0.989 (with 95% CI of [0.960, 0.997]) in multi-slice and single-slice viewing modes, respectively. Therefore, this study demonstrated the potential to use the simplest single-slice CHO to assess image quality for more realistic clinically relevant CT detection tasks.

Keywords

Channelized Hotelling observer; Computed tomography (CT); Model observer; Multi-slice viewing; Observer study

1. INTRODUCTION

Diagnostic image quality assessment is the basis for radiation dose and scanning protocol optimization in clinical CT. Task-based approaches using mathematical model-observers (MO) have become popular to characterize CT image quality, since MO provides objective

[*] McCollough.Cynthia@mayo.edu.

and quantitative image quality assessment that have been demonstrated to well-correlate with human observer (HO) performance [1]. Fourier based MO relies on the assumption of linear shift invariance and noise stationarity [2, 3]. In contrary, spatial domain channelized Hotelling observer (CHO) is more appropriate for evaluating CT images reconstructed by both filtered-back-projection (FBP) and iterative reconstruction (IR) algorithms [4–6]. The performance of CHO has been successfully validated in previous studies using low contrast object detection, localization, and classification tasks [7–9].

In most previous studies, MO and HO detection tasks were performed on 2D lesions in 2D static images (i.e. single-slice viewing mode). However, radiologists usually scroll through multiple consecutive slices (i.e. multi-slice viewing mode) to identify 3D lesions in routine practice. Some previous studies proposed methods to integrate multi-slice viewing mode into MO (MS-MO) [10, 11]. A few recent studies applied these MS-MOs to low-contrast object detection tasks with uniform water phantom images or simulated CT images, and investigated the corresponding correlation with HO [12, 13]. It was reported in [13] that MS-MO was highly correlated with HO performance in multi-slice viewing mode and SS-MO yielded fairly similar association.

Although integrating multi-slice scrolling to MO is one step closer to modeling realistic tasks, these studies were performed with uniform water background. It remains unknown if the correlation is still valid in more realistic tasks involving multi-slice scrolling with patient anatomical background. In addition, since the SS-MO in uniform background is easier to implement in practice, it would be ideal to use SS-MO in uniform background to characterize CT image quality if its correlation with the HO in more realistic tasks can be demonstrated.

In this study, we aim to investigate the correlation between CHO in a uniform water background (both SS-MO and MS-MO) and human observer performance at a multi-slice viewing mode on patient liver background for a low-contrast lesion detection task. We employed the MS-MO used in [13], and selected a simple 2-alternative forced choice (2AFC) detection task of realistic liver metastatic lesion with various conditions of size and contrast.

2. METHODS

2.1 Data preparation

An abdomen-sized water phantom was used in MO studies. The anterior-posterior (AP) and the lateral (LAT) dimensions of the phantom were 26 cm and 35 cm, respectively. It was scanned on a single-source 128-slice CT scanner (SOMATOM Definition AS+, Siemens Healthcare, Germany) following a routine abdominal CT protocol used at our institute. X-ray tube voltage was fixed at 120 kV. Data collection field-of-view (FOV) was 500 mm in diameter and scan range was 118 mm along longitudinal direction. The corresponding volume CTDI ($CTDI_{vol}$) was 13.5 mGy. We carried out 100 repeated scans for water phantom, and then used a previously validated noise insertion tool [14] to realize image noise levels at a half and a quarter of routine radiation dose.

For human observer studies, we retrospectively collected abdominal CT scans of 7 patients (without pathologically proven hepatic metastases), using routine scanning protocols. The routine scanning protocols were similar to that used in the CT scans of water phantom. A major difference was that automatic exposure control (AEC) system (CARE Dose4D, Siemens Healthcare, Germany) was turn on during abdominal CT scans. The mean $CTDI_{vol}$ of patient scans was 12.6 mGy. The same noise insertion tool was used to synthesize CT scans with a half and a quarter of routine radiation dose.

CT images were reconstructed on an off-line Siemens research image reconstruction workstation, using analytic and iterative reconstruction (IR) algorithms. For analytic reconstruction algorithm, we used weighted filtered back-projection (WFBP) with a medium sharp (B30f) kernel. For IR algorithm, we used SAFIRE (Sinogram Affirmed Iterative Reconstruction, Siemens Healthcare, Germany) with I30f kernel (comparable to B30 kernel) and a strength level of 2. Both B30f and I30f are clinically-approved reconstruction kernels. The diameter of display FOV was 380 mm and CT images were reconstructed with 512 × 512 display matrix. Slice thickness and increment were both fixed at 3 mm.

To generate 3D signals for detection tasks, the volumetric CT images of a real metastatic liver lesion were used. Briefly, we extracted the lesion from CT images of a patient with proved metastatic liver lesion, using a 3D mask manually drawn by a radiologist. Lesion contrast and lesion size were modified numerically to generate multiple experimental conditions. Lesion contrast was varied at 15 HU, 20 HU, and 25 HU. Lesion size was varied at 5 mm, 7 mm, 9 mm, and 11 mm. We used a validated MATLAB script based software toolkit [15] to insert lesions at randomly selected locations in CT images of phantom and patients. Lesion locations were validated by a senior radiologist. Therefore, we created an ensemble dataset of 12 experimental conditions (Table 1) to enable a comprehensive evaluation on the correlation between MO and HO on different image noise level, lesion attributes, and reconstruction algorithms. A volume-of-interest (VOI) centered at each lesion was extracted (Figure 1) for each trial of 2-alternative force choice (2AFC) study. Each VOI consists of 5 axial slices and each axial slice was 60 mm × 60 mm. Background images were extracted from the CT images without the inserted lesions. In each experimental condition, we generated 68 trials of lesion-present and background VOIs.

2.2 Model observer

We employed the MS-MO from the reference [13]. The formula of MS-MO is summarized as follows. First, slice-wise 2D CHO templates were created for all slices in the VOIs of training datasets:

$$\omega_{CHO,\,i} = S_{ci}^{-1}[\overline{g_{sci}} - \overline{g_{bci}}], \quad i = 1, 2, \ldots, N \quad (1)$$

Where $\omega_{CHO,i}$ denotes the 2D CHO template for the i^{th} slice, $\overline{g_{sci}}$ and $\overline{g_{bci}}$ are the averaged channel outputs of the i^{th} slices from signal-present and background VOIs, respectively, S_{ci} denotes intra-class scatter matrix (i.e. the averaged covariance matrix of the channel outputs of signal-present and background images), and N is the number of slices in each VOI.

Second, 2D CHO responses λ_i, $i = 1,2, \ldots, N$ were integrated over all slices, using a hoteling observer (HTO) model. It was formulated as follows:

$$\omega_{HTO,z} = S_z^{-1}\left(\overline{\lambda_{sz}} - \overline{\lambda_{bz}}\right) \quad (2)$$

Where $\omega_{HTO,z}$ denotes the HTO template, $\overline{\lambda_{sz}}$ and $\overline{\lambda_{bz}}$ are the mean vectors of 2D CHO response vectors of signal-present and signal-absent VOIs, respectively, and S_z is the averaged covariance matrix from λ_{sz} and λ_{bz}. The integrated response of MS-MO for a particular VOI is formulated by multi-slice integration:

$$\lambda = \omega_{HO,z}^T \lambda_z \quad (3)$$

The selection of Gabor filter parameters were the same as the reference [13]. Besides MS-MO, we also applied SS-MO to the central slices of all VOIs. MS-MO performance was quantified using area under curve (AUC) of ROC curve. Internal noise was added to the integrated response of MS-MO as follows:

$$\lambda' = \lambda + \alpha x \quad (4)$$

where α is a weighting factor, x is a normal random variable (The expected value was 0 and the standard deviation was equal to the standard deviation of λ for background images). α was determined by matching the AUC of MO with a calibration condition (Figure 2). In this study, we selected condition #10 for calibration (Table 1). The value of α was equal to 2.45 and 2.05 for MS-MO and SS-MO, respectively.

2.3 Human observer study

For each trial, signal-present and background VOIs were shown together with random placement. Four human readers were recruited to identify signal-present images on a MATLAB script based GUI (Figure 2). Before the initial reading session, readers were trained to perform detection tasks with this GUI. All experimental conditions were randomly divided into three sessions. In all reading sessions, readers were required to scroll through all slices of each VOI (i.e. multi-slice viewing mode) and identify signal-present VOIs. Human reader performance was gauged by calculating percent correct (PC) at each experimental condition.

3. RESULTS

The performance of MS-MO was highly correlated to that of HO performed with multi-slice scrolling in all experimental conditions. Figure 4 illustrates the strong association between the AUC of MS-MO and the averaged PC of HO across 8 experimental conditions: CT images were reconstructed with WFBP or IR; lesion size was varied across 5 mm, 7 mm, 9 mm, and 11 mm; lesion contrast was fixed at 15 HU; radiation dose level was routine dose.

Pearson's correlation coefficient between MS-MO performance and HO performance across all 12 experimental conditions also consolidate this observation ($\rho = 0.982$, and the corresponding 95% confidence interval (CI) was [0.936, 0.995]). Furthermore, the performance of SS-MO was still highly correlated to that of HO performed with multi-slice scrolling. Figure 5 demonstrates that the AUC of SS-MO and the averaged PC of HO still yielded a strong association across the same 8 experimental conditions in Figure 4. The corresponding Pearson's correlation coefficient was $\rho = 0.989$, with 95% CI of [0.960, 0.997].

4. CONCLUSION

CHO in a uniform water background, both single-slice and multi-slice, was shown to highly correlate with human observer performance in a low-contrast lesion detection task that involves realistic patient liver background and multi-slice scrolling. These results demonstrated the potential to use the simplest single-slice CHO in uniform background to assess image quality for more realistic CT diagnostic tasks.

Acknowledgments

This research work was supported by National Institute of Biomedical Imaging and Bioengineering of the National Institutes of Health under Award Numbers R01 EB017095 and U01 EB017185. 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. Verdun FR, Racine D, Ott JG, Tapiovaara MJ, Toroi P, Bochud FO, et al. Image quality in CT: From physical measurements to model observers. Physica Medica: European Journal of Medical Physics. 31:823–843.
2. Richard S, Li X, Yadava G, Samei E. Predictive models for observer performance in CT: applications in protocol optimization. SPIE Medical Imaging. 2011:6.
3. Wilson JM, Christianson OI, Richard S, Samei E. A methodology for image quality evaluation of advanced CT systems. Med Phys. Mar.2013 40:031908. [PubMed: 23464323]
4. Fessler JA, Rogers WL. Spatial resolution properties of penalized-likelihood image reconstruction: space-invariant tomographs. IEEE Trans Image Process. 1996; 5:1346–58. [PubMed: 18285223]
5. Li K, Garrett J, Ge Y, Chen GH. Statistical model based iterative reconstruction (MBIR) in clinical CT systems. Part II. Experimental assessment of spatial resolution performance. Med Phys. Jul. 2014 41:071911. [PubMed: 24989389]
6. Chen B, Giraldo Ramirez JC, Solomon J, Samei E. Evaluating iterative reconstruction performance in computed tomography. Med Phys. Dec.2014 41:121913. [PubMed: 25471973]
7. Yu L, Leng S, Chen L, Kofler JM, Carter RE, McCollough CH. 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. Apr.2013 40:041908. [PubMed: 23556902]
8. Leng S, Yu L, Zhang Y, Carter R, Toledano AY, McCollough CH. Correlation between model observer and human observer performance in CT imaging when lesion location is uncertain. Med Phys. Aug.2013 40:081908. [PubMed: 23927322]
9. Zhang Y, Leng S, Yu L, Carter RE, McCollough CH. Correlation between human and model observer performance for discrimination task in CT. Phys Med Biol. Jul 7.2014 59:3389–404. [PubMed: 24875060]

10. Mu C, Bowsher JE, Baydush AH, Gilland KL, DeLong DM, Jaszczak RJ. Using the Hotelling observer on multislice and multiview simulated SPECT myocardial images. IEEE Transactions on Nuclear Science. 2002; 49:661–667.
11. Platisa L, Goossens B, Vansteenkiste E, Park S, Gallas BD, Badano A, et al. Channelized Hotelling observers for the assessment of volumetric imaging data sets. J Opt Soc Am A Opt Image Sci Vis. Jun 1.2011 28:1145–63. [PubMed: 21643400]
12. Ba A, Racine D, Ott JG, Verdun FR, Kobbe-Schmidt S, Eckstein MP, et al. Low contrast detectability in CT for human and model observers in multi-slice data sets. SPIE Medical Imaging. 2015:8.
13. Yu L, Chen B, Kofler JM, Favazza CP, Leng S, Kupinski MA, et al. Correlation between a 2D Channelized Hotelling Observer and Human Observers in a Low-contrast Detection Task with Multi-slice Reading in CT. Medical physics. Jul 13.2017 44:3990–3999. [PubMed: 28555878]
14. Yu L, Shiung M, Jondal D, McCollough CH. Development and validation of a practical lower-dose-simulation tool for optimizing computed tomography scan protocols. J Comput Assist Tomogr. Jul-Aug;2012 36:477–87. [PubMed: 22805680]
15. Chen B, Yu Z, Leng S, Yu L, McCollough C. Lesion Insertion in Projection Domain for Computed Tomography Image Quality Assessment. Proceedings of SPIE--the International Society for Optical Engineering. Mar 18.2015 9412:94121R.

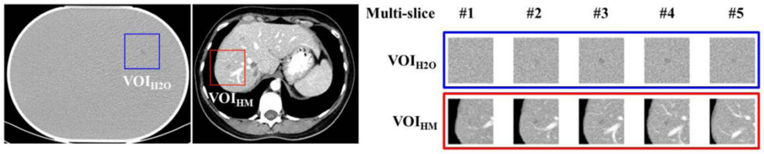

Figure 1.

Examples of the extracted VOIs from CT images of water phantom and patients, respectively. Lesion size was 9 mm, and lesion contrast was 15 HU. The display window was 400 HU/40 HU (W/L).

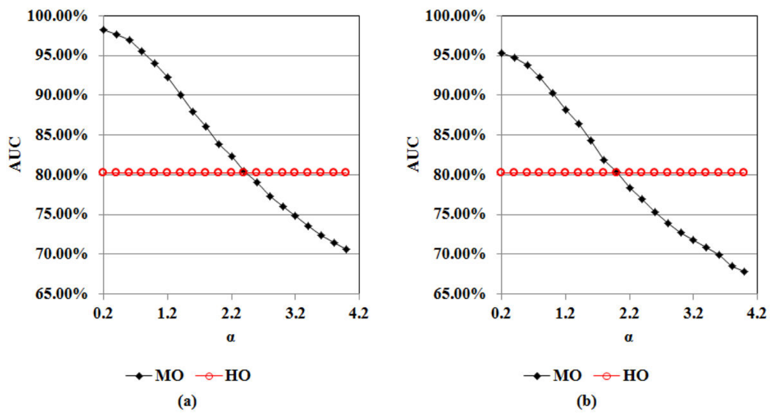

Figure 2.

AUC of MS-MO (a) and SS-MO (b) with different values of α. The weighting factor α for internal noise was determined by comparing the AUC values of MO to HO at the selected calibration condition.

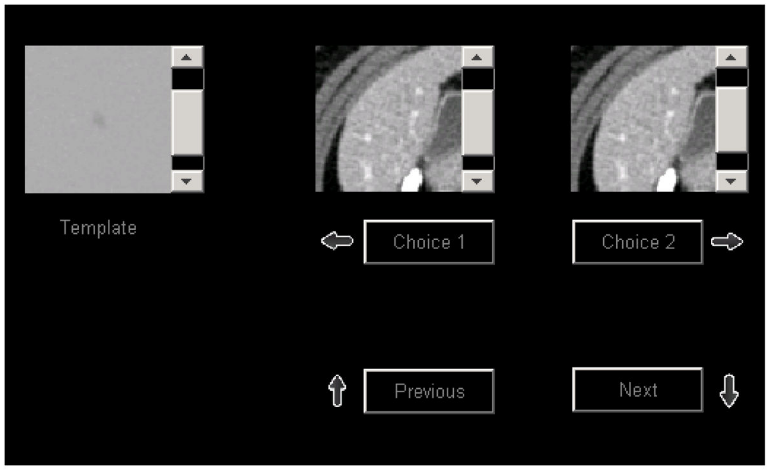

Figure 3.
GUI for human reading study

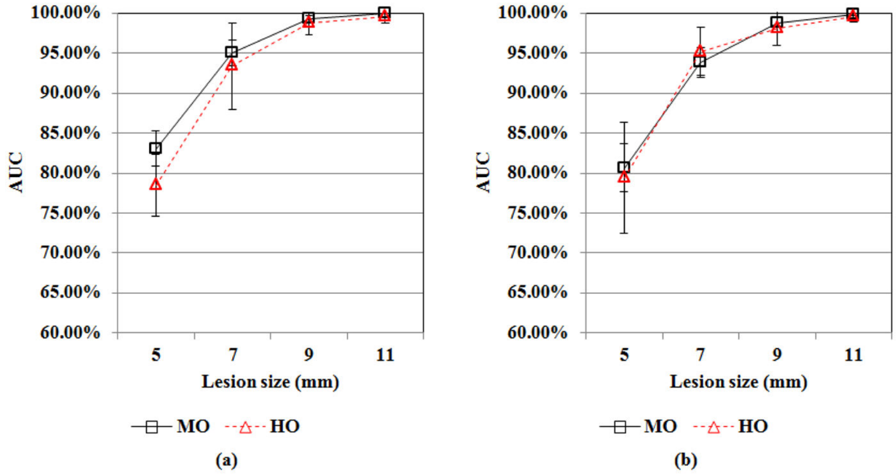

Figure 4.
Comparison between MS-MO performance and the averaged HO performance in experimental conditions with different lesion sizes and reconstruction algorithms. Lesion contrast was fixed at 15 HU and dose level was FD. CT images were reconstructed with WFBP (a) and IR (b), respectively.

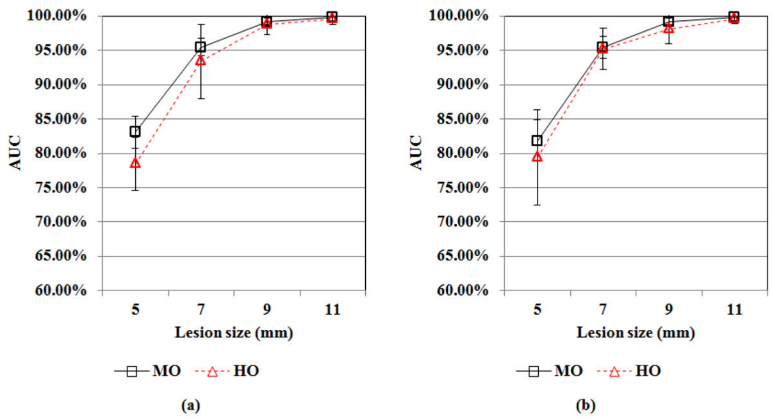

Figure 5.

Comparison between SS-MO performance and the averaged HO performance in experimental conditions with different lesion sizes and reconstruction algorithms. Lesion contrast was fixed at 15 HU and dose level was FD. CT images were reconstructed with WFBP (a) and IR (b), respectively.

Table 1

All experimental conditions in low-contrast lesion detection tasks

Condition index	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12
Lesion size (mm)	5	7	9	11	5	7	9	11	5	5	5	5
Lesion contrast (HU)	15	15	15	15	15	15	15	15	20	20	20	25
Dose level	FD	FD	FD	FD	FD	FD	FD	FD	QD	HD	FD	FD
Reconstruction algorithm	FBP	FBP	FBP	FBP	IR	IR	IR	IR	IR	IR	IR	IR