Published in final edited form as:

IEEE Nucl Sci Symp Conf Rec (1997). 2006 ; 3: 1687–1691. doi:10.1109/NSSMIC.2006.354223.

A Multiclass Model Observer for Multislice-Multiview Images

H. C. Gifford [Member, IEEE], A. Lehovich [Member, IEEE], and M. A. King [Senior Member, IEEE]

Department of Radiology, University of Massachusetts Medical School, Worcester, MA, USA

Abstract

A human-model observer for tumor detection-localization studies featuring multislice-multiview (or volumetric) image displays has been introduced. This volumetric observer, an extension of multiclass linear observers previously tested with single-slice and multislice displays, produces rating and localization data by integrating perception measurements from the different image views. A channelized NPW (CNPW) version of the observer was evaluated against humans for a background-known-exactly (BKE) detection task involving localization of Tc-99m Neotect lesions in simulated SPECT lung images. An LROC study evaluated two RBI reconstruction strategies that used different combinations of corrections for attenuation, scatter, and distance-dependent system resolution, and coronal, sagittal, and transverse slices were presented to the observers. Model-observer ranking of these strategies did not match that of the humans. Follow-up studies exploring several possible remedies for the model observer, including strategy-specific search regions and an internal-noise mechanism, showed little change. Future work will examine variations from the BKE assumption as a means of reconciling the rankings.

I. INTRODUCTION

Tumor-detection studies conducted for research purposes typically present single 2D images for human observers to evaluate. However, this methodology has potential drawbacks. Research has shown [1] that single-slice studies can bias the comparisons of test strategies intended for clinical applications. In the clinic, the physician usually has access to an entire image volume presented as multiple slices in multiple views (e.g., coronal, sagittal, and transverse slices). In contrast, a single-slice format often lacks sufficient anatomical context for separating structural noise from likely tumor sites.

We are interested in developing human-model observers for use in detection studies featuring multislice-multiview, or volumetric, image displays. Model observers have previously been proposed for such displays [2,3], but these have been limited to signal-known-exactly detection tasks. Our volumetric (or 3D) model observer is based on 2D localizing linear observers previously tested in studies with single-slice and multislice displays [4,5].

In this work, we describe a channelized NPW (CNPW) form of the volumetric observer, amounting to a particular definition of the image template that represents the 3D linear observer. An initial observer evaluation, in the setting of detection and localization of solitary pulmonary nodules (SPN) in simulated SPECT lung images, is also discussed. This evaluation involved comparison of model- and human-observer performances.

II. METHODS

A. An LROC task and the human-observer perception model

In the detection-localization task, an image $\hat{\mathbf{f}}$ is either tumor-absent or has a tumor centered on one of N pixels. The observer chooses one of $N + 1$ hypotheses:

$$H_0 : \widehat{\mathbf{f}} = \mathbf{b} + \mathbf{n} \quad (1)$$

$$H_n : \widehat{\mathbf{f}} = \mathbf{b} + \mathbf{s}_n + \mathbf{n} \quad (n=1, \dots, N), \quad (2)$$

where $\mathbf{b}$ is a known fixed background, $\mathbf{n}$ is zero-mean stochastic noise, and $\mathbf{s}_n$ is the tumor at the n^{th} location. The possible tumor locations are overlapping, meaning that the tumor size is greater than the separation between locations.

The raw LROC observer data for an image are the confidence rating λ and a localization $\mathbf{r}$. With two-dimensional (2D) images, the perception model assumes the data are acquired from the max and arg max, respectively, of a scalar perception measurement Z_n made at each suspicious location n in $\hat{\mathbf{f}}$. With a multiview display of image volumes, Z_n is the sum of perception measurements from each view. If Ω is the set of voxel indices in the observer's search region, then the data are

$$\mathbf{r} = \arg \max_{n \in \Omega} Z_n \quad (3)$$

$$\lambda = \max_{n \in \Omega} Z_n. \quad (4)$$

B. Multiclass CNPW model observers

For a $P \times 1$ 2D image $\hat{\mathbf{f}}$, a multiclass model observer computes Z_n at each pixel in a prespecified Ω according to the scalar product

$$Z_n = \mathbf{w}_n^\dagger [\widehat{\mathbf{f}} - \mathbf{b}], \quad (5)$$

where $\mathbf{w}_n$ is the n^{th} 2D observer template, and $\mathbf{b}$ is the known background. The observer template for the CNPW observer is derived by projecting the mean 2D lesion profile onto a set of C spatial frequency channels, as given by the equation

$$\mathbf{w}_n = \mathbf{U}_n [\mathbf{U}_n]^\dagger \bar{\mathbf{s}}_n. \quad (6)$$

Here, $\mathbf{U}_n$ is the $P \times C$ channel-response matrix for the n^{th} location, and $\bar{\mathbf{s}}_n$ is the mean 2D lesion profile shifted to the n^{th} location.

The shift-invariance of $\mathbf{w}_n$ is one advantage to the multiclass CNPW observer compared to the multiclass version of the channelized Hotelling (CH) observer [4]. Thus, a vector $\mathbf{Z}$ containing all the Z_n is produced from the cross-correlation

$$\mathbf{Z} = \mathbf{w} \star \star [\widehat{\mathbf{f}} - \mathbf{b}]. \quad (7)$$

A CNPW template based on a set of three square-profile channels is shown in Fig. 1A.

The obvious extension of Eq. 7 to image volumes implements a 3D cross-correlation. However, there are any number of possible definitions of a 3D template for this purpose (see, for instance, [3]). The examples we tested, based on the perception model presented in the previous section, assume that the human observer assesses a given lesion position by summing the perception measurements obtained at that position in the coronal, sagittal, and transverse 2D slices. The 3D model-observer template for this is constructed by forming the 2D templates for each of those three slices, and then combining those templates in the same manner as the axial planes $x = 0$, $y = 0$, and $z = 0$ intersect in a 3D Cartesian coordinate system. All other template voxels are zero. Figure 1B shows an example 3D template constructed from the same channels used in Fig. 1A.

This template accounts for the multiple views of the volume, but is not multislice. A human observer would be likely to also examine neighboring image slices when assessing whether a tumor is present at a given location. We model this by adding additional planes to the template. For example, with one new plane added to each side of the axial planes, the observer integrates information from three slices for each of the three orthogonal views when assessing a given location. The geometric extent of the lesion is accounted for since each of these new template planes is constructed using the corresponding slice through the mean lesion.

In this work, we discuss results obtained from a three-channel set of difference-of-Gaussian (DOG) set of channels. Details on these channels are in [5].

C. Test volumes

The setting for testing these model observers was an investigation of the effects of attenuation correction (AC) for SPECT lung imaging. The specific purpose of the observer study was to determine whether the benefits of AC compensate for the concurrent loss of the lung outlines for the detection of solitary pulmonary nodules (SPN). The imaging protocol was based on Tc-99m-labeled Neotect [6], a polypeptide that binds to somatostatin receptors and which has been used mainly to evaluate certain nodules in patients who had tested positive with other imaging tests.

A Neotect distribution in a single body geometry of the NCAT phantom [7] was discretized to 128^3 voxels (voxel width = 0.417 cm). The normal Neotect lung background was uniform. Abnormal volumes contained a single 1-cm spherical soft-tissue lesion randomly situated within the lungs. A high-count projection set of the normal phantom ($120 \times 128 \times 128$ projections and 0.417-cm pixel width) was created with the SIMIND [8] Monte-Carlo projector, while separate high-count SIMIND projections of the lung lesions were also generated and then added to the background data to form noise-free volumes with relative tumor-to-lung activity ratios of 10:1, 12:1, and 14:1. Data sets for reconstruction were produced by adding Poisson noise consistent with an acquisition of 23.2×10^6 counts, a figure obtained by averaging the counts collected over a set of nine Neotect scans from our clinic.

Volumes were reconstructed with two strategies based on the RBI reconstruction algorithm [9]. The *AIIC* strategy applied AC, scatter correction (SC), and resolution compensation (RC). The second strategy used RC only. All the corrections were applied as part of the iterative cycle. The SC was implemented through a TEW [10] scatter estimate, while an idealized attenuation map for lesion-absent data was devised by scaling the NCAT density map for the 140-keV photons of Tc-99m (linear attenuation coefficient for water = 0.15 cm^{-1}). For the lesion-present data, this map was augmented by the addition of the soft-tissue lesion at the proper lung location.

The reconstructions consisted of $70 \times 128 \times 128$ transverse slices. The number of iterations for both strategies was five, while a 3D Gaussian postfilter (FWHM = two pixels) was also applied.

These parameter values were based on previous 2D study results [11,12]. In all, 150 noisy image volumes per strategy were reconstructed, with an abnormal/normal volume pair for each of 75 lesion locations. As final postprocessing, the volumes underwent a strategy-specific upper-thresholding (see [1] for details). The volumes were then quantized to eight bits for display. Sample coronal slices taken from the noise-free reconstructions of the NCAT background for the two strategies are shown in Fig. 4.

D. The LROC study

The three co-authors read the volumes. A computer interface for viewing the images (see screenshot in Fig. 5) was programmed in IDL (ITT Visual Information Solutions, Boulder, CO, USA). The greyscale of the display was fixed, and discrete rating data was collected on a six-point scale. The 150 reconstructions per strategy were divided into a set of 50 training volumes and a set of 100 test volumes. As preliminary training, each observer read a set of twenty volumes per strategy. This was followed by five separate sessions of six retraining volumes and twenty test volumes per strategy. The reading order of the test sets and of the volumes within a given set varied with observer.

An observer's raw data from the five test sets for a given strategy were pooled, and a radius of correct localization (R_{cl}) of four voxels was assigned for the purpose of evaluating correct localizations in the data. LROC curves were fit to the data using the Swensson software [13]. The overall performance for a given strategy was represented by the area under the LROC curve (A_L) averaged over the observers.

In the model-observer study, the 3D CNPW template was estimated from the training volumes and then applied to the test volumes. Confidence ratings and localizations were determined according to Eq. 3–Eq. 4, and correct localizations were assessed with the same R_{cl} as in the human-observer study. The figure of merit was the empirical area under the LROC curve.

III. RESULTS

The individual and average performances for the human and model observers are summarized in Table I. Nominally, all three humans performed best with the RC strategy. However, the difference in average A_L (0.66 versus 0.61) between the two strategies was within experimental uncertainty.

Table I has results from two versions of the CNPW model observer. Version #1 was solely multiview, while #2 considered three slices per view. For this experiment, both observers scored higher with the AllC strategy than with the RC strategy. The additional slices for the second observer did not prove to be an advantage in this study. For this reason, the subsequent studies described in this section used version #1.

Also given in Table I is a breakdown of observer A_L 's on the basis of relative tumor contrasts. The three tumor contrasts fostered a wide range of human-observer performances, whereas model-observer performance only began to degrade at the lowest contrast. Nonetheless, this result suggests that the CNPW observer might be useful for coarsely identifying acceptable contrast ranges as part of the study design.

An analysis of the localizations made by the human observers offered additional evidence for the RC strategy outperforming the AllC strategy. We defined an observer's search accuracy as the fraction of localizations that occurred within the lungs. Table II lists the overall fraction, calculated from all of an observer's localizations with a given strategy, as well as a "high-confidence" fraction based on localizations associated with the upper half of the confidence-rating scale. This latter quantity reduced the bias effects due to the forced-localization

requirement of the study, whereby an observer may haphazardly select a location when relatively confident that the volume is normal.

The localization analysis indicates that the observers were better at defining the extent of the lungs in the RC reconstructions than in the AllC reconstructions. The loss of the lung outlines in the latter due to more-accurate activity estimation (contrast the appearance of the lungs in Figs. 4A–B) is the cause of this. Figure 5 plots some of the errant AllC localizations relative to the lungs in four transverse slices.

For a second model-observer study, we attempted to account somewhat for the human-observer localization data by defining an enlarged search region Ω_E that annexed a large portion of the heart (Fig. 7). A previous study with Ga-67 SPECT images [4] took a similar approach in using the CNPW observer to attribute performance differences between human observers to variations in search region. This type of analysis is strictly post-hoc, as Ω must be defined prior to applying the observer, so it can be difficult to make accurate comparisons when Ω may be strategy-dependent. In our experiment, the difference in model-observer A_L between using Ω and Ω_E were minimal for the AllC strategy (0.99 and 0.98, respectively). There was no effect with the RC strategy. Thus, varying the search region alone was not sufficient to model the human results.

We also investigated whether the combination of Ω_E with an internal-noise mechanism could bring the CNPW observer into agreement with the humans. The noise mechanism was a variation on that implemented in [5], where the tumor localization became a selection from a set of nonoverlapping candidate locations. For this current work, a single-parameter model randomly selected the tumor localization from among candidate locations where Z_n exceeded a percentage t of the maximum response. The associated confidence rating was the average of Z_n over the candidates. A minimum separation of eight pixels (center-to-center) was enforced between candidate locations. The threshold parameter was varied from 0.4 to 0.9 in increments of 0.1. For each parameter value, twenty LROC studies were performed, each starting with a different random-number seed. Figure 8 plots the average A_L over these studies as a function of t . While performances fall off as t decreases, there is no t such that the RC strategy with search region Ω outperformed the AllC strategy with Ω_E .

Besides the studies with the extended search and the internal-noise model, we also considered alternatives to the BKE assumption. When the background subtraction was not performed, the CNPW-observer scores were 0.48 (AllC) and 0.50 (RC). External activity spilling into the lungs at moderate levels of image smoothing was a problem for the model observer in this study. Another study modeled a quasi-BKE task for which the activity external to the lungs was subtracted and a partial spill-in compensation was performed. For that study, the scores were 0.95 (AllC) and 0.72 (RC).

IV. CONCLUSIONS

This paper has introduced a 3D CNPW model observer for tumor localization-detection tasks with image volumes. In the SPECT lung LROC study, the ranking of the reconstruction strategies with this observer disagreed with that obtained for the human observers. Variations on the model-observer study, including different search regions and the addition of internal noise, did not change this result. Future work will investigate other quasi-BKE task assumptions as a means of matching the model-observer results to those of the humans.

Acknowledgements

This work was supported by the National Institute for Biomedical Imaging and Bioengineering (NIBIB) under grant number R01-EB02798. The contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIBIB.

REFERENCES

1. Wells RG, King MA, Gifford HC, Pretorius PH. Single-slice versus multi-slice display for human-observer lesion-detection studies. *IEEE Trans. Nucl. Sci* 2000;vol. 47:1037–1044.
2. Chen M, Bowsher JE, Baydush AH, Gilland KL, DeLong DM, Jaszczak RJ. Using the Hotelling observer on multislice and multiview simulated SPECT myocardial images. *IEEE Trans Nucl. Sci* 2002;vol. 49:661–667.
3. Kim JS, Kinahan PE, Lartizien C, Comtat C, Lewellen TK. A comparison of planar versus volumetric numerical observers for detection task performance in whole-body PET imaging. *IEEE Trans Nucl. Sci* 2004;vol. 51:34–40.
4. Gifford HC, Pretorius PH, King MA. Comparison of human-and model-observer LROC studies. *Proc. SPIE* 2003;vol. 5034:112–122.
5. Gifford HC, King MA, Pretorius PH, Wells RG. A comparison of human and model observers in multislice LROC studies. *IEEE Trans. Med. Imaging* 2005;vol. 24:160–169. [PubMed: 15707242]
6. Blum JE, Handmaker H, Rinne NA. The utility of a somatostatin-type receptor binding peptide radiopharmaceutical (p829) in the evaluation of solitary pulmonary nodules. *Chest* 1999;vol. 115:224–232. [PubMed: 9925088]
7. Segars WP, Lalush DS, Tsui BMW. Modeling respiratory mechanics in the MCAT and spline-based MCAT phantoms. *IEEE Trans. Nucl. Sci* 2001;vol. 48:89–97.
8. Ljungberg M, Strand SE. A Monte-Carlo program for the simulation of scintillation camera characteristics. *Comput. Meth. Prog. Biomed* 1989;vol. 29:257–272.
9. Byrne CL. Block-iterative methods for image reconstruction from projections. *IEEE Transactions on Image Processing* 1996;vol. 5:792–794. [PubMed: 18285171]
10. Ogawa K, Harata H, Ichihara T, Kubo A, Hashimoto S. A practical method for position dependent Compton-scatter correction in single photon emission CT. *IEEE Trans. Nucl. Sci* 1991;vol. 10:408–412.
11. Gifford, HC.; Zheng, XM.; Boening, G.; Bruyant, PP.; King, MA. An investigation of iterative reconstruction strategies for lung lesion detection in SPECT; *IEEE Nuclear Science Symposium and Medical Imaging Conference Record*; 2004. p. 4241-4245.
12. Gifford HC, Zheng XM, Licho R, Schneider PB, Simkin PH, King MA. LROC assessment of SPECT reconstruction strategies for detection of solitary pulmonary nodules. *J Nucl Med* 2005;vol. 46:1489 (461P).
13. Swensson RG, Judy PF. Observer detection efficiency for visual targets at specified or unknown locations in noise backgrounds. *Journal of Mathematical Psychology* 1996;vol. 40:348.

Fig. 1.

A) A 2D CNPW-observer template based on square-profile channels. B) A 3D CNPW-observer template for coronal, sagittal, and transverse views. As shown, the horizontal plane applies to the transverse slice. The two vertical planes apply to the coronal and sagittal slices.

Fig. 2.
The three difference-of-Gaussian (DOG) channels used for this study.

Fig. 3.

Transverse slices of the NCAT density map. Lesions could appear anywhere within the lungs, indicated by the dark regions within the torso. The lung volume also represents the ideal search region Ω for the model observers.

Fig. 4. Comparison of coronal slices from the noise-free NCAT background reconstructions. A) Twelve consecutive slices for the AllC strategy. B) Corresponding slices for the RC strategy.

Fig. 5.

Computer interface for the human-observer LROC study. An observer uses the horizontal sliders to control the displays in the bottom set of windows. A localization is set by clicking somewhere in one of these windows with the mouse cursor. Confidence ratings are input using the vertical slider. During a training session (shown here) feedback in the form of the noise-free image is provided in the upper set of windows.

Fig. 6.
Example observer SPN localizations. The lung regions are shown in black.

Fig. 7.

Transverse slices of the extended search region Ω_E for the model observers. The change from Ω is the bridging of the two lung lobes in rows 2 through 4.

Fig. 8.
Effect of internal-noise parameter t on model-observer performances with the two search regions.

TABLE I

Model-observer and human-observer performances with the AIC and RC strategies. The uncertainties for the average human scores account for observer variability and uncertainties in the data fitting.

AIC				
Observer	Contrast			Overall
	Low	Mid	High	
Human #1	0.46	0.57	0.87	0.63 ± 0.07
Human #2	0.38	0.66	0.84	0.63 ± 0.07
Human #3	0.32	0.62	0.74	0.56 ± 0.07
Average	0.39	0.62	0.82	0.61 ± 0.07
CNPW #1	0.98	1.0	1.0	0.99
CNPW #2	0.98	1.0	1.0	0.99
RC				
Observer	Contrast			Overall
	Low	Mid	High	
Human #1	0.56	0.73	0.89	0.72 ± 0.06
Human #2	0.40	0.67	0.86	0.64 ± 0.06
Human #3	0.35	0.63	0.88	0.61 ± 0.06
Average	0.44	0.68	0.88	0.66 ± 0.07
CNPW #1	0.86	1.0	1.0	0.95
CNPW #2	0.80	1.0	1.0	0.93

TABLE II

The overall and high-confidence search accuracies for the human observers.

Observer	AII		RC	
	Overall	High-Conf	Overall	High-Conf
#1	0.65	0.95	1.0	1.0
#2	0.74	0.82	0.95	1.0
#3	0.73	0.95	0.90	1.0
Average	0.71	0.90	0.95	1.0