Accepted to be published in: 2022 IEEE International Symposium on Biomedical Imaging (IEEE ISBI 2022), March 28-31, 2022
ABSTRACT

In medical imaging, it is widely recognized that image quality should be objectively evaluated based on performance in clinical tasks. To evaluate performance in signal-detection tasks, the ideal observer (IO) is optimal but also challenging to compute in clinically realistic settings. Markov Chain Monte Carlo (MCMC)-based strategies have demonstrated the ability to compute the IO using pre-computed projections of an anatomical database. To evaluate image quality in clinically realistic scenarios, the observer performance should be measured for a realistic patient distribution. This implies that the anatomical database should also be derived from a realistic population. In this manuscript, we propose to advance the MCMC-based approach to achieve these goals. We then use the proposed approach to study the effect of anatomical database size on IO computation for the task of detecting perfusion defects in simulated myocardial perfusion SPECT images. Our preliminary results provide evidence that the size of the anatomical database affects the computation of IO.

Index Terms— Image-quality evaluation, SPECT, Markov chain Monte Carlo.

1. INTRODUCTION

There is wide recognition in the medical imaging community that image quality should be evaluated based on performance in clinical tasks, such as those of detection and estimation [1]. Of these, detection tasks are performed by an observer, who makes a decision on whether the signal to be detected is present or absent. The term observer typically conjures the images of a trained radiologist, but can also be mathematical observers, also referred to as model observers [2]. Of the various observers, the one that utilizes all the statistical information available regarding the task to maximize task performance is referred to as the ideal observer (IO) [3]. This observer provides the best possible performance on the detection task, as quantified using the area under the receiver operating characteristic (ROC) curve (AUC). The use of IO is recommended for optimizing system instrumentation and acquisition protocols to ensure that the measured data has the maximum possible information for the detection task [4, 1]. However, the IO requires a complete knowledge of the distribution of the image data. Such a distribution is very high dimensional, and, thus, very difficult, if not, impossible to obtain. To address this challenge, various approaches have been proposed [5, 6] but computing IO performance in clinically realistic settings remains a challenge. Thus, there is an important need for strategies to compute the IO in clinically realistic scenarios.

A seminal contribution towards IO computation was a Markov Chain Monte Carlo (MCMC)-based technique proposed by Kupinski et al. [7]. The technique demonstrated the ability to compute IO for parametric object models, in particular, a lumpy background model. This technique was later extended for a simple parametric description of the human cardiac region [8]. To improve the efficiency of the IO computation, projection data of an anatomical database were precomputed. To increase the realism of this approach, Ghaly et al. [9] proposed a strategy that used anthropomorphic extended cardiac and torso (XCAT) phantoms [10] that simulated the anatomical variability in patient populations. The above studies focused on IO computation on a patient population with anatomical parameters sampled from a uniform distribution. However, in clinical settings, the distribution of anatomical parameters is unlikely to be uniformly distributed. For example, patient heights typically follow normal distributions [11]. For accurate computation of the IO, and in general, for rigorous image-quality evaluation studies, the anatomical variability needs to be sampled from realistic patient populations. Thus, the use of uniform distributions to sample anatomical parameters may not reflect IO performance in clinically realistic settings.

Another important consideration in using MCMC-based approaches to IO computation is the size of the anatomical database. He et al. [8] indicated that this size may affect the calculation of IO. Further, they suggested that the number of anatomies required to accurately compute the IO may be a function of system resolution. Thus, the effect of the size of anatomical database on the computation of IO also needs to be investigated.

In this paper, we advance the previously developed MCMC-based methods to achieve the goal of performing IO computation with clinically realistic populations. In this
proof-of-concept study, we show the ability of the proposed method to sample from an anatomical database where the distribution of height and heart sizes are described by clinically realistic normal distributions. We then validate the use of this method to compute the IO performance. We further use the proposed method to investigate the effect of size of the anatomical database on the accuracy of the IO computation. We implement and evaluate the method in the context of computing the IO for the task of detecting cardiac defects from myocardial perfusion SPECT images.

2. THEORY

Consider a myocardial perfusion SPECT system imaging a radiotracer distribution, denoted by the infinite-dimensional vector $f$, that, we assume, lies in the Hilbert space $L_2(\mathbb{R}^3)$. Denote the SPECT imaging system by the Hilbert space operator $\mathcal{H}$. Next, denote the projection data obtained by the SPECT system by the $M$-dimensional vector $g$, that, we assume, lies in the Hilbert space $E^M$. Thus the SPECT system can be described by the mapping $\mathcal{H}: L_2(\mathbb{R}^3) \rightarrow E^M$, and the imaging-system equation is given by

$$g = \mathcal{H}f + n,$$

(1)

where $n$ denotes the $M$-dimensional noise vector. We write the object $f$ as a combination of the signal of interest, denoted by $f_s$ and the rest of the object (referred to as the background object), denoted by $f_b$. Thus

$$f = f_s + f_b.$$  

(2)

In a clinical setting, $f_s$ could denote the region with abnormal uptake, such as a lesion, while the anatomical and physiological variability in the rest of the patient would be denoted by $f_b$. Finally, let the noise-free image corresponding to the background object be denoted by $b$, i.e. $b = \mathcal{H}f_b$.

In a detection task performed on the projection data $g$, the goal is to determine, from the projection data, whether the underlying signal of interest is present or absent. Let $pr(g|H_1)$ and $pr(g|H_0)$ denote the probability distribution function of the image given that the signal is present and absent, respectively. To compute the IO, the following test statistic, referred to as the likelihood ratio, and denoted by $\Lambda(g)$, is computed:

$$\Lambda(g) = \frac{pr(g|H_1)}{pr(g|H_0)}.$$  

(3)

From this equation, we observe that computing the IO requires complete knowledge of the distribution of $g$ under both the signal present and signal-absent hypothesis.

When the to-be-detected signal $f_s$ and background $f_b$ are both known exactly, this test statistic can be computed using the knowledge of the noise statistics. We denote this test statistic by $\Lambda_{BE}(g)$. However, in a clinically realistic setting, where both the signal and the background vary, the probabilities in Eq. (3) can be difficult, if not impossible to define. A simplification is obtained when we consider the detection task where the properties of the signal are known, while the rest of the patient properties are variable. For this detection task, referred to as the signal-known-exactly/background known statistically (SKE/BKS) task, the expression for the IO is given by [7]:

$$\Lambda(g) = \int db \Lambda_{BE}(g|b)pr(b|g, H_0).$$  

(4)

If we could sample backgrounds from the posterior distribution $pr(b|g, H_0)$, then this integral can be computed through a Monte Carlo integration procedure:

$$\Lambda(g) = \frac{1}{J} \sum_{j=1}^{J} \Lambda_{BE}(g|b^j),$$  

(5)

where $b^j$ denotes the $j$th realization of the background. However, $b$ is a very high-dimensional vector (128$^3$ for SPECT). Sampling from this distribution is challenging. An alternative is to consider a parametric representation of the background [7, 8]. However, human anatomies are challenging to represent through parametric models. To address this issue, Ghaly et al. [9] proposed the use of XCAT phantoms. However, to the best of our knowledge, their anatomical database size was limited to $N_d = 54$. However, larger-sized databases may be needed to conduct rigorous image-quality evaluation studies. Moreover, in their approach, the anatomical parameters were sampled from a uniform distribution, which, as mentioned above, has limitations in modeling realistic populations.

In this study, we advance the above MCMC-based approach to sample from a more realistic patient population distribution. Similar to He et al., we parameterize the anatomy of each sample of the patient population using two parameters, namely, the radius of the left ventricle of the heart, and the body size. These are denoted by $\theta_b$ and $\theta_h$, respectively. The activity uptake in the different organs, including the heart, liver and lungs, is parameterized by a $q$-dimensional vector $\theta_{act}$. Denote $\theta = \{\theta_{act}, \theta_b, \theta_h\}$ and the $i$th element of $\theta$ in iteration $j$ by $[\theta]_i^j$. Mathematically, at each iteration, we sample from a one-dimensional proposal distribution given by:

$$[\tilde{\theta}]_i \sim \mathcal{N}(i, \sigma_i),$$  

(6)

where $\mathcal{N}$ denotes a normal distribution with mean $i$ and standard deviation $\sigma_i$. Further, $[\tilde{\theta}]_i$ denotes the newly proposed sample of the $i$th parameter. Note that for the $j$th iteration,

$$i \sim \mathcal{U}(1, q + 2),$$  

(7)

$$[\tilde{\theta}]_m = [\theta]_m, \forall m \neq i,$$  

(8)
where, \( U(1, q + 2) \) denotes sampling from a discrete uniform distribution with range between 1 to \( q + 2 \) with a step size of 1. As the anatomical database contains only discrete LV radius and body size, the anatomy proposed at each iteration should be chosen such that the LV radius and body size of that anatomy are closest to the proposed LV radius and body size. This strategy provides a way to guide the sampling process through a large-sized phantom population and allows for the prior information about the LV radius and body size to be accounted for in the calculation of acceptance ratio.

3. IMPLEMENTATION AND EVALUATION OF THE PROPOSED APPROACH

3.1. Anthropomorphic phantom model and projection-data generation

We used the XCAT phantom [10] for generating patient populations. For realistic variability in phantom, the body size, LV length and radius, and the body height were sampled from independent Gaussian distribution with mean and variance based on Emory PET dataset and listed in Ghaly et al. [12]. We only considered one transaxial slice of each anatomy that contained the centroid of the defect in defect-present case. The phantom was generated over a \( 256 \times 256 \) grid with pixel size of 2 mm. The defect-present case was generated by introducing a cold defect with 25% severity and 10% extent in the anterior wall. A total of 10000 pairs (defect present/absent) of anatomies were generated, 5000 pairs each for both male and female patients.

Projection data for the phantom population was generated with a simulated 2-D SPECT system that modeled the collimator-detector response and noise in SPECT systems. The projection data was acquired in 64 projection bins and at 3 projection angles. The pixel size of each projection bin was 8 mm. As in [9] and [8], we pre-calculated the organ-specific projection for the 10,000 anatomies for defect-absent case. The defect-only projection was also computed to be used in SKE/BKS task and to generate the defect-present projections to be used in observer study. The organs considered were the heart, lung, liver and the background. This strategy of obtaining organ-specific projections substantially reduced the computational requirements of the MCMC method.

3.2. Generating the population for observer study

We used the generated projection data from the 10,000 anatomies and defect-only projection for each anatomy to generate 2000 pairs of projection data, \( g \) using as similar strategy as in [8]. In brief, for organ \( i \) and anatomy index \( k \), denote the noise-free projection data as \( b_{i,k} \). Denote the LV radius and body size associated with anatomy index \( k \) as \( \theta_{h,k} \) and \( \theta_{b,k} \), respectively. Then the noise-free phantom projection data is given by

\[
b(\theta_{act}, \theta_{h,k}, \theta_{b,k}) = \sum_{i=1}^{4} \theta_{act,i} b_{i,k}.
\]

We refer to this dataset as the test set for computing IO performance.

We scaled the background projection data such that, on average, the total number of counts in projection data was 5000. To get the final projection data, \( g \), we added Poisson noise to this projection data. For each anatomy, we sampled the organ activity from distribution mentioned in [12].

3.3. The MCMC simulation

We initialized the activity parameters by randomly sampling from the distribution given in [12]. The initial anatomy parameter was sampled uniformly from the anatomical database. Then at each iteration of MCMC, we randomly picked a single parameter to modify. For proposing a new activity parameter \( \theta_{act,pr} \), we used a Gaussian proposal density with standard deviation set to one-tenth of the standard deviation of that particular prior organ activity distribution. However, to sample from the anatomy parameter \( k \), we made use of the distribution of LV radius and body size. Based on the previous iterations’ anatomy \( k \), we obtained the corresponding LV radius and body size. For proposing a new LV radius, we set the proposal density as a Gaussian distribution. The standard deviation of this distribution was one-fifth the standard deviation of the clinically observed LV-radius distribution. Similar strategy was used for proposing new body size. Using these proposed parameters, we selected the anatomy for which the LV radius and body size was the closest match. As the proposal density is symmetric, we accepted the newly proposed parameter set \( \tilde{\theta} \) over the previous parameter \( \theta' \) with probability

\[
\alpha(\tilde{\theta}, \theta') = \min \left( 1, \frac{\text{pr}(g|\tilde{\theta}, H_0) \text{pr}(|g, \theta'|)}{\text{pr}(g|\theta', H_0) \text{pr}(|g, \theta|)} \right).
\]

Note that, prior probabilities can be modeled again by a Gaussian distribution from the patient data.

Challenges in implementation of the MCMC technique include cases where some projection bins in \( \tilde{b} = b(\tilde{\theta}) \) and \( b' = b(\theta') \) are null. This can be addressed by discarding the contribution of such pixels while calculating the acceptance probability. We also discard the initial 400,000 burn-in iterations where non-stationarity of \( \Lambda_{BKE} \) estimates are observed to get a reliable IO estimate.

4. RESULTS

In Fig. 1(a), we show the convergence of IO estimate when the size of anatomical database, \( N_g \) was 10000. We observed that
the estimate converges after around 4 million iterations. We also divide the entire iterations into consecutive blocks and show the estimate of IO test statistics calculated from each block, as in [8]. The block size was set to 500 iterations. We observed that the distribution follows a log-normal distribution and the log of this estimate follows a normal distribution (Fig. 1(b)).

We also validate the proposed MCMC method by comparing empirical and theoretical estimate of \( \langle \hat{\Lambda}|H_0 \rangle \) and \( \langle \hat{\Lambda}|H_1 \rangle - Var(\hat{\Lambda}|H_0) \). Theoretically, both these terms should be equal to one [8]. From Fig. 2, we observed that these quantities reach 0.913 and 1.03, respectively for \( N_d = 10000 \). The closeness of these summary statistics to 1 provides evidence in the direction of validating this method. The minor deviation could be due to the unique nature of distribution of anatomical phantom population and needs further investigation. However, overall, the results in Figs. 1 and 2 provide evidence in support of validating the MCMC technique.

We next investigated the effect of the size of anatomical distributions on IO computation using the MCMC technique. For this purpose, we varied the size of anatomical database and for each setting, computed the IO performance. The AUC value as a function of anatomical database size is shown in Fig. 3. We observed that increasing the size of anatomical database resulted in an increase in AUC value. Thus, a large-sized population of the anatomical database is required to get reliable estimate of the IO performance through the MCMC approach for anthropomorphic phantoms.

![Fig. 1](image1.png)

**Fig. 1.** (a) The estimate of IO test statistics as a function of total number iterations used for MCMC simulation. (b) The distribution of block estimate of log of IO test statistic.

![Fig. 2](image2.png)

**Fig. 2.** (a) \( \langle \hat{\Lambda}|H_0 \rangle \) and (b) \( \langle \hat{\Lambda}|H_1 \rangle - var(\hat{\Lambda}|H_0) \) as a function of number of image pairs used for observer study.

![Fig. 3](image3.png)

**Fig. 3.** AUC values as a function of size of anatomical database. The error bar indicates 95% confidence interval.

**5. DISCUSSIONS AND CONCLUSION**

In this paper, an MCMC-based method was proposed in the context of computing IO performance in clinically realistic settings. We observe that the proposed method operates with the anthropomorphic XCAT phantoms, and is able to sample the anatomical descriptors of this phantom from a clinically realistic normal distribution. Our investigation of the effect of size of anatomical database on IO computation with this method showed that large sample size of the anatomical patient database are required to accurately compute the IO performance. These findings thus suggest the use of approaches that can generate such large databases.

The proposed approach has multiple applications in addition to computing the IO. This includes generating phantom populations for virtual clinical trials and other image-quality evaluation studies [13, 14, 15, 16, 17]. Another application is in simulation-guided deep learning approaches that provide the advantage of using patient populations with known ground truth, which can then be used for training [18]. The proposed method may be used to generate such a patient population.

In conclusion, we advanced a MCMC-based strategy with the eventual goal of sampling from a clinically realistic patient population. We then demonstrated the application of this strategy to sample from an anatomical database with height and heart sizes described by clinically realistic normal distributions. The strategy was validated in the context of evaluation IO performance for defect-detection tasks in myocardial perfusion SPECT. Our analysis provides evidence of the efficacy of having a large-sized anatomical database to reliably compute IO performance.
6. COMPLIANCE WITH ETHICAL STANDARDS

This is a numerical simulation study for which no ethical approval was required.

7. ACKNOWLEDGMENTS

This work was supported by the National Institute of Biomedical Imaging and Bioengineering of the National Institute of Health under grants R21-EB024647, R01 EB031051, and R56 EB028287.

8. REFERENCES

[1] Harrison H Barrett, Kyle J Myers, Christoph Hoeschen, Matthew A Kupinski, and Mark P Little, “Task-based measures of image quality and their relation to radiation dose and patient risk,” *Phys. Med. Biol.*, vol. 60, no. 2, pp. R1, 2015.

[2] Harrison H Barrett, Jie Yao, Jannick P Rolland, and Kyle J Myers, “Model observers for assessment of image quality,” *Proceedings of the National Academy of Sciences*, vol. 90, no. 21, pp. 9758–9765, 1993.

[3] Harrison H Barrett, Craig K Abbey, and Eric Clarkson, “Objective assessment of image quality. III. ROC metrics, ideal observers, and likelihood-generating functions,” *JOSA A*, vol. 15, no. 6, pp. 1520–1535, 1998.

[4] Abhinav K Jha, Kyle J Myers, Nancy A Obuchowski, Ziping Liu, Md Ashequr Rahman, Babak Saboury, Arman Rahmim, and Barry A Siegel, “Objective Task-Based Evaluation of Artificial Intelligence-Based Medical Imaging Methods:: Framework, Strategies, and Role of the Physician,” *PET clinics*, vol. 16, no. 4, pp. 493–511, 2021.

[5] Eric Clarkson and Fangfang Shen, “Fisher information and surrogate figures of merit for the task-based assessment of image quality,” *JOSA A*, vol. 27, no. 10, pp. 2313–2326, 2010.

[6] Abhinav K Jha, Eric Clarkson, and Matthew A Kupinski, “An ideal-observer framework to investigate signal detectability in diffuse optical imaging,” *Biomed. Opt. Express*, vol. 4, no. 10, pp. 2107–2123, 2013.

[7] Matthew A Kupinski, John W Hoppin, Eric Clarkson, and Harrison H Barrett, “Ideal-observer computation in medical imaging with use of Markov-chain Monte Carlo techniques,” *JOSA A*, vol. 20, no. 3, pp. 430–438, 2003.

[8] Xin He, Brian S Caffo, and Eric C Frey, “Toward realistic and practical ideal observer (IO) estimation for the optimization of medical imaging systems,” *IEEE Trans. Med. Imaging*, vol. 27, no. 10, pp. 1535–1543, 2008.

[9] Michael Ghaly, Jonathan M Links, and Eric Frey, “Optimization of energy window and evaluation of scatter compensation methods in myocardial perfusion SPECT using the ideal observer with and without model mismatch and an anthropomorphic model observer,” *J. Med. Imaging*, vol. 2, no. 1, pp. 015502, 2015.

[10] W Paul Segars, G Sturgeon, S Mendonca, Jason Grimes, and Benjamin MW Tsui, “4D XCAT phantom for multimodality imaging research,” *Med. Phys.*, vol. 37, no. 9, pp. 4902–4915, 2010.

[11] John Komlos and Joo Han Kim, “Estimating trends in historical heights,” *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, vol. 23, no. 3, pp. 116–120, 1990.

[12] Michael Ghaly, Yong Du, George SK Fung, Benjamin MW Tsui, Jonathan M Links, and Eric Frey, “Design of a digital phantom population for myocardial perfusion SPECT imaging research,” *Phys. Med. Biol.*, vol. 59, no. 12, pp. 2935, 2014.

[13] Harrison H Barrett, Lars R Furenlid, Melanie Freed, Jacob Y Hesterman, Matthew A Kupinski, Eric Clarkson, and Meredith K Whitaker, “Adaptive SPECT,” *IEEE Trans. Med. imaging*, vol. 27, no. 6, pp. 775–788, 2008.

[14] Nasrin Ghanbari, Eric Clarkson, Matthew Kupinski, and Xin Li, “Optimization of an adaptive SPECT system with the scanning linear estimator,” *IEEE Trans. Rad. Plas. Med. Sci.*, vol. 1, no. 5, pp. 435–443, 2017.

[15] Zitong Yu, Md Ashequr Rahman, Thomas Schindler, Richard Laforest, and Abhinav K Jha, “A physics and learning-based transmission-less attenuation compensation method for SPECT,” in *Proc. SPIE Med. Imag.* International Society for Optics and Photonics, 2021, vol. 11595, p. 1159512.

[16] Zitong Yu, Md Ashequr Rahman, Thomas Schindler, Robert Derenoncourt, Barry A Siegel, and Abhinav K Jha, “A Bayesian approach to tissue-fraction estimation for oncological PET segmentation,” *Phys. Med. Biol.*, vol. 66, no. 12, pp. 124002, 2021.

[17] Zitong Yu, Md Ashequr Rahman, Thomas Schindler, Robert Gropler, Richard Laforest, Richard Wahl, and Abhinav Jha, “AI-based methods for nuclear-medicine imaging: Need for objective task-specific evaluation,” *J. Nucl. Med.*, vol. 61, no. supplement 1, pp. 575, 2020.

[18] Kevin H Leung, Wael Marashdeh, Rick Wray, Saeed Ashrafinia, Martin G Pomper, Arman Rahmim, and Abhinav K Jha, “A physics-guided modular deep-learning based automated framework for tumor segmentation in PET,” *Phys. Med. Biol.*, vol. 65, no. 24, pp. 245032, 2020.