Optimizing imaging hardware for estimation tasks

Matthew A. Kupinski[a,b], Eric Clarkson[a,b], Kevin Gross[a] and John W. Hoppin[c]

[a]Optical Sciences Center
[b]Department of Radiology
[c]Program in Applied Mathematics
The University of Arizona, Tucson, AZ

ABSTRACT

Medical imaging is often performed for the purpose of estimating a clinically relevant parameter. For example, cardiologists are interested in the cardiac ejection fraction, the fraction of blood pumped out of the left ventricle at the end of each heart cycle. Even when the primary task of the imaging system is tumor detection, physicians frequently want to estimate parameters of the tumor, e.g. size and location. For signal-detection tasks, we advocate that the performance of an ideal observer be employed as the figure of merit for optimizing medical imaging hardware. We have examined the use of the minimum variance of the ideal, unbiased estimator as a figure of merit for hardware optimization. The minimum variance of the ideal, unbiased estimator can be calculated using the Fisher information matrix. To account for both image noise and object variability, we used a statistical method known as Markov-chain Monte Carlo. We employed a lumpy object model and simulated imaging systems to compute our figures of merit. We have demonstrated the use of this method in comparing imaging systems for estimation tasks.

Keywords: Image quality, estimation tasks, Fisher information

1. INTRODUCTION

Much of medical imaging research deals with developing better imaging systems or image-processing algorithms. Thus, whatever imaging specialty one works in, a quantitative measure of image quality must be defined. Task-based measures of image quality quantify the ability of an observer to perform a medically relevant task [1–3]. Tasks in medical imaging are typically tumor-detection or estimation tasks. Often, however, the task in medical imaging is a combination of a detection and an estimation task. For example, detecting a signal whose position within the patient in unknown can be posed as a detection and an estimation task, *i.e.*, detect the tumor and estimate its position.

It has been suggested that the performance of the Bayesian ideal observer be used as a performance metric for comparing imaging hardware configurations [1, 4]. However, computing the performance of the ideal observer is often computationally burdensome. Researchers have been forced to make strict assumptions about the imaging system in order to calculate the performance of the ideal observer. Examples of these assumptions include assuming flat or Gaussian backgrounds. We have recently shown it possible to compute the performance of the ideal observer using a much broader class of background (or object) models such as the lumpy or clustered lumpy object models [4]. The methods developed use Markov-chain Monte Carlo (MCMC) techniques to estimate the likelihood ratio, the ideal-observer decision variable.

For detection tasks we employ the Bayesian ideal observer which maximizes a number of figures-of-merit such as the area under the ROC curve. There is no clear analog of the ideal observer for estimation tasks. There are two inherent figures-of-merit for estimation tasks: bias and variance. Thus, if the task is to estimate a quantity θ, then we are unsure how the "ideal estimator" chooses to set the tradeoff between bias and variance. Kijewski *et al.* [5–8] advocate the use of the minimum possible variance of unbiased estimators as a figure of merit for optimizing imaging hardware. The performance of this observer is determined using the Fisher information

Corresponding author: M.A.K., E-mail: kupinski@radiology.arizona.edu, Address: Optical Sciences Center, 1630 E. University Blvd., Tucson, AZ 85721

matrix [9]. One possible definition of the ideal estimator is that unbiased observer that achieves the minimum possible variance for a particular estimation task. Future references to the ideal estimator in this paper will refer to this ideal, unbiased observer. As discussed in Section 4 there are certain theoretical disadvantages to using this definition of the ideal estimator.

In this paper we show that the performance of the ideal estimator can be computed using methods similar to those developed for signal-detection tasks. The performance of this ideal estimator can be employed to optimize imaging hardware for specific, medically-relevant estimation tasks. We will describe the method in Section 2, present a simulation in Section 3, and discuss the results of this simulation in Section 4.

2. METHOD

In a detection task, the ideal-observer test statistic, the likelihood ratio, is defined as,

$$\Lambda(\boldsymbol{g}) = \frac{p(\boldsymbol{g}|H_1)}{p(\boldsymbol{g}|H_0)}, \tag{1}$$

where $p(\boldsymbol{g}|H_i)$ is the probability of observer the image data $\boldsymbol{g}$ under the H_i hypothesis. Conventionally, H_1 signifies the signal-present hypothesis, while H_0 signifies the signal-absent hypothesis. The likelihood ratio can be rewritten as

$$\Lambda(\boldsymbol{g}) = \int d\boldsymbol{b}\, \Lambda_{\mathrm{BKE}}(\boldsymbol{g}|\boldsymbol{b}) p(\boldsymbol{b}|\boldsymbol{g}, H_0), \tag{2}$$

where the background-known-exactly likelihood ratio $\Lambda_{\mathrm{BKE}}(\boldsymbol{g}|\boldsymbol{b})$ is a known function based on the noise statistics of the imaging system (see Kupinski *et al.* [4] for derivation). If one could produce backgrounds $\boldsymbol{b}_l$ sampled from $p(\boldsymbol{b}|\boldsymbol{g}, H_0)$, then one could approximate the likelihood ratio as,

$$\hat{\Lambda}(\boldsymbol{g}) = \frac{1}{L} \sum_{l=1}^{L} \Lambda_{\mathrm{BKE}}(\boldsymbol{g}|\boldsymbol{b}_l). \tag{3}$$

We developed a method to approximate the likelihood ratio by sampling backgrounds $\boldsymbol{b}_l$ from the density $p(\boldsymbol{b}|\boldsymbol{g}, H_0)$ using MCMC techniques [4] and a certain class of background models. The technique is designed to work with lumpy object models or object models that are similar in nature, such as the clustered lumpy object model. Thus, if we could formulate other figures of merit in terms of integrals over $p(\boldsymbol{b}|\boldsymbol{g}, H_0)$ (or $p(\boldsymbol{b}|\boldsymbol{g}, H_1)$), then we could use these same techniques to estimate these other figures of merit.

We will employ the Cramér-Rao lower bound on the variance of unbiased estimators as the ideal estimator figure of merit. The Cramér-Rao bound is derived from the Fisher information matrix [9] whose elements are defined as,

$$J_{i,j} = \left\langle \frac{\partial}{\partial \theta_i} \log p(\boldsymbol{g}|\boldsymbol{\theta}) \frac{\partial}{\partial \theta_j} \log p(\boldsymbol{g}|\boldsymbol{\theta}) \right\rangle_{\boldsymbol{g}|\boldsymbol{\theta}}, \tag{4}$$

where $\boldsymbol{\theta}$ is the vector of parameters we wish to estimate. The Cramér-Rao bound for the ith element of $\boldsymbol{\theta}_i$ is determined by the diagonal elements of the inverse of the Fisher information matrix, *i.e.*,

$$\sigma_i^2 = [J^{-1}]_{i,i}, \tag{5}$$

where σ_i is a *bound* on the variance and not the variance of the ith parameter.

We have derived the Cramér-Rao bound for a broad class of estimation tasks. However, in this work we consider only one such estimation task. We consider the task of estimating the parameters of tumor such as the size and location. For example, detecting a signal with an unknown position can be posed as both a detection and an estimation task; determine whether a signal is present and estimate the position of the signal. For this situation, the likelihood is given by,

$$p(\boldsymbol{g}|\boldsymbol{\theta}) = p(\boldsymbol{g}|H_0)p(H_0) + p(\boldsymbol{g}|\boldsymbol{\theta}, H_1)p(H_1). \tag{6}$$

Note that the $\boldsymbol{\theta}$ appears only in the H_1 probability density function (PDF) $pr(\boldsymbol{g}|\boldsymbol{\theta}, H_1)$ since $\boldsymbol{\theta}$ characterizes the signal and is not necessary under the H_0 hypothesis. The Fisher information matrix (Eq. 4) requires the computation of partial derivatives of the log of the likelihood, *i.e.*, $\frac{\partial}{\partial \theta_i} \log p(\boldsymbol{g}|\boldsymbol{\theta})$. Substituting our likelihood expression (Eq. 6), we can expand this partial derivative,

$$\frac{\partial}{\partial \theta_i} \log p(\boldsymbol{g}|\boldsymbol{\theta}) \quad = \quad \frac{p(H_1)}{p(\boldsymbol{g}|H_0)p(H_0) + p(\boldsymbol{g}|\boldsymbol{\theta}, H_1)p(H_1)} \frac{\partial}{\partial \theta_i} p(\boldsymbol{g}|\boldsymbol{\theta}, H_1) \tag{7}$$

We can further marginalize the PDF $p(\boldsymbol{g}|\boldsymbol{\theta}, H_1)$ over the possible backgrounds $\boldsymbol{b}$ which results in,

$$\frac{\partial}{\partial \theta_i} p(\boldsymbol{g}|\boldsymbol{\theta}, H_1) = \int d\boldsymbol{b}\, p(\boldsymbol{b}) \frac{\partial}{\partial \theta_i} p(\boldsymbol{g}|\boldsymbol{b}, \boldsymbol{\theta}, H_1) \tag{8}$$

or equivalently,

$$\frac{\partial}{\partial \theta_i} p(\boldsymbol{g}|\boldsymbol{\theta}, H_1) = \int d\boldsymbol{b}\, p(\boldsymbol{b}) p(\boldsymbol{g}|\boldsymbol{b}, \boldsymbol{\theta}, H_1) \frac{\partial}{\partial \theta_i} \log p(\boldsymbol{g}|\boldsymbol{b}, \boldsymbol{\theta}, H_1). \tag{9}$$

Now, returning to our expression for the derivative of the log-likelihood, we can factor out the term $p(\boldsymbol{g}|\boldsymbol{\theta}, H_1)$ and apply Bayes' rule yielding,

$$\frac{\partial}{\partial \theta_i} \log p(\boldsymbol{g}|\boldsymbol{\theta}) \quad = \quad \frac{p(H_1)}{p(\boldsymbol{g}|\boldsymbol{\theta}, H_1) \left[\frac{1}{\Lambda(\boldsymbol{g})} p(H_0) + p(H_1)\right]} \int d\boldsymbol{b}\, p(\boldsymbol{b}) p(\boldsymbol{g}|\boldsymbol{b}, \boldsymbol{\theta}, H_1) \frac{\partial}{\partial \theta_i} \log p(\boldsymbol{g}|\boldsymbol{b}, \boldsymbol{\theta}, H_1) \tag{10}$$

$$\quad = \quad \frac{k\Lambda(\boldsymbol{g}))}{1 + k\Lambda(\boldsymbol{g})} \int d\boldsymbol{b}\, p(\boldsymbol{b}|\boldsymbol{g}, \boldsymbol{\theta}, H_1) \frac{\partial}{\partial \theta_i} \log p(\boldsymbol{g}|\boldsymbol{b}, \boldsymbol{\theta}, H_1), \tag{11}$$

where k is the ratio of the prior probabilities $p(H_1)/p(H_0)$. It should be noted that the term in front of the integral is a function of the likelihood ratio known as Bayes' decision rule. That is,

$$\frac{k\Lambda(\boldsymbol{g}))}{1 + k\Lambda(\boldsymbol{g})} = p(H_1|\boldsymbol{g}, \boldsymbol{\theta}) \tag{12}$$

is a monotonic transformation of the likelihood ratio.

We were able to approximate the likelihood ratio in Eq. 2 using MCMC techniques to sample backgrounds from the PDF $p(\boldsymbol{b}|\boldsymbol{g}, H_0)$. Equation 11 is a similar expression with an integral over the PDF $p(\boldsymbol{b}|\boldsymbol{g}, \boldsymbol{\theta}, H_1)$. To compute the expression shown in Eq. 11, we must use MCMC to evaluate each partial derivative of the vector $\boldsymbol{\theta}$ and the likelihood ratio $\Lambda(\boldsymbol{g})$ to produce an estimate of the partial derivative $\frac{\partial}{\partial \theta_j} \widehat{\log p(\boldsymbol{g}|\boldsymbol{\theta})}$.

To estimate the Fisher information matrix, we must employ a sample of images $\boldsymbol{g}_l$ to approximate the expectation in Eq. 4. Furthermore, we will employ our MCMC estimate of the partial derivative of the log-likelihood equation. Mathematically, this estimate is given by

$$\hat{J}_{i,j} = \frac{1}{L} \sum_{l=1}^{L} \frac{\partial}{\partial \theta_i} \widehat{\log p(\boldsymbol{g}_l|\boldsymbol{\theta})} \frac{\partial}{\partial \theta_j} \widehat{\log p(\boldsymbol{g}_l|\boldsymbol{\theta})}. \tag{13}$$

In practice, we use one Markov chain to compute each of the gradient terms and the likelihood ratio. The diagonal elements of the inverse of this estimated Fisher information matrix are used as the estimates of the Cramér-Rao bound.

3. SIMULATION

We performed a simulation study comparing three different imaging systems used to perform the same estimation task. This simulation study is very similar to the study performed in Kupinski *et al.* [4] except that the task is different. We employed a lumpy object model so that the objects being imaged were composed of a random number of Gaussian lump functions randomly placed within the field of view [10]. These objects were imaged

Imaging System	CRB (width)	CRB (amplitude)	CRB (x-pos)	CRB (y-pos)	AUC
A	3rd	3rd	2nd	3rd	1st (0.92)
B	2nd	2nd	3rd	2nd	2nd (0.88)
C	1st	1st	1st	1st	3rd (0.67)

Table 1. Summary of the results of the simulation study comparing imaging systems A, B, and C. The results are summarized in terms of rank. Imaging system A has high noise and high resolution, C has low noise and low resolution, and B is in between A and C. Also shown are previous results of an SKE ideal-observer signal detection study presented in terms of area under the ROC curve (AUC).

on three different simulated imaging systems with different noise and resolution properties. The sensitivity functions $h_m(\boldsymbol{r})$ for these imaging systems are Gaussian with varying widths and heights controlling the noise and resolution. Thus, the mean pixel value for the mth pixel is given by,

$$\overline{g}_m = \int_\infty d\boldsymbol{r}\, h_m(\boldsymbol{r}) f(\boldsymbol{r}) \tag{14}$$

$$= \sum_{n=1}^{N} \int_\infty d\boldsymbol{r}\, h_m(\boldsymbol{r}) L(\boldsymbol{r} - \boldsymbol{c}_n), \tag{15}$$

where N is the number of lumps in the background, $L(\cdot)$ is the lump function, and $\boldsymbol{c}_n$ are the centers of the lumps. Because $L(\cdot)$ and $h_m(\cdot)$ are both Gaussian functions, this integral can be computed analytically.

Exploiting the simple representation of our lumpy object model, we are able to generate a Markov chain with the stationary density being $p(\boldsymbol{b}|\boldsymbol{g}, \boldsymbol{\theta}, H_1)$, the density we wish to sample from. Each iteration of the chain changes the position of one lump slightly. There are also conditions in the code to add or remove lumps in the object model. The resultant chain of backgrounds forms our samples $\boldsymbol{b}_l$ from $p(\boldsymbol{b}|\boldsymbol{g}, \boldsymbol{\theta}, H_1)$. Thus, we approximate the gradient of the log likelihood (Eq. 11) via Monte Carlo integration. To compute this gradient we also must approximate the likelihood ratio $\Lambda(\boldsymbol{g})$ which we accomplish using the same Markov chain and the methods developed in Kupinski *et al.* [4].

The task we studied was that of estimating the position, width, and amplitude of a Gaussian signal that is present in approximately half of the sample images. We characterize the performances of our three imaging systems by the Cramér-Rao bounds on the parameters. To approximate both the likelihood ratio and the integral expression in Eq. 11 we ran the Markov chain for 300,000 iterations and used a burn-in time of 2000 iterations for each image. We used a total of 500 images for each imaging system to compute $\hat{J}$ (Eq. 13). Running in parallel on a cluster of eight computers, this computation took approximately two days to complete for one imaging system. Table 1 summarizes the results of this simulation study. Also shown in Table 1 are the results of the ideal-observer performing an SKE detection task summarized by the area under the ROC curve (AUC). Note that the performance of the ideal estimator has, except for one case, the opposite rankings of the ideal observer for an SKE detection task. This result and its implications are not yet fully understood.

4. DISCUSSION AND CONCLUSION

We have shown that computing the Cramér-Rao bounds on estimation parameters can be accomplished using Markov chain Monte Carlo techniques. This result is analogous to the techniques we developed for computing the AUC of the ideal observer performing detection tasks. We have further shown that the task of estimating parameters of the signal, such as location, can also be performed using MCMC even when the signal is not always present. We could have performed a signal-known-statistically detection-task study where the location is uncertain, yet statistically known. However, this type of study only reveals how well one can state that an image has or does not have a signal and reveals nothing about how well the observer can determine the location of the signal when it is present. This may present another method for computing observer performance in tasks where the signal is not known exactly.

There are potential down-sides to using Fisher information to characterize the performance of imaging systems. The Cramér-Rao bound is a lower bound on the variance of unbiased estimators. This bound characterizes the performance of an estimator that is unbiased and achieves this lower bound – called an efficient estimator. A biased estimator may exist that has a lower variance. If the bias is small, then using a biased estimator might be acceptable. Furthermore, it is uncertain in many tasks whether the lower bound could be reached by any estimator, *i.e.*, the efficient estimator may not exist for a certain task. It is known that if an efficient estimator exists, then the maximum-likelihood (ML) estimator is efficient. In addition, ML estimators asymptotically approach this lower bound, *i.e.*, they are asymptotically efficient. However, there remain many potential drawbacks to using Fisher information to characterize the performance of imaging systems.

The method we have presented shows promise in comparing imaging systems whose purpose is estimation. However, there are many other factors than need to be understood before the Fisher information is used to optimize imaging systems.

ACKNOWLEDGMENTS

This work was supported by NSF grant 9977116 and NIH grants P41 RR14304, KO1 CA87017-01, and R37 EB000803-12. This work was performed in collaboration with the Center for Gamma-ray Imaging (`http://gamma.radiology.arizona.edu`).

REFERENCES

1. H. H. Barrett, "Objective assessment of image quality: Effects of quantum noise and object variability," *Journal of the Optical Society of America A* **7**(7), pp. 1266–1278, 1990.
2. H. H. Barrett, J. L. Denny, R. F. Wagner, and K. J. Myers, "Objective assessment of image quality. II: Fisher information, Fourier crosstalk, and figures of merit for task performance," *Journal of the Optical Society of America A* **12**, pp. 834–852, 1995.
3. H. H. Barrett, C. K. Abbey, and E. Clarkson, "Objective assessment of image quality. III. ROC metrics, ideal observers, and likelihood-generating functions," *Journal of the Optical Society of America A* **15**, pp. 1520–1535, 1998.
4. M. A. Kupinski, J. W. Hoppin, E. Clarkson, and H. H. Barrett, "Ideal-observer computation in medical imaging using Markov-chain Monte Carlo," *JOSA A* , 2002 (to be published).
5. G. El Fakhri, S. C. Moore, and M. F. Kijewski, "Optimization of Ga-67 imaging for detection and estimation tasks: Dependence of imaging performance on spectral acquisition parameters," *Med. Phys.* **29**, pp. 1859–1866, 2002.
6. G. El Fakhri, M. F. Kijewski, and S. C. Moore, "Absolute activity quantitation from projections using an analytical approach: Comparison with iterative methods in Tc99m and I-123 brain SPECT," *IEEE Trans. Nucl. Sci.* **48**, pp. 768–773, 2001.
7. S. C. Moore, S. P. Muller, and M. F. Kijewski, "Lower bounds on the bias of tumor activity estimates from respiratory motion in combined PET and CT.," *J. Nucl. Med.* **42**, p. 385, 2001.
8. M. F. Kijewski, S. C. Moore, and S. P. Muller, "Nonuniform attenuation correction for estimation tasks in Tc99m brain SPECT.," *J. Nucl. Med.* **39**, p. 135, 1998.
9. H. L. Van Trees, *Detection, estimation, and modulation theory (Part I)*, Academic Press, New York, 1968.
10. J. P. Rolland and H. H. Barrett, "Effect of random background inhomogeneity on observer detection performance," *Journal of the Optical Society of America A* **9**, pp. 649–658, 1992.