

Emission Tomography from Compressed Data

Thomas J. Kragh and Alfred O. Hero

University of Michigan, Ann Arbor, MI 48109-2122

<tkragh, hero>@eecs.umich.edu

Abstract

In emission tomographic modalities such as SPECT or PET various regions of the detector space yield different amounts of information about the emission source. This paper develops a framework for exploring the tradeoffs between binning the detector space to reduce memory storage (complexity) and the associated loss in performance for image reconstruction tasks and detection tasks. We use high rate vector quantization theory to establish just how much relevant information can be preserved after compression of the emission measurement data. We illustrate our results for one dimensional deconvolution and two dimensional lesion detection in PET.

1. Introduction

While the framework in this paper is more generally applicable, the goal of this work is to quantify the fundamental tradeoff between performance and memory storage for emission tomography scanners whose measurement domain, e.g., photon count positions on detectors, are binned to a finite number of cells. Our motivation is to explore measurement-space quantization strategies for accelerating image reconstruction when the standard uniform detector binning method, i.e. scalar quantization, is impractical. For example, the Compton Scatter Single Photon Emission Tomography (C-SPECT) scanner being developed at the University of Michigan has an 8-dimensional measurement space generated by a pair of position, time and energy sensitive detectors [8,9]. For such high dimensional measurement space, fine uniform quantization over each measurement axis becomes impractical since the number of bins, i.e. quantization cells, increases exponentially in the number of dimensions [2]. To perform image reconstruction one must compute the transition probabilities $p(y|x)$ for an emission at position x in the image-domain to be detected with attribute y , consisting of position, energy and other attributes of the detected photons. In traditionally binned tomographic reconstruction, this enumeration of $p(y|x)$ over

N discrete image pixels and M detection bins results in an $M \times N$ system matrix of discretized transition probabilities.

For PET or mechanically collimated SPECT systems, the storage of the system matrix is feasible since the matrix sparsity is high (above 95% in most cases). However, for imaging modalities with non-sparse system matrices (such as C-SPECT), this is a major problem. For a "typical" 2-dimensional C-SPECT problem involving $O(10^6)$ detections and $O(100^2)$ reconstructed image pixels, the storage requirement for the system matrix can easily be in the range of dozens of gigabytes and in the terabyte range for the intended fully 3-dimensional implementation. An alternative which circumvents this curse of dimensionality is to perform list-mode processing, where the transition probability and associated likelihood function contribution are computed individually for each photon count detected in the continuous (unquantized) detector space [2, 12]. While this avoids the exponential complexity increase encountered by uniform binning strategies, list-mode reconstruction suffers from a linear complexity increase as a function of the number of counts, which is a major drawback for the new generation of highly sensitive scanners that will acquire many millions of counts over the scan time interval.

In this paper we formulate the detector binning problem in terms of rate constrained vector quantization (VQ). Here the rate of the encoder corresponds to the number of detector bins, i.e., the number of rows of the system matrix, and the vectors correspond to the vector of attributes (positions, energy, etc.) of the measured counts. Using a high rate analysis of VQ distortion, we derive asymptotic expressions for the loss in image reconstruction performance and the loss in detection performance for emission tomographic systems as a function of complexity, as measured by the number of cells in the optimal VQ. Using these expressions we trace out tradeoff curves which are used to quantify the loss in performance due to reducing the number of measurement bins. Under some standard assumptions on asymptotic cell shapes, our high rate analysis yields expressions for the optimal point density specifying the concentration of cells which minimizes performance loss. As the asymptotically optimal rate constrained VQ may itself be difficult to im-

plement in high dimensions, we also give asymptotic expressions for lattice quantizers which can be used to specify optimal binning in each of the measurement dimensions for a given complexity constraint.

2 VQ for Emission Tomography

Let $\underline{y}_i \in \mathbb{R}^k$ denote a vector of measured attributes, e.g. position, of the i -th detected photon in an emission tomography scanner. The sequence of attributes $\{\underline{y}_i\}_i$ detected over a certain time period correspond to a spatial Poisson process with intensity $\mu(\underline{y}) = \int p(\underline{y}|\underline{x})\lambda(\underline{x})d\underline{x}$ where \underline{x} is the emission position of the detected photon in the image-domain, λ is the image intensity, and $p(\underline{y}|\underline{x})$ is the probability of detecting at \underline{y} a photon emitted at \underline{x} . The intensity μ is assumed to be supported on a bounded set $S \subset \mathbb{R}^k$. As is customary we will use the voxelized approximation $\mu(\underline{y}) = \sum_{j=1}^N p(\underline{y}|\lambda_j)\lambda_j$ where λ_j is the integral of λ over the j -th voxel. The general objective is to make inferences about the voxelized intensity $\{\lambda_j\}_j$ based on the set of detected count attributes $\{\underline{y}_i\}_i$. As explained in the introduction, practical implementations require binning each of the measurements \underline{y}_i into one of a finite number of cells. Our aim is to capture the intrinsic loss incurred by binning by formulating binning as a vector quantization problem.

An M -point, k -dimensional vector quantizer (VQ) [4, 11] consists of a partition $\{S_1, \dots, S_M\}$ of a subset $S \subset \mathbb{R}^k$ into M unique quantization cells, and a codebook $C = \{\underline{y}_1, \dots, \underline{y}_M\}$ consisting of M quantizer values in \mathbb{R}^k . A k -dimensional vector \underline{y} with probability density $p(\underline{y})$ is quantized to one of M codebook values by a quantization rule $Q(\underline{y}) = \underline{y}_l : \underline{y} \in S_l, l = 1, \dots, M$.

The quantization rule is typically designed to minimize a function of the average error or *distortion* between a source \underline{y} and a quantized source $Q(\underline{y})$. The optimal VQ minimizes this distortion over all possible partitions and codebooks over \mathbb{R}^k . This minimum distortion can be used to quantify the fundamental tradeoff between performance (distortion) and complexity (rate), giving the so-called rate-distortion curve. For high but finite rate simple asymptotic expressions for the minimum distortion and the asymptotically optimal VQ have been obtained for many different types of distortion measures such as MSE [11], weighted MSE [10], and various detection criteria [5]. A distinguishing feature of the problem considered here is that the domain of the distortion measure is different from the domain of the measurements to which the VQ is applied. In particular, in tomographic reconstruction one is interested in quantizing photon counts in the measurement domain in order to minimize the distortion in an estimate of the image intensity.

2.1 Image Reconstruction

Here one is interested in estimating the intensity parameters $\theta = \{\lambda_j\}_j$. We can characterize the increase in MSE of an estimator of θ due to quantization by the trace of the difference between the continuous- and quantized-measurement Fisher information Matrix (FIM) of the spatial Poisson process [14],

$$\begin{aligned} \Delta \text{tr} &= \text{trace}(\mathbf{F}) - \text{trace}(\hat{\mathbf{F}}) \\ &= T \sum_{j=1}^N \sum_{l=1}^M \left[\int_{S_l} \frac{a_j^2(\underline{y})}{\mu(\underline{y})} d\underline{y} - \frac{a_{lj}^2}{\mu_l} \right] \geq 0 \quad (1) \end{aligned}$$

where $a_j(\underline{y})$ is the probability density of \underline{y} given the emission occurred in voxel j , $a_{lj} = \int_{S_l} a_j(\underline{y})d\underline{y}$, and $\mu_l = \int_{S_l} \mu(\underline{y})d\underline{y}$. For a fixed number M of VQ cells $\{S_l\}_{l=1}^M$, the search for the VQ which minimizes the loss in trace-FIM is related to A-optimal design in the theory of optimal experiment design [3]. The minimization of the trace difference is also indirectly related to minimization of the Cramèr-Rao Bound on the variance of an estimator of θ .

When the number of cells M is large and the corresponding quantization cells are small, and assuming that the density $p(\underline{y}|\lambda)$ is smooth over the extent of each cell, the difference term in (1) can be developed in a Taylor series about \underline{y}_l up to terms of order $o(\|\underline{y} - \underline{y}_l\|^2)$ [7],

$$\Delta \text{tr} \approx \frac{m(H)}{M^{2/k}} \int_S p(\underline{y}|\lambda) \frac{\sum_{j=1}^N \left\| \nabla_{\underline{y}} \frac{d}{d\lambda_j} \ln p(\underline{y}|\lambda) \right\|^2}{\zeta(\underline{y})^{2/k}} d\underline{y}, \quad (2)$$

where $d/d\lambda_j \ln p(\underline{y}|\lambda) = a_j(\underline{y})/p(\underline{y}|\lambda)$, the gradient operator $\nabla_{\underline{y}} = [\partial/\partial y_1, \dots, \partial/\partial y_k]^T$ is a k -dimensional column vector, $\|\bullet\|$ is the Euclidean norm, and $\zeta(\underline{y})$ is the ‘‘point density’’

$$\zeta(\underline{y}) \cong \frac{1}{MV(S_l)} \text{ if } \underline{y} \in S_l, \quad (3)$$

which defines the average number of cells per unit volume about the l th measurement space partition S_l . The derivation of the expression (2) applies to the restricted class of VQ's that have asymptotic cell shapes that are smooth deformations of a symmetric reference cell H , e.g. a hypercube in \mathbb{R}^k , having normalized moment of inertia (NMI) $m(H)$ [7]. The asymptotic expression (2) generalizes to weighted distortion criteria studied in [10] upon replacing the norm $\|\underline{z}\|^2 = \underline{z}^T \underline{z}$ with $\|\underline{z}\|_B^2 = \underline{z}^T B \underline{z}$, where B is a positive definite sensitivity matrix.

Using Hölder's inequality, the optimal point density function $\zeta^*(\underline{y})$ which minimizes the distortion in (2) takes the form

$$\zeta^*(\underline{y}) = \left\{ p(\underline{y}|\lambda) \sum_{j=1}^N \left\| \nabla_{\underline{y}} \frac{d}{d\lambda_j} \ln p(\underline{y}|\lambda) \right\|^2 \right\}^{\frac{k}{k+2}} \frac{1}{c_{k,N}} \quad (4)$$

where $c_{k,N}$ is a unit normalization factor (the integral over S of the term in $\{\bullet\}$). Substituting (4) into (2) gives an expression for least-possible loss in the FIM-trace of any M -point quantizer in \mathbb{R}^k whose cells are a scaled version of a symmetric reference cell H . In addition if the reference cell is the minimum NMI tessellating polytope in \mathbb{R}^k , then substituting $m(H) = m_k^*$ into (2) as well gives an expression for the least-possible trace-FIM distortion of any M -point quantizer in \mathbb{R}^k .

3 VQ for Minimum Discrimination Loss

A frequently more relevant metric in medical imaging is performance for a diagnostic task such as lesion detection [1, 6]. In this setting the objective is to decide between one of the following hypotheses

$$\begin{aligned} H_0 : p(\underline{y}) &= p_0(\underline{y}) = p(\underline{y}) \\ H_1 : p(\underline{y}) &= p_\epsilon(\underline{y}) = (1 - \epsilon)p(\underline{y}) + \epsilon p_\Delta(\underline{y}), \end{aligned}$$

where $p_\Delta(\underline{y})$ is a perturbation density reflecting how a lesion or other anomaly manifests changes in the measurements. For simplicity we have omitted the explicit dependence of these densities on λ . Here we will focus on the challenging case where $\epsilon \ll 1$ using the asymptotic analysis technique of Gupta [5] which we summarize below.

The performance of any detection algorithm is characterized by the probabilities of decision error. For n i.i.d. measurements $\{y_i\}_{i=1}^n$ Sanov's theorem specifies the false alarm probability α_n and the miss probability β_n for large n :

$$\alpha \cong e^{-n\mathcal{D}(p_\lambda \| p_0)} \quad (5)$$

$$\beta \cong e^{-n\mathcal{D}(p_\lambda \| p_1)}, \quad (6)$$

where $\mathcal{D}(q \| p)$ is the relative entropy (RE) between densities q and p , also known as discrimination and Kullback-Liebler divergence, p_λ is the *tilted* density, which is the geometric mixture of $p_0(\underline{y})$ and $p_1(\underline{y})$,

$$p_\lambda(\underline{y}) = \frac{p_0^{1-\lambda}(\underline{y})p_1^\lambda(\underline{y})}{\int_S p_0^{1-\lambda}(\underline{y})p_1^\lambda(\underline{y})d\underline{y}}$$

and $\lambda \in [0, 1]$ is implicitly defined by the Neyman-Pearson threshold $T = \mathcal{D}(p_\lambda \| p_1) - \mathcal{D}(p_\lambda \| p_0)$.

The RE between two continuous densities q, p is $\mathcal{D}(q \| p) = \int_S q(\underline{y}) \ln q(\underline{y})/p(\underline{y}) d\underline{y}$ while the RE between the quantized versions of these two densities $P = \{P_1, \dots, P_M\}$, $Q = \{Q_1, \dots, Q_M\}$ is $\mathcal{D}(Q \| P) = \sum_{i=1}^M Q_i \ln Q_i/P_i$. Using these representations, for small ϵ and $m = 0, 1$ we have the losses in RE due to quantization

$$\begin{aligned} \Delta \mathcal{D}(p_\lambda \| p_m) &= \mathcal{D}(p_\lambda \| p_m) - \mathcal{D}(P_\lambda \| P_m) \\ &= q_m(\lambda)\epsilon^2 \Delta D + o(\epsilon^3) \end{aligned} \quad (7)$$

where $q_m(\lambda) = \lambda^2, (1 - \lambda)^2$ if $m = 0, 1$, respectively, and

$$\Delta D = \frac{1}{2} \left[\int_S \frac{p_\Delta^2(\underline{y})}{p(\underline{y})} d\underline{y} - \sum_{i=1}^M \frac{P_{\Delta i}^2}{P_i} \right]. \quad (8)$$

Clearly the two losses ($m=0,1$) both increase in ΔD and thus choosing a VQ to minimize ΔD guarantees (for small ϵ) that the Sanov error exponents (5) and (6) will undergo minimum loss. Analogously to our study of the trace of the FIM, a similar asymptotic high rate analysis of ΔD results in the following expression for the loss in discrimination

$$\Delta D \approx \frac{m(H)}{2M^{2/k}} \int_S p(\underline{y}) \frac{\left\| \nabla_{\underline{y}} \frac{d}{d\epsilon} \ln p_\epsilon(\underline{y})|_{\epsilon=0} \right\|^2}{\zeta(\underline{y})^{2/k}} d\underline{y}. \quad (9)$$

Again, we can solve for the discrimination-optimal point density by applying Hölder's inequality to (9)

$$\zeta^*(\underline{y}) = \left\{ p(\underline{y}) \left\| \nabla_{\underline{y}} \frac{d}{d\epsilon} \ln p_\epsilon(\underline{y})|_{\epsilon=0} \right\|^2 \right\}^{\frac{k}{k+2}} \frac{1}{d_k} \quad (10)$$

where d_k is a unit normalization constant. This expression can be substituted into (9) to obtain the minimum high rate loss in discrimination ΔD .

3.1 Discrimination Loss for Lattice Quantizer

A lattice quantizer is a sub-optimal VQ that has complexity advantages with respect to non-lattice VQ. For a lattice quantizer, the M codevectors are uniformly spaced over the lattice support domain $S_L \supseteq S$, resulting in a constant point density $\zeta(\underline{y}) = 1/V(S_L)$. Substituting this constant point density into (9) results in an asymptotic high rate discrimination loss of

$$\begin{aligned} \Delta D &\approx \frac{m(H)}{2} \left(\frac{V(S_L)}{M} \right)^{2/k} \\ &\int_S p(\underline{y}) \left\| \nabla_{\underline{y}} \frac{d}{d\epsilon} \ln p_\epsilon(\underline{y})|_{\epsilon=0} \right\|^2 d\underline{y}. \end{aligned} \quad (11)$$

4 Discussion

The expressions (2) and (9) for the high rate loss in FIM and in discrimination, respectively, are similar to the Bennett's integral formula for the high rate MSE distortion [4, 11] for recovery of the continuous measurements from their quantized values. In particular, asymptotically the loss falls off at a rate of $M^{-2/k}$ with a constant term that depends on the measurement distributions, the quantizer cell shape $m(H)$, and point density $\zeta(\underline{y})$. It is important to observe that for the trace-FIM and discrimination distortion measures, the rate constant is proportional to the

integral of the gradient of the score function associated with estimating θ in the case of image reconstruction, and for estimating ϵ in the case of detection of an epsilon perturbation. The score function contains the information necessary to discriminate between values of θ . If this function is zero for some region S_o of the measurement space S then quantization over S_o is of no value for estimating θ . On the other hand, if the score function is a non-zero constant over S_o then it suffices to know that y came from S_o and quantization over S_o is again of no value. Thus the asymptotic expressions are consistent with intuition that one should only refine the quantization over those regions where the score function is not constant.

5 Numerical Examples

5.1 Loss in FIM for 1D Deconvolution

Let $\mu(y) = h(y) * \lambda(y)$ be a continuous measurement intensity, where λ is the unknown image intensity, h is a blurring or point spread function, and $*$ denotes convolution. The task is to estimate the N components of a discretized version of λ from M quantized measurements. A plot of the observed measurement density $p(y)$, the point-density function $\zeta(y)$, as well as the quantizer cell locations for a $M = 32$ level quantizer is shown in figure 1, and numerical values for a $M = 150$ level quantizer are given in table 1. Note that the trace-FIM quantizer places cells near regions of change in the measurement density rather than in proportion to the density as for the standard MSE-optimal quantizer.

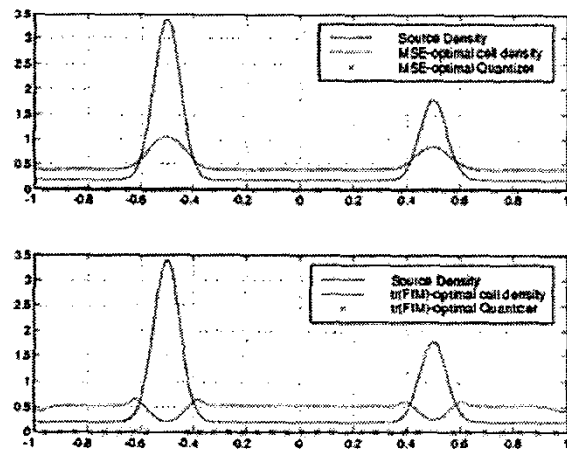


Figure 1. Comparison of optimal point densities and bin locations for a 32-bin, MSE-optimal and trace-FIM-optimal 1D quantizer.

Quantizer	$\frac{1}{N}\text{tr}(\hat{F})$	$\frac{1}{N}\text{tr}(\hat{F}^{-1})$	$\text{cond}\hat{F}$
uniform	28.5283	3.350×10^6	3.00×10^{10}
mse	27.1602	∞	3.25×10^{17}
trace-FIM	28.6227	6.128×10^6	1.02×10^{12}

Table 1. Numerical values for the trace of the Fisher information and inverse Fisher information matrices, for three different quantizer point densities, for an $M = 150$ level quantizer. Third column is the condition number which indicates a virtually singular FIM for the standard MSE optimal quantizer.

5.2 Discrimination Loss Calculation for 2D PET

Let the unknown parameters under hypotheses H_0, H_1 be θ and $\theta + \Delta\theta$, where $\theta \in \mathbb{R}^N$ is some non-random parameter vector and $\Delta\theta$ is a perturbation of θ . We will model the emission source density $\lambda(\underline{x})$ as an 100×100 -pixel, 2-dimensional (planar) image over a 50×50 cm field of view. The associated parameter vector θ corresponds to a tomographic slice through a simulated anthropomorphic phantom [13] while the perturbation parameter $\Delta\theta$ corresponds to a 3cm circular lesion in the bottom of the left lung, as shown in Figure 2.

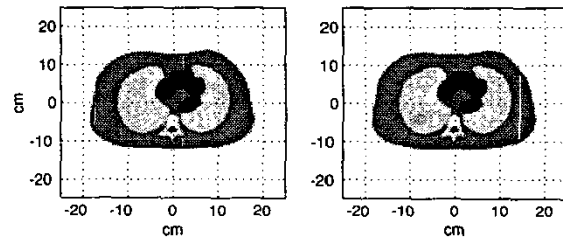


Figure 2. Simulated 2D PET emission source (left), emission source with 3cm circular lesion (right)

The total detected counts was set to 10^6 , with 10^3 counts from the lesion, for a perturbation parameter of $\epsilon = 10^{-3}$. Figure 3 shows the calculated discrimination for lattice quantized sinogram measurements along with the limiting case for continuous measurements. Figure 4 shows the calculated discrimination loss for lattice quantized sinogram measurements as well as the asymptotic discrimination loss (11). The best-possible discrimination loss for an optimal point-density VQ (10) is also shown for comparison.

6 Conclusions

This paper has presented a framework for studying the fundamental tradeoff between memory storage complexity

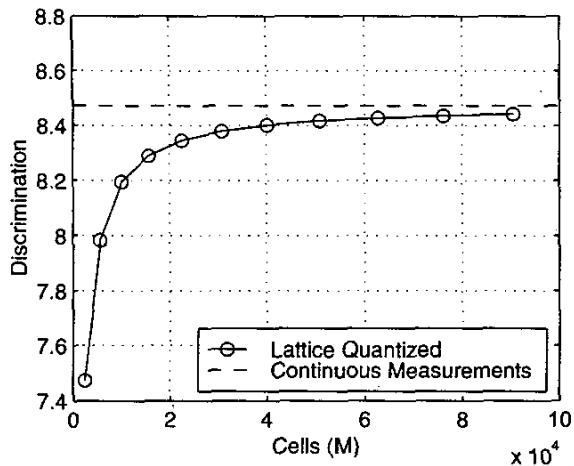


Figure 3. Discrimination versus number of sinogram cells

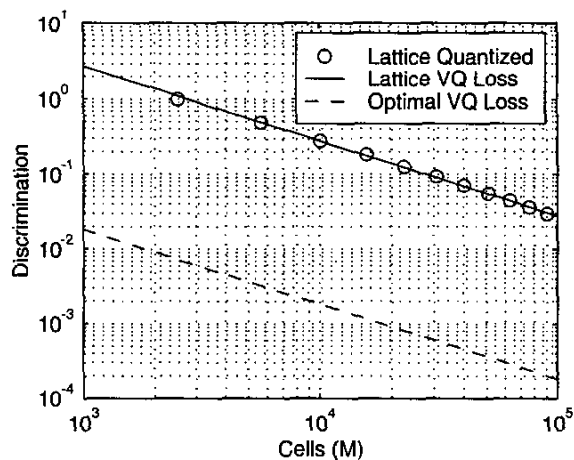


Figure 4. Calculated discrimination loss versus number of sinogram cells for a Lattice VQ, along with the asymptotic discrimination loss for both a Lattice and Optimal VQ.

and performance for emission tomography. This framework relies on the tools developed in high rate vector quantization theory. There are several open problems that are worthy of study. One such problem is the implementation of optimal vector quantization strategies for unknown image intensities. A proposed approach using adaptive VQ and the LBG algorithm is discussed in [7]. Another issue is how close one may come to attaining predicted distortion limits with a fixed rate lattice quantizer. Finally, the FIM and the Sanov error exponents are large n approximants to reconstruction and detection performance. It would be worthwhile to explore other tractable but non-asymptotic distortion measures.

References

- [1] 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. *J. Opt. Soc. Am. A*, 12(5):834–52, May 1995.
- [2] H. H. Barrett, T. White, and L. C. Parra. List-mode likelihood. *J. Opt. Soc. Am. A*, 14(1):2914–23, Nov. 1997.
- [3] V. V. Federov. *Theory of Optimal Experiments*. Academic Press, New York, 1972.
- [4] R. Gray and D. Neuhoff. Quantization. *IEEE Trans. Inform. Theory*, 44(6):2325–2383, Oct. 1998.
- [5] R. Gupta. *Quantization Strategies for Low-Power Communications*. PhD thesis, University of Michigan, Ann Arbor, MI, 1999.
- [6] C. Hua, S. J. Wilderman, T. J. Kragh, N. H. Clinthorne, and W. L. Rogers. The roles of observer models in evaluating Compton camera performance. In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, Oct. 1999.
- [7] T. J. Kragh. *Tradeoffs and Limitations in Statistically Based Image Reconstruction Problems*. PhD thesis, University of Michigan, Ann Arbor, MI, 2002.
- [8] J. W. LeBlanc. *A Compton Camera for Low Energy Gamma Ray Imaging in Nuclear Medicine Applications*. PhD thesis, University of Michigan, Ann Arbor, MI, 1999.
- [9] J. W. LeBlanc, N. H. Clinthorne, C.-H. Hua, E. Nygard, W. L. Rogers, D. K. Wehe, P. Weilhammer, and S. J. W. and. C-SPRINT: a prototype Compton camera system for low energy gamma ray imaging. *IEEE Tr. Nuc. Sci.*, 46(3):943–9, June 1998.
- [10] J. Li, N. Chaddha, and R. Gray. Asymptotic performance of vector quantizers with a perceptual distortion measure. *IEEE Tr. Info. Theory*, 45(4):1082–91, May 1999.
- [11] S. Na and D. Neuhoff. Bennett's integral for vector quantizers. *IEEE Trans. Inform. Theory*, 41(4), July 1995.
- [12] L. C. Parra and H. H. Barrett. List-mode likelihood - EM algorithm and noise estimation demonstrated on 2D-PET. *IEEE Tr. Med. Im.*, 17(2):228–35, 1998.
- [13] W. P. Segars. *Development of a new dynamic NURBS-based cardiac-torso (NCAT) phantom*. PhD thesis, The University of North Carolina, Chapel Hill, NC, May 2001.
- [14] D. L. Snyder and M. I. Miller. *Random Point Processes in Time and Space*. Springer-Verlag, New York, 1991.