



Published in final edited form as:

Proc SPIE. 2007 January 1; 6510: 651002-. doi:10.1117/12.716908.

Statistical Characterization of Radiological Images: Basic Principles and Recent Progress

Harrison H. Barrett and

College of Optical Sciences and Department of Radiology University of Arizona, Tucson AZ 85724

Kyle J. Myers

NIBIB/CDRH Laboratory for the Assessment of Medical Imaging Systems Food and Drug Administration, Rockville, MD 20850

Abstract

This paper surveys our current understanding of the statistical properties of radiological images and their effect on image quality. Attention is given to statistical descriptions needed to compute the performance of ideal or ideal-linear observers on detection and estimation tasks. The effects of measurement noise, random objects and random imaging system are analyzed by nested conditional averaging, leading to a three-term expansion of the data covariance matrix. Characteristic functionals are introduced to account for the object statistics, and it is shown how they can be used to compute the image statistics.

Keywords

Image quality; covariance; detection; Hotelling observer; Wiener estimator

1. INTRODUCTION

Image quality is an inherently statistical concept. It must be defined in terms of the average performance of some observer on some task of interest, and this performance is limited by many stochastic effects. For a fixed object and imaging system, the image is random because of Poisson noise in the radiation flux and excess noise in the detector. In addition, the object being imaged is never known exactly and must be treated as random, and in many cases the characteristics of the imaging system itself are random. The effect of these sources of randomness on task performance depends on the nature of the task and on the observer, but also on how the data are acquired and processed before being presented to the observer.

It is the goal of this paper to present a unified view of the statistics of radiological images and to survey recent progress in their mathematical characterization. Attention will focus on statistical properties needed to compute the performance of ideal or ideal-linear observers on detection and estimation tasks. Much of the discussion will follow the approach and notation of Barrett and Myers¹ and related papers.²⁻⁵

The treatment begins in Sec. 2 with a survey of the sources of randomness in radiological images, illustrating the general principles with examples from SPECT, digital radiography and ultrasound. In Sec. 3 we briefly discuss some mathematical descriptions of random images,

and in Sec. 4 we relate them to task performance. These general principles are applied to radiographic detectors in Sec. 5, where we also suggest some ways of improving task performance by rapid online processing of the detector signals. Sec. 6 focuses on the role of object randomness, with emphasis on models referred to as lumpy and clustered lumpy backgrounds. Sec. 7 shows how object statistics propagate through an imaging system and influence image statistics, and Sec. 8 summarizes the paper.

2. SOURCES OF RANDOMNESS

2.1. Mathematical basics and notation

Objects of interest in imaging are functions of two or three spatial variables, time and possibly other variables such as photon energy or direction of emission. For simplicity we consider only the spatial dependence and write the object as $f(\mathbf{r})$, where $\mathbf{r} = (x, y, z)$, or more abstractly as \mathbf{f} . If we wish to think in terms of vector spaces, the function \mathbf{f} is a vector in an infinite-dimensional Hilbert space.

Digital images, on the other hand, are vectors in a finite-dimensional vector space. If an imaging system acquires M measurements, $\{g_m, m = 1, \dots, M\}$, they can be arranged as an $M \times 1$ data vector denoted \mathbf{g} . The vector \mathbf{g} can refer to raw data or to a processed or reconstructed image in tomography.

Thus the imaging system can be construed as an operator \mathcal{H} that maps an infinite-dimensional vector to an M -dimensional one. Abstractly,

$$\mathbf{g} = \mathcal{H}\mathbf{f} + \mathbf{n}, \quad (2.1)$$

where \mathbf{n} is an $M \times 1$ random vector describing the noise in the measurements.

If the system (including the reconstruction algorithm if used) is linear, this mapping can be expressed as¹

$$g_m = \int d^3r f(\mathbf{r}) h_m(\mathbf{r}) + n_m, \quad (2.2)$$

where $h_m(\mathbf{r})$ expresses the sensitivity of the m^{th} measurement to radiation originating at point \mathbf{r} . Thus $h_m(\mathbf{r})$ can be called a point-response function (PRF).

The image \mathbf{g} is random because of the random noise \mathbf{n} , but also because the object \mathbf{f} is random. In addition, the PRF, or equivalently the system operator \mathcal{H} , can be random.

2.2. Examples

SPECT imaging systems use photon-counting detectors, and it is often said that the data are Poisson-distributed and uncorrelated. This statement is true only for a fixed object and a fixed imaging system, and only if we consider the raw projection data. Complicated correlations are generated by the reconstruction algorithm, and the object (the distribution of a radiopharmaceutical in SPECT) can vary from patient to patient, and from day to day for one patient. Patient motion is also a source of randomness, especially for long-exposure studies like SPECT.

The system in SPECT is always random if we think of scattering and attenuation of the radiation in the patient's body as part of the system. There may also be unknown mechanical errors or

other uncertainties in the system configuration that can be described as random. Moreover, any reconstruction algorithm must use an assumed or measured matrix representation of the system operator \mathcal{H} , and uncertainties in this matrix can be treated as random, even when they remain the same for multiple patients.

In digital radiography, the random noise results from a complicated cascade of processes in the detector; it is not Poisson, not white, and not stationary. In fact, it is not even Gaussian, in general, as we shall see below. The object is random because of patient-to-patient variations and patient motion, and the image-forming system is random if we include scatter within the patient's body in \mathcal{H} .

Ultrasound images suffer from Gaussian electronic noise and also from a form of coherent noise called speckle. The object is random because it varies from patient-to-patient, and also because ultrasound essentially measures the sound-wave reflectivity from complicated rough interfaces. Indeed, it is this roughness that leads to the speckle noise in the first place. The system in ultrasound is random because the wave incident on an interface and reflected from it must propagate through a complicated medium with spatially varying speed of sound, and this medium produces random phase distortions on the wave reaching the transducer.

3. CHARACTERIZATION OF RANDOM IMAGES

3.1. Random data vectors

In principle, random images can be characterized by a multivariate probability density function $\text{pr}(\mathbf{g})$, but the size and complexity of this description should not be underestimated. As noted above, \mathbf{g} is an $M \times 1$ random vector, and M can be huge. For a 1000×1000 detector used in direct imaging, $M = 10^6$, and for CT M can be 10^8 or more.

An equivalent – and equally huge – description is the characteristic function, defined by

$$\psi_{\mathbf{g}}(\boldsymbol{\xi}) \equiv \langle \exp(-2\pi i \boldsymbol{\xi}^t \mathbf{g}) \rangle = \int d\mathbf{g} \exp(-2\pi i \boldsymbol{\xi}^t \mathbf{g}) \text{pr}(\mathbf{g}), \quad (3.1)$$

where the angle brackets in the first form denote statistical expectation, and the integral in the second form is over all M components of \mathbf{g} . The frequency-like variable $\boldsymbol{\xi}$ (not to be confused with a spatial frequency) is another $M \times 1$ vector and $\boldsymbol{\xi}^t$ is its transpose; thus $\boldsymbol{\xi}^t \mathbf{g}$ is a scalar product in an MD space. The characteristic function can be viewed either as the expectation of the Fourier kernel or the MD Fourier transform of the PDF.

Lower-dimensional – and hence more tractable – descriptions of the random vector can be derived from either the PDF or the characteristic function. The mean $\bar{\mathbf{g}}$ is an $M \times 1$ vector with components given by

$$\bar{g}_m \equiv \int d\mathbf{g} \text{pr}(\mathbf{g}) g_m = \int dg_m \text{pr}(g_m) g_m = \frac{-1}{2\pi i} \left[\frac{\partial \psi_{\mathbf{g}}(\boldsymbol{\xi})}{\partial \xi_m} \right]_{\boldsymbol{\xi}=0}. \quad (3.2)$$

The integral in the first form is over all M components of \mathbf{g} , but the second form shows that only the univariate marginal PDF $\text{pr}(g_m)$ is needed to compute the mean component \bar{g}_m ; the third form shows how the mean can be obtained by differentiating the characteristic function.

Second moments can be defined similarly, either in terms of bivariate PDFs or second derivatives of the characteristic function. As we shall see below, the most important second moment is the covariance matrix \mathbf{K}_g , given in component form as

$$[\mathbf{K}_g]_{mm'} = \langle (g_m - \bar{g}_m)(g_{m'} - \bar{g}_{m'}) \rangle. \quad (3.3)$$

A more compact way of writing this result is in terms of outer products. If \mathbf{a} and \mathbf{b} are both $M \times 1$ vectors, their outer product $\mathbf{a}\mathbf{b}'$ is an $M \times M$ matrix with components $a_m b_{m'}$. Thus

$$\mathbf{K}_g = \langle (\mathbf{g} - \bar{\mathbf{g}})(\mathbf{g} - \bar{\mathbf{g}})' \rangle. \quad (3.4)$$

Note that the diagonal elements of the covariance matrix are the variances of the components.

3.2. Objects as random processes

An object $f(\mathbf{r})$ is a random function, also called a random process, and it is a vector in an infinite-dimensional Hilbert space. Thus a PDF $\text{pr}(\mathbf{f})$ would have to be infinite-dimensional. If, however, we fix the position vector, say at $\mathbf{r} = \mathbf{r}_0$, then $f(\mathbf{r}_0)$ is a scalar random variable, and the mean at that position can be computed from knowledge of the univariate density $\text{pr}[f(\mathbf{r}_0)]$; doing this for all \mathbf{r}_0 generates a nonrandom function $\bar{f}(\mathbf{r})$ which is the mean of the random process $f(\mathbf{r})$.

Similarly, from knowledge of the bivariate density $\text{pr}[f(\mathbf{r}), f(\mathbf{r}')]]$ we can define the autocovariance function, which is the continuous analog of the covariance matrix:

$$K_f(\mathbf{r}, \mathbf{r}') = \langle [f(\mathbf{r}) - \bar{f}(\mathbf{r})][f(\mathbf{r}') - \bar{f}(\mathbf{r}')]' \rangle, \quad (3.5)$$

This function of six variables can be regarded as the kernel of a covariance operator denoted as \mathcal{K}_f .

Marginal PDFs, means, covariances and all other statistical properties can be derived from the infinite-dimensional counterpart of the characteristic function defined in (3.1). For an MD random vector \mathbf{g} , the characteristic function is a function of another MD vector $\boldsymbol{\zeta}$; for a random function, M is infinite and the frequency-like vector $\boldsymbol{\zeta}$ must be replaced with a function denoted $s(\mathbf{r})$ or \mathbf{s} . Thus the characteristic function becomes a characteristic *functional* (function of a function), defined by analogy to (3.1) as

$$\Psi_{\mathbf{f}}(\mathbf{s}) \equiv \langle \exp(-2\pi i \mathbf{s}^\dagger \mathbf{f}) \rangle, \quad (3.6)$$

where \dagger denotes adjoint (analogous to transpose) and hence $\mathbf{s}^\dagger \mathbf{f}$ is the scalar product, $\int d^3r s(\mathbf{r}) f(\mathbf{r})$.

As we shall see in Sec. 6, there are many random processes for which the characteristic functional is known analytically.

3.3. Stationarity and Fourier descriptions

Largely for historical reasons, there is a tendency in imaging to work in terms of Fourier transforms of objects and images rather than the space-domain descriptions used above. Fourier transforms are especially natural (and can be justified in terms of group theory¹) for problems with translational symmetry, but imaging systems virtually never exhibit such symmetry.

In electrical engineering, from which so much of imaging theory springs, a linear electrical filter is translationally invariant if its response to an impulse is independent of the time that the impulse arrives. This condition requires merely that the components of the filter (resistors, capacitors, inductors) are themselves independent of time, so there is no preferred origin on the time axis. In that case the response to a general signal $f(t)$ is obtained by convolving it with the impulse response (and adding noise if appropriate), and the filter is said to be shift-invariant.

Similarly, the noise statistics in electrical engineering are invariant to shifts along the time axis if the physical mechanisms that generate the noise (currents and temperatures) are independent of time. A temporal random process $f(t)$ is said to be stationary (in the wide sense) if its mean is independent of time and its autocovariance function $K(t, t')$ depends only on the time difference $t - t'$.

In virtually every imaging system, however, there is a preferred *spatial* origin: the optical axis in a lens system, the center of the detector in digital radiography or the axis of rotation in tomography. Thus neither shift invariance nor stationarity can hold exactly, and great care must be exercised in invoking these properties even as approximations.

The general linear mapping described by the integral in (2.2) is never exactly a convolution, but there are two circumstances in which it is tempting to approximate it by one. First, in tomographic imaging where g_m is the grey level in the m^{th} voxel of a linear reconstruction, one might replace the integral in (2.2) by a sum ($\int d^3r \rightarrow \sum_n$) and assume that $h_m(\mathbf{r}) = h(\mathbf{r}_n - \mathbf{r}_m)$, in which case (2.2) is a discrete 3D convolution. Second, in digital radiography we might take $f(\mathbf{r})$ to be the 2D radiation distribution on the detector, rather than the transmission of the 3D object through which the radiation passed, and assume that the PRF is the same for all detector pixels, so that $h_m(\mathbf{r}) = h(\mathbf{r} - \mathbf{r}_m)$. In that case g_m in (2.1) can be regarded as a noisy sample of a 2D convolution integral. Both of these approximations are commonly made, at least implicitly, in the imaging literature, but the resulting errors are seldom discussed.

The conditions for spatial stationarity of the noise are also very difficult to meet. Interesting objects virtually never have a mean that is independent of location or an autocovariance function $K_f(\mathbf{r}, \mathbf{r}')$ that is just a function of $\Delta\mathbf{r} \equiv \mathbf{r} - \mathbf{r}'$. If we assume that these unrealistic conditions do hold, however, we can define a power spectral density or Wiener spectrum as the Fourier transform of the autocovariance function:

$$S_f(\rho) = \int d^3\Delta r K_f(\Delta\mathbf{r}) \exp(-2\pi i \rho \cdot \Delta\mathbf{r}). \quad (3.7)$$

This function of three variables describes the fluctuations of $f(\mathbf{r})$ in the 3D spatial-frequency domain, but only for stationary random processes.

A more general description, applicable to all random processes, is the Wigner distribution function, defined by

$$W_f(\mathbf{r}_0, \rho) = \int d^3\Delta r K_f(\mathbf{r}_0 + \frac{1}{2}\Delta\mathbf{r}, \mathbf{r}_0 - \frac{1}{2}\Delta\mathbf{r}) \exp(-2\pi i \rho \cdot \Delta\mathbf{r}). \quad (3.8)$$

This function of six variables describes the fluctuations jointly in space and spatial frequency, and it reduces to the power spectral density for stationary random processes. It is especially useful for approximately stationary random processes where $K_f(\mathbf{r}_0 + \frac{1}{2}\Delta\mathbf{r}, \mathbf{r}_0 - \frac{1}{2}\Delta\mathbf{r})$ is a slowly varying function of \mathbf{r}_0 .

For digital images, stationarity in the discrete image domain means that the covariance matrix is Toeplitz or block Toeplitz. An $M \times M$ matrix \mathbf{K} is Toeplitz if $K_{mm'}$ depends only on $m - m'$, where $1 \leq m \leq M$ and $1 \leq m' \leq M$. This assumption might hold for a 1D detector illuminated by a uniform flux of x rays in digital radiography. For a 2D detector, the matrix is block Toeplitz if $K_{mm'}$ depends only on $\mathbf{r}_m - \mathbf{r}_{m'}$, where \mathbf{r}_m is the 2D location of the m^{th} detector element.

In an attempt to apply Fourier methods to imaging, Toeplitz matrices are often approximated as circulant matrices by imposing a completely unphysical wrap-around as illustrated in Fig. 1.

The advantage of the wrap-around approximation is that circulant matrices are diagonalized by discrete Fourier transforms (DFTs). For 2D images we can define a 2D matrix \mathbf{D} such that $\mathbf{D}\mathbf{g} = \mathbf{G}$, where \mathbf{G} is the 2D DFT of \mathbf{g} . The covariance matrix of \mathbf{G} is given by [cf. (3.4)]

$$\mathbf{K}_G = \langle (\mathbf{G} - \overline{\mathbf{G}})(\mathbf{G} - \overline{\mathbf{G}})^\dagger \rangle = \mathbf{D}\mathbf{K}_g\mathbf{D}^\dagger, \quad (3.9)$$

where the dagger denotes the adjoint (complex conjugate transpose) of the matrix. If \mathbf{K}_g is circulant, then the covariance matrix of the DFT components is diagonal:

$$[\mathbf{K}_G]_{mm'} = \text{Var}(G_m) \delta_{mm'}. \quad (3.10)$$

The diagonal elements, which are the variances of the DFT components, are universally referred to in the digital radiography literature as a noise power spectrum. Moreover, it is always tacitly assumed that the off-diagonal elements of \mathbf{K}_G are zero by virtue of the circulant model, but the authors know of no example in the literature where this assumption has been verified.

3.4. Multiply stochastic random vectors

As noted in Sec. 2, \mathbf{g} is random because of random noise, random objects and possibly a random system. A general way of treating these multiple stochastic effects is the method of nested conditional averages, which we shall illustrate by considering just noise and object variability.

Let $\mathbf{T}(\mathbf{g})$ be some arbitrary function of \mathbf{g} . Its average is

$$\langle \mathbf{T}(\mathbf{g}) \rangle_{\mathbf{g}} = \langle \langle \mathbf{T}(\mathbf{g}) \rangle_{\mathbf{g}|\mathbf{f}} \rangle_{\mathbf{f}} \equiv \overline{\overline{\mathbf{T}}}, \quad (3.11)$$

where the double overbar denotes an average with respect to both measurement noise and object randomness. We can also define a single-bar average over just measurement noise:

$$\overline{\mathbf{T}}(f) \equiv \langle \mathbf{T}(\mathbf{g}) \rangle_{\mathbf{g}|\mathbf{f}}. \quad (3.12)$$

The covariance matrix on \mathbf{g} with both measurement noise and object variability is defined by

$$\mathbf{K}_g = \langle \langle [\mathbf{g} - \bar{\mathbf{g}}][\mathbf{g} - \bar{\mathbf{g}}]^t \rangle_{\mathbf{g}|\mathbf{f}} \rangle_{\mathbf{f}}. \quad (3.13)$$

Adding and subtracting $\bar{\mathbf{g}}(\mathbf{f})$ in each factor gives

$$\mathbf{K}_g = \langle \langle [\mathbf{g} - \bar{\mathbf{g}}(\mathbf{f}) + \bar{\mathbf{g}}(\mathbf{f}) - \bar{\mathbf{g}}][\mathbf{g} - \bar{\mathbf{g}}(\mathbf{f}) + \bar{\mathbf{g}}(\mathbf{f}) - \bar{\mathbf{g}}]^t \rangle_{\mathbf{g}|\mathbf{f}} \rangle_{\mathbf{f}}. \quad (3.14)$$

Since $\langle [\mathbf{g} - \bar{\mathbf{g}}(\mathbf{f})] \rangle_{\mathbf{g}|\mathbf{f}} = 0$, we see that

$$\mathbf{K}_g = \langle \langle [\mathbf{g} - \bar{\mathbf{g}}(\mathbf{f})][\mathbf{g} - \bar{\mathbf{g}}(\mathbf{f})]^t \rangle_{\mathbf{g}|\mathbf{f}} \rangle_{\mathbf{f}} + \langle \langle [\bar{\mathbf{g}}(\mathbf{f}) - \bar{\mathbf{g}}][\bar{\mathbf{g}}(\mathbf{f}) - \bar{\mathbf{g}}]^t \rangle_{\mathbf{f}} \rangle_{\mathbf{f}}. \quad (3.15)$$

The first term in this expression is just the noise covariance matrix \mathbf{K}_n averaged over \mathbf{f} . The second term represents object variability as seen in the mean image. Thus we define

$$\mathbf{K}_g \equiv \bar{\mathbf{K}}_g^{(noise)} + \mathbf{K}_g^{(obj)}. \quad (3.16)$$

This division of the overall covariance does not require any assumptions about the form of either $\text{pr}(\mathbf{g}|\mathbf{f})$ or $\text{pr}(\mathbf{f})$. In particular, it does not require that the noise be object-independent, and it does not require that either the noise or the object be Gaussian.

A similar decomposition with three terms arises if we also consider the system randomness, in which case we obtain⁵

$$\mathbf{K}_g \equiv \bar{\mathbf{K}}_g^{(noise)} + \bar{\mathbf{K}}_g^{(obj)} + \mathbf{K}_g^{(sys)}. \quad (3.17)$$

The first term in this equation vanishes if there is no measurement noise, the second vanishes if the object is not random, and the third vanishes if the system is not random. For more details, especially on the third term, see Ref. ⁵.

4. EFFECTS OF RANDOMNESS – TASK PERFORMANCE

An emerging consensus in medical imaging holds that image quality must be defined by the performance of some observer on some task of clinical or scientific interest, where the task refers to the information that is desired from the image and the observer is the means of extracting the information. This approach is known as objective or task-based assessment of image quality.¹⁻⁵

Tasks can be classification, estimation, or a combination of the two. In a classification task the goal is to label the object that produced the image as being in one of two or more classes. Equivalently, we can say that the task is to discriminate among two or more hypotheses. If there are only two classes or hypotheses, the classification task is said to be binary. An example of a binary classification task is detection of a tumor or other specified signal in a medical

image. In this case the classes are signal-absent and signal-present; the null hypothesis H_0 is that the signal is absent and the alternative hypothesis H_1 is that it is present.

In an estimation task, the goal is to assign numerical values to one or more parameters of interest. When a patient has a known tumor, for example, we might want to estimate its volume or its position. A joint detection/estimation task would be to detect the tumor and, if the decision is made that it is present, estimate its volume and location.

A binary classification task can always be performed by computing some scalar-valued function of the data, denoted $t(\mathbf{g})$ and referred to as a test statistic. If $t(\mathbf{g})$ is a linear function of \mathbf{g} , it is referred to as a linear discriminant function.

The test statistic is usually constructed so that it has higher values, on average, when H_1 is true than when H_0 is true, and in that case the decision can be made by comparing the test statistic to a threshold τ and deciding that H_1 is true whenever $t(\mathbf{g}) > \tau$. Of course this decision is not always correct, and we can have false positives (where $t(\mathbf{g}) > \tau$ but the signal is actually not present) or true positives (where again $t(\mathbf{g}) > \tau$ but now the signal is present). Increasing τ reduces the number of false-positive decisions but also reduces the number of true detections.

A generally applicable figure of merit (FOM) for binary classification tasks is the area under the receiver operating characteristic (ROC) curve, which is a plot of the true-positive fraction (TPF) vs. the false-positive fraction (FPF) as the threshold τ is varied. A common surrogate FOM, however, is a scalar separability measure or signal-to-noise ratio (SNR) defined by

$$\text{SNR}^2 \equiv \frac{[\langle t \rangle_1 - \langle t \rangle_0]^2}{\frac{1}{2}\text{Var}_1(t) + \frac{1}{2}\text{Var}_0(t)}, \quad (4.1)$$

where $\langle t \rangle_j$ is the expectation of the test statistic when hypothesis H_j is true ($j = 0, 1$) and $\text{Var}_j(t)$ is the corresponding variance. If the test statistic is normally distributed under both hypotheses, as it often is for linear discriminants, the area under the ROC curve is a simple monotonic function of the SNR.¹

For an estimation task where P parameters are to be estimated, we can assemble them into a $P \times 1$ vector θ , and an estimate of this parameter vector, derived from an image vector \mathbf{g} , is denoted $\hat{\theta}(\mathbf{g})$. A common figure of merit is the ensemble mean-square error defined by

$$\text{EMSE} \equiv \left\langle \left\| \hat{\theta}(\mathbf{g}) - \theta \right\|^2 \right\rangle, \quad (4.2)$$

where the angle brackets denote an average over all sources of randomness in \mathbf{g} and also over some prior probability density on θ itself.

The most common joint classification/estimation task is detection and localization, and the corresponding FOM is the area under the localization ROC (LROC) curve. This concept has recently been extended by Clarkson⁶ to the estimation ROC (EROC) curve.

4.1. Ideal and ideal-linear observers for binary classification

The ideal observer for a binary classification task can be defined as one that maximizes the area under the ROC curve (AUC). The test statistic for this observer is the likelihood ratio, defined by^{7,8}

$$\Lambda(\mathbf{g}) = \frac{\text{pr}(\mathbf{g}|H_1)}{\text{pr}(\mathbf{g}|H_0)}. \quad (4.3)$$

Equivalently, the ideal observer can use the logarithm of the likelihood ratio, given by

$$\lambda(\mathbf{g}) = \ln \Lambda(\mathbf{g}) = \ln \text{pr}(\mathbf{g}|H_1) - \ln \text{pr}(\mathbf{g}|H_0). \quad (4.4)$$

Comparing $\lambda(\mathbf{g})$ to $\ln \tau$ leads to the same decision as comparing $\Lambda(\mathbf{g})$ to τ and thus produces the same ROC curve and the same AUC.

The ideal test statistics $\Lambda(\mathbf{g})$ or $\lambda(\mathbf{g})$ are usually complicated nonlinear functions of the data \mathbf{g} , and computing them requires knowledge of the complete MD probability law of the data under both hypotheses. As we stressed above, M is huge in imaging problems, but nevertheless, there are many circumstances where the ideal observer is computationally tractable. The likelihood ratio can be computed if the data are multivariate normal (described by an MD Gaussian PDF), and in virtually any situation where the measurement noise dominates and the signal to be detected is nonrandom. It can also be computed by a method known as Markov-chain Monte Carlo or MCMC^{1,9} if a parametric model for the random object is available (see Sec. 6).

In many realistic cases, however, the PDFs $\text{pr}(\mathbf{g}|H_j)$ are unknown or too complicated to be useful. A common recourse is then the ideal linear or Hotelling observer,^{10–12} for which the test statistic is the linear discriminant that maximizes the SNR defined in (4.1). This optimal linear discriminant is given by

$$t_{Hot}(\mathbf{g}) = [\Delta\langle \mathbf{g} \rangle]^t \left[\frac{1}{2} \mathbf{K}_{\mathbf{g}|H_0} + \frac{1}{2} \mathbf{K}_{\mathbf{g}|H_1} \right]^{-1} \mathbf{g}, \quad (4.5)$$

where the superscript t denotes transpose, the angle brackets denote averages over all sources of randomness (background object, signal to be detected, system and measurement noise) and $\Delta\langle \mathbf{g} \rangle \equiv \langle \mathbf{g} \rangle_1 - \langle \mathbf{g} \rangle_0$ is the difference between the mean data vectors under the two hypotheses.

The Hotelling observer performs only linear operations on data, and it requires knowledge of the mean data vectors and covariance matrices under the two hypotheses rather than the full PDFs needed by the ideal observer. The Hotelling observer is ideal, in the sense of maximizing the area under the ROC curve, if the data are normally distributed with the same covariance matrix under the two hypotheses.

Note that the Hotelling discriminant, as we have defined it here, requires true ensemble means and covariances, in contrast to the familiar Fisher discriminant which uses means and covariances derived from samples. An $M \times M$ sample covariance is singular unless the number of samples exceeds M , so the Fisher discriminant is virtually never applicable to image data where each sample is an image and M is huge. In practice, we might use sample images to estimate the object term in the covariance expansion (3.17), but fundamentally the covariance matrix is defined by an average over an infinite ensemble.

The computational difficulties in estimating and inverting a large covariance matrix are often avoided by the use of a small set of features rather than raw image data. If the features are derived by passing the image data through a few linear operators called channels, then the ideal

linear discriminant operating in this much smaller data space is called the channelized Hotelling observer (CHO). The channels can be chosen so that they mimic the spatial-frequency-selective channels found in the human visual system, or they can be chosen to minimize the loss in detectability SNR attendant to the dimensionality reduction. Channelized linear observers that are designed to predict the performance of human observers are said to be *anthropomorphic*, while those that attempt to predict the performance of the true Hotelling observer operating on the image data are called *efficient*.

For much more on these topics, see Barrett and Myers¹ and the now copious literature on CHOs.

4.2. Ideal and ideal-linear observers for estimation

The EMSE defined in (4.2) is minimized by an ideal estimator which computes^{7,8}

$$\widehat{\theta}(\mathbf{g}) = \int d\theta \text{pr}(\theta|\mathbf{g})\theta = \frac{\int d\theta \text{pr}(\mathbf{g}|\theta) \text{pr}(\theta)\theta}{\int d\theta \text{pr}(\mathbf{g}|\theta) \text{pr}(\theta)}. \quad (4.6)$$

The first form shows that the ideal estimator computes the posterior mean of the parameter (*i.e.*, the mean with respect to the PDF on the parameter after the data are acquired). The second form, obtained by use of Bayes' theorem, shows that it equivalently computes the mean with respect to the likelihood $\text{pr}(\mathbf{g}|\theta)$ weighted by the prior $\text{pr}(\theta)$. As with the ideal observer for classification, the requisite PDFs are difficult to know, and the estimate is generally a nonlinear function of the data (unless the data are Gaussian).

The linear estimator that minimizes the EMSE is the generalized Wiener estimator (not to be confused with a Wiener filter), given by^{1,7,8}

$$\widehat{\theta}_w(\mathbf{g}) = \bar{\theta} + \mathbf{K}_{\theta\mathbf{g}} \mathbf{K}_{\mathbf{g}}^{-1} [\mathbf{g} - \langle \mathbf{g} \rangle], \quad (4.7)$$

where $\bar{\theta}$ is the prior mean of the parameter vector, $\langle \mathbf{g} \rangle$ is again the data vector averaged over all sources of randomness, and $\mathbf{K}_{\theta\mathbf{g}}$ is the $P \times M$ cross-covariance between the parameter vector being estimated (θ) and the data vector \mathbf{g} .

As with the Hotelling observer for a detection task, the Wiener estimator requires that we know the overall data covariance matrix, and it also requires the cross-covariance $\mathbf{K}_{\theta\mathbf{g}}$.

5. MEASUREMENT NOISE

We have seen that the covariance matrix plays a key role in linear estimators and classifiers, and that it can be divided into three terms as in (3.17). This section examines the first term in more detail; the following two sections do the same for the other two terms.

5.1. Photon counting

A prime example of a photon-counting detector is the Anger scintillation camera used for SPECT. In that application, the raw data form an $M \times 1$ vector \mathbf{g} where M is the number of projection angles \times the number of pixels in each projection. Conditional on a specific object and imaging system, the data are described rigorously by

$$\Pr(\mathbf{g}|\mathcal{H}, \mathbf{f}) = \prod_{m=1}^M \frac{[(\mathcal{H}\mathbf{f})_m]^{g_m}}{g_m!} \exp[-(\mathcal{H}\mathbf{f})_m], \quad (5.1)$$

where $(\mathcal{H}\mathbf{f})_m$ is the conditional mean of the data for one pixel at one projection angle. This Poisson law is exact even when we consider the detector quantum efficiency, scatter in the object or detector, background radiation, and the effects of the position-estimation algorithm and the energy window. We emphasize, however, that it is conditional on an object and a system.

The corresponding conditional covariance matrix is given by $\text{diag}\{\mathcal{H}\mathbf{f}\}$, where $\text{diag}\{\mathbf{x}\}$ denotes a diagonal matrix with vector \mathbf{x} along the diagonal. The diagonal elements in this case are just the means of the Poisson random variables.

With random objects and/or systems, the PDF is no longer Poisson, but nevertheless the first term in (3.17) remains diagonal; it has elements given by

$$\left[\overline{\mathbf{K}_g}^{(noise)} \right]_{mm} = \langle \langle \text{diag}\{\mathcal{H}\mathbf{f}\} \rangle_{\mathcal{H}} \rangle_{\mathcal{H}} = \text{diag}\{\overline{\mathcal{H}\mathbf{f}}\}, \quad (5.2)$$

Thus the diagonal elements of the noise covariance term are just the average object $\overline{\mathbf{f}}$ imaged through the average system $\overline{\mathcal{H}}$.

5.2. Integrating detectors

Integrating detectors as used in CT and digital radiography have much more complicated noise statistics than photon-counting detectors. The initial x-ray interaction with the detector is randomly either a Compton scatter or a photoelectric absorption which occurs at a random 3D position in the material. The amount of energy deposited at this initial interaction site is therefore random, but there can also be multiple subsequent interactions. If the initial interaction is Compton, the scattered photon can be scattered again or photoelectrically absorbed, and if it is photoelectric, the ensuing fluorescent x-ray photons can be absorbed or scattered. Thus the energy deposition is a complex spatial random process, which then acts as the energy source for production of secondary particles (optical photons in the case of scintillation detectors, electron-hole pairs) in the case of semiconductor detectors). The number of such secondaries produced at a given location is a stochastic function of the energy deposited there, and the secondary particles then propagate randomly to the external readout elements (photodetectors or electrodes). The readout elements themselves contribute additional randomness as a result of dark current or electronic noise.

The result of this complex sequence of events is that each absorbed x-ray photon contributes a pulse of random amplitude and random spatial distribution to the final, integrated image. The random variations in pulse amplitude increase the variance of the image, by the reciprocal of a quantity called the Swank factor, and the spatial distribution causes the noise term in the covariance matrix to have off-diagonal terms.

There is a large literature, much of it originating in the SPIE Medical Physics Conference, on analyzing the image statistics of x-ray images obtained with integrating detectors. Often the results are based on assumptions of spatial stationarity, but an alternative description based on nonstationary point process is also possible. The conditional mean vector and covariance matrix for fixed object and system can be calculated with all random effects taken into account,

and the subsequent average over random object and system is relatively straightforward because the conditional noise covariance is a linear functional of the x-ray fluence.¹³

There is probably no field of imaging for which the statistics of the measurement noise have been so thoroughly investigated and are so fully understood as in digital radiography. A recent publication that illustrates the state of the art and discusses practical computational issues is Satarivand *et al.*¹⁴

5.3. Photon counting with integrating detectors

If an integrating detector is operated at a rapid frame rate with a weak enough photon flux, it is possible to resolve the individual regions of energy deposition. From the pixel values associated with individual events, it is possible to estimate the coordinates of the initial interaction and the energy of the photon. Researchers at the University of Arizona^{15, 16} have been using this approach with integrating arrays of cadmium zinc telluride (CZT) semiconductor detectors for over a decade, and more recently it has been shown by several groups^{17–22} that the method also works well with scintillation detectors.

Fig. 2 shows an example from the work of Miller *et al.*,²² who used a lens with a high numerical aperture to couple the light from a columnar CsI scintillator to a low-noise CCD detector. Shown on the left is a single frame of image data with a ^{99m}Tc gamma-ray source. The image on the right was obtained with an algorithm that parses the data frame for clusters of pixels associated with a single absorbed gamma ray, reads out an $N \times N$ array of pixel values for each cluster, and applies a centroiding algorithm to estimate the center of the cluster. Note that all of the readout noise has disappeared, the blur from nonlocal energy deposition is gone and a substantial improvement in spatial resolution has been obtained. In essence, the low-resolution integrating detector has been converted to a high-resolution photon-counting detector by post-detection processing.

Further improvement in performance can be obtained by performing maximum-likelihood (ML) processing^{16, 21, 22} on each cluster of pixels. Fig. 3 shows an example of 3D position estimation in a CZT detector, and Fig. 4 shows the improvement in energy resolution that can be obtained.

Though these examples are for radioisotope sources in SPECT, recent advances in CCD detectors offer considerable promise that similar results will be obtainable for digital mammography and perhaps even CT. Current CCD devices can read out about 10^8 pixels per second, and the use of multiple CCD chips and multiple readout lines per chip can push this number to 10^9 or even 10^{10} pixels per second. At such rates a mammographic image could be acquired in a minute or less with well-resolved clusters of pixels for individual absorbed x rays.

Additional interesting statistical issues arise when the clusters do overlap, even with rapid frame rates. In that case we cannot identify the pixels associated with individual x-ray interactions, yet we do not have so many interactions contributing to each pixel that we can invoke the central-limit theorem and assume Gaussian statistics. Some initial investigations of the ideal observer for this intermediate case have been presented,²³ but further work is needed.

6. OBJECT STATISTICS

Mathematical models for describing random objects include parametric shape models; wavelet models and other filter-bank models; cuspy, kurtotic models; normal and log-normal texture models; filtered noise; and lumpy and clustered lumpy models. These approaches are reviewed in Chap. 8 of Barrett and Myers¹; here we concentrate on the lumpy models as a way of illustrating the mathematical description of random objects.

6.1. Generation of lumpy and clustered lumpy backgrounds

The term “lumpy background” originated in the work of Rolland,²⁴ who generated a random object by superimposing “lumps” of a predetermined shape. The resulting object was thus given by

$$f(\mathbf{r}) = \sum_{n=1}^N l(\mathbf{r} - \mathbf{r}_n), \quad (6.1)$$

where N is the number of lumps in the image and \mathbf{r}_n defines the center of the n^{th} lump. The random process is thus defined by $N + 1$ random parameters, the N vectors \mathbf{r}_n (which may be either 2D or 3D) and N itself.

The lump function $l(\mathbf{r})$ was a Gaussian spatial distribution in Rolland’s work, but other authors have used different functions. Examples of backgrounds with Gaussian lumps are shown in Fig. 5. By proper choice of parameters, these backgrounds can be made to resemble the textures seen in radionuclide images.

A useful extension of Rolland’s model is the clustered lumpy background devised by Bochud *et al.*²⁵ to represent the texture in mammograms. In that work, a cluster of identical blobs forms a “superblob”, and the final model is a superposition of superblobs:

$$f(\mathbf{r}) = \sum_{k=1}^{N_s} \sum_{n=1}^{N_k} l_k(\mathbf{r} - \mathbf{r}_{nk} - \mathbf{R}_k), \quad (6.2)$$

Examples are shown in Fig. 6.

6.2. Statistical properties of lumpy and clustered lumpy backgrounds

The free parameters in a lumpy background are the number of lumps N and the lump centers $\{\mathbf{r}_n\}$. The number of lumps is almost always taken to be a Poisson random variable, and centers are statistically independent, identically distributed, and described by a PDF given (for a 3D object) by

$$\text{pr}(\mathbf{r}_n) = \frac{b(\mathbf{r}_n)}{\int_S d^3r b(\mathbf{r})}, \quad (6.3)$$

where $b(\mathbf{r})$ is the mean number of lumps per unit volume and S is the support of the object. The denominator in (6.3) is the mean number of lumps in S , denoted \bar{N} .

The autocovariance function $K_f(\mathbf{r}, \mathbf{r}')$ of this object model can be computed for any choice of lump shape and $b(\mathbf{r})$. If $b(\mathbf{r})$ is a specified nonrandom function, the covariance properties are determined completely by that function and the shape of the lump function, but for more flexibility $b(\mathbf{r})$ can also be a random process.

If $b(\mathbf{r})$ is independent of position, the process is stationary within the support S , and if the lumps are Gaussian $K_f(\mathbf{r} - \mathbf{r}')$ is a Gaussian spatial function. It does not follow, however, that $f(\mathbf{r})$ is a Gaussian random process; the univariate PDF $\text{pr}[f(\mathbf{r})]$ can be far from Gaussian if the lumps are widely spaced. In the limit of large \bar{N} , however, many lumps overlap at each point

(analogously to the pixel clusters in an integrating detector) and the object PDF approaches a Gaussian by the central-limit theorem.

A complete description of the statistics of a lumpy background is the characteristic functional. For nonrandom $b(\mathbf{r})$, it is given by¹

$$\Psi_{\mathbf{f}}(\mathbf{s}) = \exp \left\{ -\bar{N} + \int_S d^3 r_n b(\mathbf{r}_n) \exp \left[-2\pi i \int_S d^3 r s(\mathbf{r}) l(\mathbf{r} - \mathbf{r}_n) \right] \right\}. \quad (6.4)$$

All statistical properties of the lumpy background are contained in this characteristic functional, and a similar expression can be given for the clustered lumpy background.

7. FROM OBJECT STATISTICS TO IMAGE STATISTICS

7.1. Linear transformation of means and covariances

For lumpy and clustered lumpy backgrounds, and many other random object models, we can presume that the mean $\bar{f}(\mathbf{r})$ and the autocovariance function $K_f(\mathbf{r}, \mathbf{r}')$ (or abstractly, $\bar{\mathbf{f}}$ and \mathcal{K}_f) are known analytically. We now show how to use this information to compute the mean vector and covariance matrix for the image.

If the imaging system, including the reconstruction algorithm if used, is linear and nonrandom, the mean image vector is given by

$$\bar{\mathbf{g}}_m = \int d^3 r h_m(\mathbf{r}) \bar{f}(\mathbf{r}) \text{ or } \bar{\mathbf{g}} = \mathcal{H} \bar{\mathbf{f}}, \quad (7.1)$$

where here the single overbar indicates an average over measurement noise only.

Similarly, the object term in the covariance expansion (3.17) is given by

$$\left[\overline{\mathbf{K}_{\mathbf{g}}^{(obj)}} \right]_{mm'} = \int d^3 r \int d^3 r' h_m(\mathbf{r}) K_f(\mathbf{r}, \mathbf{r}') h_m(\mathbf{r}') \text{ or } \overline{\mathbf{K}_{\mathbf{g}}^{(obj)}} = \mathcal{H} \mathcal{K}_f \mathcal{H}^\dagger, \quad (7.2)$$

where \mathcal{H}^\dagger is the adjoint or backprojection operator for the system.

Both (7.1) and (7.2) can be averaged further over \mathcal{H} if the system is random.

7.2. Linear transformation of the characteristic functional

Random processes for which the characteristic functional is known analytically include lumpy and clustered lumpy backgrounds, any Poisson random process (stationary or not), any Gaussian random process (complex or real, arbitrary covariance), fully developed speckle, certain texture models related to non-Gaussian speckle, and the image-plane irradiance in adaptive optics. These statistical descriptions of the object can be used to compute properties of the image if we can show how characteristic functionals are transformed by imaging systems.

Consider first a noise-free linear mapping $\mathbf{g} = \mathcal{H} \mathbf{f}$. The characteristic function of the random vector \mathbf{g} is given by

$$\psi_{\mathbf{g}}(\boldsymbol{\xi}) \equiv \langle \exp[-2\pi i \boldsymbol{\xi}^\dagger \mathbf{g}] \rangle = \langle \exp[-2\pi i \boldsymbol{\xi}^\dagger (\mathcal{H}\mathbf{f})] \rangle = \langle \exp[-2\pi i (\mathcal{H}^\dagger \boldsymbol{\xi})^\dagger \mathbf{f}] \rangle, \quad (7.3)$$

where the last step follows from the definition of the adjoint.¹ Thus

$$\psi_{\mathbf{g}}(\boldsymbol{\xi}) = \Psi_{\mathbf{f}}(\mathcal{H}^\dagger \boldsymbol{\xi}). \quad (7.4)$$

This is a very powerful result: If we know the characteristic *functional* for \mathbf{f} , we immediately have the characteristic *function* for $\mathcal{H}\mathbf{f}$, and therefore we know *all statistical properties* of the random vector (image) \mathbf{g} .

This result can be modified to include noise. If $\mathbf{g} = \mathcal{H}\mathbf{f} + \mathbf{n}$ and \mathbf{n} is independent of \mathbf{f} , the characteristic function for the data is

$$\psi_{\mathbf{g}}(\boldsymbol{\xi}) = \psi_{\mathbf{n}}(\boldsymbol{\xi}) \Psi_{\mathbf{f}}(\mathcal{H}^\dagger \boldsymbol{\xi}), \quad (7.5)$$

where $\psi_{\mathbf{n}}(\boldsymbol{\xi})$ is the characteristic function for the noise. In many cases, signal-independent noise is Gaussian and $\psi_{\mathbf{n}}(\boldsymbol{\xi})$ is easy to write down.

For Poisson noise, it is found that²⁶

$$\psi_{\mathbf{g}}(\boldsymbol{\xi}) = \Psi_{\mathbf{f}}[\mathcal{H}^\dagger \boldsymbol{\Gamma}(\boldsymbol{\xi})]. \quad (7.6)$$

where $\boldsymbol{\Gamma}$ is an operator defined by

$$[\boldsymbol{\Gamma}(\boldsymbol{\xi})]_m = \frac{-1 + \exp(-2\pi i \xi_m)}{-2\pi i}. \quad (7.7)$$

To summarize this section, if we know the mean and autocovariance function for an object random process and the system is nonrandom and linear, we can compute the mean image and the object term in the covariance by (7.1) and (7.2) respectively. If we know the characteristic functional for the object model, as we do in many cases, we can get the characteristic function for the data with signal-independent or Poisson noise from (7.5) or (7.6).

8. SUMMARY AND CONCLUSIONS

Image quality is all about image statistics, and images are random because of measurement noise, random objects and random imaging systems. For objective assessment of image quality with ideal observers, we need to know the probability density functions of the data for all object classes or all values of a parameter to be estimated. For ideal linear observers we need to know mean data vectors and covariance matrices.

This paper has surveyed the current state of our understanding of these statistical descriptors. The object was viewed as a nonstationary random process and the image as a finite random vector. The general method of nested conditional averaging was introduced and shown to lead

to an exact three-term decomposition of the covariance matrix, with terms accounting for measurement noise, object variability and system variability.

The noise term in the covariance was discussed for photon-counting detectors as in SPECT and integrating detectors as in digital radiography. It was shown that integrating detectors can function as photon counters if the radiation flux is low enough, and that maximum-likelihood methods can be applied in that case to improve spatial resolution, estimate the depth of interaction and energy of each absorbed photon and eliminate readout noise. Though most obviously applicable to radionuclide imaging, the prospects for applying these estimation methods in digital radiography are good.

The effect of the object randomness on the image statistics can be studied starting with either the autocovariance function of the object random process or with the characteristic functional – an infinite-dimensional generalization of the familiar characteristic function for random variables. Characteristic functionals are known analytically for many complicated, realistic object models, and they can be used to compute image statistics for an arbitrary linear system and either Poisson or Gaussian noise

Acknowledgments

This research was supported by the National Institutes of Health under grants R37 EB000803 and P41 EB002035.

References

1. Barrett, HH.; Myers, KJ. Foundations of Image Science. John Wiley and Sons; New York: 2004.
2. Barrett HH. Objective assessment of image quality: effects of quantum noise and object variability. *J Opt Soc Am A* 1990;7:1266–1278. [PubMed: 2370589]
3. Barrett HH, Denny JL, Wagner RF, Myers KJ. Objective assessment of image quality: II. Fisher information, Fourier crosstalk, and figures of merit for task performance. *J Opt Soc Am A* 1995;12:834–852.
4. Barrett HH, Abbey CK, Clarkson E. Objective assessment of image quality: III. ROC metrics, ideal observers and likelihood-generating functions. *J Opt Soc Am A* 1998;15:1520–1535.
5. Barrett HH, Myers KJ, Devaney N, Dainty JC. Objective assessment of image quality: IV. Application to adaptive optics. *J Opt Soc Am A* 2006;23:3080–3105.
6. Clarkson, EW. Estimation ROC curves and their corresponding ideal observers. *Proc. SPIE*; 2007. p. 03
7. Melsa, JL.; Cohn, DL. Decision and Estimation Theory. McGraw-Hill; New York: 1978.
8. Van Trees, HL. Detection, Estimation, and Modulation Theory. Vol. 1. Wiley; New York: 1968.
9. Kupinski MA, Hoppin JW, Clarkson E, Barrett HH. Ideal-observer computation in medical imaging with use of Markov-chain Monte Carlo. *J Opt Soc Am A* 2003;20:430–438.
10. Hotelling H. The generalization of Student's ratio. *Ann Math Stat* 1931;2:360–378.
11. Smith WE, Barrett HH. Hotelling trace criterion as a figure of merit for the optimization of imaging systems. *J Opt Soc Am A* 1986;3:717–725.
12. Fiete RD, Barrett HH, Smith WE, Myers KJ. The Hotelling trace criterion and its correlation with human observer performance. *J Opt Soc Am A* 1987;4:945–953. [PubMed: 3598746]
13. Barrett HH, Wagner RF, Myers KJ. Correlated point processes in radiological imaging. *Proc SPIE* 1997;3032:110–124.
14. Satarivand M, Hajdok G, Cunningham IA. Computational engine for development of complex cascaded models of signal and noise in x-ray imaging systems. *IEEE Transactions in Medical Imaging* 2005;24(2):211–222.
15. Barber HB, Barrett HH, Dereniak EL, Hartsough NE, Perry DL, Roberts PCT, Rogulski MM, Woolfenden JM, Young ET. A gamma-ray imager with multiplexer read-out for use in ultra-high-resolution brain SPECT. *IEEE Trans Med Imaging* 1993;40:1140–1144.

16. Marks DG, Barber HB, Barrett HH, Tueller J, Woolfenden JM. Improving performance of a CdZnTe imaging array by mapping the detector with gamma rays. *Nucl Instrum and Meth in Phys Res, A* 1999;428:102–112.
17. Taylor, S.; Barrett, H.; Zinn, K. Small animal SPECT using a lens-coupled CCD camera. Poster presentation, Academy of Molecular Imaging; March, 2004;
18. Beekman F, de Vree G. Photon-counting versus an integrating CCD-based gamma camera: important consequences for spatial resolution. *Physics in Medicine and Biology* 2005;50(12):N109–N119. [PubMed: 15930598]
19. Teo, K.; Shestakova, I.; Sun, M.; Barber, WC.; Hasegawa, BH.; Nagarkar, VV. Evaluation of a emccd detector for emission-transmission computed tomography. *IEEE NSS/MIC Conf. Rec*; 2005. p. 3050-3054.
20. Meng L. An Intensified EMCCD Camera for Low Energy Gamma Ray Imaging Applications. *Nuclear Science, IEEE Transaction on August*;2006 53:2376–2384.
21. Miller BW, Barber H, Barrett H, Shestakova I, Singh B, Nagarkar V. Single-photon spatial and energy resolution enhancement of a columnar CsI(Tl)/EMCCD gamma-camera using maximum-likelihood estimation. *Proc SPIE* 2006;6142:61421T.
22. Miller BW, Barber HB, Barrett HH, Chen L, Taylor SJ. Photon-counting gamma camera based on columnar CsI(Tl) optically coupled to a back-illuminated CCD. *Proc SPIE* 2007;6510:22.
23. Chen L, Barrett HH. Non-Gaussian noise in X-ray and gamma-ray detectors. *Proc SPIE* 2005;5745:366–376.
24. Rolland JP, Barrett HH. Effect of random background inhomogeneity on observer detection performance. *J Opt Soc Am A* 1992;9:649–658. [PubMed: 1588452]
25. Bochud FO, Verdun FR, Hessler C, Valley JF. Detectability on radiological images: The effect of the anatomical noise. *Proc SPIE* 1995;2436:156–164.
26. Clarkson E, Kupinski MA, Barrett HH. Transformation of characteristic functionals through imaging systems. *Opt Express* 2002;10(13):536–539. [PubMed: 19436394]

$$\begin{bmatrix} c & b & a & 0 & 0 & 0 & 0 & 0 & 0 \\ d & c & b & a & 0 & 0 & 0 & 0 & 0 \\ e & d & c & b & a & 0 & 0 & 0 & 0 \\ 0 & e & d & c & b & a & 0 & 0 & 0 \\ 0 & 0 & e & d & c & b & a & 0 & 0 \\ 0 & 0 & 0 & e & d & c & b & a & 0 \\ 0 & 0 & 0 & 0 & e & d & c & b & a \\ 0 & 0 & 0 & 0 & 0 & e & d & c & b \\ 0 & 0 & 0 & 0 & 0 & 0 & e & d & c \end{bmatrix}
 \quad
 \begin{bmatrix} c & b & a & 0 & 0 & 0 & 0 & e & d \\ d & c & b & a & 0 & 0 & 0 & 0 & e \\ e & d & c & b & a & 0 & 0 & 0 & 0 \\ 0 & e & d & c & b & a & 0 & 0 & 0 \\ 0 & 0 & e & d & c & b & a & 0 & 0 \\ 0 & 0 & 0 & e & d & c & b & a & 0 \\ 0 & 0 & 0 & 0 & e & d & c & b & a \\ a & 0 & 0 & 0 & 0 & e & d & c & b \\ b & a & 0 & 0 & 0 & 0 & e & d & c \end{bmatrix}$$

Figure 1.

Left: a 9×9 Toeplitz matrix that might represent the covariance matrix of a 9-element linear detector in digital radiography. Right: the corresponding circulant matrix obtained by pretending that the distances are computed modulo 9.

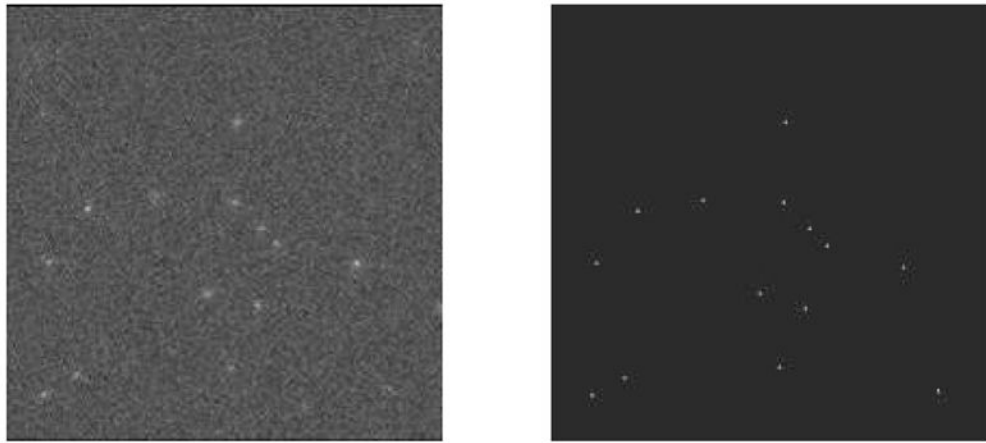


Figure 2.

Illustration of photon counting with a lens-coupled CCD detector. Left: A single frame of data showing clusters of pixels excited by individual 140 keV photons, superimposed on considerable readout noise. Right: Image obtained by event detection and centroiding for position estimation.

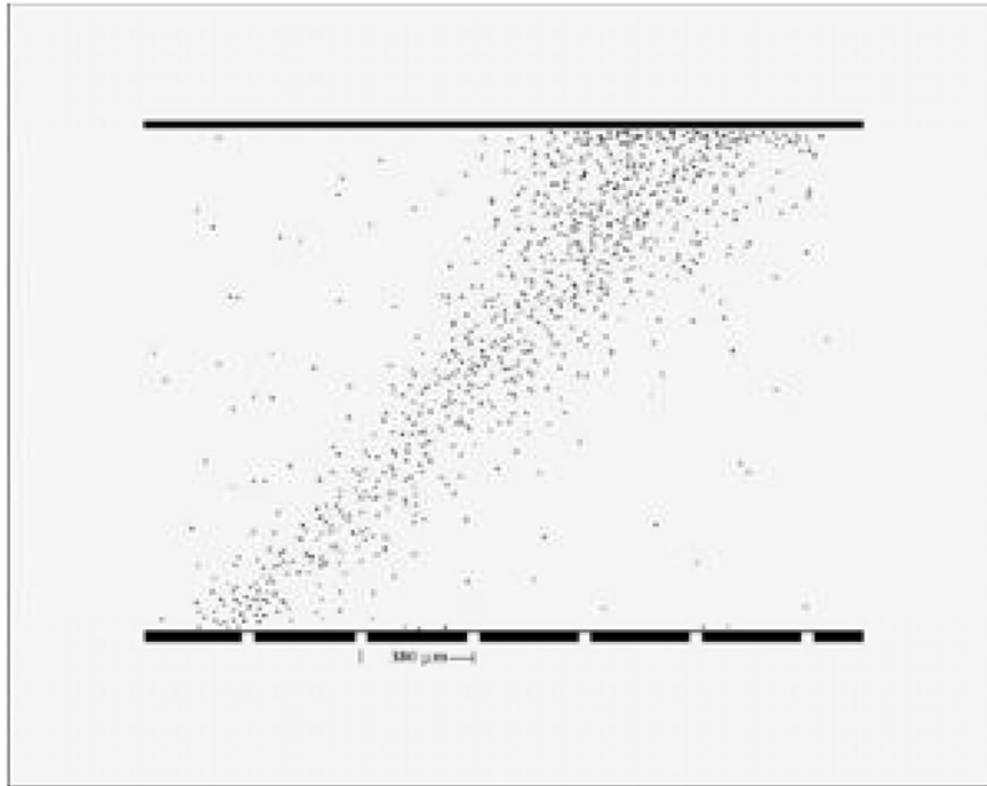


Figure 3.

Illustration of 3D position estimation in an integrating CZT detector array.¹⁶ The detector used a 2 mm thick slab of CZT, a 64×64 array of pixel electrodes on a $380 \mu\text{m}$ pitch; readout was by a multiplexer with a 1 ms frame time. The illumination for this figure was a collimated beam of 140 keV photons incident at 45 degrees to the detector face.

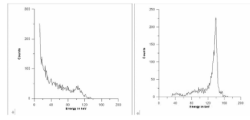


Figure 4. Illustration of improvement in energy resolution obtained by ML processing with an integrating CZT array.¹⁶ Left: Pulse-height spectrum for a single pixel. Right: ML energy estimate. The detector in this case used a 48×48 array of pixels on a $125 \mu\text{m}$ pitch.

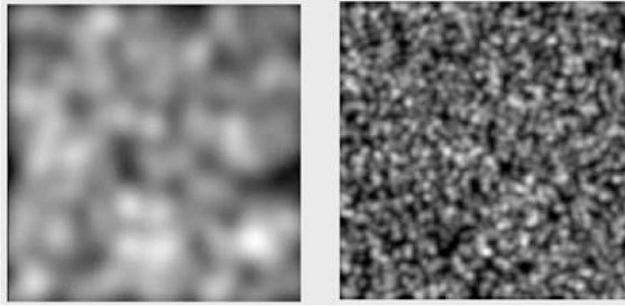


Figure 5.
Illustration of lumpy backgrounds with Gaussian lumps

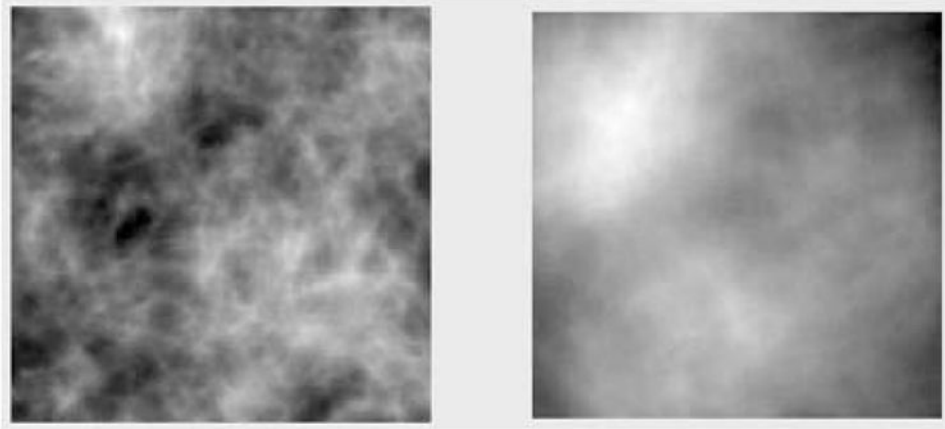


Figure 6.
Illustration of clustered lumpy backgrounds