On Bayesian Image Reconstruction from Projections: Uniform and Nonuniform *A Priori* Source Information

ZHENGRONG LIANG, MEMBER, IEEE, RONALD JASZCZAK, MEMBER, IEEE, AND KIM GREER

Abstract—**A method for Bayesian image reconstruction from projections is presented. The method incorporates *a priori* uniform or nonuniform source distribution probabilistic information and data fluctuations of a Poisson nature. The *a priori* uniform and nonuniform source distributions are modeled in terms of *a priori* source probability density functions. Maximum *a posteriori* probability solutions are given, as determined by a system of equations. Iterative Bayesian imaging algorithms for the solutions are derived, by treating the *a priori* source information as additional constraints to supplement the data likelihood information, via an expectation maximization technique. Comparisons of the *a priori* uniform and nonuniform Bayesian algorithms to the maximum likelihood algorithm are carried out using computer generated noise-free and Poisson randomized projections. Improvement in image reconstruction from projections with the Bayesian algorithms is demonstrated. Superior results are obtained using the *a priori* nonuniform source distribution.**

INTRODUCTION

IT has been recognized that Bayesian analysis has the advantage, in image processing, of considering *both* the statistical nature of data measurements and the *a priori* source distribution information in terms of probability density functions [1]–[4]. Most previous Bayesian work in image processing assumed a Gaussian *a priori* probability distribution to characterize the source distribution and data noise. The Gibbs, Poisson, and Gamma *a priori* source distributions have also been considered in the literature [2], [3], [5], [6]. Recently we proposed a statistical model of *a priori* source distribution [7], [8], intended to reflect the intrinsic probabilistic information of the source distribution. By incorporating *a priori* uniform and nonuniform probability distributions into the statistical model, two image information functions were obtained [9]. Under a principle of maximum *a priori* probability, the two image information functions strongly resemble the entropy forms defined by Shannon [10] and Kullback *et al.* [11], respectively. Since the assumptions made in the statistical model are very general, it is valid and applicable in many image processing problems. However, an appropriate estimate of the *a priori* source information must be developed for each specific application. The principle objectives of this paper are as follows:

1) to develop a Bayesian image processing (BIP) algorithm which incorporates the *a priori* image information described above;

2) to demonstrate the validity of the *a priori* image information in selected image processing problems;

3) to compare the BIP algorithms and the maximum likelihood (ML) algorithm [12], [13] in the problem of image reconstruction from computer generated projections.

It is noted that the iterative BIP approach via the expectation maximization (EM) technique [14] presented in this paper uses the *a priori* source information as additional constraints to supplement the data likelihood information. A BIP approach via the recursive Picard technique [15], which treats the data likelihood information as additional constraints to supplement the *a priori* image information, is in preparation [16].

A PRIORI IMAGE INFORMATION AND BAYESIAN ALGORITHMS

The source distribution region is, as usual in digital image processing, divided into J source elements (or voxels). Each voxel has an average value over its volume, $\{\phi_j\}, j = 1, 2, \cdots, J$. In nuclear isotope imaging, ϕ_j stands for the Gamma photon emission from voxel j per unit time at time = 0 (at time = 0 means that the decay of the nuclear isotope should be considered and corrected, if necessary); in X-ray imaging, it represents the attenuation density of voxel j; and in optical picture processing, it is the radiance value of voxel j. For simplicity, in the following sections, ϕ_j is referred to generally as the density of voxel j or "quanta" in voxel j.

If the voxel densities $\{\phi_j\}$ are quantized into density units (or photon "quanta"), then ϕ_j represents the number of quanta in voxel j. If the total number of quanta $N = \Sigma_j \phi_j$ can be assumed to be fixed, the image density distribution can then be characterized as a multinomial random process in which the N quanta distribute randomly over the J voxels. The form of this distribution is well known in the field of applied mathematics (see [17], [18], for example). In this paper, we use this distribution to model the *a priori* source information and apply it to the

Manuscript received April 7, 1988; revised March 15, 1989. This work was supported by the National Cancer Institute, DHHS, under Grant CA33541, and by the Duke–North Carolina NSF Engineering Research Center (CDR-8622201).

The authors are with the Department of Radiology, Nuclear Medicine Division, Duke University Medical Center, Durham, NC 27710.

IEEE Log Number 8928108.

problem of image reconstruction from projections. Let p_j represent the *a priori* probability of a quantum falling in voxel j. The *a priori* source distribution probability is then expressed as [6]–[8]

$$P(\Phi) = \frac{N!}{\prod_j \phi_j!} \prod_{j=1}^{J} [p_j]^{\phi_j} \tag{1}$$

where Φ is the voxel vector $\Phi = \{\phi_j\}$.

The *a priori* probability $P(\Phi)$ of (1) reflects a statistical random process of image density distribution given the *a priori* probabilistic distribution information $\{p_j\}$. The multinomial law of (1) has been widely used in discrete data analysis with different interpretations about the variables $\{\phi_j\}$ and $\{p_j\}$ [15], [17]–[20]. When $p_j = 1/J =$ constant, (1) reduces to a normalized version of the image distribution function that Frieden used [21]

$$W(\Phi) = N! \Big/ \prod_j \phi_j!. \tag{2}$$

Therefore, (1) is a more general formulation than Frieden's. The distribution function $W(\Phi)$ has been widely used in image processing problems [22]–[25]. An interpretation of $P(\Phi)$ and $W(\Phi)$ can be made as follows: the Boltzmann law of (2) represents the number of ways that a typical distribution $\{\phi_j\}$ can occur assuming that the occurrence of a quantum in a voxel does not affect the possible location of any other quantum [19], [22], [26]; by referring to the multinomial law of (1), (2) may be interpreted as an *a priori* source distribution probability assuming that each quantum has the same probability (or $p_j = 1/J$) of occurring in any voxel [7]; using this assumption and (1), a constant of proportionality (or normalization) can be derived as follows:

$$\prod_j [p_j]^{\phi_j} = (1/J)^{\Sigma_j \phi_j} = J^{-N}. \tag{3}$$

The number of ways [i.e., $W(\Phi)$] that the quanta $\{\phi_j\}$ may be distributed within the J voxels multiplied by this constant of proportionality can then be interpreted as a probability. Since the underlying assumptions of the quantization of image density and the random distribution of the quanta made in (1) are very general, $P(\Phi)$ can also be applied to many fields, if the *a priori* probability $\{p_j\}$ can be appropriately estimated [6]–[9], [27], [28].

By the definition of probability

$$p_j = \lim_{N \to \infty} \phi_j / N \tag{4}$$

(i.e., with very large number N, there are ϕ_j quanta falling into voxel j), it can be assumed that the *a priori* probability may be approximated as:

$$p_j = \bar{\phi}_j / N \tag{5}$$

where $\bar{\phi}_j$ represents the *a priori* mean number of quanta in voxel j (or *a priori* mean value of voxel j). The estimation of $\{\bar{\phi}_j\}$ is quite important for optimal solution $\{\phi_j\}$ given the measured data [2], [6]–[9], [25], [29], [30] and will be discussed in detail in a later section and in Appendix A.

If a maximum *a priori* probability principle applies, then maximizing the function $P(\Phi)$ is equivalent to maximizing the log function

$$\begin{aligned} H(\Phi) &= \ln P(\Phi) = \ln (N!) \\ &\quad + \sum_j [\phi_j \ln (p_j) - \ln (\phi_j!)] \\ &= \ln (N!) + \sum_j [\phi_j \ln (\bar{\phi}_j) \\ &\quad - \phi_j \ln (N) - \ln (\phi_j!)]. \end{aligned} \tag{6}$$

By use of the Stirling formula

$$\ln (K!) \approx K \ln (K) - K + \tfrac{1}{2} \ln (2\pi K). \tag{7}$$

For large K, we can neglect the term $[1/2 \ln (2\pi K)]$. Using this approximation and the constraint of $N = \Sigma_j \phi_j$, function (6) becomes

$$H(\Phi) = - \sum_j \phi_j \ln (\phi_j / \bar{\phi}_j). \tag{8}$$

Function (8) is the image information function containing the *a priori* image density distribution probabilistic information (5). It strongly resembles the entropy form defined by Kullback and Leibler [11] and has the defined contents of image density $\{\phi_j\}$ and the *a priori* mean information $\{\bar{\phi}_j\}$. Kullback *et al.* introduced the expression given by (8) in terms of probability p_j (instead of ϕ_j) as a measure of the "information for discrimination" between two probability distributions $\{p_j\}$ and $\{\bar{p}_j\}$ [11]. The information expression of Kullback *et al.* is widely used in information theory and is referred to as physical relative (or cross) entropy in statistical mechanics. Discussions and interpretations of the relative entropy expression of Kullback *et al.* have been given in the literature [25], [26], [31]. If the *a priori* probability distribution $\{p_j\}$ is uniform, i.e., $\bar{\phi}_j = \bar{\phi} = N/J$, function (8) reduces to

$$H(\Phi) = - \sum_j \phi_j \ln (\phi_j) + N \ln (N/J). \tag{9}$$

This expression is the same as Frieden's image entropy function except for a constant [21]. Function (9) strongly resembles the entropy form defined by Shannon [10] plus a constant and has the defined contents of image density $\{\phi_j\}$ and the assumption of uniform *a priori* probability distribution $\{p_j = 1/J\}$. The expression $-\Sigma_j p_j \ln (p_j)$ was introduced by Shannon as a measure of the missing information, or uncertainty, in a probability distribution $\{p_j\}$ and is generally used to represent the physical entropy in statistical mechanics.

The information functions (8) and (9) reflect two extreme cases: without any other constraints, function (9) specifies a uniform image and function (8) the mean image that one wishes to reconstruct. The mean values $\{\bar{\phi}_j\}$ in function (8) will be estimated and updated (see Appendix A), as shown below in the EM iterative procedure, from initial uniform to nonuniform in a very general man-

ner. Because of the nature of this approach, it can be applied to many image processing problems.

A Bayesian analysis, derived using the image information functions (8) and (9), is given below using Bayes' Law [32]

$$P(\Phi|\boldsymbol{Y}) = P(\boldsymbol{Y}|\Phi)P(\Phi)/P(\boldsymbol{Y}) \tag{10}$$

and a maximum *a posteriori* probability solution Φ^* is determined by maximizing the Bayesian function $g(\Phi)$

$$g(\Phi) = \ln P(\Phi|\boldsymbol{Y}) = \ln P(\boldsymbol{Y}|\Phi) + \ln P(\Phi) - \ln P(\boldsymbol{Y}) \tag{11}$$

where Φ is the source vector as defined previously and $\boldsymbol{Y}$ is the data vector. The measurement probability $P(\boldsymbol{Y}|\Phi)$ in (10) reflects the statistical nature of data fluctuations. The Poisson nature of photon measurements is reflected in $P(\boldsymbol{Y}|\Phi)$. If each data element Y_i obeys Poisson statistics with the mean $\Sigma_j R_{ij}\phi_j$ and all the data elements $\{Y_j\}$ are uncorrelated, the measurement probability is [2], [12], [32]

$$P(\boldsymbol{Y}|\Phi) = \prod_{i=1}^{I} \exp\left(-\sum_j R_{ij}\phi_j\right)\left(\sum_j R_{ij}\phi_j\right)^{Y_i} \Big/ Y_i! \tag{12}$$

where R_{ij} is the probability of receiving a quantum from voxel j for measurement point i [12], [13]. In image reconstruction from projections, R_{ij} is often referred to as an element of the projection matrix [1], [4]. If the projection matrix is properly normalized, each element represents the probability. Similarly in image restoration applications, $\{R_{ij}\}$ is referred to as the point spread function (PSF) of the imaging system [33] if the PSF is normalized.

It might be easier to understand the image distribution of (1) from (12). Let $R_{ij} = \delta_{ij}$ (i.e., assuming that each photon is detected and localized, $Y_i = \phi_j$), (12) reduces to

$$P(\boldsymbol{Y} = \Phi|\Phi) = \prod_i e^{-\phi_j}(\phi_j)^{\phi_j}/(\phi_j!). \tag{13}$$

This measurement is the actual image that obeys Poisson process and maximizes the data likelihood of (12). Equation (13) is accurate only if the voxel size can be as small as possible. For voxels having finite size, (13) may be expressed as

$$P(\Phi|\bar{\Phi}) = \prod_j e^{-\bar{\phi}_j}(\bar{\phi}_j)^{\phi_j}/(\phi_j!). \tag{14}$$

Equation (14) specifies a process that is not observed directly. In this process, $\bar{\phi}_j$ represents the average of actual image density over the volume of voxel j (or the *a priori* mean value of voxel j) and ϕ_j represents the average of estimated density over voxel j that contributes to observed data [2], [29] as defined before. If we impose the constraint of $N = \Sigma_j \phi_j$ being fixed upon (14), both (1) and (14) are equivalent. In other words, (1) is more restrictive than (14).

The likelihood function of the data distribution is expressed as [12]

$$L(\boldsymbol{Y}|\Phi) = \ln P(\boldsymbol{Y}|\Phi) = \sum_i \left[-\sum_j R_{ij}\phi_j + Y_i \ln\left(\sum_j R_{ij}\phi_j\right) - \ln(Y_i!)\right]. \tag{15}$$

Considering the *a priori* image information (8) and the data likelihood function (15), the Bayesian function becomes

$$g(\Phi) = L(\boldsymbol{Y}|\Phi) + H(\Phi) - \ln P(\boldsymbol{Y}). \tag{16}$$

Since both the probability functions of (8) and (15) are not accurate unless both the underlying assumptions and $\{\bar{\phi}_j\}$ and $\{R_{ij}\}$ are accurate, the Bayesian function is modified as

$$\tilde{g}(\Phi) = L(\boldsymbol{Y}|\Phi) + \xi H(\Phi) - \ln P(\boldsymbol{Y}) \tag{17}$$

where ξ is an adjustable parameter. This modification is a deliberate ploy to allow an explicit tradeoff in the reconstruction between the likelihood function and the *a priori* probability function. If the accuracy of both the *a priori* source probability and the measurement probability have equal confidence levels, $\xi = 1$, otherwise $\xi \neq 1$. In this paper, $0 < \xi < 1$ is assumed for the purpose of treating the *a priori* source information $H(\Phi)$ as additional constraints to supplement the data likelihood information $L(\boldsymbol{Y}|\Phi)$.

Since $\tilde{g}(\Phi)$ is strictly concave, the Bayesian solution Φ^* of (17) which maximizes the modified Bayesian function $\tilde{g}(\Phi)$ is uniquely determined by a system of equations

$$\sum_i R_{ik}\left(Y_i \Big/ \sum_j R_{ij}\phi_j^*\right) - \sum_i R_{ik} = \xi Z_k(\Phi^*), \qquad k = 1, 2, \cdots, J \tag{18}$$

where

$$Z_k(\Phi^*) = -\frac{\partial H(\Phi)}{\partial \phi_k}\bigg|_{\Phi=\Phi^*}. \tag{19}$$

An analytical determination of ξ in (18) is discussed in Appendix B and an experimental estimate of ξ is given in a later section. If one neglects the *a priori* source information or assumes that it is a constant, then $P(\Phi) =$ constant and $Z_k(\Phi^*) = 0$, and (18) determines the maximum likelihood solution [17], [18], [29], [32].

Many iterative numerical techniques [29], [34] can be employed to determine the Bayesian solution Φ^*. In this paper, the EM technique [12]–[14] is used to determine Φ^* iteratively. By treating the *a priori* source information as an additional constraint and following the mathematical derivations shown in [35], a BIP algorithm is given by [7]

$$\phi_k^{(n+1)} = \phi_k^{(n)} \frac{\sum_i R_{ik}\left(Y_i \Big/ \sum_j R_{ij}\phi_j^{(n)}\right)}{\sum_i R_{ik} + \xi_k^{(n)} Z_k^{(n)}} \tag{20}$$

and

$$Z_k^{(n)} = Z_k(\phi_k^{(n)} + \lambda d_k^{(n)}) \tag{21}$$

where $\lambda \approx 1$ and $d_k^{(n)} = \phi_k^{(n)} - \phi_k^{(n-1)}$ are assumed for easy computation and the initial estimate $\phi_k^{(0)}$ is, as usual, chosen as constant (N/J). The function $Z_k^{(n)}$ is defined for the nonuniform and uniform *a priori* information as follows:

a) for the nonuniform *a priori* probabilistic information of (8)

$$Z_k^{(n)} = \ln(\phi_k^{(n)} + \lambda d_k^{(n)}) - \ln(\bar{\phi}_k) + 1; \tag{22}$$

b) for the uniform *a priori* probabilistic information of (9)

$$Z_k^{(n)} = \ln(\phi_k^{(n)} + \lambda d_k^{(n)}) + 1. \tag{23}$$

If there is no *a priori* source probability information available, i.e., $Z_k^{(n)} = 0$, the BIP algorithm (20) reduces to the ML algorithm [12], [13], which has been shown to be extremely valuable in certain problems [12], [13], [25], [30], [36], [37].

The adjustable parameter ξ, appearing in (18) and (20), is defined as a function of iterative index n and source voxel index k as

$$\xi_k^{(n)} = \frac{An^{\nu}}{B + n^{\tau}} \sum_i R_{ik} \tag{24}$$

where A, B, ν, and τ are constants. The dependence of $\xi_k^{(n)}$ on index n allows one to impose the effect of the *a priori* information $H(\Phi)$ smoothly. After the quantity $[An^{\nu}/(B + n^{\tau})]$ reaches its maximal value, $\xi_k^{(n)}$ is fixed to this value for all further iterations. Since the first term $\Sigma_i R_{ik}$ in the denominator of (20) may change from voxel to voxel, $\xi_k^{(n)} \approx \Sigma_i R_{ik}$ of (24) is assumed to ensure that $H(\Phi)$ is imposed relatively uniformly on all voxels. The functional dependence of $\xi_k^{(n)}$ on the iterative index n is quite important in obtaining the optimal solution. If the *a priori* source information is accurately known, $\xi_k^{(n)}$ should be a monotonic sigmoidal function of index n, and A, B, and $\nu = \tau$ can vary significantly without affecting the solution. If the *a priori* information is only approximately estimated, the choice of A, B, ν, and τ is quite important for optimal results, and their values can vary only in certain intervals for a particular imaging situation. The theoretical determination of these constants is quite complicated and requires more computation time as shown in Appendix B. A few examples for choosing these constants experimentally have been given [1], [2], [29]. In general, the less accurate the *a priori* source information, the smoother the dependence of $\xi_k^{(n)}$ on n, and the smaller the maximal value of $\xi_k^{(n)}$. With incorrect *a priori* information, artifacts in the BIP reconstruction have been observed [2], [6]–[9], [29], [35]. Note that although the approximation of $\phi_k = \phi_k^{(n)} + \lambda d_k^{(n)}$ in (21) is assumed for easy computation, other approximations can be used [29].

The BIP algorithm of (20), given the uniform and nonuniform *a priori* information (23) and (22), respectively, is applied to computer generated noise-free and Poisson randomized projections in the following section.

Computer Simulation Results

The BIP algorithms of (20) with (22) and (23), respectively, were tested and compared to the ML algorithm [12]–[13] in two different imaging situations: 1) computer generated noise-free data; and 2) Poisson randomized data.

Fig. 1 shows the simulated phantom $\{S_j\}$ containing four hot disks of 4 quanta each and a cold disk of 2 quanta, superimposed upon a uniform background of 3 quanta. The five disks each have a radius of five voxels. The elliptical phantom is contained in a rectangular region divided into 64×64 voxels. The density outside the elliptical phantom is zero. To illustrate the effect of suppressing the unpredictable noise with Bayesian analysis, the deterministic effects of finite spatial resolution, attenuation, scattering, and collimator divergence were ignored, since the inclusion of them is largely a matter of introducing a more complicated R_{ij} matrix [38]. The noise-free projections were calculated using the summation $\{\Sigma_j R_{ij} S_j\}$ for a parallel beam geometry consisting of 64 equally-spaced projection angles in the interval [0, 180] degrees, with 64 equally-spaced rays per projection. The R_{ij} was chosen as the intersection length of projection ray i and voxel j (i.e., the element of the projection matrix mentioned before [1], [4]), and is used in the normalization of (20). The calculated noise-free projection rays $\{\Sigma_j R_{ij} S_j\}$ were used as the mean values of Poisson randomized projection rays $\{Y_i\}$ which were generated by a Poisson random number generator [17]. The projection rays with zero mean value were set to zero. The Poisson randomized projection rays $\{Y_j\}$ with nonvanishing means are in the range of [0, 120] counts. The summation of the noise-free projection rays is 356341.94. The total counts of Poisson randomized projection rays is 356946.

If the mean density distribution $\{\bar{\phi}_j\}$ is assumed to be the phantom of Fig. 1 (i.e., the *a priori* mean density is accurately known), only one iteration was needed in the case of noise-free projections and a few iterations in the case of noisy projections to produce very good quality images with the *a priori* nonuniform BIP algorithm of (20) and (22). For these results, the following values of A, B, ν, and τ were used: $A = 1$ to 10, $B = 1$ to 100, $\nu = \tau = 1$ to 3. In practice, there is no such *a priori* information available. This paper provides a possible way to estimate the means $\{\bar{\phi}_j\}$ and the adjustable parameter ξ in carrying out the Bayesian solution Φ^* via the BIP algorithm (20). The means are estimated using prior iterated Bayesian results starting from a uniform initial estimate, and the adjustable parameter was chosen much smoother as shown below. Other ways to determine the means and the adjustable parameter can be more general at the expense of computation time (see Appendixes A and B).

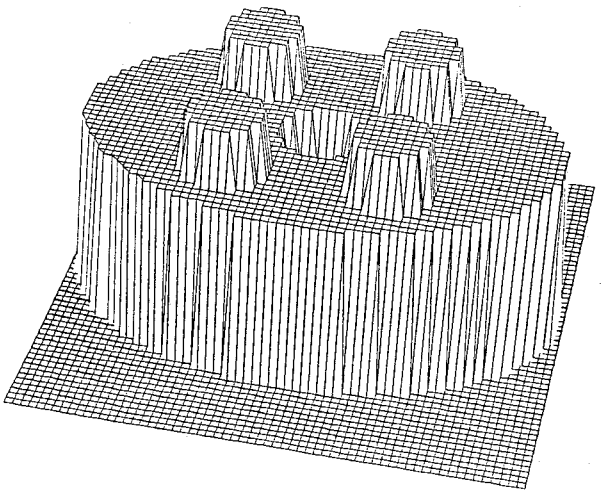

Fig. 1. The computer simulated elliptical phantom containing four hot spots of 4 quanta each and a cold spot of 2 quanta, superimposed upon a uniform background of 3 quanta. Outside the elliptical region, the density is zero.

A. Results from Noise-Free Projections

Fig. 2 shows the image reconstructed by applying the nonuniform BIP algorithm [(20) and (22)] to the noise-free projections after 10 iterations. In the case of noise-free data (very strong *a priori* information), convergence to the phantom is ensured if the projection matrix R has at least as many rows (projection rays) as columns (voxels) and is full rank. Thus, it is not necessary to utilize additional constraints. It is valuable to consider other constraints only if they can speed up the convergence. For that purpose the mean values $\{\bar{\phi}_j\}$ were updated, from a uniform initial estimate, by [2], [6]–[9], [29]

$$\bar{\phi}_j = \phi_j^{(n)} + \mu d_j^{(n)} \tag{25}$$

where the factor $\mu > 0$ serves to advance the mean and therefore speed up convergence via overrelaxation. The value $\mu = 2$ was used for Fig. 2. The constants of $A = 1$, $B = 100$, $\nu = 0.5$, and $\tau = 1$ were chosen in carrying out the results of the BIP algorithms for finite iterations (≤ 500). After 250 iterations, $\xi_k^{(n)} = 0.045\ \Sigma_i R_{ik}$ was assumed. The *a priori* uniform BIP algorithm [(22) and (23)] produces the similar image of Fig. 2 for the noise-free data after 10 iterations. Fig. 3 shows the image obtained using the ML algorithm [12], [13] after 10 iterations for the noise-free data. The BIP algorithms showed more rapid convergence. If $\nu = 1$ and $\tau = 0.5$ were used, artifacts in the BIP images were observed.

The three algorithms were run up to 500 iterations; monotonic convergence was observed. The convergence process can be quantitatively specified by a root-mean-square-error test function [1], [9]

$$\Psi_0 = \left[\sum_j \left(\phi_j^{(n)} - S_j\right)^2 \Big/ \sum_j \left(S_j - \bar{S}\right)^2 \right]^{1/2} \tag{26}$$

where $\bar{S}$ is the mean of $\{S_j\}$. Fig. 4 shows the convergence curves of the three algorithms using the test function (26) for the case of noise-free projections. It is also useful to introduce a modified root-mean-square-error test function for images having sharp boundaries and reconstructed from noisy projections [29], [39]

$$\Psi_1 = \left[\sum_j \left(\bar{\phi}_j^{(n)} - \bar{S}_j\right)^2 \Big/ \sum_j \left(\bar{S}_j - \bar{S}\right)^2 \right]^{1/2} \tag{27}$$

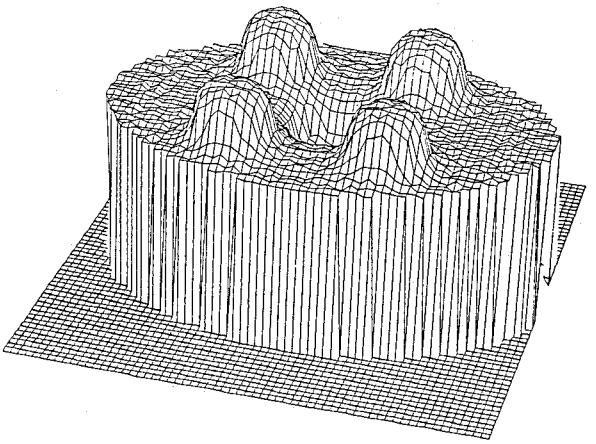

Fig. 2. The image reconstructed using the *a priori* nonuniform BIP algorithm [(20) and (22)] after 10 iterations for the noise-free projections.

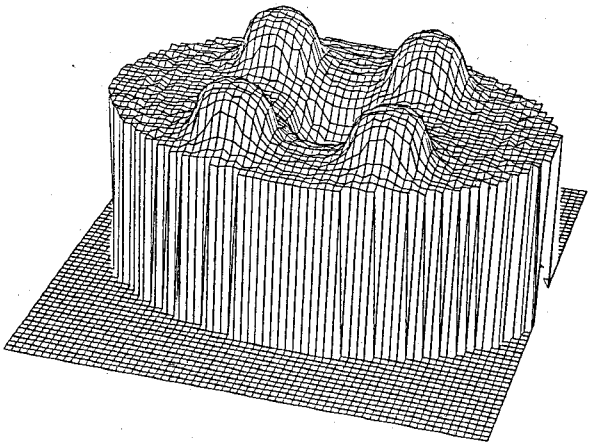

Fig. 3. The image obtained by using the ML algorithm [(20) with $Z_k^{(n)} = 0$] after 10 iterations for the noise-free projections.

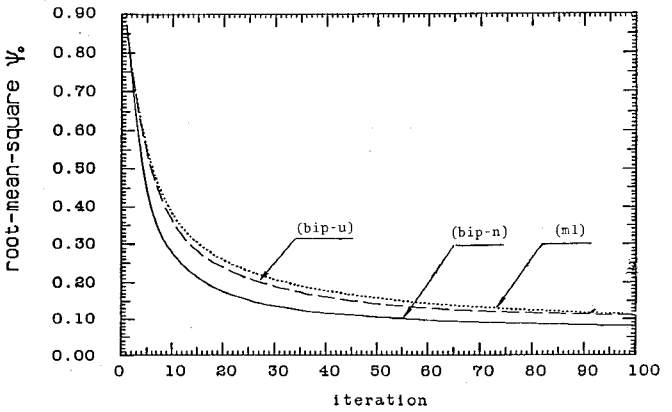

Fig. 4. The test function of root-mean-square-error (26) in the case of noise-free projections for the ML (ml, dotted line), uniform BIP of (20) and (23) (bip-u, broken line), and nonuniform BIP (bip-n, solid line) algorithms.

where the weighting process is expressed as

$$\bar{S}_j = \sum_i w_{ij} S_i \Big/ \sum_i w_{ij}, \qquad i = j-2, j-1, j, j+1, j+2 \quad (28)$$

and

$$w_{j-2,j} = 0.2,\ w_{j-1,j} = 0.5,\ w_{j,j} = 1.0,\ w_{j+1,j} = 0.5,\ w_{j+2,j} = 0.2 \quad (29)$$

and similarly for $\bar{\phi}_j^{(n)}$. Note that the $\bar{\phi}_j^{(n)}$ in (27) is different from $\bar{\phi}_j$ in (25). The smoothing of (28) and (29), for each iteration, is used to minimize the effect of a small shift between neighboring voxels. For example, if a voxel j has the strength $\phi_j \approx S_j >> 0$ and is shifted one position $j+1$ where $S_{j+1} = 0$, the test function (26) will give a large value that is not consistent with visual observation.

Fig. 5 shows the results of the test function (27) in the case of Fig. 4. Note that smoothing the iterated ML results [or $\bar{\phi}_j^{(n)}$ of (28) and (29)] to suppress noise and artifacts is not the same as averaging the prior iterated BIP results locally to estimate the *a priori* means $\{\bar{\phi}_j\}$. The solid line in Fig. 4 represents the test function (26) for the nonuniform BIP algorithm and the dotted line in Fig. 5 represents the same test function for the smoothed ML algorithm. The difference between the two methods is more apparent in the presence of noisy data.

B. Results from Poisson Randomized Projections

Figs. 6 and 7 show the images reconstructed using the nonuniform BIP and ML algorithms, respectively, for the Poisson randomized projections after 20 iterations. In the noisy situation, the mean value $\bar{\phi}_j$ of voxel j was estimated by averaging over its nearby voxels, utilizing *a priori* local correlation information as discussed in Appendix A [2], [6]–[9], [25], [29], [30]. In one-dimensional notation, the average over nearby voxels is expressed as

$$\bar{\phi}_j = \sum_i T_{ij} \phi_j^{(m)} \Big/ \sum_i T_{ij}, \qquad i = j-1, j, j+1 \quad (30)$$

where m represents the iterative index. If $m = n$, the average is done after each iteration. In carrying out the nonuniform BIP result of Fig. 6, m was chosen to update the mean values after every 5 iterations to save computation time. $T_{ij} = 1$ was used in Fig. 6 for easy calculation; the means $\{\bar{\phi}_j\}$ were smoothed locally in this choice of T_{ij}. Other choices can be more general than $T_{ij} = 1$ and will still be easy to compute. For example, a Gaussian form could be used. With finite full-width-half-maximum (FWHM), a Gaussian form T_{ij} has nonuniform weights over the nearby neighborhood. More elaborate $\{T_{ij}\}$ that consider the structure of the local correlation are discussed in Appendix A and [40]. The overrelaxation of (25), $\phi_j^{(n)} + \mu d_j^{(n)}$, can be averaged by $\{T_{ij}\}$ to estimate the means in the case of data containing noise. In this approach, the rate of convergence is accelerated while the

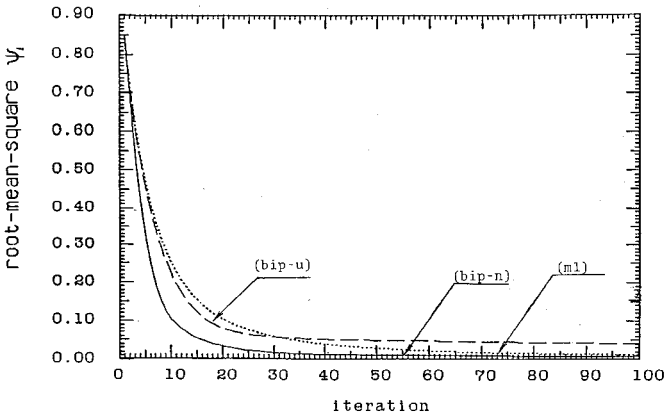

Fig. 5. The test function of modified root-mean-square-error (27) in the case of noise-free projections.

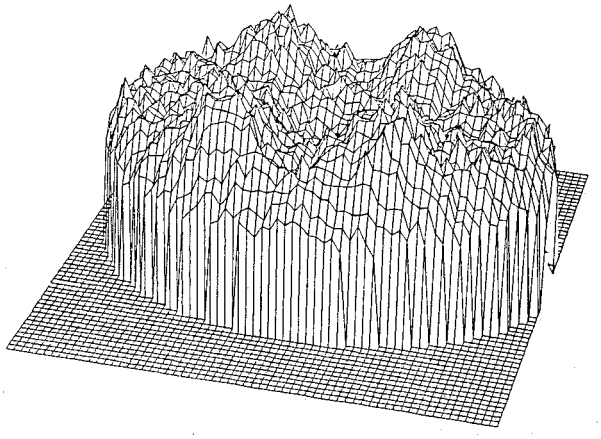

Fig. 6. The image produced by the nonuniform BIP algorithm after 20 iterations using the projections with Poisson noise.

$\{T_{ij}\}$ preserve the *a priori* concept that there is a local correlation between neighboring voxels. After 20 iterations, the ML image begins to be noisy. The uniform BIP algorithm of (20) and (23) has better convergence performance, compared to ML, but the image noise can still be observed. The image noise is significantly suppressed with the nonuniform BIP algorithm, as shown in Fig. 6. No divergence was observed up to 100 iterations with the nonuniform BIP algorithm.

Figs. 8 and 9 show the convergence of the three algorithms using the test functions (26) and (27), respectively, for the case of noisy projections. The curves of the modified test function (27) match more closely the visual observation of the iterated images because it includes not only the amplitude but also the spatial components of the images as mentioned after (29). The solid line in Fig. 8 and the dotted line of Fig. 9 contrast the difference between averaging for the *a priori* mean with the BIP algorithm and smoothing the iterated results with the ML algorithm using the same test function (26).

A complete discussion on stopping criteria is beyond the scope of this paper [36], [37]. Based upon the choices of the constants of the adjustable parameter $\xi_k^{(n)}$ of (24), and the average update $\{\bar{\phi}_j\}$ of (30), both the test func-

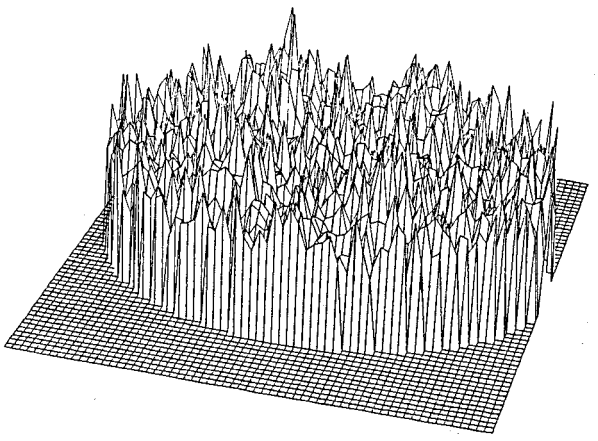

Fig. 7. The reconstructed image using ML algorithm after 20 iterations for the projections with Poisson noise.

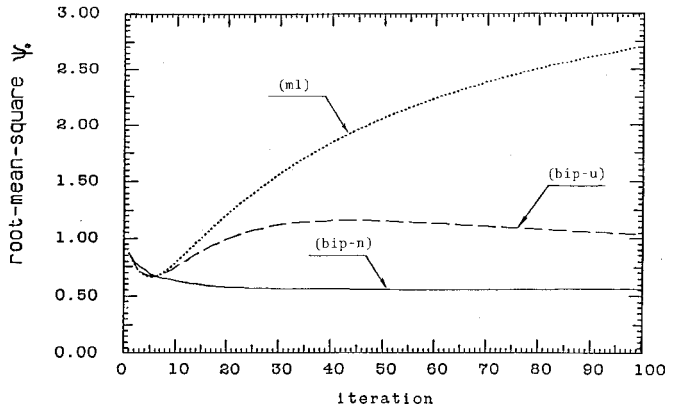

Fig. 8. The test function of root-mean-square-error (26) in the case of noisy projections.

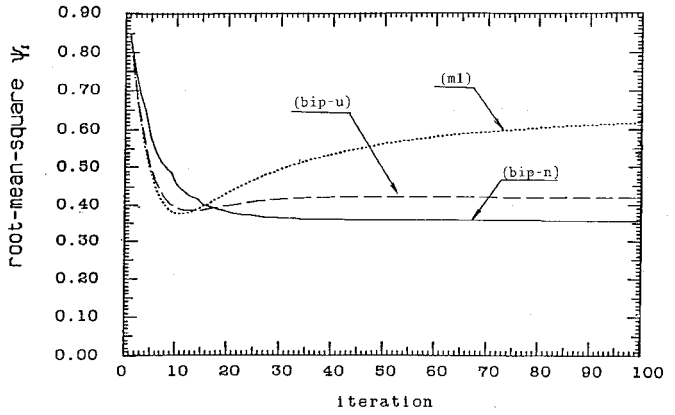

Fig. 9. The test function of modified root-mean-square-error (27) in the case of noisy projections.

tions (26) and (27) indicate that the solution of the nonuniform BIP algorithm did not change much between 20 and 100 iterations. For this reason, we did not run the algorithm beyond 100 iterations. Of course, different choices of $\xi_k^{(n)}$ and $\{\bar{\phi}_j\}$ will result in different curves for Figs. 8 and 9. For example, if we let $T_{ij} = \delta_{ij}$, the solid lines would be close to the broken lines. If we use very small $\xi_k^{(n)}$, the solid lines would match the dotted lines, respectively. Similarly, assuming $\xi_k^{(n)}$ is very large, the biases in estimating the *a priori* source information would result in artifacts for the BIP solutions. A suitable stopping criterion must be chosen to terminate the iteration process. The stopping rules based upon χ^2 distribution and hypothesis testing [23], [24], [37] may be applicable for the algorithm of (20). A few practical criteria that can be used to terminate the iterations are currently under investigation. It is reasonable to expect that stopping criteria can be empirically developed for specific imaging situations if there is *a priori* information available.

It is noted that the iterative EM scheme and the estimation of the mean values were implemented in a very general manner, and the BIP algorithm of (20)–(24) can be applied to many image processing problems.

Conclusion

A Bayesian image reconstruction method has been presented. It uses the *a priori* uniform or nonuniform image density distribution probabilistic information as additional constraints to supplement the data likelihood information. The likelihood information considers the measurement fluctuations of a Poisson nature. The *a priori* image information was modeled as a multinomial random process based on general assumptions of image density quantization [see (1)]. The estimation of the *a priori* mean image density information was carried out in a very general manner such that it can be used in many image processing problems. Since the deterministic effects of finite spatial resolution, attenuation, scattering, and collimator divergence can be accounted for in the same way for both the ML and BIP approaches, the advantage of incorporating the *a priori* source information with BIP to suppress noise and artifacts is obvious. At a given iteration number, superior images were obtained with the nonuniform *a priori* image density distribution information. Application of the nonuniform BIP algorithm to Monte Carlo simulation and experimental phantom imaging data is being reported elsewhere [41], [42].

Appendix A
Estimation of the *A Priori* Means $\{\bar{\phi}_j\}$

The *a priori* source distribution of (1) provides an easy way to consider more restrictive *a priori* information via the probability distribution $\{p_j\}$. In general, the *a priori* probability distribution $\{p_j\}$ would be uniform from a voxel to its neighboring voxels in a homogeneous region, and nonuniform at the boundaries of those homogeneous regions due to physical interaction and structural correlation [43], [44]. This *a priori* source information of local correlation may be interpreted as either continuity if the neighbors change rather smoothly, or discontinuity otherwise [28], [40]. In this paper, (5) provides an appropriate interpretation via the means $\{\bar{\phi}_j\}$. Mathematically, the neighbors are classified as first-order, second-order, and so on. The further away the neighbors, the weaker the correlation. The first-order correlation for voxel (i, j) (in two dimensions) is among the first-order neighbors $(i,$

$j \pm 1)$ and $(i \pm 1, j)$. The weight factor $\{T_{ij}\}$ of (30) serves the purpose of considering the local correlation. In general, T_{ij} is nonuniform over the first-order neighbors of voxel (i, j). Similar interpretations are valid for the second-order neighbors $(i \pm 1, j \pm 1)$ and higher orders. A simple example for explaining T_{ij} is given below: first we compute the 8 sums over the first- and second-order neighbors

$$\begin{aligned} s_1 &= \phi(i-1, j+1)/\sqrt{2} + \phi(i, j+1) \\ &\quad + \phi(i+1, j+1)/\sqrt{2} \\ s_2 &= \phi(i, j+1) + \phi(i+1, j+1)/\sqrt{2} \\ &\quad + \phi(i+1, j) \end{aligned} \tag{A.1}$$

and so on; then we average $\phi(i, j)$ with the s_i that is closest to $\phi(i, j)$. In this paper, we incorporate the *a priori* local correlation information into the EM iterative process by weighting the prior BIP results by T_{ij} [the simple choice is (30)] starting from a uniform estimated initial. If even more restrictive *a priori* information (for example, specific source pattern information) were available, a different approach would then be used [8], [29], [40].

Appendix B
Determination of the Adjustable Parameter ξ

An analytical determination of $\xi_k^{(n)}$ in (20) may be carried out using the following two constraints.

1) Descent Step Constraint: This approach is the same as that of the steepest descent method [34]. First, let us rewrite the denominator of (20) as $\Sigma_i R_{ik}(1 + x^{(n)} \cdot Z_k^{(n)})$, where $x^{(n)} = An^{\nu}/(B + n^{\tau})$. By substituting (20) into (17) and maximizing $\tilde{g}(\Phi^{(n)}, x^{(n)})$ with respect to $x^{(n)}$, the parameter $x^{(n)}$ is then determined by

$$\frac{\partial \tilde{g}(\Phi^{(n)}, x^{(n)})}{\partial x^{(n)}} = 0. \tag{B.1}$$

2) Mean Square Error Constraint: By imposing the mean square error constraint upon the updated voxel values $\{\phi_j^{(n+1)}\}$

$$\sum_i \left[Y_i - \sum_j R_{ij} \phi_j^{(n+1)}(x^{(n)}) \right] = I\sigma^2 \tag{B.2}$$

the value of $x^{(n)}$ is then given by (B.2), where σ^2 is the known random noise power of projections. The constraint of (B.2) can also be used to determine the ξ in (18).

Acknowledgment

The authors are grateful to the reviewers for their critical evaluations and many helpful suggestions.

References

[1] E. Levitan and G. T. Herman, "A maximum *a posteriori* probability expectation maximization algorithm for image reconstruction in emission tomography," *IEEE Trans. Med. Imaging*, vol. MI-6, pp. 185–192, 1987.
[2] Z. Liang and H. Hart, "Bayesian image processing of data from constrained source distributions: I. Non-valued, uncorrelated and correlated constraints," *Bull. Math. Biol.*, vol. 49, pp. 51–74, 1987.
[3] S. Geman and D. E. McClure, "Bayesian image analysis: An application to single photon emission tomography," in *Proc. Stat. Soc. Amer., Comput. Sect.*, Washington, DC, 1985, pp. 12–18.
[4] K. M. Hanson and G. W. Wecksung, "Bayesian approach to limited-angle reconstruction in computed tomography," *J. Opt. Soc. Amer.*, vol. 73, pp. 1501–1509, 1983.
[5] K. Lange, M. Bahn, and R. Little, "A theoretical study of some maximum likelihood algorithms for emission and transmission tomography," *IEEE Trans. Med. Imaging*, vol. MI-6, pp. 106–114, 1987.
[6] Z. Liang, "Image enhancement by estimated *a priori* information," *Proc. SPIE, Med. Imaging II*, vol. 914, pp. 684–689, 1988.
[7] ——, "Statistical models of *a priori* information for image processing," *Proc. SPIE, Med. Imaging II*, vol. 914, pp. 677–683, 1988.
[8] Z. Liang and H. Hart, "Bayesian reconstruction in emission computerized tomography," *IEEE Trans. Nucl. Sci.*, vol. 35, pp. 788–792, 1988.
[9] Z. Liang, "Statistical modeling of *a priori* information for image processing problems: A mathematical expression of images," in *Proc. Stat. Soc. Amer., Interface Sect.*, VA, Apr. 1988, pp. 824–831.
[10] C. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 623–656, 1948.
[11] S. Kullback and R. Leibler, "On information and sufficiency," *Ann. Math. Statist.*, vol. 22, pp. 79–86, 1951.
[12] L. A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *IEEE Trans. Med. Imaging*, vol. MI-1, pp. 113–122, 1982.
[13] K. Lange and R. Carson, "EM reconstruction algorithms for emission and transmission tomography," *J. Comput. Assist. Tomogr.*, vol. 8, pp. 306–316, 1984.
[14] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Stat. Soc.*, vol. 39, pp. 1–38, 1977.
[15] E. Isaacson and H. B. Heller, *Analysis of Numerical Methods*. New York: Wiley, 1966.
[16] Z. Liang, R. Jaszczak, and K. Greer, "On Bayesian image reconstruction from projections: II. Maximum *a priori* source distribution probability," in preparation.
[17] B. Carnahan, H. A. Luther, and J. O. Wikes, *Applied Numerical Methods*. New York: Wiley, 1978.
[18] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. New York: McGraw-Hill, 1984.
[19] R. Kikuchi and B. H. Soffer, "Maximum entropy image restoration: I. The entropy expression," *J. Opt. Soc. Amer.*, vol. 67, pp. 1656–1665, 1977.
[20] B. R. Frieden, "Dice, entropy and likelihood," *Proc. IEEE*, vol. 73, pp. 1764–1770, 1985.
[21] ——, "Restoring with maximum likelihood and maximum entropy," *J. Opt. Soc. Amer.*, vol. 62, pp. 511–518, 1972.
[22] ——, "Statistical models for the image restoration problems," *Comput. Vision, Graph. Image Processing*, vol. 12, pp. 40–58, 1980.
[23] S. F. Gull and G. J. Daniell, "Image reconstruction from incomplete and noisy data," *Nature*, vol. 272, pp. 686–690, 1978.
[24] S. F. Burch, S. F. Gull, and J. Skilling, "Image restoration by a powerful maximum entropy method," *Comput. Vision, Graph. Image Processing*, vol. 23, pp. 113–128, 1983.
[25] M. I. Miller and D. L. Snyder, "The role of likelihood and entropy in incomplete-data problems: Applications to estimate point-process intensities and Toeplitz constrained covariances," *Proc. IEEE*, vol. 75, pp. 892–907, 1987.
[26] E. T. Jaynes, "Prior probabilities," *IEEE Trans. Syst., Sci., Cybern.*, vol. 4, pp. 227–241, 1968.
[27] Z. Liang and R. Jaszczak, "Statistical models of *a priori* information for image processing: II. Finite distribution range constraints," in *Proc. SPIE, Appl. Digital Image Processing XI*, 1988, pp. 53–58.
[28] Z. Liang, "Statistical models of *a priori* information for image processing: Neighboring correlation constraints," *J. Opt. Soc. Amer.*, vol. 5, pp. 2026–2031, 1988.
[29] ——, Ph.D. dissertation, The City University of New York, 1986.
[30] D. L. Snyder, M. I. Miller, *et al.*, "Noise and edge artifacts in maximum-likelihood reconstruction for emission tomography," *IEEE Trans. Med. Imaging*, vol. MI-6, pp. 228–238, 1987.

[31] J. Shore and R. W. Johnson, "Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 26–37, 1980.
[32] C. Rao, *Advanced Statistical Methods in Biometric Research*. New York: Wiley, 1952.
[33] H. C. Andrews and B. R. Hunt, *Digital Image Restoration*. Englewood Cliffs, NJ: Prentice-Hall, 1977.
[34] D. G. Luenberger, *Introduction to Linear and Non-linear Programming*. Reading, MA: Addison-Wesley, 1973.
[35] H. Hart and Z. Liang, "Bayesian image processing in two dimensions," *IEEE Trans. Med. Imaging*, vol. MI-6, pp. 201–208, 1987.
[36] Y. Vardi, L. A. Shepp, and L. Kaufman, "A statistical model for positron emission tomography," *J. Amer. Stat. Assoc.*, vol. 80, pp. 8–35, 1985.
[37] E. Veklerov and J. Llacer, "Stopping rule for the MLE algorithm based on statistical hypothesis testing," *IEEE Trans. Med. Imaging*, vol. MI-6, pp. 313–324, 1987.
[38] Z. Liang, R. Jasczak, *et al.*, "Bayesian reconstruction of SPECT in parallel, fan, and cone beam geometries," *Ann. Nucl. Med. Soc.*, vol. 29, p. 871, 1988.
[39] Z. Liang, R. Jaszczak, and H. Hart, "Study and performance evaluation of statistical methods in image processing," *Comput. Med. Biol.*, vol. 18, pp. 395–408, 1988.
[40] Z. Liang and H. Hart, "Source continuity and boundary discontinuity considerations in Bayesian image processing," *Med. Phys.*, vol. 15, pp. 754–756, 1988.
[41] R. Jaszczak, Z. Liang, *et al.*, "Cone beam SPECT: An acquisition geometry to increase detection efficiency," in *Proc. European Nucl. Med.*, 1988.
[42] R. Jaszczak, Z. Liang, *et al.*, "Validation of Bayesian reconstruction for cone beam SPECT: Use of *a priori* non-uniform source information," Workshop on The Formation, Processing and Evaluation of Medical Images, NATO, ASI, Portugal, 1988.
[43] J. W. Goodman, *Statistical Optics*. New York: Wiley, 1985.
[44] B. R. Frieden, "Maximum-probable restoration of photon-limited images," *Appl. Opt.*, vol. 26, pp. 1755–1764, 1987.