INFORMATION GAIN FROM COUNT CORRECTIONS IN SPECT IMAGE RECONSTRUCTION AND CLASSIFICATION[1]

L. Shao*, A.O. Hero**, W.L. Rogers***, N.H. Clinthorne***
*Nuclear Medicine, Dept. of Radiology
Hospital of the University of Pennsylvania, Phila., PA 19104
**Dept. of Electrical Engineering and Computer Science and ***Division of Nuclear Medicine
The University of Michigan, Ann Arbor, MI 48109

ABSTRACT

We present a method for quantifying the impact of count correction side information on the capability of a SPECT projective tomography system to perform image (emitter) reconstruction and feature classification. The method involves computing the image or feature related information gain which results from the presence of count correction side information at the detector. For image reconstruction this information gain is computed using Shannon's mutual information (MI) as in [1], while for feature classification we use the cut-off rate of the SPECT information channel. For reconstruction the gain is proportional to the information divergence between the spatially dependent probability of deletion of a γ-ray originating at a particular emitter location and the spatially independent average deletion probability. For classification the gain is proportional to the difference between the arithmetic mean and the geometric mean of the average number of γ-ray deletions for each of the image classes. Results of analysis and numerical study are presented which indicate that the information gain associated with using count correction data is much more significant for reconstruction of the emitters than for classification of the emitter density when the total detected fluence is low.

I. INTRODUCTION

In SPECT systems, the projection data is acquired by restricting the γ-rays of a radioactive source into certain paths. This is accomplished by a collimation device, such as a parallel-hole collimator or a coded aperture, to improve the system resolution. However, collimation results in the elimination of many γ-rays paths and thus negatively affects system sensitivity. For example, for a parallel-hole collimator the γ-ray detection efficiency is on the order of 10^{-4} for a spatial resolution of 1cm [2]. Generally, the lost or deleted γ-rays are due to such factors as absorption into the aperture septa and non-interaction with the detector scintillator. A natural question is: if some form of count correction side information could be used at the detector what performance gains are possible? In particular, since the mean number of deletions generally depends on the particular source distribution, it is reasonable to expect that count correction side information consisting of the number of deleted γ-rays can improve the estimation capabilities of the system. It is the purpose of this paper to study the impact of this count correction side information on the image reconstruction and feature classification performance of SPECT.

To quantify the effect of count deletions on image reconstruction and classification, we will give expressions for the gain in source-to-detector information transfer which occurs as a result of incorporating count corrections at the detector. We use information transfer as a performance criterion since it enables us to characterize the intrinsic task specific information content of the projections data without having to specify a particular decoding algorithm to reconstruct or classify the image. Since low information content necessarily implies poor performance of the decoding algorithm, the information transfer is a fundamental quantity determining the estimation capabilities of the overall system [3, 4]. For image reconstruction Shannon's measure of mutual information between the emitter locations and the detector data with and without count corrections side information will be compared. A result of [1] is that the impact of this side information is significant when the detected count rate is relatively low and the probability that a γ-ray is deleted, called the deletion probability, is highly dependent on the location of the γ-ray emitter. This can be a relevant regime for dynamic tracer studies where image features can be assumed to be stationary only over short time intervals. For the classification problem, we introduce a cutoff rate approximation to mutual information developed in [4]. For the special case of Poisson noise limited tumor detection we obtain numerical results which indicate that, unlike the image reconstruction problem, the classification information gain due to count corrections is small and is relatively insensitive to the mean count rate.

II. MATHEMATICAL MODEL

There are many kinds of γ-ray detector geometries and collimation configurations, but the basic configuration is shown in Fig. 1. It consists of an object containing radioactive γ-ray emitters, a set of detector surfaces, and a collimation device. The γ-rays which pass through the collimation device are registered on the detector surfaces to form projections from which an image or image feature can be estimated.

[1]This research was supported in part by the National Cancer Institute, DHHS, under PHS grant CA32856

0018-9499/90/0400-0652$01.00 © 1990 IEEE

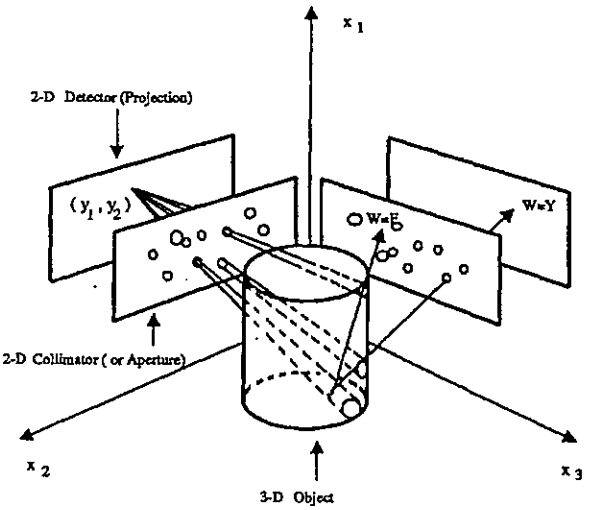

Figure 1: *A general 3-D γ-ray detector system*

Assume that the system has no scattering, no attenuation and the emitter distribution does not vary over time. The relevant statistical quantities which characterize the process of acquiring detector data during a finite interval of time $[0,T]$ are: 1) the emitter spatial positions, $X_1,...,X_n, \in \mathcal{X}$, from which gamma-rays are emitted over $[0,T]$; 2) the emission times $\{t_i^x\}_{i=1}^n$ of gamma-rays, which are specified by the points of increase of the emission counting process $N_x = \{N_x(t) : t \in [0,T]\}$; 3) the positions on the detector, $Y_1,...,Y_m, \in \mathcal{Y}, m \leq n$, of incident gamma-rays which originate at spatial positions in $\mathcal{X}$ and are transmitted through the aperture; 4) the detection times $\{t_i^y\}_{i=1}^m$ at which the incident γ-rays are detected, which are specified by the point of increase of the detection counting process $N_y = \{N_y(t) : t \in [0,T]\}$; 5) the failure, denoted F, of a gamma-ray generated at position x to be detected. The sequence of detected γ-ray positions Y_i and deletions F is denoted by $W_1, W_2, ..., W_n$, the ideal detection process. Unlike the detection process $Y_1,...,Y_m$, the ideal detection process includes the count correction side information related to the number of deletions.

III. INFORMATION GAINS

A. Mutual Information and Cutoff Rate

The general formula for the mutual information between two random quantities U and V is:

$$I(U;V) \triangleq E\left[\ln \frac{dP_{UV}(U,V)}{dP_U(U)dP_V(V)}\right]. \quad (1)$$

where $dP_U/d\mu_u$, $dP_V/d\mu_v$, and $dP_{UV}/(d\mu_u \times d\mu_v)$ are marginal and joint densities relative to measures μ_v, μ_u, and $\mu_u \times \mu_v$, respectively. The mutual information (1) can also be expressed in terms of entropies:

$$I(U;V) = H(U) - H(U/V) = H(V) - H(V|U). \quad (2)$$

where $H(V)$ and $H(U)$ are the entropies of V and U, respectively, and $H(V|U)$ and $H(U|V)$ are conditional entropies. A useful measure of the fidelity of the information carrying channel is the channel capacity C which is the maximum achievable rate of information transmission:

$$C \triangleq \max_{dP_U} I(U;V). \quad (3)$$

For image reconstruction the mutual information satisfies the following properties [3]: 1) it provides a lower bound on the minimum achievable mean square error; and 2) it is a composite measure of system resolution and sensitivity.

A quantity related to the channel capacity is the channel cutoff rate [5]:

$$R_0 \triangleq \max_{dP_U} \left[-\ln \int_{\mathcal{V}} d\mu_v \left(\int_{\mathcal{U}} \frac{dP_U(u)}{d\mu_u}\left(\frac{dP_{V|U}(v|u)}{d\mu_v}\right)^{\frac{1}{2}}\right)^2\right]. \quad (4)$$

For many digital communications channels the cutoff rate provides a practical limit on the information transfer through the channel while the channel capacity is a theoretical limit [6]. The relation between these measures is

$$0 \leq R_0 \leq C.$$

In many problems the cutoff rate is simpler to calculate than either the mutual information or the channel capacity.

In (1) U and V could be any source and measurement random variables in which we are interested. In particular, for image reconstruction without count corrections: $U = (\underline{X}, n)$ are the positions of the n emitters and $V = (\underline{Y}, N_y)$ are the detected γ-ray positions Y_i and detection times t_i^y, $i = 1,...,m$; while when count correction side information is available $V = (\underline{W}, N_x)$. On the other hand, for the classification problem U is a random variable with values in $\{h_1,...,h_m\}$ where h_k denotes the k-th possible image class. To evaluate (1) we need to find the associated conditional distribution and marginal distribution $dP_{V|U}$ and dP_V. For the system shown in Fig. 1 and stationary Poisson counting statistics, $dP_{V|U}$ and dP_V are functions of the following quantities [1]: the source probability density $f_X(x)$ which is the density of emitter locations; the fluence or projection probability density $f_Y(y)$ which is the density of γ-ray locations incident on the detector; the conditional probability density of Y_i given X_i, $f_{Y|X}(y|x)$; the probability of deletion of a γ-ray emitted at position x, $p_F(x) \triangleq P(F|X_i = x)$; the average probability of deletion $\bar{p}_F = \int_{\mathcal{X}} p_F(x) f_x(x) dx$; the detector fluence density $[1 - \bar{p}_F] f_Y(y)$ which integrates to the average detector fluence per emission, i.e. the detector efficiency $[1 - \bar{p}_F]$; the mean number Λ of emissions over $[0,T]$; the distributions of the Poisson distributed total number of emissions and detections: $P_{N_x(T)}(n) = \Lambda^n/n!e^{-\Lambda}$ and $P_{N_y(T)}(m) = ([1-\bar{p}_F]\Lambda)^m/m!e^{-[1-\bar{p}_F]\Lambda}$, respectively.

B. Information Gain for Reconstruction

By using the definition (1), the mutual information with count corrections and without count corrections can be obtained by setting $U = (\underline{X}, n), V = (\underline{W}, N_x)$ and $U = (\underline{X}, n), V = (\underline{Y}, N_x)$, respectively:

$$I_0 \triangleq I((\underline{X}, n); (N_x, \underline{W})) = E\left[\ln \frac{dP(\underline{X}, n|N_x, \underline{W})}{dP(\underline{X}, n)}\right] \quad (5)$$

$$I_{ed} \triangleq I((\underline{X}, n); (N_y, \underline{Y})) = E\left[\ln \frac{dP(\underline{X}, n|N_y, \underline{Y})}{dP(\underline{X}, n)}\right] \quad (6)$$

where I_{ed} denotes mutual information between emitter and detector data. To evaluate and compare I_0 and I_{ed} the following relations have been derived [1]:

$$I_o = H(n) + \Lambda(1 - \bar{p}_F) I(X_i; Y_i) + \Lambda D(p_F(x)||\bar{p}_F) \quad (7)$$

$$I_{ed} \leq I_1 \triangleq H(m) + \Lambda(1 - \bar{p}_F) I(X_i; Y_i). \quad (8)$$

where $I(X_i; Y_i)$ is the mutual information for a single emission-detection pair:

$$I(X_i; Y_i) = \int_{\mathcal{Y}} dy f_Y(y) \int_{\mathcal{X}} dx f_{X|Y}(x|y) \ln \frac{f_{X|Y}(x|y)}{f_X(x)}; \quad (9)$$

and $D(p_F(x)||\bar{p}_F)$ is the *information divergence* between the conditional probability of deletion $p_F(x)$ and the average probability of deletion $\bar{p}_F$:

$$D(p_F(x)||\bar{p}_F) = \int_{\mathcal{X}} f_X(x) \left[p_F(x) \ln \frac{p_F(x)}{\bar{p}_F}\right] dx. \quad (10)$$

In (7) $H(n) = h_{Poisson}(\Lambda)$ is the entropy of the Poisson-Λ number of emissions; and in (8) $H(m) = h_{Poisson}([1 - \bar{p}_F]\Lambda)$ is the entropy of the number of detections.

Using (7) and (8) we have a lower bound on absolute information gain due to the count corrections:

$$I_0 - I_{ed} \geq I_0 - I_1 = H(n) - H(m) + \Lambda D(p_F(x)||\bar{p}_F). \quad (11)$$

Define the relative gain per emission $Gain \triangleq [I_o - I_1]/\Lambda$. Then:

$$\begin{aligned} Gain &\geq \frac{I_o - I_1}{\Lambda} \\ &\geq \frac{H(n) - H(m)}{\Lambda} + D(p_F(x)||\bar{p}_F) \quad (12) \\ &\geq 0. \end{aligned}$$

Consider the lower bound (12) on the gain. The lower bound is composed of the sum of two terms: 1) the scaled difference $[H(n) - H(m)]/\Lambda$ between the entropies of the number of emissions and number of detections; and 2) the information divergence $D(p_F(x)||\bar{p}_F)$, given in (10). While each of these terms are non-negative, $H(n)$ and $H(m)$ are at most logarithmic functions of Λ [1] and therefore for large Λ the information divergence term dominates the gain lower bound. The information divergence term can be interpreted as an (asymmetric) measure of the distance between the conditional failure probability $p_F(x)$ and the average failure probability $\bar{p}_F$ in the sense that (10) is zero if the two probabilities are equal while (10) increases as $\int p_F(x)|p_F(x) - \bar{p}_F|dx$ increases [7]. Furthermore, $D(p_F(x)||\bar{p}_F)$ has the interpretation of information, $I(X_i; F)$, between the failure event "F" and the emitter location X_i. Therefore, if "F" conveys no information about emitter position, in the sense that $D(p_F(x)||\bar{p}_F) = 0$, the resultant information gain bound reduces to the difference $[H(n) - H(m)]/\Lambda$.

C. Information Gain for Classification

Let $\mathcal{U} = \{h_1, ..., h_M\}$ denote M image classes, each having probability $P_1, ..., P_M$, respectively, where $\sum_i P_i = 1$. The mutual information $I(U, V)$ can be interpreted as the information transfer from the object class in $\mathcal{U}$ to the observations in $\mathcal{V}$. For a more suggestive notation we use h to denote the random variable $U \in \{h_1, ..., h_m\}$. Using $dP_{UV} = dP_{V|U} dP_U$, from (1) the mutual information is:

$$I(h; V) = E\left[\ln \frac{dP_{V|U}(V|h)}{\sum_{i=1}^{M} dP_{V|U}(V|h_i) P_i}\right]. \quad (13)$$

For no count corrections $V = (\underline{Y}, N_y)$ and (13) is [4]:

$$\begin{aligned} I(h; V) &= \sum_{k=1}^{M} P_k \left[P(m = 0|h_k) \ln \frac{P(m = 0|h_k)}{P(m = 0)}\right] \quad (14) \\ &+ E\left[\ln \frac{\left[\prod_{i=1}^{m} f_Y^h(Y_i)\right] P^h_{N_y(T)}(m)}{\sum_{k=1}^{M} P_k \left[\prod_{i=1}^{M} f_Y^{h_k}(Y_i)\right] \times P^{h_k}_{N_y(T)}(m)}\right]. \end{aligned}$$

where $P^{h_k}_{N_y(T)}(m)$ is the probability that the m detected emissions come from the k-th image class $h_k \in \mathcal{U}$. An analogous expression can be derived for the case where count corrections are available. Note that the denominator of (14) contains the non-commuting $\sum$ and $\prod$ operators which makes the analytic computation of (14) intractable even for the case of detection ($M = 2$). In the sequel we use the cutoff rate to characterize information transfer from $U = h$ to V.

The cutoff rate can be approximately related to the minimum achievable probability of classification error through the Fano bound [4]. Fano's lemma states that if $\mathcal{U}$ is a discrete source with an alphabet of M symbols $\{h_1, ..., h_M\}$, then any decision rule choosing elements of $\mathcal{U}$ based on observing a random variable V has total probability of error P_e which satisfies:

$$B(P_e) + P_e \ln(M - 1) \geq H(h) - I(h; V), \quad (15)$$

where $B(p)$ is the binary entropy function:

$$B(p) = -p \ln p - (1 - p) \ln(1 - p), \quad p \in [0, 1]. \quad (16)$$

In (15) $H(h)$ is the entropy of $h = U$. It easily seen by evaluating the second derivative that $B(p) + p \ln(M - 1)$ is strictly concave downward over $p \in [0, 1]$, and therefore (15) can be turned into simultaneous upper and lower bounds on P_e [4].

Case of Detection ($M = 2$)

For $M = 2$, the function $B(p) + p\ln(M-1) = B(p)$ is strictly concave and symmetric about its maximum at $p = 1/2$. Therefore the left hand side of (15) has an inverse over $P_e \in [0, 1/2]$. Define the inverse function $B^{-1}(c)$ of $B(p)$ over $p \in [0, 1/2]$, $0 \le c \le \max_p B(p)$:

$$B^{-1}(c) \triangleq \{p \in [0, 1/2] : B(p) = c\}. \tag{17}$$

With this we have the lower bound:

$$P_e \ge B^{-1}(B(P_e) - I(h; V)). \tag{18}$$

Now, recall from the previous section that the cutoff rate R_0 is a lower bound on the capacity C of any communication channel: $R_0 < C$. It has been observed [6, 5] that for some channels R_0 can be interpreted as an upper bound on the mutual information $I(h, V)$ for low complexity encoding of source h, that is, R_0 is the maximum mutual information which can be achieved by any practical coding of the source symbols. In the case of SPECT, the coding of the source symbols h_0 =*normal scan* versus h_1 =*abnormal scan* is determined by the emitter distributions under h_0 and h_1. While it is not possible to justify using R_0 in SPECT as an upper bound on $I(h, V)$ on theoretical grounds, numerical studies performed in [4] indicate that R_0 is close to an achievable upper bound on $I(h; V)$ in the sense that:

$$P_e^{min} \approx B^{-1}(B(P_1) - R_0). \tag{19}$$

where P_e^{min} is the minimum probability of detection error achieved by implementing a MAP detection algorithm.

Using methods in [4, Chapter 6] and Appendix it can be shown that the detection cutoff rates with count loss corrections and without count loss corrections are:

$$R_0 = -\ln\frac{1}{2}\left(1 + e^{-\frac{\Lambda^{h_1}+\Lambda^{h_0}}{2}+\frac{1}{2}(\Lambda^{h_0}\bar{p}_F^{h_0}+\Lambda^{h_1}\bar{p}_F^{h_1})+K_I}\right). \tag{20}$$

$$R_1 = -\ln\frac{1}{2}\left(1 + e^{-\frac{\Lambda^{h_1}+\Lambda^{h_0}}{2}+\sqrt{\Lambda^{h_0}\Lambda^{h_1}}\sqrt{\bar{p}_F^{h_0}\bar{p}_F^{h_1}}+K_I}\right), \tag{21}$$

respectively, where:

$$K_I = \sqrt{(1-\bar{p}_F^{h_0})\Lambda^{h_0}(1-\bar{p}_F^{h_1})\Lambda^{h_1}}\int_{\mathcal{Y}} dy\sqrt{f_Y^{h_0}(y)f_Y^{h_1}(y)}$$

and Λ^{h_k}, $f_Y^{h_k}$, and $\bar{p}_F^{h_k}$ are the emitter count rate, the fluence density, and the average deletion probability for image class h_k, $k = 0, 1$. The integral in K_I is inversely proportional to the *Hellinger distance* and it is equal to one if and only if the two densities $f_Y^{h_1}$ and $f_Y^{h_0}$ are identical. It follows that R_0 and R_1 are measures of the distance between the overall fluence densities $(1-\bar{p}_F^{h_k})\Lambda^{h_k} f_Y^{h_k}$, $k = 0, 1$. If these two fluence densities are identical then $R_0 = R_1 = 0$; the observations contain no information which would permit distinguishing between object classes h_0 and h_1. On the other hand, if these densities are distinct then R_0 and R_1 increase to ln2 as Λ_1 and Λ_0 get large; perfect discrimination becomes possible when there is ln2 = 1 bit of information available. The information gain in terms of cutoff rate for detection is:

$$Gain = R_0 - R_1. \tag{22}$$

Note from (20) and (21) that R_0 and R_1 differ only in the middle term of the exponential. For R_0 this term is the arithmetic mean $(\Lambda^{h_0}\bar{p}_F^{h_0} + \Lambda^{h_1}\bar{p}_F^{h_1})/2$, while for R_1 this term is the geometric mean $\sqrt{\Lambda^{h_0}\bar{p}_F^{h_0}\Lambda^{h_1}\bar{p}_F^{h_1}}$, of the average number of deletions for image classes h_0 and h_1. Since the geometric mean is always less than the arithmetic mean, unless the average number of deletions are identical under h_0 and h_1, generally $Gain > 0$ as expected. The magnitude of $Gain$ is increasing in the difference between the arithmetic and geometric means. Hence information gain due to count corrections depends on the degree to which the average number of deletions differ and can therefore be used to discriminate between h_0 and h_1.

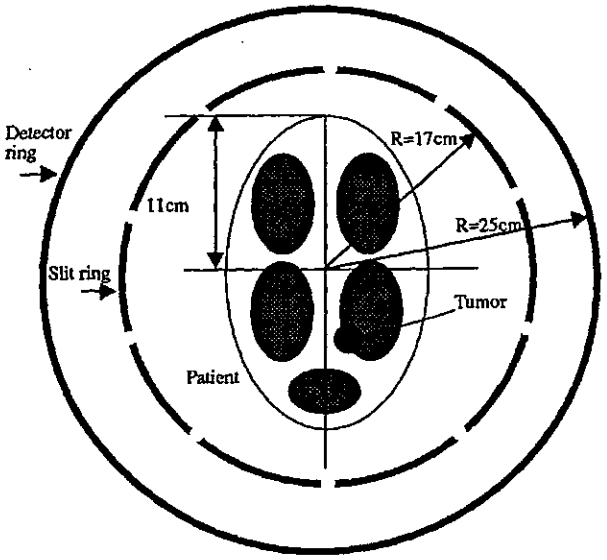

Figure 2: *Ring geometry and brain phantom with tumor*

IV. NUMERICAL RESULTS

We present results for the 2-D ring geometry shown in Fig 2 to illustrate the effect of count rate Λ on the reconstruction and classification information gains (12) and (22).

The reconstruction gain for the 2-D ring geometry is shown in Fig. 3. For this geometry, a 10-slit single width aperture ring was investigated with $340mm$ ring diameter, slit width $3.4mm$, and field of view $220mm$ in diameter. The tumor in the brain phantom at the center of the ring in Fig. 2 has a 2cm diameter and its intensity is twice the level of the dark elliptical regions indicated. The average detector efficiency for this example is $1 - \bar{p}_F = 1.45 \times 10^{-4}$

and therefore the actual mean detected count rate is 4 orders of magnitude less than the values of Λ indicated on the abscissa of Fig. 3. Also shown are the two additive components $D(F) \triangleq D(\bar{p}_F(x)||\bar{p}_F)$ and $[H(n) - H(m)]/\Lambda$ which compose the curve for *Gain*. When Λ is small reconstruction *Gain* is dominated by the latter component while when Λ is large *Gain* is dominated by the former component. It is evident that as Λ increases we have a diminishing return on the incorporation of count corrections in the sense that the reconstruction gain curve falls from about 2.4% of the ideal detection mutual information I_0/Λ for $\Lambda = 47dB$ down to less than 0.1% of I_0 for $\Lambda = 97dB$.

In Fig. 4 we consider the reconstruction and classification information quantities I_0 and R_0, respectively, for the example of Fig. 2. Recall that for reconstruction the source to be estimated is the emission process $X_1, \ldots X_n$ while for classification the source is the image class h. Therefore, in order to perform a meaningful comparison between these information quantities, what is actually plotted in Fig. 4 are the source entropy normalized quantities $I_0/[\Lambda H(X_i) + H(n)]$ and $R_0/H(h)$ which each have the interpretation of information transfer per bit of source entropy. Note that the source entropy increases with Λ for the reconstruction problem, while the source entropy is independent of Λ for the classification problem. This accounts for the monotonic decrease versus monotonic increase of the reconstruction versus classification information quantities. The normalized cutoff rate for classification increases to 1 and is almost 5 orders of magnitude greater than the normalized mutual information for reconstruction. This behavior reflects the property that it is easier to accurately detect localized features of an image than it is to perform an accurate global reconstruction. As contrasted with the relative reconstruction gain, which can be as high as 0.5% over this region of Λ, for $\Lambda > 80dB$ the classification gain as a proportion of the ideal detection cutoff rate R_0 was not observed to be greater than 0.0001%.

V. CONCLUSION

Emitter image reconstruction suffers in performance when count rate Λ is small and there is much to be gained from count corrections. For large count rate, the information gain decreases as a proportion of the mutual information obtainable using count corrections. For classification, the information gain also decreases as the count rate increases but the gain is a much lower proportion of the mutual information (cutoff rate) as compared to the case of reconstruction. This is due to the fact that there is inherently higher image-class information content in the projections than there is emitter-locations information content. It is concluded that for sufficiently high count rates, count corrections have less potential for improving classification performance as compared to reconstruction performance. Future work will include extending these results to classification problems more general than binary detection;

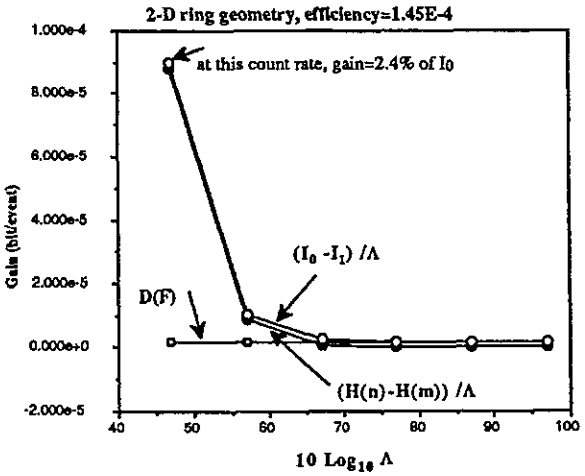

Figure 3: *Information gain for reconstruction in 2-D ring case*

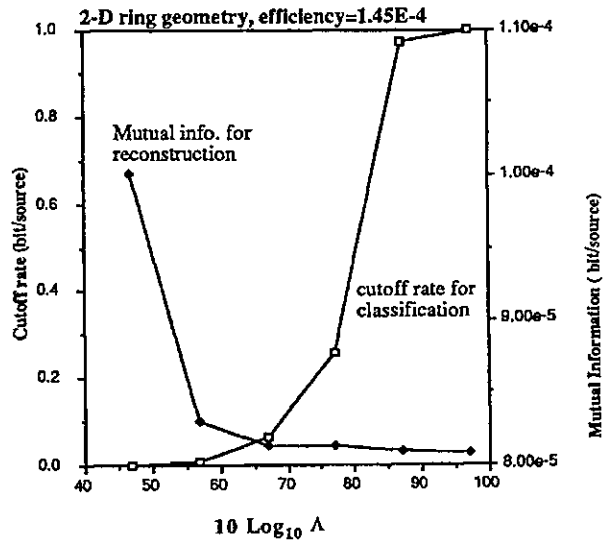

Figure 4: *MI for reconstruction and Cutoff rate performance for classification*

joint classification/reconstruction; and reconstruction of the fixed (static) mean emitter distribution.

APPENDIX

The following is the derivation of cutoff rate for classification with count correction. From [4, Eqn. 6.54], we have the simplified general cutoff rate expression:

$$R_g = -\ln\frac{1}{2}\left(1 + e^{-\frac{\Lambda_1+\Lambda_0}{2}+\sqrt{\Lambda_1\Lambda_0}\int_{\mathcal{V}} d\mu_v \sqrt{\frac{dP_V^{h_1}(v)}{d\mu_v}\frac{dP_V^{h_0}(v)}{d\mu_v}}}\right) \tag{23}$$

where $dP_V^{h_i}/d\mu_v$ is a density with respect to measure μ_v, and Λ_i is the detected photon rate, $i = 0, 1$. Now, for the cutoff rate with count correction, V becomes W and Λ_0 becomes Λ^{h_0}, Λ_1 becomes Λ^{h_1}. Since $dP_W(w)$ consists of continuous part $dP_Y(y)$ and discrete part $dP_F(f)$, (23) can be written as:

$$R_1 = -\ln\frac{1}{2}\left(1 + e^{-\frac{\Lambda^{h_1}+\Lambda^{h_0}}{2}+\sqrt{\Lambda^{h_1}\Lambda^{h_0}}\int_{\mathcal{W}} d\mu_w \sqrt{\frac{dP_W^{h_1}(w)}{d\mu_w}\frac{dP_W^{h_0}(w)}{d\mu_w}}}\right) \tag{24}$$

and

$$\begin{aligned}
&\int_{\mathcal{W}} d\mu_w \sqrt{\frac{dP_W^{h_1}(w)}{d\mu_w}\frac{dP_W^{h_0}(w)}{d\mu_w}} \\
&= \int_{\mathcal{Y}} dy \sqrt{\tilde{f}_Y^{h_1}(y)\tilde{f}_Y^{h_0}(y)} + \sqrt{\bar{p}_F^{h_1}\bar{p}_F^{h_0}} \\
&= \int_{\mathcal{Y}} dy \sqrt{(1-\bar{p}_F^{h_1})f_Y^{h_1}(y)(1-\bar{p}_F^{h_0})f_Y^{h_0}(y)} + \sqrt{\bar{p}_F^{h_1}\bar{p}_F^{h_0}}
\end{aligned} \tag{25}$$

where $\tilde{f}_Y^{h_i}(y)$ is a non-normalized fluence distribution, $i = 0, 1$ [4]. Substitute (24) and (25) into (23) to obtain (21).

References

[1] A. O. Hero and L. Shao, "Information loss associated with deleted counts in single photon emission computed tomography," *IEEE Trans. on Medical Imaging*, to appear Sept., 1990.

[2] W. L. Rogers, N. H. Clinthorne, L. Shao, P. Chiao, J. Stamos, and K. F. Koral, "SPRINT II, a second generation single photon ring tomograph," *IEEE Trans. on Medical Imaging*, vol. 7, no. 4, pp. 291–297, Dec. 1988.

[3] L. Shao, A. O. Hero, W. L. Rogers, and N. H. Clinthorne, "The mutual information criterion for SPECT aperture evaluation and design," *IEEE Trans. on Medical Imaging*, vol. 8, no. 4, pp. 322–336, Dec. 1989.

[4] L. Shao, *Mutual information optimization and evaluation of single photon emission computed tomography.* PhD thesis, The University of Michigan, Ann Arbor, MI 48109, Oct. 1989.

[5] J. L. Massey, "Capacity, cutoff rate, and coding for a direct-detection optical channel," *IEEE Trans. Communications*, vol. COM-29, no. 11, pp. 1615–1621, Nov. 1981.

[6] R. Blahut, *Applications of Information Theory.* Prentice Hall, 1987.

[7] I. Csiszár and J. Korner, *Information Theory: Coding Theorems for Discrete Memoryless Systems.* Academic Press, Orlando FL, 1981.