USE OF PRIOR INFORMATION IN CODED-APERTURE IMAGING

T. A. Gooley[1], H. H. Barrett[123], T. J. Roney[2], W. E. Smith[4]

ABSTRACT

We investigate the effects of various types of prior information for use in medical image reconstruction. The setting in which this is done is reconstruction of an image (**f**) based on measured data (**g**), where the two are related via the linear imaging equation $\mathbf{g} = \mathbf{Hf}$, where **H** represents the point spread function of the imaging system. The system considered here is a multiple-pinhole coded-aperture imaging system. Because the dimension of **g** is much less than the dimension of **f**, we have a severely under-determined set of equations.

A coded-aperture imaging system is simulated and reconstructions are performed. The object that we reconstruct is a digitized picture of the left ventricle and the surrounding heart muscle. This gives rise to a "nearly binary" picture. Comparisons of the various types of prior information are made.

[1] Program in Applied Mathematics, University of Arizona, Tucson, AZ. 85721

[2] Optical Sciences Center, University of Arizona, Tucson, AZ. 85721.

[3] Dept. of Radiology, University of Arizona, Tucson, AZ. 85724.

[4] The Institute of Optics, University of Rochester, Rochester, N.Y. 14627.

1. Introduction

The use of maximum likelihood estimation in emission tomography has been studied extensively since Shepp and Vardi (1982) and, independently, Lange and Carson (1984) introduced the expectation maximization (EM) algorithm (Dempster et. al. (1977)) as a means of finding the maximum likelihood estimate (MLE) of the average intensity of radioactivity within a pixel. Since then, it has been demonstrated by several that the method of maximum likelihood estimation is not well suited to the emission tomography problem (Snyder & Miller (1985), Veklerov & Llacer (1987), Geman & McClure (1985), Liang & Hart (1987a, 1987b)) due to noise in the data and also the unbounded nature of the likelihood function.

To ameliorate these problems, several approaches have been taken, all of which have the same basic characteristic of "smoothing" the estimate by not forcing strict agreement with the data (strict agreement with the data is one of the pitfalls of the MLE) (Snyder & Miller (1985), Veklerov & Llacer (1987), Geman & McClure (1985a, 1985b), Liang & Hart (1987a, 1987b, 1987c), Miller et. al. (1986), Hart & Liang (1987), Levitan & Herman (1987)).

Among these approaches is utilization of prior information about the object which is being estimated (or reconstructed). There have been studies that have used prior information of varying degrees which, at least aesthetically, appear to yield reconstructions superior to those of the MLE (Geman & McClure (1985a, 1985b), Liang & Hart (1987a, 1987b, 1987c), Hart & Liang (1987), Levitan & Herman (1987)).

In this study, we propose to use prior information that is quite restrictive (and perhaps therefore very problem-or object-dependent) and various ways of incorporating this prior knowledge into our reconstruction schemes. The application is reconstruction of medical images in emission tomography. We simulate a multiple-pinhole coded-aperture imaging system for use in the reconstruction of a digitized version of a photograph of the left ventricle and its surrounding heart muscle (Figure 4a). This particular object is well approximated as being binary, and this knowledge (which is quite restrictive) is one of the pieces of prior information that is used (others are discussed below). Simulations are performed and reconstructions using varying degrees of prior information are made.

Section II contains a brief explanation of what is involved in a multiple-pinhole coded-aperture imaging system, and the problem of estimation (or reconstruction) of an object in the emission tomography setting is developed. Section III sets the problem up in a statistical framework and shows how the prior information is incorporated into the reconstruction schemes. Section IV contains the results of the reconstruction schemes that are introduced in the previous section. Section V suggests directions for further work and conclusions.

2. Emission tomography and coded-aperture imaging

In the field of nuclear medicine, the goal of emission tomography is to estimate the activity distribution of a radioactive pharmaceutical that has been taken up in varying amounts by a specific target organ (e.g. brain, heart, liver) in the human body. In single-photon emission computed tomography (SPECT), the radioactive pharmaceutical is absorbed by the target organ and radiation is subsequently emitted from the organ in the form of gamma rays. The activity distribution in the organ is then estimated (as a function of position) by knowledge of detected photon emissions.

Suppose that a bank of (discrete) detectors is arranged around the source (the

radiation-emitting organ) in some fashion, and denote the average value at the ith detector element by $\overline{g}_i$. If the object is then denoted by $f(x,y)$, we can represent $\overline{g}_i$ as

$$\overline{g}_i = \int h_i(x,y) f(x,y) dx\, dy, \tag{2.1}$$

where the integral is taken over the region of support of $f(x,y)$. Here $h_i(x,y)$ represents the point-spread function of the imaging system associated with the ith detector and can be thought of as giving the response in (discrete) data space to a point source in (continuous) object space. Now suppose that the object $f(x,y)$ is expressed in terms of a set of basis functions, $\{u_j,(x,y)\}$, as

$$f(x,y) = \sum_j f_j u_j(x,y)\,, \tag{2.2}$$

where f_j represents the weight of the jth basis function. We then have

$$\overline{g}_i = \int h_i(x,y) \left\{ \sum_j f_j u_j(x,y) \right\} dx\, dy = \sum_j \left\{ \int h_i(x,y) u_j(x,y) dx\, dy \right\} f_j = \sum_j h_{ij} f_j, \tag{2.3}$$

where it has been assumed that the integration and summation signs can be interchanged. This equation can be expressed in matrix form as $\overline{\mathbf{g}} = \mathbf{Hf}$, where the dimension of $\overline{\mathbf{g}}$ is equal to the number of detector elements and the dimension of $\mathbf{f}$ is equal to the number of basis functions. The notation $\overline{\mathbf{g}}$ has been used to signify that the $\overline{\mathbf{g}}_i$ given by (2.3) is to be interpreted as an average value. In the work reported here, $f(x,y)$ is discretized into a matrix of square pixels so that f_j represents the average activity distribution over pixel j, and $u_j(x,y)$ represents a rectangle function occupying the area corresponding to pixel j.

The number of elements in the vector f corresponds to the number of pixels in the discretized object. Inherent to any data collection configuration is the effect of noise. For this reason, what is actually observed is not $\overline{\mathbf{g}}_i$ as given by (2.3), but additive noise in addition to this $\overline{\mathbf{g}}_i$. There is also a noise factor that results from the error introduced by discretization of $f(x,y)$. Because of the noise, we now write $\mathbf{g} = \mathbf{Hf} + \mathbf{n}$, where $\mathbf{n}$ is a noise vector, and where it is understood that the $\mathbf{g}$ given here is the data that are actually observed and not the average $\mathbf{g}$, or $\overline{\mathbf{g}}$, as written in $\overline{\mathbf{g}} = \mathbf{Hf}$.

The way that the point-spread function (and hence the matrix $\mathbf{H}$, which is also referred to as the projection operator) is determined, and therefore the manner in which the data are collected, is dependent upon the sort of imaging system used. In the current work, we shall simulate what we call a multiple-pinhole coded-apertures imaging system. Fresnel zone plate coded-aperture were used in nuclear medicine by Barrett (1972), and Rogers, et. al., (1972) in 1972. See also (Mertz & Young (1961), Dicke (1968), Tipton et. al. (1976)). A schematic of a particular two-dimensional multiple-pinhole coded-aperture system is shown below in Figure 1.

The system depicted here is the one that was simulated for the current work. Here we have discretized the object space into a 39 x 39 set of square pixels. Each pixel then takes on a certain value which represents the average activity distribution within that pixel, and the goal once again is to estimate this number. The radiation emitted from object space can get to detector space only through the openings, which are pinholes in

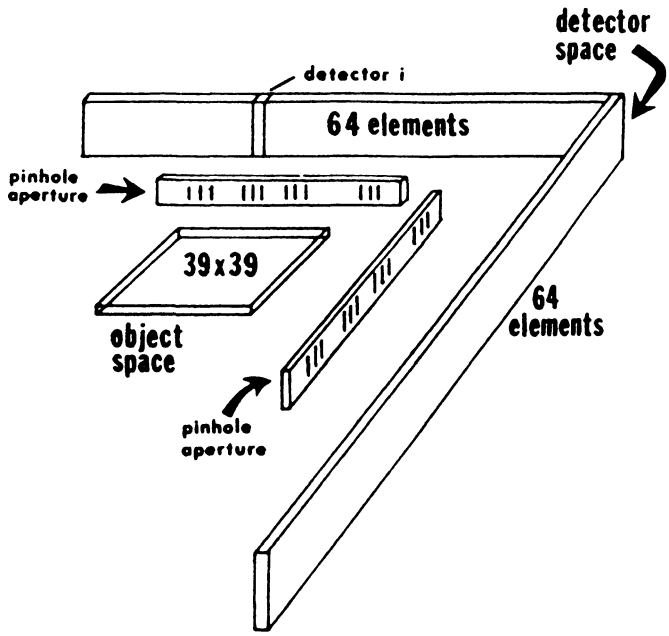

Figure 1. A schematic of the two-dimensional multiple-pinhole coded-aperture imaging system simulated for this study

this study, in the aperture plane. The location and number of pinholes (or openings) is referred to as the code of the aperture.

In Figure 1 we have depicted what is known as an orthogonal-view system. The detectors are modular scintillation cameras, which have been designed and built in our research group (Aarsvold et. al. (1988), Milster (1987)). In the simulations done here, we have used 64 detector elements in each view for a total of 128 detector elements and a 39 × 39 discretized version of the object for a total of 1521 pixels or basis functions. This means that in the equation $\mathbf{g} = \mathbf{Hf} + \mathbf{n}$, $\mathbf{g}$ and $\mathbf{n}$ are (128 × 1) column vectors, $\mathbf{f}$ is a (1521 × 1) column vector, and $\mathbf{H}$ is then a (128 × 1521) matrix. The problem then, is given the data $\mathbf{g}$, estimate (reconstruct) the discretized object $\mathbf{f}$ based upon $\mathbf{g}$ and also based upon any prior information that one might have concerning the object.

3. Use of the prior information

As have been done by many others, we view the reconstruction problem statistically (Shepp & Vardi (1982), Lange & Carson (1984), Snyder & Miller (1985), Veklerov & Llacer (1987), Geman & McClure (1985a, 1985b), Liang & Hart (1987a, 1987b, 1987c), Miller et. al. (1986), Hart & Liang (1987), Levitan & Herman (1987), Jennison & Jubb (1987)). Specifically, we view the collection of the data as a Poisson process, so that each g_i is an independently distributed random variable with mean equal to the ith component of the vector $\mathbf{Hf}$. We assume that there are m detector elements which gives the likelihood function of the data as

$$p(\mathbf{g}|\mathbf{f}) = \prod_{i=1}^{m} p(\mathbf{g}_i|\mathbf{f}) = \prod_{i=1}^{m} \frac{\exp[-(\mathbf{Hf})_i][(\mathbf{Hf})_i]^{g_i}}{(g_i)!}, \tag{3.1}$$

where we unabashedly fail to distinguish between a random variable and the value of a random variable with our notation. The principle of maximum a posteriori (map) estimation would then tell us to find the $\mathbf{f}$ that solves

$$\max_{\mathbf{f}} p(\mathbf{f}|\mathbf{g}), \text{ or equivalently, } \max_{\mathbf{f}} p(\mathbf{g}|\mathbf{f})p(\mathbf{f}), \tag{3.2}$$

where $p(\mathbf{f})$ is the prior distribution on the object $\mathbf{f}$. For various specifications of $p(\mathbf{f})$, see (Geman & McClure (1985a, 1985b), Liang & Hart (1987a, 1987b, 1987c), Hart & Liang (1987), Levitan & Heman (1987), Jennison & Jubb (1987)). If then, we have any prior information about the object that we wish to use, we need to specify this in terms of some $p(\mathbf{f})$ or else incorporate it into our reconstruction scheme in some other manner.

The prior information that will be used in this work is as follows:

(1) We assume that the object is nearly binary. By this it is meant that each pixel must be very near one of two distinct values.

(2) Neighboring pixels tend to behave alike. In other words, if two pixels are defined to be neighbors (e.g. pixels which sit adjacent to each other), they would tend to have the same (or very near the same) average activity throughout the pixel. This will not always be the case, for example, near the inner and outer boundaries of the heart muscle around the left ventricle.

(3) The activity in each pixel is non-negative.

Motivated by the first assumption, we model the marginal distribution of each pixel as the sum of two normal distributions, where each normal distribution has a mean equal to one of the two binary values, say a or b. Due to the second assumption, we would expect that if pixel i took on a certain value f_i, then a neighboring pixel, say pixel j, would take on a value f_j the same as, or very near f_i. Pictorially, this is depicted in a simple two-pixel example by Figure 2 below. The conditional distribution of pixel 1 given pixel 2, $p(f_1|f_2)$, reflects this neighbor dependence.

What we would like to do then, is choose the value of pixel k, f_k, based upon knowledge of values of the neighbors of pixel k. We therefore choose f_k from the conditional distribution of pixel k given values of the neighbors of pixel k, $p(f_k|\partial f_k)$, where the notation ∂f_k denotes the given values that the neighbors of pixel k assume. This distribution is approximated by the expression

$$p(f_k|\partial f_k) \simeq p(f_k|f_k \in A)p(f_k \in A) + p(f_k|f_k \in B)p(f_k \in B), \tag{3.3}$$

where $p(f_k|f_k \in A)$ is the probability that pixel k takes on value f_k given that it came from class A, and $p(f_k \in A)$ is the probability that k comes from class A. A similar explanation holds for class B. Note that this expression is simply a weighted sum of the two normal distributions of mean a and b, respectively, where the weights $(p(f_i \in \cdot))$ carry the dependence of pixel k upon its neighbors.

Once f_k is chosen, this value along with the current values for all other pixels forms an estimate of $\mathbf{f}$, say $\widehat{\mathbf{f}}$. This procedure of choosing f_k based upon neighboring values is repeated over and over for each pixel, thus generating a series of estimates. The "best" estimate then is defined as the one that solves

$$\max_{\mathbf{f} \in F} p(\mathbf{g}|\mathbf{f}), \tag{3.4}$$

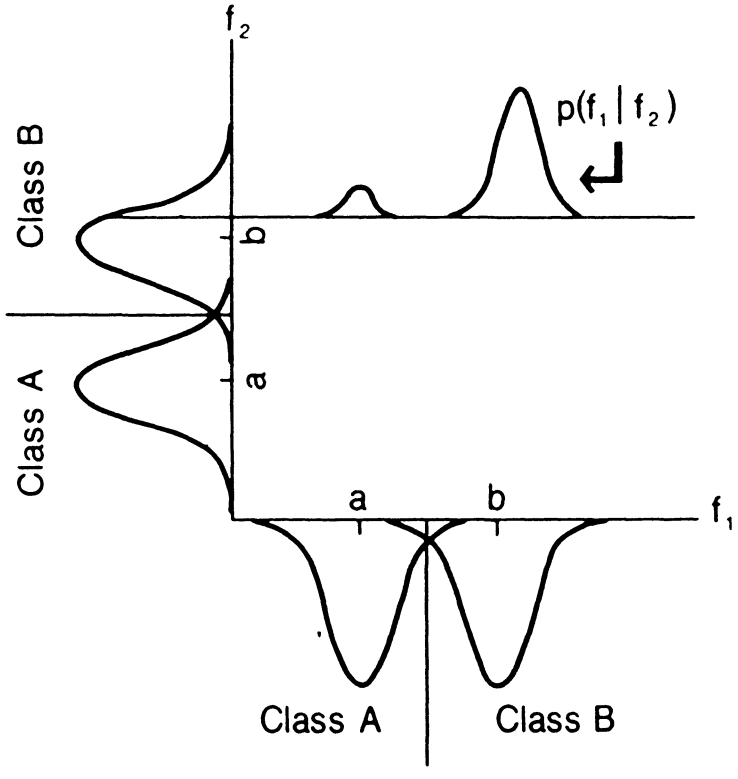

Figure 2. The f_1 and f_2 axes contain the possible values that pixel 1 and pixel 2, respectively, can take on. These possible values follow a distribution that is the sum of two normal laws, say $N(a, \sigma_a^2)$ and $N(b, \sigma_b^2)$. Each pixel is classified as belonging to class A or class B according to which side of the dividing line it falls on. The conditional distribution of f_1 given f_2 reflects the amount of correlation between pixel 1 and pixel 2.

where F denotes the class of estimates that is consistent with the above sampling procedure. The maximum in (3.4) is estimated using the simulated annealing algorithm. Simulated annealing is a Monte Carlo procedure that estimates the global extremum of a function. For details on this algorithm, see (Geman & Geman (1984), Mitra et. al. (1986), van Laarhoven & Aarts (1987), Smith et. al. (1985), Paxman et. al. (1985)).

Notice that by using the sampling procedure explained above, we are encouraging (forcing actually, at least to the extent that the variances put on the normal distributions allow) nearly binary objects. We are also encouraging neighbors to behave alike by choosing a pixel's value from this "conditional" distribution since it is chosen to agree in some sense with its neighbors. Results of this technique are shown in Figure 4b.

To investigate the effect of the binary restriction and the agreeing-neighbors relation, we sampled f_k from a uniform distribution as opposed to this "conditional" distribution since the uniform distribution by itself does nothing to encourage binary objects or neighbors to behave alike. The solution to (3.4) is then sought, where F now

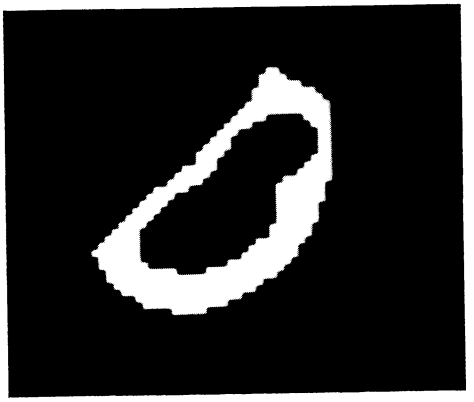

Figure 4a. The object to be reconstructed. As stated earlier, this is a digitized version of the left ventricle and the surrounding heart muscle.

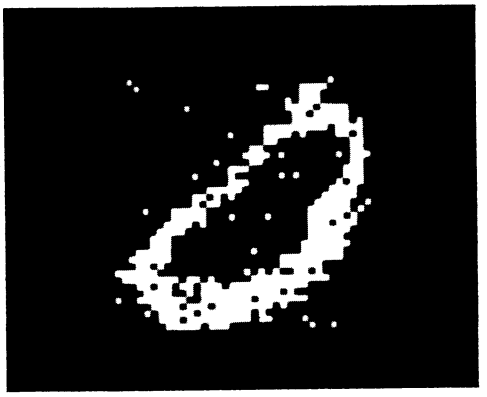

Figure 4b. The estimate of the solution of (3.4), where F is taken to be the class of estimates consistent with the "conditional" sampling scheme. In Figure 3, the boxes indexed $(2,1)$ and $(2,2)$ would be checked, i.e., the smoothness and binary constraints are introduced only through the sampling scheme.

becomes the class of estimates consistent with the uniform sampling procedure. Results of this are shown in Figure 4d.

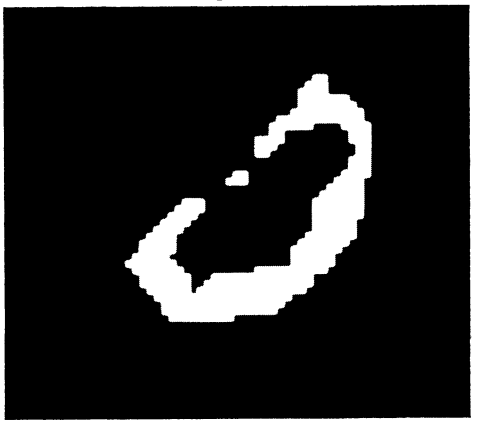

Figure 4c. A 3×3 median-filtered version of 4b.

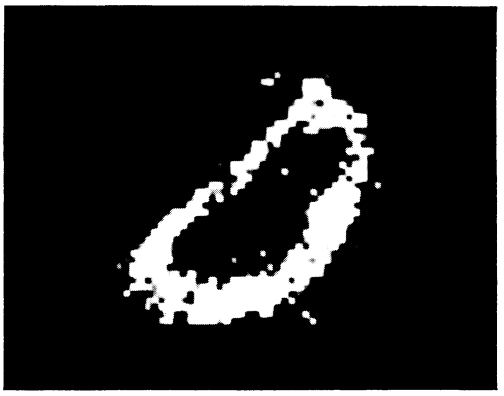

Figure 4d. The estimate of the solution to (.4), where F is taken to be the class of estimates consistent with the uniform sampling scheme. In Figure 3, none of the boxes would be checked, i.e., neither the smoothness nor the binary constraint is introduced as prior information.

Equation (3.4) is now modified by including a term for a prior $p(\mathbf{f})$. The form of $p(\mathbf{f})$ used gives a "prize" for the object being nearly binary and is simply taken as the product of all marginals $p(f_i)$, i.e.

$$p(\mathbf{f}) = \prod_{i=1}^{n} p(\mathbf{f}_i). \tag{3.5}$$

Recall that each $p(f_i)$ is represented as the sum of two normal distributions so that

$$p(f_i) = K_1 \exp\left\{-\frac{(f_i - a)^2}{2\sigma_a^2}\right\} + K_2 \exp\left\{-\frac{(f_i - b)^2}{2\sigma_b^2}\right\}, \tag{3.6}$$

where K_1, K_2 are normalizing constants, and σ_a^2, σ_b^2 are the variances of the normal distributions of mean a and b, respectively. This form of $p(\mathbf{f})$ obviously does not assume any correlation or smoothness in $\mathbf{f}$ since independence is necessarily assumed for $p(\mathbf{f})$ to be formulated in the above manner.

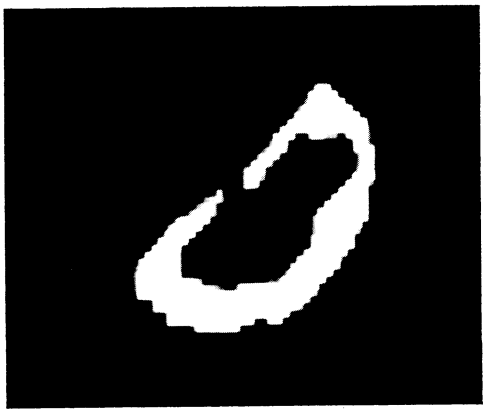

Figure 4e. A 3 × 3 median-filtered version of 4d.

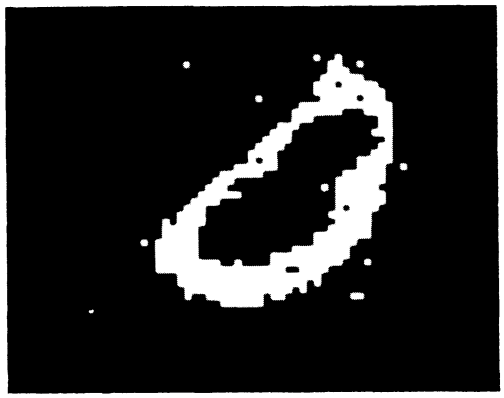

Figure 4f. The estimate of the solution to (3.7), where F is taken to be the class of estimates consistent with the "conditional" sampling scheme. In Figure 3, the boxes indexed $(1,2)$, $(2,1)$, and $(2,2)$ would be checked, i.e., smoothness and binary constraints are introduced through the sampling scheme, while the binary constraint is reinforced by specification of $p(\mathbf{f})$.

We therefore attempt to introduce the smoothness or correlation in the reconstruction by resorting to the "conditional" sampling technique. The problem to solve then becomes

$$\max_{\mathbf{f}\in F} p(\mathbf{g}|\mathbf{f})p(\mathbf{f}), \tag{3.7}$$

where $p(\mathbf{f})$ is given by (3.5), (3.6) and F is the class of estimates consistent with the appropriate sampling scheme. Results of this procedure, where F is taken to be the class of estimates consistent with the "conditional" sampling scheme, are shown in Figure 4f.

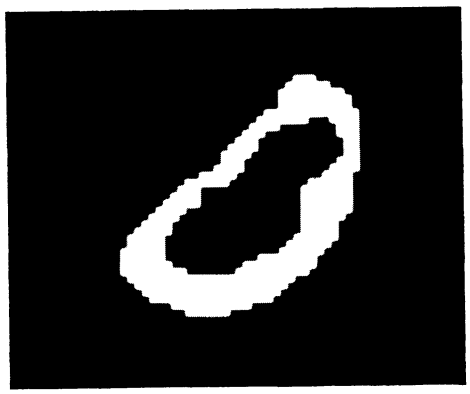

Figure 4g. A 3×3 median-filtered version of 4f.

Now note that the binary structure of the reconstruction is appearing through the specification of $p(\mathbf{f})$ as well as through our "conditional" sampling procedure, and that this sampling also tends to encourage neighbors to behave similarly. We then, as done above, sample f_k from a uniform distribution as opposed to the "conditional" distribution, which would seem to take away the similar-neighbors constraint yet maintain the binary aspect because of the $p(\mathbf{f})$ term. This means that F is then taken as the class of estimates consistent with the uniform sampling scheme. These results are shown in Figure 4h.

Figure 4h. The estimate of the solution of (3.7), where F is taken to be the class of estimates consistent with the uniform sampling scheme. In Figure 3, the box indexed (1,2) would be checked, i.e., the binary constraint is introduced only through specification of $p(\mathbf{f})$.

We have attempted in this section, to incorporate some quite restrictive prior information into our reconstruction scheme in a somewhat different manner. Two ways of introducing our prior knowledge were considered: (1) by specification of $p(\mathbf{f})$; (2) by choosing estimates from a restricted class of estimates. This may be summarized in the 2×2 table shown below in Figure 3.

	smoothness	binary
p(f)	(1,1)	(1,2)
sampling f	(2,1)	(2,2)

Figure 3. The rows of this table represent the two ways in which the prior information has been introduced into our reconstruction schemes. The columns represent the type of prior information that we introduce.

Depending on which of the schemes discussed earlier is used, the appropriate indexed box (or boxes) in this table may then be checked accordingly.

4. Results

In this section, reconstructions using the schemes discussed in Section 3 are shown.

We have not proposed a figure of merit to compare these reconstructions by other than simply viewing them and then deciding which one looks most like the original object. Using this criterion, it appears that the reconstruction scheme that utilizes the greatest amount of prior information that we have suggested here is the one that

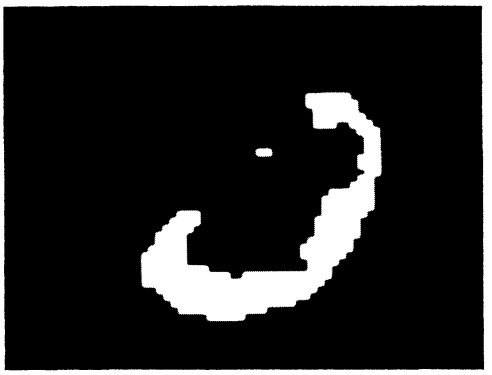

Figure 4i. A 3 x 3 median-filtered version of 4h.

most resembles the original, i.e. Figure 4f (and the filtered version of 4f, Figure 4g). As expected, when the binary constraint is not enforced, the reconstruction exhibits intermediate grey levels as seen in Figure 4d (and the filtered version of 4d, Figure 4e). Figures 4b and 4h appear to be quite similar, but one could perhaps argue that 4b has fewer "isolated" pixels. This may be a result of the "conditional" sampling scheme used to obtain Figure 4b, as this sampling scheme was not used in the reconstruction Figure 4h.

5. Conclusions and further directions

As stated at the end of Section 3, we looked at two ways of introducing some rather restrictive prior information into our reconstruction schemes. These two ways were summarized in the table shown in Figure 3. We have not, by any means, even begun to exhaust all the possibilities for scenarios within this 2 x 2 table. There are many choices of $p(\mathbf{f})$ which could be used to reflect certain prior beliefs. For example, Jennison and Jubb (1987) attempt to capture the smoothness of an (binary) image by modeling the prior density as a Markov Random Field which has the effect of penalizing restorations which have a large number of discrepant neighbors. The initial results of the current study are, nonetheless, encouraging. By looking at the various reconstructions presented in Section 4, it appeared that the ones utilizing the most prior information were the ones that most resembled the original object.

In the future, we plan to do some studies which involve certain figures of merit to help us judge the worth of the different schemes. In particular, we plan to do psychophysical studies to see how the human observer would perform given a certain task. After all, our ultimate goal is to present a physician with a picture from which he/she can make the most accurate diagnosis can be made. It is far from obvious that the most aesthetically pleasing picture to the eye is the one which would yield the best diagnosis, especially if the object has a great deal of structure.

More realistic simulations must also be undertaken. Particularly, a better job (i.e., more realistic) of simulating the projection operator $\mathbf{H}$ needs to be considered. Finally, we would like to do all this analysis with real data collected from a multiple-pinhole coded-aperture imaging system.

References

Aarsvold, J. N., Barrett, H. H., Chen, J., Landesman, A. C., Milster, T. D., Patton, D. D., Roney, T. J., Rowe, R. K., Seacat, R. H. III, & Strimbu, L. M. (1988). Modular scintillation cameras: a progress report. *Proceedings SPIE, Medical Imaging II: Image Data Management and Display*, 319–325.

Barrett, H. H. (1972). Fresnel zone plate imaging in nuclear medicine. *Journal of Nuclear Medicine* **13**, 382.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistics Society (London)* B **39**, 1–38.

Dicke, R. H. (1968). Scatter-hole cameras for x-rays and gamma rays. *Astrophysics Journal* **153**.

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**, 721–741.

Geman, S., & McClure, D. E. (1985a). Statistical methods for tomographic image reconstruction. *Technical Report, Division of Applied Mathematics & Center for Intelligent Control Systems.*

Geman, S., & McClure, D. E. (1985b). Bayesian image analysis: and application to single photon emission tomography. *Proceedings of the American Statistical Association, Statistical Computing Section*, 12–18.

Hart, H., & Liang, Z. (1987). Bayesian image processing in two dimensions. *IEEE Transactions on Medical Imaging* **6**, 201–208.

Jennison, K., & Jubb, M. (1987). Statistical image restoration and refinement. *Proceedings of XII Information Processing in Medical Imaging*, to appear.

Lange, K., & Carson, R. (1984). EM reconstruction algorithms for emission and transmission tomography. *Journal of Computer Assisted Tomography* **8**, 306–316.

Levitan, E., & Herman, G. T. (1987). A maximum a posteriori probability expectation maximization algorithm for image reconstruction in emission tomography. *IEEE Transactions on Medical Imaging* **6**.

Liang, Z., & Hart, H. (1987a). Bayesian image processing of data from constrained source distributions-I. non-valued, uncorrelated and correlated constraints. *Bulletin of Mathematical Biology* **49**, 51–74.

Liang, Z., & Hart, H. (1987b). Bayesian image processing of data from constrained source distributions-II. valued, uncorrelated and correlated constraints. *Bulletin of Mathematical Biology* **49**, 75–91.

Liang, Z., & Hart, H. (1987c). Bayesian image processing of data from constrained source distributions-fuzzy pattern constraints. *Physics in Medicine and Biology* **32**, 1481–1494.

Mertz, L., & Young, N. O. (1960). Fresnel transformation of images. *Proceedings of the International Conference on Optical Instrumentation.* 305.

Miller, M. I., Snyder, D. L., & Moore, S. M. (1986). An evaluation of the use of sieves for producing estimates of radioactivity distributions with the EM algorithm for PET. *IEEE Transactions on Nuclear Science* **33**, 492–495.

Milster, T. D. (1987). *Design and construction of a modular gamma camera.* Ph.D. Thesis, University of Arizona, Tucson.

Mitra, D., Romeo, F., & Sangiovanni-Vincentelli, A. (1986). Convergence and finite-time behavior of simulated annealing. *Advances in Applied Probability* **18**, 747–771.

Paxman, R. G., Barrett, H. H., Smith, W. E., & Milster, T. D. (1985). Image reconstruction from coded data: II Code design. *Journal of the Optical Society of America* A **2**, 501–509.

Rogers, W. L., Han, K. S., Jones, L. W., & Beierwaltes, W. H. (1972). Application of a Fresnel zone plate to gamma-ray imaging. *Journal of Nuclear Medicine* **21**, 371–378.

Shepp, L. A., & Vardi, Y. (1982). Maximum likelihood reconstruction for emission tomography. *IEEE Transactions on Medical Imaging* **1**, 113–121.

Smith, W. E., Paxman, R. G., & Barrett, H. H. (1985). Image reconstruction from coded data: I. Reconstruction algorithms and experimental results. *Journal of the Optical Society of America* A **2**, 491–500.

Snyder, D. L., & Miller, M. I. (1985). The use of sieves to stabilize images produced with the EM algorithm for emission tomography. *IEEE Transactions on Nuclear Science* **32**, 3864–3872.

Tipton, M. D., Dowdy, J., & Stokely, E. M. (1976). Background suppression of multiple pinhole-coded aperture scintigrams. *Proceedings of the International Conference on Medical Physics AAPM, 4th.*

van Laarhoven, P. J. M., & Aarts, E. H. L. (1987). *Simulated Annealing: Theory and Applications.* D. Reidel, Dordrecht, Holland.

Veklerov, E., & Llacer, J. (1987). Stopping rule for the MLE algorithm based on statistical hypothesis testing. *IEEE Transactions on Medical Imaging* **6**, 313–319.