LBL-32488
UC-405

Lawrence Berkeley Laboratory
UNIVERSITY OF CALIFORNIA

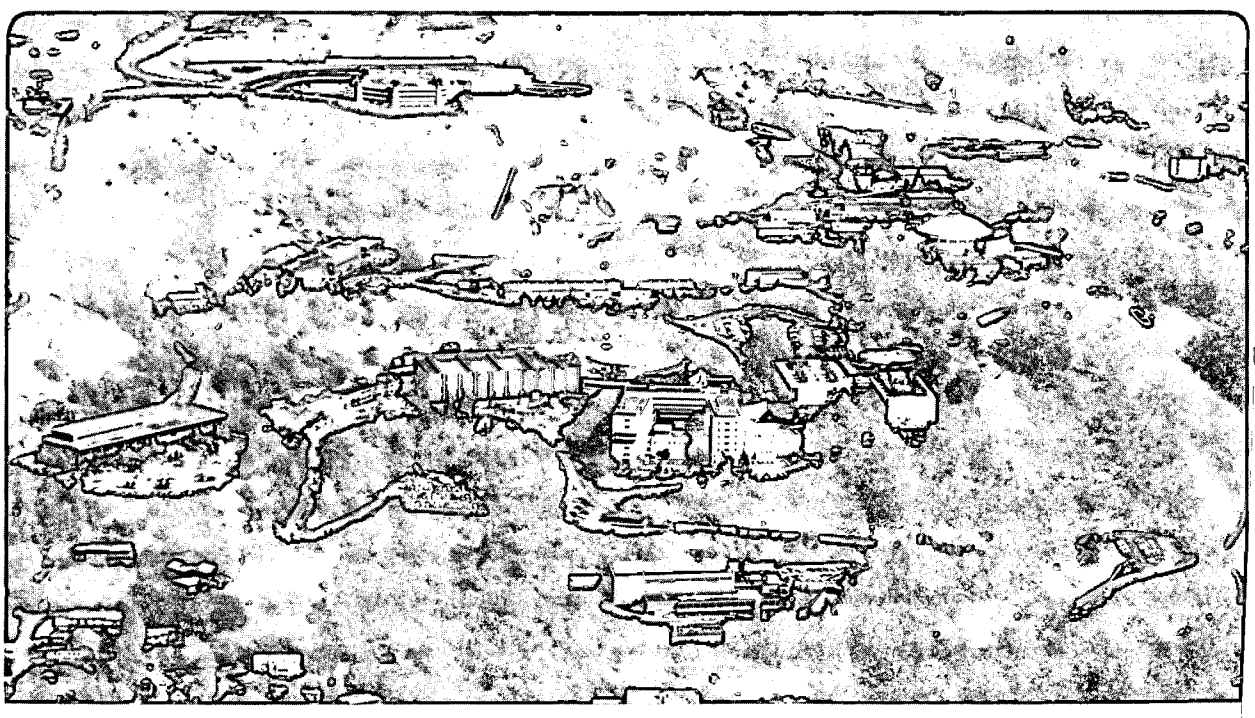
Engineering Division

Presented at the SPIE Mathematical Methods in Medical Imaging,
San Diego, CA, July 22–23, 1992,
and to be published in the Proceedings

The Use of Geometric Prior Information in Bayesian Tomographic Image Reconstruction: A Preliminary Report

J. Llacer, B.M. ter Haar Romeny, and M.A. Viergever

June 1992



LOAN COPY
Circulates
for 4 weeks
Bldg. 50 Library.
Copy 2
LBL-32488

Prepared for the U.S. Department of Energy under Contract Number DE-AC03-76SF00098

DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor the Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or the Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or the Regents of the University of California.

LBL-32488
UC-405

The use of geometric prior information in Bayesian tomographic image reconstruction: a preliminary report.

Jorge Llacer*, Bart M. ter Haar Romeny# and Max A. Viergever#
*Lawrence Berkeley Laboratory, University of California, Berkeley, California, USA,
#University Hospital Utrecht, 3584 CX Utrecht, The Netherlands

ABSTRACT

In this paper we examine the possibility of using pure geometrical information from a prior image to assist in the reconstruction of tomographic data sets with lower number of counts. The situation can arise in dynamic studies, for example, in which the sum image from a number of time frames is available, defining desired regions-of-interest (ROI's) with good accuracy, and the time evolution of uptake in those ROI's needs to be obtained from the low count individual data sets. The prior information must be purely geometrical in such a case, so that the activity in the ROI's of the prior does not influence the estimated uptake from the individual time frames. It is also desired that the prior does not impose any other conditions on the reconstructions, i.e., no smoothness or deviation from a known set of values is desired.

We attack this problem in the framework of Vision Response Functions (VRF's), based on the work done by J.J. Koenderink in Utrecht. We show that there are assemblies of VRF's that can be presented in a form that is invariant with respect to rotations and translations and that some functions of those invariants can convey the desired geometric prior information independent of the level of activity in the ROI's, except at very low levels.

Preliminary results based on a one-dimensional reconstruction problem will be presented. Using the zero crossings of the Gaussian derivative form of the Laplacian of a prior image at different scales, a variant of the EM algorithm has been found that allows the reconstruction of low count data sets with those priors. At this time, this involves using a modified Conjugate Gradient (CG) maximization method for the M-step of the algorithm. The results show that the distorted shapes of reconstructions of data sets with low counts are effectively corrected by the method, although many questions exist at this time about basic and computational issues.

1. - MOTIVATION FOR PRESENT STUDY

Reconstruction from Positron Emission Tomography (PET) data by Maximum Likelihood Estimator (MLE) methods has reached a certain degree of maturity. Images obtained by the appropriate use of a cross-validation stopping rule, followed by slight filtering, have been shown to provide reasonably unbiassed images with one particular characteristic, when compared with Filtered Backprojection (FBP) images: the expected error in a particular region of a reconstruction is approximately proportional to the square root of the average number of counts in that region, while FBP images show an expected error which is high and almost independent of that average number of counts[1].

It is felt that improvements on reconstructions will have to come, then, from the use of *a priori* information in a Bayesian context. Those improvements, if they can be attained, are still necessary. Figure 1 shows the reconstruction of a slice of a data set obtained from a normal volunteer in a Fluoro-deoxiglucose (FDG) study of the brain. The reference image contained approximately 30 million (30M) counts, corresponding to a very good data set, while the FBP and MLE (1.4M) images correspond to the images that one can expect to obtain in one single time frame of a dynamic study of a patient. It is quite clear that the 1.4M reconstructions contain distortions of many structures which are expected to be imaged correctly in the reference data set. Interestingly enough, both FBP and MLE reconstructions show nearly identical distortions. The same effect has been repeatedly observed in Bayesian reconstructions with entropy prior information, or with the method of sieves. It appears, then, that reconstructions whose information is only due to the measured tomographic data, will have similar distortions, regardless of the method of reconstruction.

Dynamic studies in which data are obtained from a patient over a number of time frames are quite common. The reconstruction of the data set obtained by adding all the time frames is often used by the physician for the purpose of defining one or more Regions-of-Interest (ROI's). Then, the data at different time frames are reconstructed and examined in order to obtain the time dependence of the uptake of a radioisotope in the ROI's, for example. The motivation of the present study, then, is to attempt to answer the following question:

- Can the knowledge obtained from the reference image help with the reconstruction of the low count images?

Since the low count images may have varying rates of radioisotope uptake in different parts of the image, only the geometrical information (not the uptake information) of the reference image is desired. The prior information that we extract from the reference image must, therefore, be substantially independent of the amount of activity in a given structure, except, of course, when there is little or no activity in it. In this study we will look at the invariant forms of assemblies of vision response functions (VRF's) as prior information for reconstruction. The only properties of the representation chosen to convery prior information which make sense are the ones that are independent of the chosen coordinates, but rather depend on the object itself. These properties are called invariant (under a certain group of transofmrations). Because of the non-projective origin of the data, we may constrain ourselves to the group of orthogonal transformations, i.e., translations and rotations. In some cases we like an additional invariance: invariance under general intensity transformations, which occurs when contrast or brightness adjustments take place. In our case, this means invariance of geometric information under radioisotope concentration changes, as indicated above.

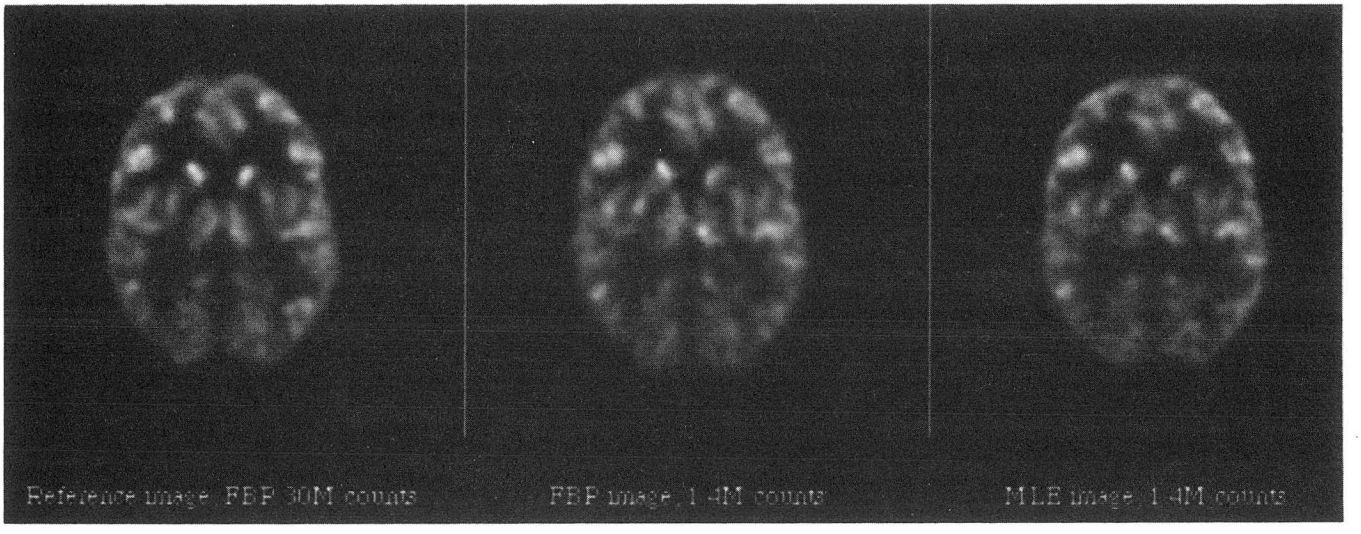


Fig. 1 - Reconstruction of a data set with 30 million counts (30M) and of a data set with 1.4M counts of the same patient showing that distortions of anatomical structures are distorted in both the FBP and MLE reconstruction results.

2. BAYESIAN RECONSTRUCTION

2.1 General formulation

In this report we shall use the following notation:

$\boldsymbol{a}$ - the vector of radioisotope activities to be estimated.

$\boldsymbol{p}$ - the vector of measured projection data

$\boldsymbol{f}$ - the matrix of transition probabilities

The MLE formulation of the reconstruction problems seeks to maximize the likelihood function $P(\boldsymbol{p}|\boldsymbol{a})$, by finding an estimate $\boldsymbol{a}$ for which vector $\boldsymbol{p}$ has the highest probability of being the result of the measurement. A Bayesian target function would answer to the different requirement of obtaining the image $\boldsymbol{a}$ of highest probability *given* a measurement $\boldsymbol{p}$, i.e., maximizing $P(\boldsymbol{a}|\boldsymbol{p})$. The latter target function, if it can be defined, is the correct one to maximize in the case of noisy data: Maximizing $P(\boldsymbol{p}|\boldsymbol{a})$ will result in an image whose projection into data space follows the noisy data as close as possible, except when stopped by an appropriate criterion[2,3]. Maximizing the second target function, $P(\boldsymbol{a}|\boldsymbol{p})$, need not have that characteristic.

By using Bayes' rule,

$$P(a|p) = P(p|a)P(a) / P(p). \tag{1}$$

The Bayesian target function incorporates the likelihood function, or conditional probability $P(p|a)$, but it also incorporates $P(a)$, the probability distribution of the image a, or image "prior". (Note that the probability $P(p)$ is unity). If we know that the image has a certain set of characteristics that can be put in the form of a probability distribution, maximization of (1) can be expected to result in a better reconstruction than obtained by maximizing $P(p|a)$ alone.

Two basic types of priors that have been investigated by different groups: 1) non-informative priors, which result in a regularization of the MLE process, and 2) informative priors that use knowledge specific to the image being reconstructed. Among the latter, the work initiated by the Geman brothers in 1984[4] , and followed by a number of researchers, uses Gibbs energy functions to enforce a certain degree of smoothness in neighbourhoods separated by boundaries. The recent work by Leahy and Yan[5] and by C.T. Chen and X. Ouyang[6] are current examples of that work: using anatomical boundaires extracted from MR images, they have demonstrated improved results in simulated PET reconstructions. In the work reported here we will examine the use of geometrical information obtained from sum images of PET studies with multiple time frames as a way to assist in the reconstruction of data from single time frames.

2.2 The EM algorithm

The EM algorithm[7] will be used as a basis for the work to be described here. The resulting two-step process for Maximum Likelihood reconstruction of PET data was initially reported by Shepp and Vardi[8] and it will be described here in a form similar to the one that we are using for our MLE-PET reconstruction work[1]. We will consider the application of absorption and detector non-uniformity corrections, but, to simplify notation, we will leave out the correction for randoms background. We will also leave out considerations relating to conservation of counts.

Expanding the above notation into a more complete form, we will use:

$p_j, j = 1,\ldots,J$	projection data, counts detected in tube j.
$c_j, j = 1,\ldots,J$	absorption and gain corrections vector
$a_i, i = 1,\ldots,I$	activity in a pixel, the image being estimated
$a_i^{(k)}, i = 1,\ldots I$	the activity estimated at the kth iteration
f_{ji}	transition matrix elements, uncorrected
$f'_{ji} = f_{ji} / c_j$	transition matrix elements, corrected
$h_j = \sum_{i=1}^{I} \frac{f_{ji} a_i}{c_j} = \sum_{i=1}^{I} f'_{ji}\, a_i$	the projection of an image estimate.

An expression for the corrected sensitivity of the detector system for a pixel will also be defined as

$$q'_i = \sum_{j=1}^{J} \left(\frac{f_{ji}}{c_j} \right). \tag{2}$$

For the development of the EM algorithm we will make use of a vector $\boldsymbol{x}$, the "complete" data set, whose elements x_{ji} are the number of disintegrations from pixel i that are detected in tube j. The relationship between the complete data set and the "incomplete" set $\boldsymbol{p}$ is given by

$$p_j = \sum_i x_{ji} \quad . \tag{3}$$

We will use the EM algorithm to estimate $\boldsymbol{a}$ by maximizing the log likelihood of $\boldsymbol{x}$ given an activity distribution $\boldsymbol{a}$. The complete data set $\boldsymbol{x}$ has elements which are Poisson distributed with means $f'_{ji}\, a_i$, while the detected counts or incomplete data set vector $\boldsymbol{p}$ has elements which are Poisson variables with mean $\sum f'_{ji}\, a_i$. We consider only the case in which the data $\boldsymbol{p}$ are uncorrected for attenuation, etc., i.e., they are purely Poisson in nature.

The target function to be maximized is, from Ref. 8,

$$L(\boldsymbol{a}) = \log \mathrm{P}(\boldsymbol{x}|\boldsymbol{a}) = \sum_j \sum_i \left[-f'_{ji}\, a_i + x_{ji} \log\left(f'_{ji}\, a_i\right) - \log\left(x_{ji}!\right) \right] \tag{4}$$

where P is the conditional probability of a complete data set $\boldsymbol{x}$ having resulted from an activity distribution $\boldsymbol{a}$.

E-step: Shepp and Vardi show in Ref. 8 that, for Poisson random variables x_{ji}, the conditional expected value of x_{ji} given the data elements p_j for a current estimate of the parameter $a_i^{(k)}$ is

$$x_{ji}^{(k)} = E\left\{ x_{ji} | \mathbf{p}, \mathbf{a}^{(k)} \right\} = \frac{a_i^{(k)} f'_{ji}\, p_j}{\sum_{i'} f'_{ji'}\, a_{i'}^{(k)}} \tag{5}$$

A physical interpretation of this estimate can be seen by considering that if a tube has p_j counts and the pixel values are $a_i^{(k)}$, the expected fraction of p_j that was emitted by the ith pixel is $f'_{ji}\, a_i^{(k)} / \sum_{i'} f'_{ji'}\, a_{i'}^{(k)}$.

We can then compute the expected value of the target function Eq. (4) given the incomplete data $\boldsymbol{p}$ and the current estimate $\boldsymbol{a}^{(k)}$

$$E\left\{ L(\boldsymbol{a}|\boldsymbol{p}, \boldsymbol{a}^{(k)}) \right\} = \sum_j \sum_i \left[-f'_{ji}\, a_i + E\left\{ x_{ji} | \boldsymbol{p}, \boldsymbol{a}^{(k)} \right\} \log\left(f'_{ji}\, a_i\right) - E\left\{ \log\left(x_{ji}!\right) \right\} \right] \tag{6}$$

Replacing from Eq. (5), we have

$$E\left\{ L(\mathbf{a}|\mathbf{p}, \mathbf{a}^{(k)}) \right\} = \sum_j \sum_i \left[-f'_{ji}\, a_i + \frac{p_j f'_{ji}\, a_i^{(k)}}{\sum_{i'} f'_{ji'}\, a_{i'}^{(k)}} \log\left(f'_{ji}\, a_i\right) + \text{terms independent of } a_i \right] \tag{7}$$

M-step: For the MLE case, we now maximize the expectation Eq. (7) with respect to the parameters $\boldsymbol{a}$, which will lead to the new values $\boldsymbol{a}^{(k+1)}$ Setting the partial derivatives with respect to a_i equal to zero,

$$\frac{\partial E\{\ \}}{\partial a_i} = \sum_j \left[-f'_{ji} + \frac{p_j f'_{ji}\, a_i^{(k)}}{\sum_{i'} f'_{ji'}\, a_{i'}^{(k)}} \frac{1}{a_i} \right] = 0 \tag{8}$$

and solving for a_i, we get the next estimate of $a_i^{(k+1)}$:

$$a_i^{(k+1)} = \frac{a_i^{(k)}}{q'_i} \sum_j \left\{ \frac{p_j f'_{ji}}{\sum_{i'} f'_{ji'}\, a_{i'}^{(k)}} \right\} \tag{9}$$

In Bayesian reconstructions, with the target function containing additional terms that depend exclussively on the estimate $\boldsymbol{a}$, the E step will be identical to that of the MLE, as given in (5), above. The M step, however, will require the maximization of a target function different from (4), which may result in the need to obtain the maximum by an iterative method. The M step will be further discussed below after the appropriate prior form for our problem has been established.

3. - INFORMATIVE PRIORS

3.1 Vision Response Functions (VRF's)

We want to develop priors that contain pure geometrical information about an image, but are to some extent inmune to noise. It appears that Vision Response Functions, as proposed and developed by J.J. Koenderink and a substantial group of followers and contributors at the University of Utrecht over the last several years, may provide the tools for the development of such priors. An excellent summary of properties of the VRF's has been provided by B.M. ter Haar-Romeny et al[9]. From the practical point of view of using VRF's as descriptors of prior information in an environment with noisy, poorly resolved images, as in Emission Tomography, it appears that the three characteristics indicated below make VRF's appealing:

1 - VRF's make use of the concept of scale. Desirable features in an image may appear at a particular range of scales, while noise and image distortions may be detectable at a different range of scales.

2 - The basic operation in using VRF's is the convolution of an image with derivatives of Gaussian functions. Because of the averaging effect of the operation, extraction of derivatives, for example, does not result in increased noise, even at higher orders.

3 - The VRF's are "soft" local operators. The value of a particular pixel in the convolution of an image with a VRF will be influenced by a local neighbourhood whose size depends on the chosen scale, with soft edges.

3.2 Invariants as priors in tomographic reconstruction

It is possible to combine Gaussian derivative operators in such a manner that image features are detected independently of orientation and translation. These combinations are called "invariants" and they form the principal group of operators in which we are interested. As was discussed in Section 1, we are interested in geometric descriptors that are insensitive to the intensity of a feature. If we are describing, for example, a particular anatomical structure in a human brain whose radioisotope uptake as a function of time is substantially different from that of surrounding tissue, we would not want the descriptor of the anatomical structure to change shape as a function of time, except, of course, when the uptake is very small or zero.

For that reason, we need to restrict ourselves to invariants that have some characteristic that is independent of image intensity. An initial search up to second order derivatives has yielded two invariants whose zero crossings are by nature independent of intensity transformations:

a) The laplacian L_{ii} and b) the umbilicity $U = \frac{L_{ll}^2}{L_{ij}L_{ji}} - 1$.

We note that we use the convention that repeated indices imply summation over all possible index values, i.e., the dimensions,

$$L_{ii} = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}. \tag{10}$$

Furthermore, applying the L_{ii} operator implies convolving with the Laplacian of a two-dimensional Gaussian function of a given standard deviation σ, corresponding to the chosen scale.

The loci of zeros of the L_{ii} operator, when applied to a two dimensional image, are closely related to edges: If one defines an edge as the locus of points of inflection defined by the vanishing of the second derivative along the local gradient direction, then the zeros of the L_{ii} operator accurately describe edges provided that lines of equal intensity (isophotes) are sufficiently straight in the neighborhood[10].

The loci of zeros of the U operator can be best understood by considering an image as a 3-dimensional entity, in which the z-dimension is the intensity of the image. Curvature is define as the reciprocal of the radius of a circle that fits in the neighborhood of the point considered. The directions of maximum and minimum curvature (the principal curvatures) at a point are at right angles with respect to each other. I both curvatures are positive, we have a "mountain", if both are negative, we have a "valley". If the principal curvatures are of opposite sign, we have a saddle point. Then, Gaussian curvature is the product of the two principal curvatures, so that regions with positive Gaussian curvature are "mountains" or "valleys" (elliptical areas), and regions with negative Gaussian curvature are saddle points (hyperbolic areas). Then, the loci of $U = 0$, separates the elliptical from the hyperbolic areas and can be considered as a "patch classifier"[10].

The behavior of the L_{ii} and U operators can be further understood by considering the following simulated PET example. Fig. 2, top left, shows a phantom in which an elliptical ring of 100% activity contains eight ellipses in a 25% background. The 128 x 128 image contains Poisson noise corresponding to a total of 20 million (20M) counts. The left ellipses are all at 100% intensity, while the right ellipses have intensities of 30, 50, 70 and 90%. The other two images in Fig. 2 show the result of applying the L_{ii} operator to the 20M count image at two different small scales. The zero level corresponds to middle grey.

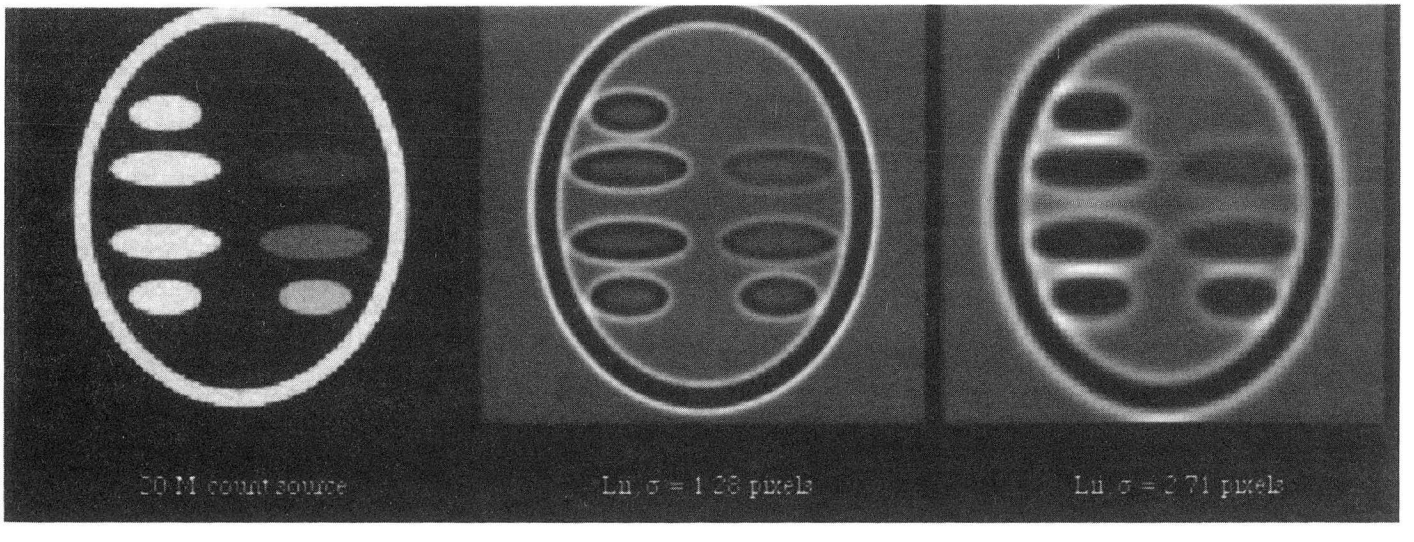


Fig. 2 - Mathematically generated phantom with 8 elliptical regions and the result of applying the L_{ii} operator at two different scales. The top right ellipse of the set may not be easily seen in the printed page.

Fig. 3 shows the image of Fig. 2 with overlays corresponding to the zero crossings of the L_{ii} operator for three different scales. At σ = 1.28 pixels, it is evident that the zero crossings are picking up a considerable amount of noise, which has largely disappeared at σ = 2.11 pixels. It is interesting to note that the contours obtained for the right hand ellipses preserve the edge shape quite well, in spite of having decreased intensity when compared to the left hand ellipses. This is particularly significant, since this is a required characteristic for the image descriptors in which we are interested. The zero crossings of the L_{ii} operator in two dimensions always result in closed contours.

Fig. 4 shows the results of applying the operator U to the image of Fig. 2 for a scale of σ=1.28 pixels, where noise effects are prominent. The zero crossings of that operator at σ=1.28 and 2.11 are also shown. Patch separation generated by the U operator at σ=2.11 is not too different (for this simple phantom) from the approximate edge detection of the L_{ii} operator. This is so because the different ellipses are actually "mountains" and they are separated from the valleys, where the two principal curvatures will not generally have the same sign. The zero crossings of the U operator are also insensitive to intensity level, except for noise effects, a desirable characteristic for our purposes.

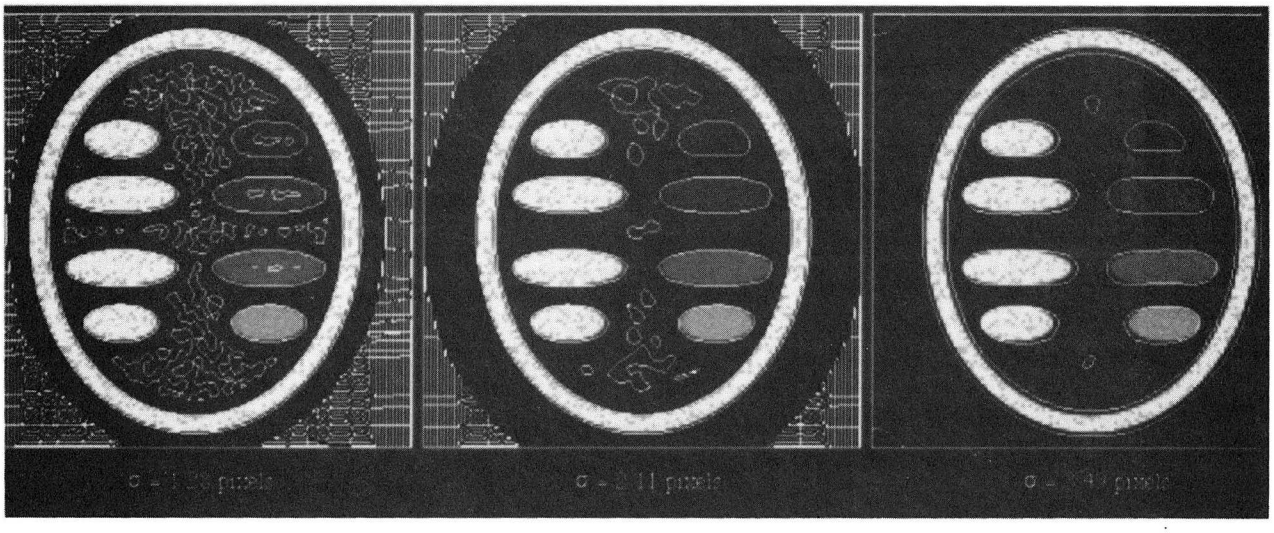


Fig. 3 - Overlays of the zero crossings of the L_{ii} operator on the image of Fig. 1. Notice that contour shape is practically independent of intensity of the elliptical region, except for noise effects.

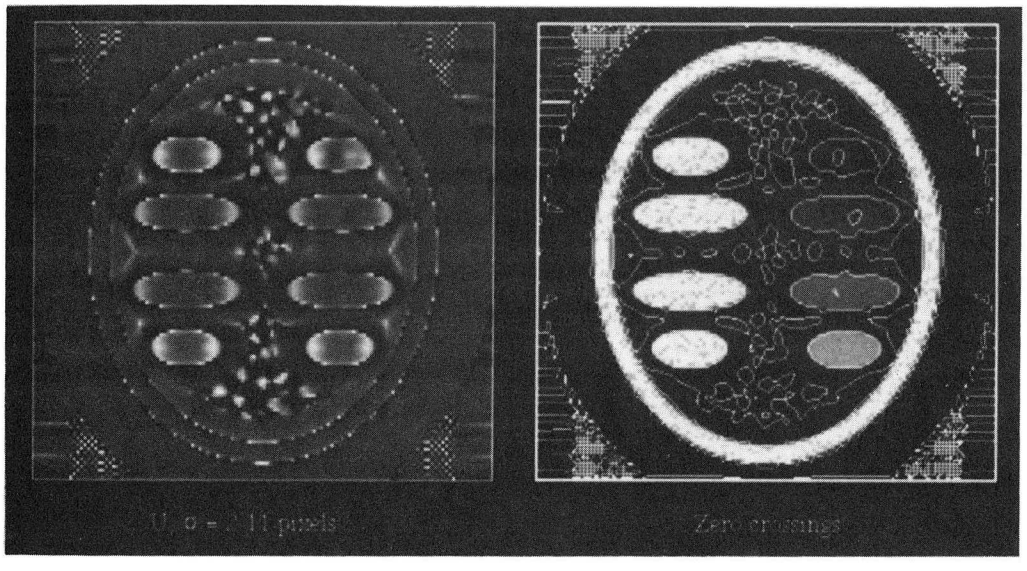


Fig. 4 - The Umbilicity operator applied to the image of Fig. 2 and the zero crossings of that operator at a scale of σ=2.11 pixels.

3.3 Examples from human FDG studies

Data from PET data of FDG human brain studies have been reconstructed and processed by the L_{ii} and the U operators at different scales and the zero crossings have been found. Images have been obtained first from a data set with excellent statistics (30M counts) and then from a low count data set (1.4 M counts). Fig. 5 shows the zero crossings of the L_{ii} and U operators superimposed on the reconstructed images for data sets in the upper part of a brain. The top images correspond to a 30M count data set (very good statistics), while the bottom images are for the 1.4M count data. One scale is displayed (σ = 1.28 pixels) for the L_{ii} operator and one scale (σ = 1.65 pixels) for the U operator. The chosen scales seem to convey a maximum of geometrical information for the structures which dominate those brain images, without being affected significantly by small scale noise. The zero crossings of the 1.4M images are substantially different from those of the 30M images.The images of Fig. 5 illustrate the question that we are trying to answer: Can a knowledge of the priors shown in the top row of the figure (30M count images) assist in reconstruction of the images at the bottom row (1.4M counts)?

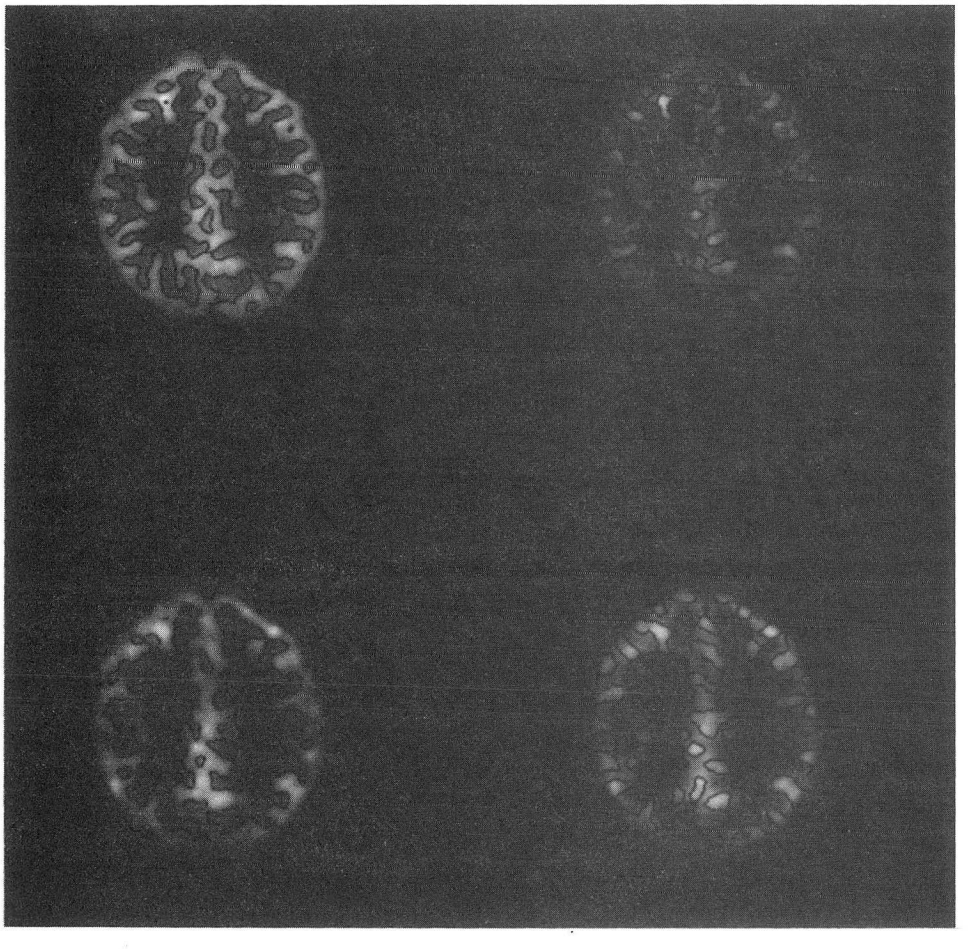


Fig. 5 - Reconstructions of human brain FDG studies with zeros of the Laplacian (left) and Umbilicity (right) operators supersimposed. Top: 30M count data, bottom: 1.4M count data.

4. ONE-DIMENSIONAL RECONSTRUCTION PROBLEM

4.1 Preliminary objectives

In order to attempt to answer the above question, it was felt that starting with two-dimensional images would be too complex and time comsuming. Instead one-dimensional reconstruction would be attempted. The initial objectives have been: 1) to find out whether the EM formulation could be used in defining an algorithm to solve a Bayesian target function that includes zero-crossings of invariant operators, 2) to see whether it is possible to force the zero crossings of a reconstruction with low counts into concordance with the zero crossings of a high count prior and 3) to find out what happens to the expected error in a region of interest (and to other regions) when such concordance is achieved.

4.2 Problem definition and MLE reconstructions

For the first experiments, we have chosen one-dimensional images of 256 pixels. An instrument with a point response function (prf) of a |1/x| shape convolved with a small Gaussian resolution kernel was conceived as a way to imitate the response of a tomograph. The prf generated, with a resolution kernel of σ = 1.5 pixels, is shown in Fig. 6. A source consisting of some structures which mimic in some way a cut through a human brain reconstruction was generated, as shown also in Fig. 6. The result of convolving the source with the prf is also shown there.

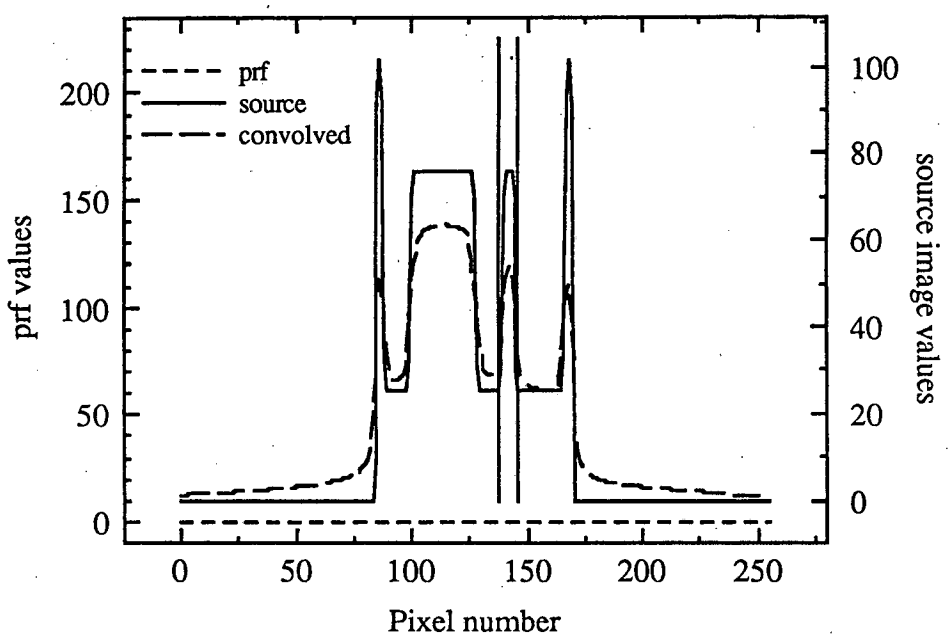


Fig. 6 - One-dimensional problem description. System point response is shown, as well as a source phantom and the source convolved with the prf.

A reference data set containing the Poisson statistics of a total of 500 K counts in the image was then obtained, as well as 10 independent data sets (t0 through t9) containing 10 K counts each. A standard MLE algorithm was prepared for the one-dimensional reconstruction, with a cross-validation stopping rule[1]. The 500K data set and the 10K data sets were then reconstructed. No post-filtering operation was carried out on the MLE results, which show some amount of ringing. A specific ROI was selected for evaluation of the current approach to Bayesian reconstruction, shown in Fig. 6 by vertical lines. The locations of the L_{ii}=0 points (which always appear in pairs in one-dimensional images) was chosen for conveying the prior information from the 500K results into the reconstruction of the 10K data.

4.3 Bayesian reconstructions with a limited Region-of-interest (ROI)

4.3.1 Algorithm development

We will first define a description for a prior based on an energy function of the form $(1/K)e^{-\alpha V(a)}$. With such an energy function, the target function to be maximized will be, from (4),

$$B(a) = \sum_j \sum_i \left[-f'_{ji} a_i + x_{ji} \log\left(f'_{ji} a_i\right) - \log\left(x_{ji}!\right) \right] - \alpha V(a) \tag{11}$$

As indicated earlier, the E step of the EM algorithm is identical to the one for the MLE algorithm, i.e., the conditional expected value of x_{ji} given the data elements p_j for a current estimate of the parameter $a_i^{(k)}$ is

$$x_{ji}^{(k)} = E\{x_{ji} | \mathbf{p}, \mathbf{a}^{(k)}\} = \frac{a_i^{(k)} f'_{ji} p_j}{\sum_{i'} f'_{ji'} a_{i'}^{(k)}} \tag{12}$$

but, from (7) and (11), the M step will consist in maximizing

$$E\{B(a|p, a^{(k)})\} = \sum_j \sum_i \left[-f'_{ji} a_i + \frac{p_j f'_{ji} a_i^{(k)}}{\sum_{i'} f'_{ji'} a_{i'}^{(k)}} \log\left(f'_{ji} a_i\right) \right] - \alpha V(a) \tag{13}$$

with respect to the a_i elements.

4.3.1.1 The distance function

The following is a first attempt at defining a distance function that can be used as the potential function V in (13). Consider that we have a prior image A which, when operated on by a Gaussian invariant operator L, has zeros located a Z_m. Then we consider an estimate $\boldsymbol{a}$ obtained during the reconstruction process whose zero crossings of the same operator L are located at z_n.

Next we consider a Region-of-Interest for zero interations ($\mathrm{ROI_Z}$) that contains those zeros Z_m that we want the reconstruction to match. We also define an interaction distance D_{int} such that, if a z_n is within D_{int} from a Z_m in $\mathrm{ROI_Z}$, that z_n will contribute to the potential function V. In a more formal manner, a zero z_n of $L(\boldsymbol{a})$ will contribute to V if

$$|z_n - Z_m| \le D_{\mathrm{int}} \tag{14}$$

for some zero Z_m of $L(A)$ in $\mathrm{ROI_Z}$.

Then, we define the potential function

$$V(a) = \frac{1}{n_z} \sum_{k=1}^{n_z} D(z_k) \tag{15}$$

where n_z is the number of zeros that fulfill the condition (14) and $D(z_k)$ is the distance $|z_k - Z_m|$.

With those definitions, the function to maximize for the M step is, from (13) and (15),

$$E\{B(a|p, a^{(k)})\} = \sum_j \sum_i \left[-f'_{ji} a_i + \frac{p_j f'_{ji} a_i^{(k)}}{\sum_{i'} f'_{ji'} a_{i'}^{(k)}} \log\left(f'_{ji} a_i\right) \right] - \alpha \frac{1}{n_z} \sum_{k=1}^{nz} D(z_k). \tag{16}$$

Since the distance terms cannot be written explicitely as functions of the estimate elements a_i, the maximization of (16) cannot be carried out analytically, as is the case with the MLE problem. We will report here the attempts at maximizing (16) by a conjugate gradient method, which appears to have succeeded.

4.3.1.2 Conjugate Gradient maximization

For the Conjugate Gradient (CG) method, we need an expression for the function to be maximized and its partial derivatives with respect to the parameters to be estimated. A modified form of the CG algorithm given in *Numerical Recipes in C*[11] has been used. The modification has consisted in limiting the search for parameter values to $a_i > 0$ for all i, so that the logarithm term in (16) can be evaluated.

The function F to be maximized is precisely equation (16), given above. F will have to be evaluated a large number of times for different values of parameters a_i. For every set of parameters, evaluation of the Bayesian term requires obtaining the zero crossings of $L(\boldsymbol{a})$ and obtaining from them the potencial function of (15).

For the partial derivatives $\frac{\partial F}{\partial a_i}$, the term obtained from the likelihood is easy:

$$\left.\frac{\partial F}{\partial a_i}\right|_{likelihood} = \sum_{j=1}^{J}\left[-f_{ji} + \frac{p_j f'_{ji}\, a_i^{(k)}}{\sum_{i'} f'_{ji'}\, a_{i'}^{(k)}}\frac{1}{a_i}\right] \tag{17}$$

For the Bayesian term, let us consider first some estimate $a(x)$ in a continuous one-dimensional space to which a Gaussian derivative operator L is applied

$$W(x) = \int_{-\infty}^{\infty} dx' L(x-x')a(x'). \tag{18}$$

Next we look for the disturbance caused to $W(x)$ by the application of an increment Δa at some coordinate x_1,

$$\Delta W(x) = \int_{-\infty}^{\infty} dx' L(x-x')\Delta a\boldsymbol{\delta}(x'-x_1) = \Delta a L(x-x_1). \tag{19}$$

Let x_0 be a zero crossing of $W(x)$, i.e.,

$$W(x_0) = \int_{-\infty}^{\infty} dx' L(x_0-x')a(x') = 0. \tag{20}$$

With the increment Δa, there will be change in zero crossing described by Δx_0, so that

$$\int_{-\infty}^{\infty} dx'\left\{L\left[(x_0+\Delta x_0)-x'\right]a(x')\right\} + \Delta a L\left[(x_0+\Delta x_0)-x_1\right] = 0 \tag{21}$$

If we do a first order expansion $L(x_0+\Delta x_0) = L(x_0) + \Delta x_0 L'(x_0)$, and substitute in (21), we obtain the relation

$$\frac{\Delta x_0}{\Delta a} = -\frac{L(x_0 - x_1)}{\int_{-\infty}^{\infty} dx' L'(x_0 - x')a(x') + \Delta a L'(x_0 - x_1)} \tag{22}$$

and, when taking the limit as $\Delta a \to 0$,

$$\frac{\partial x_0}{\partial a} = -\frac{L(x_0 - x_1)}{\int_{-\infty}^{\infty} dx' L'(x_0 - x')a(x')}. \tag{23}$$

The interpretation of (23) is the following: the change in a zero crossing initially at x_0 when the function $a(x)$ is changed infinitesimaly at x_1 is given by the value of the operator L at a distance $(x_0 - x_1)$ from the origin, divided by the convolution of the derivative of L with the function $a(x)$ evaluated at x_0. The numerator of (23) relates the change in zero crossing to the distance between the zero which we are evaluating and the pixel that we are modifying. The denominator brings in information about the function $a(x)$ that we are dealing with.

Then, returning to the distance function D of (15) for a particular zero at z_k, in a discrete one-dimensional coordinate system,

$$\frac{\partial D(z_k)}{\partial a_i} = \frac{\partial D(z_k)}{\partial z_k}\frac{\partial z_k}{\partial a_i} \tag{24}$$

where the subscript i in a_i indicates that the change in parameter vector $\boldsymbol{a}$ occurs at coordinate x_i.

The expression $\frac{\partial D(z_k)}{\partial z_k}$ is the change in distance between the zero z_k of the estimate and the nearest Z_m of the prior, for an infinitesimal positive change in zero crossing of z_k. For the simple distance function that we have defined for F in (15), the values of the partial derivative are given by

$$T_{km} = \begin{cases} +1 \text{ for } z_k > Z_m \\ 0 \text{ for } z_k = Z_m \\ -1 \text{ for } z_k < Z_m \end{cases} \tag{25}$$

The complete expression for the partial derivatives of the function F to be maximized is then represented by

$$\frac{\partial F}{\partial a_i} = \sum_{j=1}^{J}\left[-f_{ji} + \frac{p_j f'_{ji} a_i^{(k)}}{\sum_{i'} f'_{ji'} a_{i'}^{(k)}}\frac{1}{a_i}\right] + \alpha\left\{\frac{1}{n_z}\sum_{k=1}^{n_z} T_{km}\frac{L(z_k - x_i)}{\int_{-\infty}^{\infty} dx' L'(z_k - x')a^{(k)}(x')}\right\} \tag{26}$$

where the zeros accepted in the summation over k fulfill the requirement (14). The denominator of the second term is a convolution which is carried out discretely.

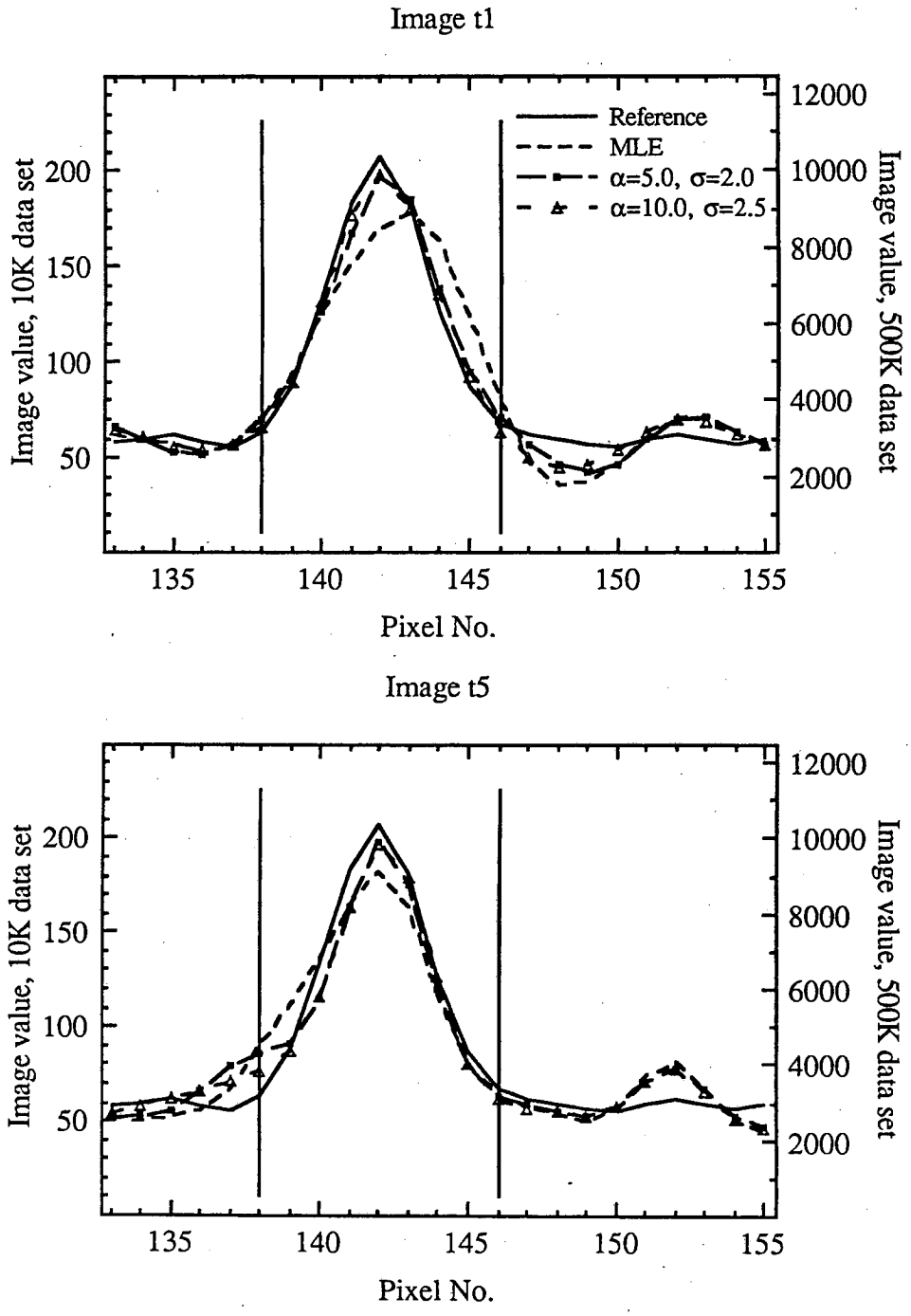


Fig. 7. Two examples from the 10 Bayesian reconstructions carried out. With the application of the prior information, the shape of the reconstructed peak is closer to the reference image.

4.3.2 Results of reconstructions.

Bayesian reconstructions by the EM algorithm, in which the E step was that of (12) and the M step was obtained by maximizing (16) by the modified CG method were carried out to the same number of iterations needed by the MLE/cross-validation method. The starting point for Bayesian reconstructions was the result of the 1st iteration of the MLE, so as to avoid problems with the denominator of (23). Reconstructions with $\alpha=0$ (in Eq. 16) gave results that were indistinguishable from MLE results to several digits of accuracy, indicating that the CG method used in the M step did not introduce significant errors. The results reported here are for reconstructions with the operator L_{ii} , for $\sigma=2.0$, $\alpha=5.00$ and for $\sigma=2.5$ and $\alpha=10.0$. The choices are empirical, at this time.

All ten 10K data sets showed improvement in the shape of the peak in the chosen ROI. We present two examples here in Fig. 7. Shown are: the 500K reconstructed image used as prior, the MLE reconstruction results which is generally distorted, and then the Bayesian results for the two sets of σ and α parameters indicated. In all cases, the Bayesian reconstructions with increasing value of α bring the peak closer to the shape and position of the prior image. The use of $\sigma=2.5$ instead of $\sigma=2.0$ also increases shape control at the "skirts" of the reconstruction.

It has been observed that the application of the Bayesian prior to to the selected ROI results is a slight decrease in the number of counts in the integral under the peak of that region. The lost counts are displaced to adjacent areas. In addition, the variance from the mean value of that integral does not improve when the Bayesian constraint is applied. This seems to point to the need

of using a Bayesian prior for the complete image. With all the features of the image being constrained and the conservation of counts insured by the inclusion of a Lagrance multiplier in (11), it is expected that there will be no count loss and, we hope, the variance will decrease.

In order to apply the Bayesian prior to the complete image, it appears that the next step is to devise a method for controlling the shape of wide features, whose edges are important at small scales, but the features are large. A multi-scale approach appears necessary. Work is continuing in that direction.

4.4 Questions raised and future work

The preliminary work reported above raises a number of questions, both relating to the one-dimensional simulations described above and to the problem of two-dimensional reconstructions. This will be an attempt at cataloguing the areas that need further investigation:

4.4.1. - One-dimensional simulation work.

One-dimensional work is interesting as a way of investigating algorithms and their behavior. There are two principal areas in which further work is needed: a) conceptual and b)computational.

a) Conceptual:

- When using more than one scale σ in the Bayesian term, the term corresponding to each σ should be normalized so that α can be a dimensionless quantity. Likewise, the interaction distance between zeros D_{int} of (14) should be a distance related to the scale parameter, although it is not clear at this time how that should be done.

- What is the effect of each σ, separately? and what range of σ's should be used, and with which step size?

- In reconstructing the complete image by the Bayesian method, it appears necessary to start iterations at a large scale and decrease scale as the iterative method continues and smaller detail is developed in the reconstructed image. How does one relate σ to the detail in the iterations?

- How many iterations should be used for the Bayesian reconstructions?. The results given above have been for the same number of iterations determined by the cross-validation stopping rule for the MLE, but it is not clear what the results would be if the same stopping rule were used for the Bayesian reconstructions.

b) Computational:

- The use of the Conjugate Gradient method for the M-step is very slow. It will be important to speed up the reconstruction process by, either speeding up the M-step considerably, or by dispensing completely with the EM method and using another form of iterative algorithm (like the Successive Substitutions method), which may be considerably faster.

4.4.2 - Two dimensional work.

In two dimensions, the more fundamental questions can be asked:

- What invariants should be used?

- Which scales are important?

- Is there a way to define the Bayesian parameter α a priori?

- Is the distance function used in the above 1-dimensional work adequate, or should other distance functions be investigated?

- and finally, what benefits are obtained from that work?

5. CONCLUSIONS

The work reported above is just the beginning of an attempt at using geometric information based of Vision Response Functions as priors for tomographic image reconstruction. The preliminary reconstructions with 1-dimensional images look promising in that it has been possible to correct distortions in reconstructions from data sets with a small number of counts using prior information from data sets with high number of counts. However, more questions have been raised than answered at this time. It appears possible, however, to work towards answering those questions in a systematic way and that is the direction that will be taken.

6. ACKNOWLEDGMENTS

The principal author would like to acknowledge lenghtly and interesting discussions with Luc Florac, Alfons Salden and the students in the Computer Vision Group, University Hospital Utrecht, and the help of scientific and administrative personnel of that Institution during his sixth month stay in Utrecht. The work of the principal author has been supported, in part, by the Director, Office of Energy Research, Office of Health and Environmental Research, Physical and Technological Division, of the U.S. Department of Energy under Contract DE-AC03-765SF00098, and by a grant from NWO, the Scientific Research Organization of the Netherlands. The work of the Utrecht Computer Vision Group is sponsored, in part, by a grant from the Dutch Ministry of Economic Affairs and the compaines Philips, Agfa-Gevaert and KEMA.

7. REFERENCES

1. - J. Llacer, E. Veklerov, D.J. Coakley, E.J. Hoffman and J. Nunez, "Optimization of Maximum Likelihood Estimator Images for PET Studies of the Human Brain: 1 - Algorithm Implementation and 2 - Statistical Analysis of Human Brain FDG Studies", LBL-30737, submitted for publication.

2. - E. Veklerov and J. Llacer, "Stopping Rule for the MLE Algorithm based on statistical hypothesis testing", *IEEE Trans. Med. Imaging*, Vol. MI-6, No. 4, pp. 313-319, 1987.

3. - K.J. Coakley and J. Llacer, "The use of Cross-validation as a Stopping Rule in Emission Tomography Image Reconstruction:, *SPIE Proceedings, Medical Imaging V*, Vol. 1443, pp. 226-233, 1991.

4. - S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images", IEEE Trans. Pattern Anal. Machine Intell., No. 6, pp. 721-741, 1984.

5. - R. Leahy and X. Yan, "Incorporation of Anatomical MR data for Improved Functional Imaging in PET", *Information Processing in Medical Imaging*, Lecture Notes in Computer Science, 511, Springer, pp. 105 - 119, 1991.

6. - C. T. Chen and X. Ouyang, "Sensor Fusion in Image Reconstruction", presented at the 1990 IEEE Nuclear Science Symposium.

7. - A.P.Dempster, N.M.Laird and D.B. Rubin, "Maximum Likelihood from imcomplete data via the EM algorithm", J. Royal Stat. Soc., Vol. B39, pp. 1-37, 1977.

8. - L. A. Shepp and Y. Vardi, "Maximum Likelihood reconstruction for emission tomography", IEEE Trans. Med. Imaging, Vol. MI-1, No. 2, pp. 113-121, 1982.

9. - B.M. ter Haar Romeny, Luc M.J. Florack, J.J. Koenderink and M.A. Viergever, "Scale Space: Its Natural Operators and Differential Invariants", *Information Processing in Medical Imaging*, Lecture Notes in Computer Science, 511, Springer, pp. 239-255, 1991.

10. - B.M. ter Haar Romeny and L.M.J. Florack, "A Multiscale Geometric Model of Human Vision", in *Perception of visual information*, Hendee and Wells, Eds., Springer, Berlin, 1991. Report 3DCV 90-14, University of Utrecht Computer Vision Research Group.

11. - W.H. Press, B.P.Flannery, S.A. Teukolsky and W.T. Vetterling, *Numerical Recipes in C*, Cambridge University Press, 1990.

LAWRENCE BERKELEY LABORATORY
UNIVERSITY OF CALIFORNIA
TECHNICAL INFORMATION DEPARTMENT
BERKELEY, CALIFORNIA 94720