Nonlinear Backprojection for Tomographic Reconstruction [1]

Blanca I. Andía[2], Ken D. Sauer[2] and Charles A. Bouman[3]
[2] Department of Electrical Engineering, 275 Fitzpatrick
University of Notre Dame, Notre Dame, IN 46556-5637
[3] School of Electrical Engineering, Purdue University
West Lafayette, IN 47907-0501

Abstract

This work focuses on a tomographic image reconstruction method which will be referred to as nonlinear backprojection (NBP). Rather than explicitly statistically modeling the forward process and the unknown image, we train an optimal nonlinear backprojection operator which can be implemented non-iteratively. Under appropriate assumptions, the method forms its estimate by applying nonlinear £lters to sinogram data, followed by conventional backprojection. The nonlinear £lters are designed through off-line training. Nonlinear backprojection shows promising results relative to both £ltered backprojection and MAP Bayesian methods.

I. INTRODUCTION

Image reconstruction from projections can be approached through deterministic or statistical methods. The most common technique is deterministic £ltered backprojection (FBP), which is quite effective for complete projection data with high SNR. Control of noise artifacts is restricted primarily to the selection of the low-pass element of the backprojection £lter, whose design has been well studied [1]. The other principal methods that have been introduced are statistical, choosing the estimate which best matches the probabilistic behavior of the data. To accomplish this, a probabilistic model for the forward physical process is introduced. When the maximum-likelihood (ML) criterion is applied, the reconstruction problem is often approached via the expectation-maximization (EM) algorithm [2] or modi£cations thereof, most commonly ordered subsets EM (OSEM) [3]. When Bayesian estimation is used, an *a priori* model that re¤ects beliefs about the image cross section is necessary [4, 5, 6]. Bayesian methods result in substantial improvements in many limited data or low SNR problems. The penalty paid for statistical methods is primarily the cost of numerical optimization to £nd the maximum *a posteriori* probability (MAP) or ML estimate. The *a posteriori* mean is typically still more complex.

In this paper, we take a new statistical direction. Rather than follow typical inverse problem formulation, we attempt to learn optimal or quasi-optimal reconstruction operators directly. This new direction does not require the explicit modeling of the forward process or unknown image. It will be referred to as nonlinear backprojection (NBP), since in its current form, it consists of nonlinear operations on the sinogram, followed by backprojection. Some interesting potential advantages of this method are

a. Elimination of iterative estimation, which should save computation time relative to common Bayesian techniques.

b. Elimination of the need to specify a statistical model for the forward process and for the images to be reconstructed.

Point b. permits the implicit incorporation of potentially greater complexity in the statistical behavior of images than the simple Markov random £eld models found in most Bayesian formulations. Various nonlinear system effects, such as scatter and beam-hardening, may also be implicitly included, provided they appear also in the data used during the learning process.

NBP consists of a training phase and an application phase. To permit a better understanding of the NBP training process, a well-known statistical structure is imposed, i.e. the projection data (sinogram) is modeled by a *mixture model* [7]. Sinogram data conditioned on the image are often modeled as Poisson. However, in the current problem, we are interested in the unconditioned distribution of sinogram data, which must include stochastic behavior of the image ensemble. The choice of using a mixture model provides a ¤exible and semi-parametric way to model unknown distribution shapes. Mixture models can be used where there is no a priori information about group structure in the data, but we wish to cluster the data into a number of groups [8]. A one-dimensional case of mixture parameter estimation for actual sinogram data (from the results of Fig. 2) is shown. Here, the histogram is approximated by a mixture model of 4 classes.

The goal of the training phase is the inference of parameters of the mixture model of the data and the optimization of a set of linear two-dimensional sinogram £lters. The process of £nding the number of classes in the mixture and their corresponding parameters is called clustering. The classes and their descriptive parameters are later used to classify the sinogram samples of the training set. Also each class will have an associated two-dimensional sinogram £lter.

The £lters can then be designed by minimizing, as a function of their coef£cients, the mean square error (MSE) between the training images and the images reconstructed by backprojecting the nonlinearly £ltered sinogram samples. After this off-line training the cluster model and £lters are applied to the reconstruction from arbitrary sinograms in a non-iterative, nonlinear £ltered backprojection.

[1] This work was supported by National Science Foundation grant CCR97-07763.

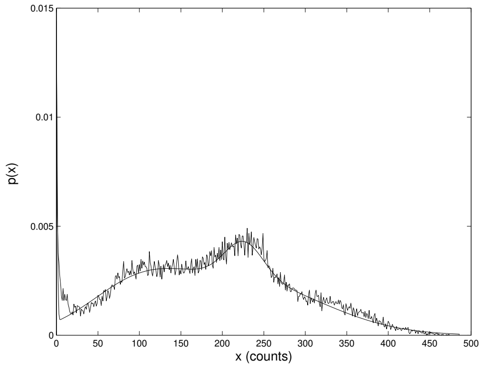


Figure 1: Mixture model of 4 classes approximating one-dimensional empirical probability density function of projection data.

Previously published works have similarly optimized estimates of the sinogram before conventional backprojection [9, 10]. NBP differs from previous methods in the replacement of explicit statistical models for the likelihood function and *a priori* models for the image by the training of a very general model for the data and adaptive two-dimensional £lters. A preliminary exploration of NBP was presented in [11].

An analogy between nonlinear backprojection and neural networks can also be observed. A neural network is a massively parallel distributed processor that has a natural propensity for storing *experiential* knowledge and making it available for use [12]. It resembles natural neural systems in that knowledge acquired through a learning process is stored in interneuron connection strengths. The similarity between nonlinear backprojection and neural networks is in their need of a "learning" process that comes from examples. Both acquire, through this training, a non-explicitly described transformation optimized with respect to the characteristics of the training set. Rather than apply the full generality of large-scale network necessary for the sort of reconstruction problems we wish to solve, we develop a non-linear, non-iterative estimator with a limited number of unknown parameters and commensurately simpler training.

II. Theoretical Background of NBP

Without limits on its form, the design and training of an optimal tomographic inverse operator appears infeasibly complex. We therefore impose a Gaussian mixture model on windows of the data in the sinogram to allow inference of a limited number of parameters while maintaining the capability of approximating varied distributions. These parameters will be used to classify the training data.

An important assumption that simpli£es the training process is the following: the minimum mean square error (MMSE) estimate of a point in the reconstruction, given the proper classi£cation of all data, will be a linear function of the sinogram data in the appropriate windows (i.e. the set of sinogram windows which have an effect on the reconstruction of such point). Under the data-dependence of classi£cations, this requires a nonlinear £ltering of the sinogram, with the nonlinear £lter as a weighted sum of the linear £lters optimized for each class. The £ltering is similar to methodology previously applied in image restoration and compression [13] and interpolation [14]. Afterwards, the reconstruction is obtained by backprojecting the £ltered sinogram. The behavior of the nonlinear £lter depends only on the local characteristics of the sinogram. At the cost of higher complexity, the £lter's characteristics could also depend on the pixel-domain structure, though the work done so far does not include this option.

Let Y be a vector containing all the sinogram data, X_k be a pixel in the unknown image cross section, and Y_j be the vector of sinogram samples taken from the j^{th} window, where each j corresponds to a single point in the sinogram (the window has a unique position along both the radial (t) and angular (θ) coordinates). This window will normally be a two-dimensional region centered at sinogram datum j. Also let $\{j\}$ be the set of indices of data that in¤uence pixel X_k; thus, in the discretized case, the choice of indices $\{j\}$ approximately describes a column of the projection matrix corresponding to the given pixel.

Under a standard mixture density for multi-dimensional data, we model each window of sinogram data, Y_j, as being generated by a discrete class, C_j, where $0 \leq C_j < M$. We will have M £lters of the size of the window used for Y_j, denoted by F_c where $0 \leq c < M$, one for each class. With this notation in mind, we will make two key assumptions expressed in the following two equations, which will make an optimal estimate computationally feasible.

$$E[X_k|\{Y_j\},\{C_j\}] = \sum_j F_{C_j}^T Y_j \; , \tag{1}$$

$$P[C_j = c|\{Y_j\}] = p(c|Y_j) \tag{2}$$

Equation (1) states that given the class information, the minimum mean square error (MMSE) pixel estimate of X_k may be obtained by applying an appropriate £lter to each sinogram window that in¤uences pixel X_k (i.e. the set of windows $\{Y_j\}$), and then backprojecting the result. Equation (2) states that the distribution of each class is dependent on only the projection data in the associated window. These two assumptions simplify the estimation process while, it is hoped, preserving suf£cient generality to improve on current methods.

Using (1) and (2), we compute the MMSE estimator of X_k given the data in the set of windows $\{Y_j\}$ by,

$$\begin{aligned} E[X_k|\{Y_j\}] &= E[E[X_k|\{Y_j\},\{C_j\}]|\{Y_j\}] \\ &= E\left[\sum_j F_{C_j}^T Y_j \middle| \{Y_j\}\right] \\ &= \sum_j E\left[F_{C_j}^T|\{Y_j\}\right] Y_j \\ &= \sum_j \left(\sum_c F_c^T p(c|Y_j)\right) Y_j \; . \end{aligned} \tag{3}$$

Thus the estimator is formed by applying a spatially varying £lter, $\sum_c F_c^T p(c|Y_j)$ to the set of sinogram windows $\{Y_j\}$.

Since the £lter depends on the data in the window through $p(c|Y_j)$, this weighted superposition of many conditionally optimal £lters is nonlinear, and may be thought of as an adaptive £lter. Finally, in summing over the nonlinearly £ltered sinogram windows, pixel X_k is reconstructed via conventional backprojection.

The distribution $p(c|Y_j)$ must be estimated and the £lters F_c designed as the training phase of NBP. Under a Gaussian mixture model for Y_j, where $p(Y_j|c)$ is a multivariate Gaussian distribution, and $\sum_{c=0}^{M-1} \pi_c = 1$, we have

$$p(c|Y_j) = \frac{p(Y_j|c)\pi_c}{\sum_c p(Y_j|c)\pi_c} \, . \tag{4}$$

The parameters of $p(Y_j|c)$ and the probabilities π_c may be estimated using the EM algorithm within an agglomerative clustering algorithm [15, 16, 17].

To design the £lters F_c, we £rst rewrite (3) in the form

$$E[X_k|\{Y_j\}] = \sum_c F_c^T \left(\sum_j p(c|Y_j) Y_j \right) = \sum_c F_c^T \bar{Y}_c \, , \tag{5}$$

where $\bar{Y}_c = \sum_j p(c|Y_j) Y_j$, forms the component of the pixel estimate that will result from the £ltering by the c^{th} £lter.

By de£ning the matrices F and $\bar{Y}$ as

$$F = \begin{bmatrix} F_0 \\ F_1 \\ \vdots \\ F_{M-1} \end{bmatrix}$$

$$\bar{Y} = \begin{array}{c} \\ \\ \\ \end{array} \begin{bmatrix} & \vdots & & \\ \bar{Y}_0^T & \bar{Y}_1^T & \cdots & \bar{Y}_{M-1}^T \\ \bar{Y}_0^T & \bar{Y}_1^T & \cdots & \bar{Y}_{M-1}^T \\ & \vdots & & \end{bmatrix} \begin{array}{l} \\ \text{(pixel k)} \\ \text{(pixel k+1)} \\ \\ \end{array}$$

the collection of optimal £lters F^* may be computed as the least squares solution to

$$F^* = \arg\min_F E\left[(X - \bar{Y}F)^T \left(X - \bar{Y}F \right) \right] \, . \tag{6}$$

The computation of the expectation in (6) is impractical, especially since we wish to avoid explicitly modeling pixels X_k as well as the conventional likelihood linking X and Y. Therefore, it is replaced by an ensemble average across a set of training images (as realizations of X) and pseudo-random projection data from forward projections of these X (as realizations of Y). Following the training of cluster parameters and £lters, the NBP system is ready to be applied to arbitrary sinogram data sets outside the training set.

III. Implementational Considerations

In the current implementation, the vector for clustering and classi£cation of each projection point consisted of 9 data from a 3×3 window in the sinogram, including both adjacent angles as well as neighboring rays. These sets of vectors are fed into a clustering program [14] that obtains the underlying mixture distribution for the data, i.e., it divides the training data into clusters parameterized by their means, variances and *a priori* probabilities. Such ML estimation problems for mixtures' parameters are degenerate in the absence of constraints [18, 19]. Experimental work has shown an advantage in constraining covariance matrices to be diagonal and equal in modestly-sized training sets, which reduces the number of parameters and improves robustness.

The use of such small windows is possible because of the use of projection data that is pre-£ltered in the radial variable. The pre-£ltering used a multiplication of the ramp £lter as dictated by £ltered backprojection theory explained in [20] and a raised cosine roll-off £lter for noise attenuation. This preserves nonlocal properties of backprojection. The raised cosine roll-off £lter has cut-off signi£cantly higher than would be desirable for FBP.

It is important to note that no £ltering is applied to the training data during the clustering and classi£cation process. A £lter such as that used for pre-£ltering includes signi£cant high-frequency noise emphasis, therefore it reduces the accuracy of classi£cation. Pre-£ltering is applied to the data before NBP's non-linear £ltering and backprojection both during the training phase and the application phase.

Another important issue which is currently under study is that of choice of data discriminants for classi£cation. Simple Euclidean distance among clusters is found not to be generally indicative of the difference among corresponding £lters. Finding optimal discriminants would result in a clustering taking greatest advantage of the nonlinear £lters' adaptability. These discriminants could also lower the dimensionality of the data used for clustering and result in lower computational cost. Toward this end, the windows of projection data are Karhunen-Loéve (KL) transformed before being clustered. Such a distance preserving transform will have no effect on the optimal clusterings, though some effect on our sub-optimal clusters is likely. However, the KL transform produces a set of uncorrelated components and dimensionality reduction could be achieved by selecting components according to their variances. The KL transform is obtained by generating an ensemble-averaged correlation matrix from the windows of data that belong to the training set. The DC component, for example, does not appear to provide relevant information for the £lter design problem, and is therefore eliminated from classi£cation criteria. The type of classi£cation for these experiments is "soft", as given in (3), i.e the testing data is not assigned to a single class (single £lter), and £ltering is a weighted combination of the linear £lters, with weights obtained from the *a posteriori* probabilities of classes.

For fewer than 100 classes in C, the least-squares problem of (6) is solved directly via matrix inversion routines using MATLAB. That is, the solution for (6) can be expressed as, $F^* = (\bar{Y}^T \bar{Y})^{-1} \bar{Y}^T X$. For larger numbers, iterative descent methods are used to solve F^*, with £lters that contain unity in their center coef£cient and zero elsewhere as initial conditions. While gradient methods may be applied, we have found that signi£cantly faster convergence is achieved using Iterative Coordinate Descent (ICD) in the domain of the £lter coef£cients, similar to the image reconstruction methods of [21]. Since this training is off-line, though, its cost is of limited importance.

IV. Experimental Results

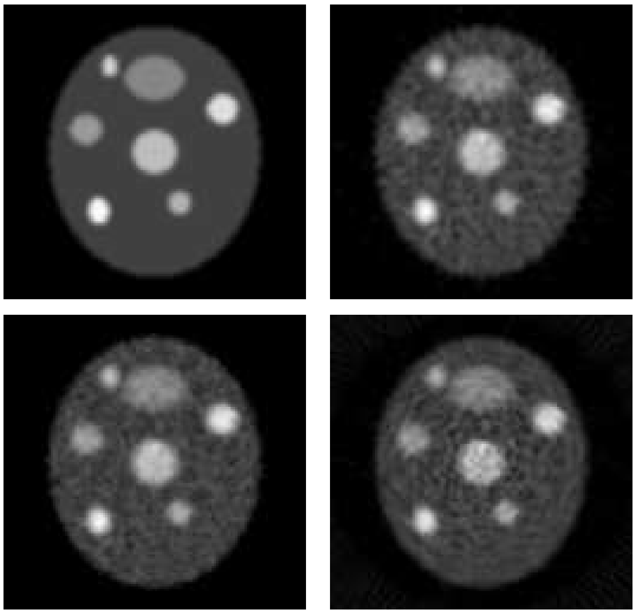
For Fig. 2, the training was realized on a set of phantoms whose projection data consisted of 64 uniformly spaced projections of 128 rays each from a £eld of diameter 200 mm. These phantoms have been created with limited resolution to mimic more realistic medical problems. For comparison, we include in all cases below results using FBP and MAP estimation under variations of the Generalized Gaussian Markov Random Field (GGMRF) as the *a priori* image model, with density function of the form

$$g_x(x) = \frac{1}{z} \exp \left\{ -\frac{1}{q\sigma^q} \sum_{\{j,k\} \in \mathcal{N}} b_{j,k} |x_j - x_k|^q \right\}$$

where $\mathcal{N}$ is the set of all neighboring pixel pairs, $b_{j,k}$ is the coef£cient linking pixels j and k, σ represents the scale of the prior image, and $1 \leq q \leq 2$ is a parameter which controls the smoothness of the reconstruction. This model includes a Gaussian MRF for $q = 2$, and an absolute-value potential function with $q = 1$. In general, smaller values of q allow sharper edges to form in reconstructed images. Prior information may also be available in the form of constraints on the reconstructed solution. We assume that the set of feasible reconstructions Ω is convex, and we choose Ω to be the set of non-negative reconstructions. Parameter selections for the GGMRF include both heuristic choices and ML estimation of the scale as in [22].

The £lters obtained in the training process are applied to the phantom of Fig. 2 which was not part of the training set. A comparison of the original phantom, the FBP reconstruction, the MAP reconstruction and the NBP reconstruction is presented. The measure of performance used for these results is the MSE, calculated only within the circular region inscribed in the square image. NBP provides good edge preservation; however, it introduces some artifacts across the image which do not appear in the FBP or MAP reconstructions. Numerically, the MSE results are comparable, but NBP does less well.

Fig. 3(a) shows the set of linear £lters that are used on Fig. 2. Each slice of the graph represents the 3×3 £lter values for a speci£c class, concatenated into a one-dimensional array. The clustering determined 30 quasi-optimal classes; therefore there are 30 linear £lters with widely varying frequency responses. The number of clusters which is optimal under the Rissanen criterion applied in [14] is likely greater than 30; since here we are interested in the simplicity of a relatively small numbers of classes, this number is used as the maximum throughout the results below. Fig. 3(b) shows the prior probabilities for each of the classes corresponding to these £lters.



a	b
c	d

Figure 2: Comparison of tomographic reconstruction techniques on projection data generated by a synthetic phantom image. The projection data has Poisson noise added. Total Counts = 1,310,720. (a) Original Phantom; (b) FBP reconstruction with Gaussian £lter (MSE = 1.7616); (c) MAP reconstruction with Gaussian prior model (MSE = 1.708); (d) NBP reconstruction using 30 training £lters (MSE = 2.2786);

The issue of £nding discriminants for the projection data is under study. Some of the experimental work is shown in Fig. 4. Fig. 4(a) provides a measure of the dispersion in the space of £lters. Notice that clusters 10, 22 and 27 are the most distant clusters from the origin; they also are among the ones with the lowest *a priori* probability of occurrence as shown in Fig. 3(b). Fig. 4(b) shows the sinogram of the noisy phantom in Fig. 2 and the locations having highest *a posteriori* probability of the clusters corresponding to £lters farthest from the origin. Many of these locations are near high intensity edges.

The clusters most distant from the origin tend to generate £lters which are most nearly impulsive, as shown in Fig. 5(a). This suggests that higher-pass £ltering of such data can reduce MSE provided it is applied selectively. The least impulsive £lters from the chart of Fig. 5(a) appear to be relatively low-pass, as can be seen in Fig. 3(a). Fig. 5(b) illustrates a relatively uniform pattern of application for these £lters under MMSE training.

In Fig. 6 FBP, MAP with Gaussian *a priori* image models, and NBP are compared for a 120×128 SPECT heart imaging data having approximately 150,000 total counts. NBP for this case was trained using synthetic phantoms of similar characteristics to apparently typical cross sections. There is a better subjective quality of the NBP result here with good edge preservation and good noise smoothing qualities. The NBP image is competitive in quality with the iterative MAP reconstructions using either the ML scale parameter or the

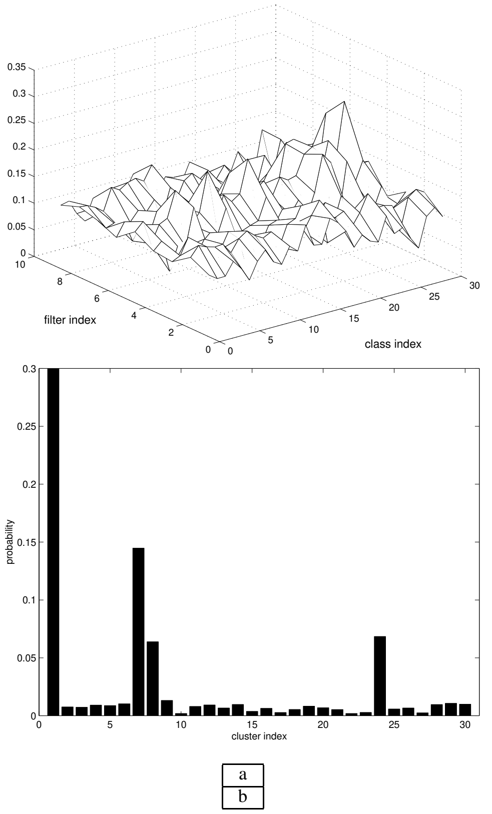


Figure 3: Resulting £lters after training on synthetic phantom data. (a) Linear £lters used on projection data generated by phantom images. Conventional FBP with little noise attenuation would result from £lters each having a single pulse at the center index; (b) Maximum likelihood estimation of probabilities of occurrence of these £lters' respective classes.

larger value resulting in less regularization. The £lters for this case (Fig. 7) demonstrate signi£cantly more variety than those of the synthetic phantom.

Fig. 9 represents the testing of the £lters resulting from three training processes on different sets of synthetic heart data, to determine NBP's sensitivity to accurate choice for training data. All data sets consist of 120 projections and 128 rays each from a £eld diameter of 456 mm, similar to the ones used for Fig. 6. The £rst training set resembled data from the heart of Fig. 9, the second training set resembled cross sections from the male heart of Fig. 6, and the third training set included both sets. The results of these training processes are applied to the real SPECT heart data from a female. The results and comparisons to other methods are shown in Fig. 9. The results show that the training on male, female or the combination of both data provides similar results. NBP reconstruction provides comparable results to the MAP reconstruction with $p = 2.0$ but with potentially better resolution in addition to reduced computation. As in the previous SPECT data, these £lters are signi£cantly more varied than those of the synthetic phantom. Fig. 9(b) and (c) show a potential liability of the edge-preserving MAP reconstruction: contouring in low SNR data. Finally, Fig. 8 shows the set of £lters obtained after training on a combined set of phantom heart male and female

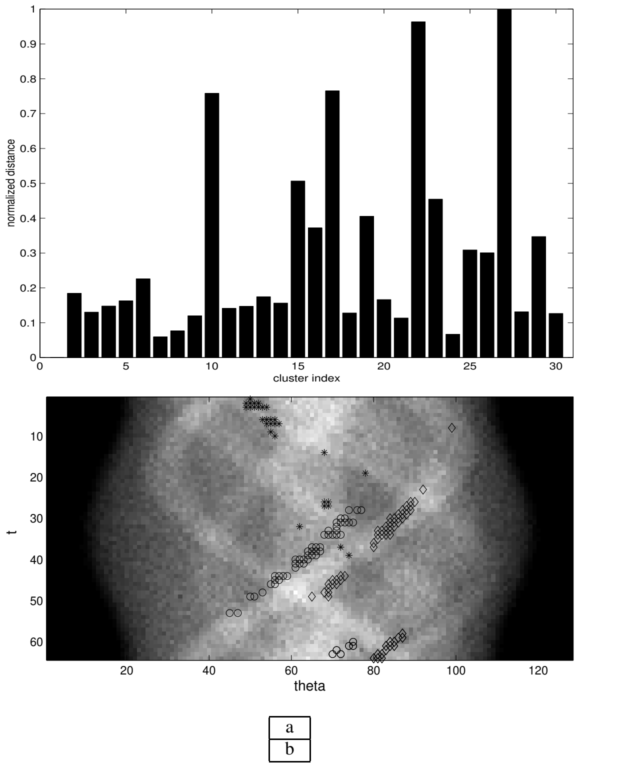


Figure 4: (a) Normalized distance of each cluster from the center point; (b) Points in the sinogram where the £lters corresponding to the most distant clusters are applied ($*$ 10, $\circ$ 22, $\diamond$ 27).

images. These £lters are signi£cantly more varied than those of Fig. 3.

V. Conclusion

Nonlinear backprojection aims to achieve a tomographic reconstruction method with improved visual quality relative to FBP reconstructions and comparable to iterative reconstructions such as Bayesian MAP estimates or OSEM approximations to ML. The goal is costs near those of FBP, and signi£cantly below that of conventional statistical reconstruction methods. The results here show quality competitive with statistical estimates; however, veri£cation across a broader variety of imaging conditions will be needed.

Training of the £lters is computationally costly, but it is an off-line process and is therefore not an important factor in the cost of reconstruction. Classi£cation speed, however, is of fundamental importance. Fast methods for data classi£cation will be a key aspect of future work. Finding adequate discriminants to reduce the dimension of the feature vector may improve ef£ciency by both simplifying the classi£cation process and reducing the number of necessary £lters.

VI. References

[1] G. T. Herman. *Image Reconstruction from Projections: The Fundamentals of Computerized Tomography*. Academic Press, New York, 1980.

[2] L. A. Shepp and Y. Vardi. Maximum likelihood reconstruction for emission tomography. *IEEE*

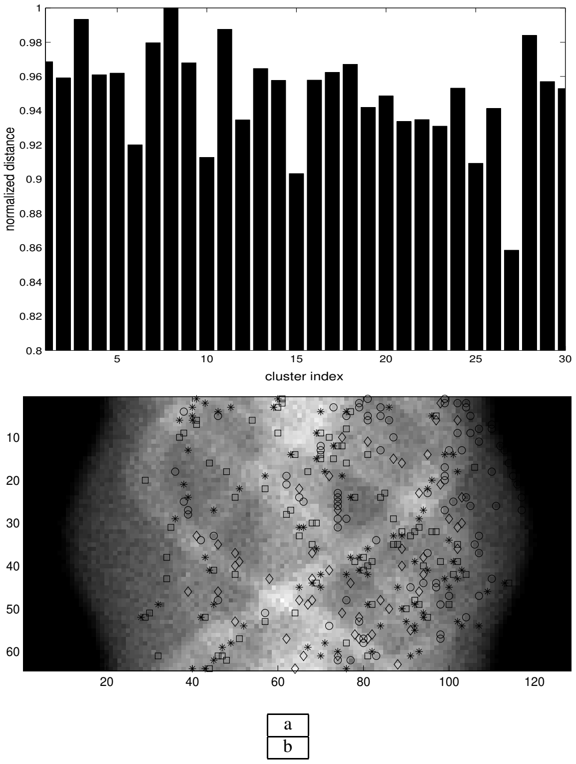


Figure 5: Establishing the most distinct £lters. (a) Normalized distance of each £lter to the impulsive (all-pass) £lter; (b) Points in the sinogram where the most distant £lters are applied ($*$ 3, $\circ$ 8, $\diamond$ 11, $\square$ 28).

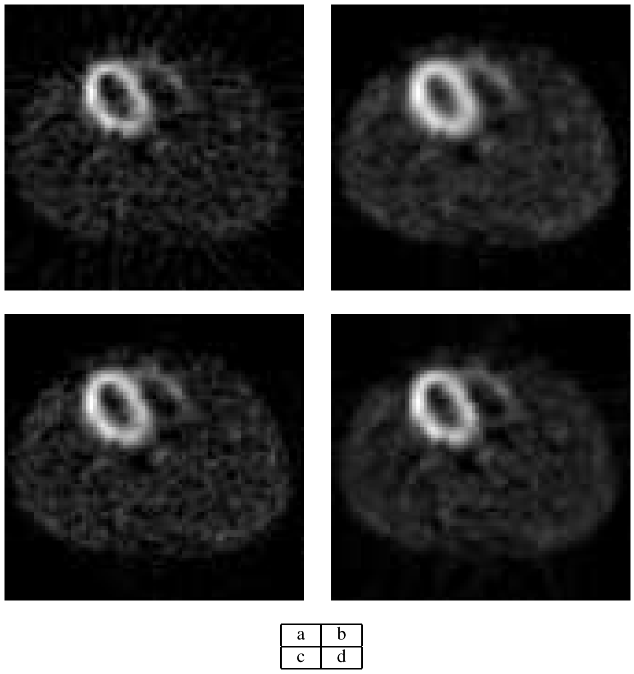


Figure 6: Comparison of tomographic reconstruction techniques on SPECT data generated by a human heart. Total Counts $\approx$ 150,000. (a) FBP reconstruction; (b) MAP reconstruction with Gaussian prior model ($q = 2.0$), ML value of $\sigma = 0.028\ \mathrm{mm}^{-1}$; (c) MAP reconstruction with Gaussian prior model, $\sigma = 0.058\ \mathrm{mm}^{-1}$; (d) NBP reconstruction using 30 training £lters. (SPECT data courtesy of T.-S. Pan, M.A. King and the University of Massachusetts.)

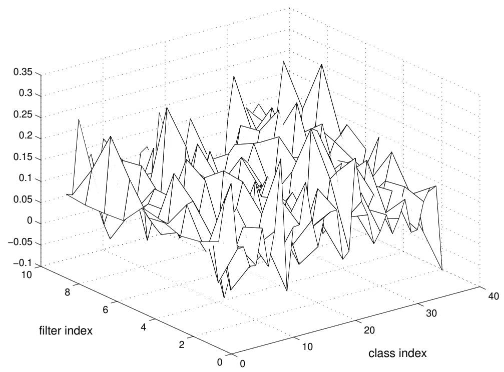


Figure 7: Nonlinear £lters used for heart image of Fig. 6. The £lters were obtained by training on a synthetic heart image.

Transactions on Medical Imaging, 1(2):113–122, Oct. 1982.

[3] H.M. Hudson and R.S. Larkin. Accelerated image reconstruction using ordered subsets of projection data. *IEEE Transactions on Medical Imaging*, 13(4):601–609, Dec. 1994.

[4] Gabor T. Herman and Arnold Lent. A computer implementation of a Bayesian analysis of image reconstruction. *Information and Control*, 31:364–384, 1976.

[5] Kenneth M. Hanson and George W. Wecksung. Bayesian approach to limited-angle reconstruction in computed tomography. *Journal of Optical Society of America*, 73(11):1501–1509, Nov. 1983.

[6] S. Geman and D. McClure. Bayesian images analysis: An application to single photon emission tomography. In *Proc. Statist. Comput. sect. Amer. Stat. Assoc.*, pages 12–18, Washington, DC, 1985.

[7] A. Webb. *Statistical Pattern Recognition*. Oxford University Press, Inc., New York, 1999.

[8] Geoffrey J. McLachlan and Kaye E. Basford. *Mixture Models: Inference and Applications to Clustering*. Marcel Dekker, Inc., New York and Basel, 1988.

[9] J. Prince and A. S. Willsky. A hierarchical algorithm for limited-angle reconstruction. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 1468–1471, May 23-26 1989.

[10] R. D. Nowak K. E. Timmermann and K. J. Jones. Improved emission tomography via multiscale sinogram analysis. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 3417–3420, Mar. 15-19 1999.

[11] Blanca I. Andia, Ken Sauer, and Charles A. Bouman. Nonlinear, noniterative Bayesian tomographic image reconstruction. In *International Conference on Image Processing*, volume 2, pages 668–671, Oct. 1999.

[12] S. Haykin. *Neural Networks: A Comprehensive Foundation*. Macmillan, New York, NY, 1994.

[13] K. Popat and R. W. Picard. Cluster-based probability

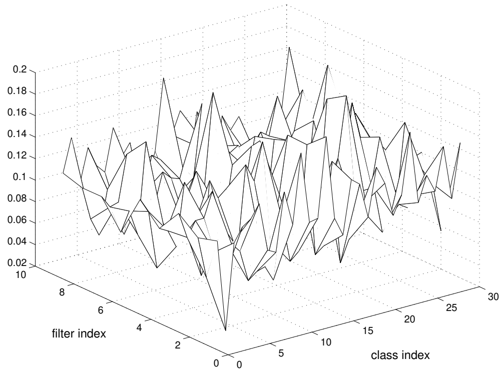


Figure 8: Nonlinear £lters used for heart image of Fig. 9. The £lters were obtained by training on a set of phantom heart images, with equal numbers of data sets resembling male and female SPECT heart data.

model and its application to image and texture processing. Technical Report 351, MIT Media Laboratory, Cambridge, MA.

[14] B. Atkins. *Classi£cation Based Methods in Optimal Image Interpolation*. PhD thesis, Purdue University, West Lafayette, IN, Dec. 1998.

[15] C. A. Bouman and M. Shapiro. A multiscale random £eld model for Bayesian image segmentation. *IEEE Transactions on Image Processing*, 3(2):162–177, Mar. 1994.

[16] L. E. Baum and T. Petrie. Statistical inference for probabilistic functions of £nite state Markov chains. *Annals Mathematical Statistics*, 37:1554–1563, 1966.

[17] Murray Aitkin and Donald B. Rubin. Estimation and hypothesis testing in £nite mixture models. *Journal of the Royal Statistical Society*, B-47(1):67–75, 1985.

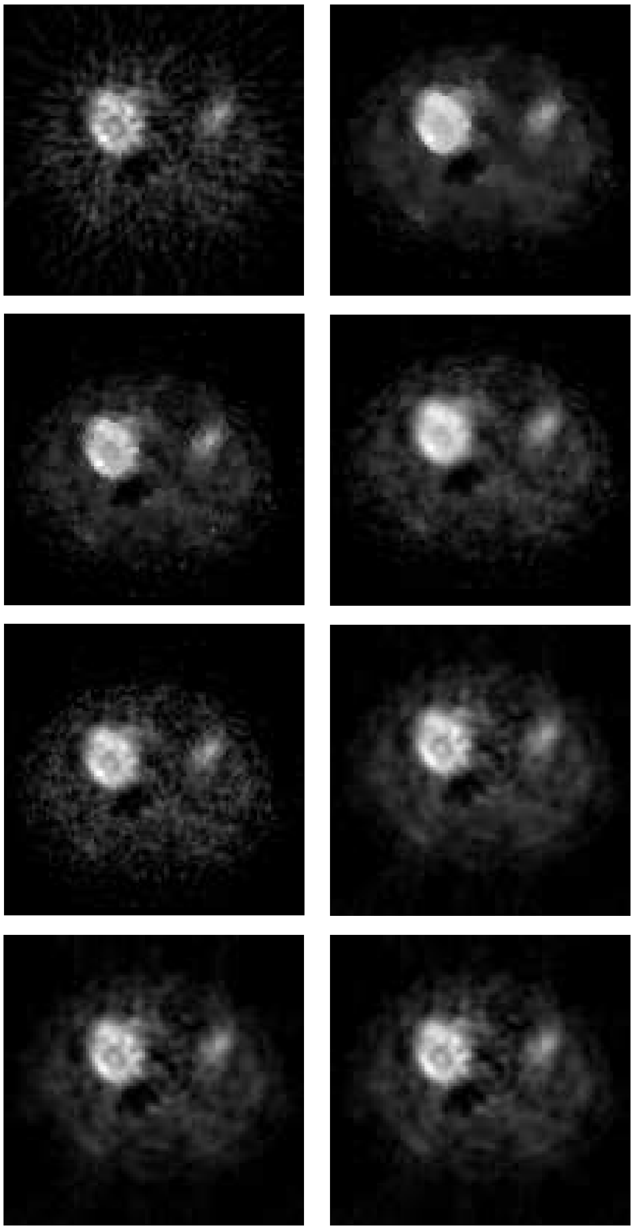
[18] J. Kiefer and J. Wolfowitz. Consistency of the maximum likelihood estimator in the presence of in£nitely many incidental parameters. *Annals Mathematical Statistics*, 27:887–906, 1956.

[19] R. A. Redner and H. F. Walker. Mixture densities, maximum likelihood and the EM algorithm. *SIAM Review*, 26(2):195–239, Apr. 1984.

[20] S. R. Deans. *The Radon Transform and Some of its Applications*. Wiley, New York, 1983.

[21] C.A. Bouman and K. Sauer. A uni£ed approach to statistical tomography using coordinate descent optimization. *IEEE Transactions on Image Processing*, 5(3):480–492, Mar. 1996.

[22] S. S. Saquib, C. A. Bouman, and K. Sauer. ML parameter estimation for Markov random £elds, with applications to Bayesian tomography. *IEEE Transactions on Image Processing*, 7:1029–1044, July 1998.



a	b
c	d
e	f
g	h

Figure 9: Comparison of tomographic reconstruction techniques on SPECT data generated by a human female heart. Total Counts $\approx$ 150,000. (a) FBP reconstruction; (b) MAP reconstruction with edge-preserving generalized Gaussian prior model (q = 1.1) and ML value of $\sigma = 0.018$ mm^{-1}; (c) MAP reconstruction with $q = 1.1$ and $\sigma = 0.035$ mm^{-1}; (d) MAP reconstruction with Gaussian prior model ($q = 2.0$), $\sigma = 0.028$ mm^{-1}; (e) MAP reconstruction with Gaussian prior, $\sigma = 0.058$ mm^{-1}; (f) NBP reconstruction using 30 £lters trained on synthetic female heart data. (g) NBP reconstruction using 30 £lters trained on synthetic male heart data. (h) NBP reconstruction using 30 £lters trained on synthetic female and male heart data.