BAYESIAN MULTIRESOLUTION ALGORITHM FOR PET RECONSTRUCTION *†

Thomas Frese[†], *Charles A. Bouman*[†], *Ned C. Rouze*[‡], *Gary D. Hutchins*[‡] *and Ken Sauer*[††]

[†]Purdue University, School of Electrical and Computer Engineering, West Lafayette, IN 47907-1285, {frese, bouman}@ecn.purdue.edu

[‡]Indiana University School of Medicine, Department of Radiology, Indianapolis, IN 46202-5111., {nrouze,gdhutchi}@iupui.edu

[††]University of Notre Dame, Department of Electrical Engineering, Notre Dame, IN 46556-5637, sauer@nd.edu

ABSTRACT

We introduce a spatially non-homogeneous adaptive image model and multiresolution reconstruction algorithm for Bayesian tomographic reconstruction. In contrast to existing approaches, the proposed image model is formulated in a multiresolution wavelet domain and relies on training data to incorporate the expected characteristics of typical reconstructions. The actual tomographic reconstruction is performed in the space domain to simplify enforcement of the positivity constraint. We apply the proposed algorithm to simulated data and to data acquired using the IndyPET dedicated research scanner. Our experimental results indicate that our algorithm can improve reconstruction quality over fixed resolution Bayesian methods.

1 INTRODUCTION

One of the major challenges in Bayesian tomographic reconstruction is the design of edge-preserving prior models. Existing prior models are largely based on spatially homogeneous Markov random field (MRF) implementations. A disadvantage of such models is that Gaussian MRF's tend to oversmooth edges while non-Gaussian MRF's can artificially enhance edges and remove important detail.

In this report, we present a prior model designed to preserve sharp edges without eliminating fine detail in the reconstruction. In contrast to existing models for iterative Bayesian tomographic reconstruction, our approach is formulated in the wavelet domain where it exploits dependencies of wavelet coefficients across scales. Training data are used to incorporate the expected characteristics of typical reconstructions. The model presented here extends an earlier version[1] by introducing a recursive algorithm for optimizing the reconstruction at all resolutions simultaneously.

In comparison to commonly used wavelet models[2, 3], the dependency structure of the proposed model is more general. Specifically, the model does not restrict the dependencies to interactions within a quadtree structure. Instead, the wavelet coefficients at each location are assumed to depend on an arbitrary but fixed window of scaling coefficients at the same scale. A disadvantage of this more general approach is that MAP optimization must be performed iteratively. For tomographic reconstruction this is not a limitation since the forward model requires that the optimization be performed iteratively in any case.

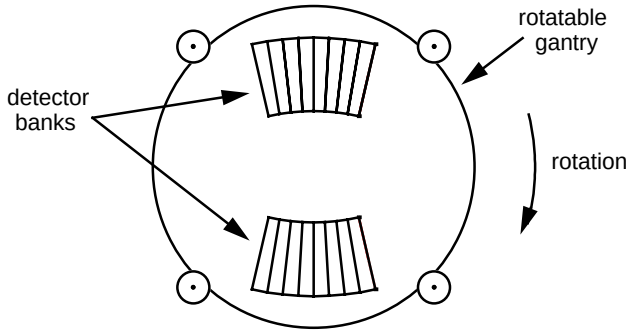


Figure 1: IndyPET small animal scanner.

Our model assumes linear dependencies between the wavelet coefficients at different scales, thereby simplifying optimization in a Bayesian framework. To account for the non-Gaussian statistics of typical reconstructions, the parameters of the model are spatially adapted to the local characteristics of the image. The actual tomographic reconstruction is performed in the space domain. In comparison to performing the reconstruction in the wavelet domain, the space-domain optimization simplifies enforcement of the positivity constraint.

We apply the proposed algorithm to data acquired using the IndyPET[4] positron emission tomography (PET) scanner. This scanner has been developed as a high resolution, high sensitivity dedicated research scanner for use in small animal, intermediate sized animal and small field of view human (brain, breast) imaging applications. As sketched in Fig. 1, the distinguishing feature of the scanner is the use of two, approximately planar, detector banks with adjustable separation which are mounted on a rotatable gantry. Each detector bank consists of 8 BGO detector blocks. In comparison to using a complete ring of detectors, this geometry results in reduced parallax and more uniform resolution throughout the

*THIS WORK WAS SUPPORTED BY THE NATIONAL SCIENCE FOUNDATION UNDER GRANT MIP97-07763.

†TO APPEAR IN ICIP 2000

field of view. Furthermore, the detector separation can be adjusted for each subject, which greatly increases the sensitivity for smaller subjects.

2 MODEL AND ALGORITHM

Our approach is to model the conditional distribution of the wavelet coefficients at each resolution given the information at all coarser scales. Let x be the N-dimensional vector of the image pixel values in raster order. Assuming an L resolution wavelet decomposition, let $z^{(n)}$ denote the wavelet coefficients at scale n for $1 \leq n \leq L$. Furthermore, let $x^{(n)}$ be the corresponding scaling coefficients where $0 \leq n \leq L$ such that $x^{(0)} = x$. The notation is illustrated in Fig. 2 for the case of a 1-D wavelet decomposition. We can now write the wavelet transform of x as $z = Wx$ where $z = (z^{(1)}, \cdots, z^{(L)}, x^{(L)})$. Using this notation, we can write the distribution $\log p_z(z)$ as

$$\log p_z(z) = \sum_{n=1}^{L} \log p(z^{(n)}|x^{(n)}) + \text{const} \tag{1}$$

where we assume $x^{(L)}$ to be uniformly distributed and use the fact that the scaling coefficients $x^{(n)}$ contain the same information as $z^{(n+1)}, \ldots, z^{(L)}, x^{(L)}$.

To obtain a practical model, we assume the wavelet coefficients at different locations to be conditionally independent given the scaling coefficients. Furthermore, we assume the wavelet coefficients at each location to depend only on a small window of scaling coefficients. Let s denote a spatial location at a given scale n. Then $x_s^{(n)}$ is the scaling coefficient and $z_s^{(n)}$ is the vector of the wavelet coefficients at location s and scale n. For the 2-D case, $z_s^{(n)}$ has three components corresponding to the high-low, low-high, and high-high coefficients of a separable wavelet decomposition. We now define $x_{\partial s}^{(n)}$ as a window of scaling coefficients centered at location s. Incorporating our assumptions, we can then write

$$\log p_z(z) = \sum_{n=1}^{L} \sum_{s \in S^{(n)}} \log p(z_s^{(n)}|x_{\partial s}^{(n)}) + \text{const} \tag{2}$$

where $S^{(n)}$ is the set of all locations s at scale n. For our implementation, we assume a square window ∂s of size 3×3 pixels.

To model the conditional distributions $\log p(z_s^{(n)}|x_{\partial s}^{(n)})$, we assume the wavelet coefficients z to be jointly Gaussian with spatially varying parameters. In this case, the conditional distributions are of the form

$$\log p(z_s^{(n)}|x_{\partial s}^{(n)}) = -\frac{1}{2}\|z_s^{(n)} - A_s^{(n)} x_{\partial s}^{(n)} - b_s^{(n)}\|^2_{B_s^{(n)}} + c \tag{3}$$

where $\|\varepsilon\|_B^2 = \varepsilon^t B \varepsilon$ and c is a constant. Consequently, the model is parameterized by $A_s^{(n)}$, $b_s^{(n)}$ and $B_s^{(n)}$.

Assuming an orthonormal wavelet transform W, we can now write the model as a function of x

$$\log p_x(x) = \log p_z(Wx) \tag{4}$$

$$= -\frac{1}{2} \sum_{n=1}^{L} \sum_{s \in S^{(n)}} \|z_s^{(n)} - A_s^{(n)} x_{\partial s}^{(n)} - b_s^{(n)}\|^2_{B_s^{(n)}} + c' \tag{5}$$

where $z_s^{(n)} = (Wx)_s^{(n)}$. The model (5) is used as the prior distribution for the Bayesian reconstruction of $x = x^{(0)}$. In addition to this fine scale reconstruction, we can perform coarse scale reconstructions of the scaling coefficients $x^{(l)}$ at scale l. To compute the coarse scale reconstruction of $x^{(l)}$, we apply the model only to the wavelet coefficients at the coarser scales $n = l+1, \ldots, L$. We thus define the prior distribution $\log p_{x^{(l)}}$ for the reconstruction at scale l as

$$\log p_{x^{(l)}}(x^{(l)}) = \log p(z^{(l+1)}, \ldots, z^{(L)}, x^{(L)}) \tag{6}$$

$$= -\frac{1}{2} \sum_{n=l+1}^{L} \sum_{s \in S^{(n)}} \|z_s^{(n)} - A_s^{(n)} x_{\partial s}^{(n)} - b_s^{(n)}\|^2_{B_s^{(n)}} + c_l \ . \tag{7}$$

Assuming noisy projection measurements y and a forward model $\log p(y|x^{(l)})$, the Bayesian MAP estimate $\hat{x}^{(l)}$ of the scaling coefficients $x^{(l)}$ at scale l is the solution to the optimization problem

$$\hat{x}^{(l)} = \arg \max_{x^{(l)} \geq 0} \{\log p(y|x^{(l)}) + \log p_{x^{(l)}}(x^{(l)})\} \ . \tag{8}$$

We use a multiresolution algorithm to perform the image reconstruction and to adaptively select the parameters of the linear model. The basic concept of the multiresolution algorithm is to compute a sequence of Bayesian MAP estimates from coarse to fine scale. The algorithm starts with the reconstruction of the scaling coefficients $x^{(L)}$ at the coarsest scale L and then successively performs the reconstructions at the finer scales $l = L-1, \ldots, 0$. At each step in this sequence, the current reconstruction is used to initialize the model parameters at the next finer scale and to re-adapt the parameters at the coarser scales. Let $\theta_s^{(l)}$ denote the vector of model parameters at location s and scale l

$$\theta_s^{(l)} = [A_s^{(l)}, b_s^{(l)}, B_s^{(l)}] \ . \tag{9}$$

After computing the MAP reconstruction $\hat{x}^{(l)}$ at scale l, we update all the parameters $\theta_s^{(n)}$ for $n \geq l$ based on the wavelet decomposition of $\hat{x}^{(l)}$. The new parameters $\theta_s^{(n)}$ are then used to reconstruct $x^{(l-1)}$ at the next finer scale.

The parameters $\theta_s^{(n)}$ are computed using a nonlinear prediction method derived from recent work in image interpolation[5]. Specifically, we use a trainable predictor that first classifies $x_{\partial s}^{(n)}$ into a class $k_{n,s}$. The parameters $\theta_s^{(n)}$ are then selected as the parameter vector $\theta_{k_{n,s}}^{(n)}$ associated with class $k_{n,s}$. The classifiers and associated parameter vectors are obtained during a training phase. Additional methods are employed to adjust for the overall scaling of the image x. The amount of regularization imposed by the reconstruction algorithm can be adjusted through change of a global scaling parameter. A more detailed description of the structure of the nonlinear predictor can be found in [1].

3 SPACE DOMAIN OPTIMIZATION

The positivity constraint, $x^{(l)} \geq 0$, is an essential component of the MAP optimization equation (8). However, enforcement of positivity can be very difficult in the wavelet domain, particularly for general wavelet transforms. To avoid this problem, we perform the MAP optimization in the space domain

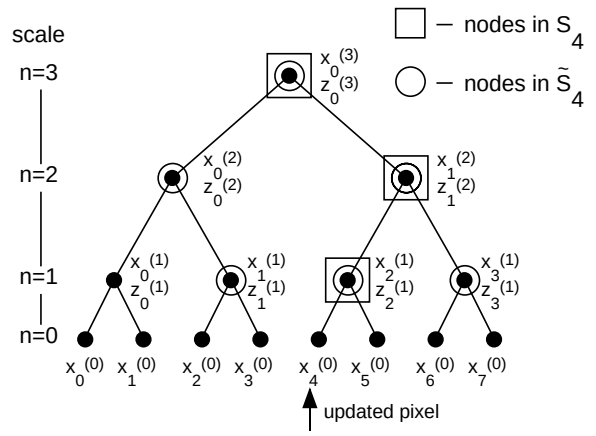


Figure 2: Optimization example for 1-D Haar wavelet decomposition and window $\partial s = \{s-1, s, s+1\}$. The nodes in circles and squares illustrate the sets $S_4^{(n)}$ and $\tilde{S}_4^{(n)}$ for the update of pixel $x_4^{(0)}$.

where enforcement of positivity is simple. Another advantage of space domain optimization is that it simplifies the forward model in tomography.

We perform the MAP optimization using an iterative coordinate descent (ICD)[6] strategy. To simplify the presentation, we will assume reconstruction at the finest scale $l = 0$. The coarse scale reconstructions are performed analogously by treating the reconstruction scale as the finest scale. Let e_i denote the unit vector in direction x_i. Since (5) is quadratic, we can write the prior as a function of the perturbation γ of pixel x_i as

$$\log p_x(x + \gamma e_i) = \alpha_1 \gamma + \frac{1}{2}\alpha_2 \gamma^2 + \log p_x(x) \qquad (10)$$

where α_1 and α_2 are the first and second partial derivatives of $\log p_x(x)$ with respect to x_i. Since (10) expresses the prior distribution as a function of only a single pixel in the space-domain, the MAP optimization can be performed using the standard ICD algorithm[6].

To obtain the derivatives α_1 and α_2, let $\varepsilon_s^{(n)}$ denote the error in the wavelet domain

$$\varepsilon_s^{(n)} = z_s^{(n)} - A_s^{(n)} x_{\partial s}^{(n)} - b_s^{(n)} \ . \qquad (11)$$

Let $W_{si}^{(n)}$ and $D_{si}^{(n)}$ denote the coefficients of the wavelet decomposition such that

$$\frac{\partial z_s^{(n)}}{\partial x_i} = W_{si}^{(n)} \qquad \frac{\partial x_s^{(n)}}{\partial x_i} = D_{si}^{(n)} \ . \qquad (12)$$

The notation in (12) is for the 1-D case, where both $\partial z_s^{(n)}/\partial x_i$ and $\partial x_s^{(n)}/\partial x_i$ are scalars. We can now write the derivative of $\varepsilon_s^{(n)}$ with respect to x_i as

$$d\varepsilon_{si}^{(n)} = \frac{\partial \varepsilon_s^{(n)}}{\partial x_i} = W_{si}^{(n)} - A_s^{(n)} D_{(\partial s)i}^{(n)} \qquad (13)$$

where $D_{(\partial s)i}^{(n)}$ is the vector with elements $D_{ki}^{(n)}$ for $k \in \partial s$. To compute α_1 and α_2, we need to consider only the locations (s, n) for which $d\varepsilon_{si}^{(n)} \neq 0$. Then we define the sets $S_i^{(n)}$ as

$$S_i^{(n)} = \{s : W_{si}^{(n)} \neq 0 \text{ or } D_{si}^{(n)} \neq 0\} \ . \qquad (14)$$

The derivatives $d\varepsilon_{si}^{(n)}$ are nonzero only at locations whose prediction window ∂s includes nodes in $S_i^{(n)}$. Let us define the sets $\tilde{S}_i^{(n)} = \{s : d\varepsilon_{si}^{(n)} \neq 0\}$, then

$$\tilde{S}_i^{(n)} = \{k : s \in \partial k \text{ for some } s \in S_i^{(n)}\} \qquad (15)$$

$$= \bigcup_{s \in S_i^{(n)}} \partial s \ . \qquad (16)$$

Figure 2 illustrates the sets $S_i^{(n)}$ and $\tilde{S}_i^{(n)}$ for the case of a 1-D Haar wavelet decomposition.

Based on (5), (13) and (16) we can compute α_1 and α_2 as

$$\alpha_1 = \sum_{n=1}^{L} \sum_{s \in \tilde{S}_i^{(n)}} (\varepsilon_s^{(n)})^t B_s^{(n)} d\varepsilon_{si}^{(n)} \qquad (17)$$

and

$$\alpha_2 = \sum_{n=1}^{L} \sum_{s \in \tilde{S}_i^{(n)}} (d\varepsilon_{si}^{(n)})^t B_s^{(n)} d\varepsilon_{si}^{(n)} \ . \qquad (18)$$

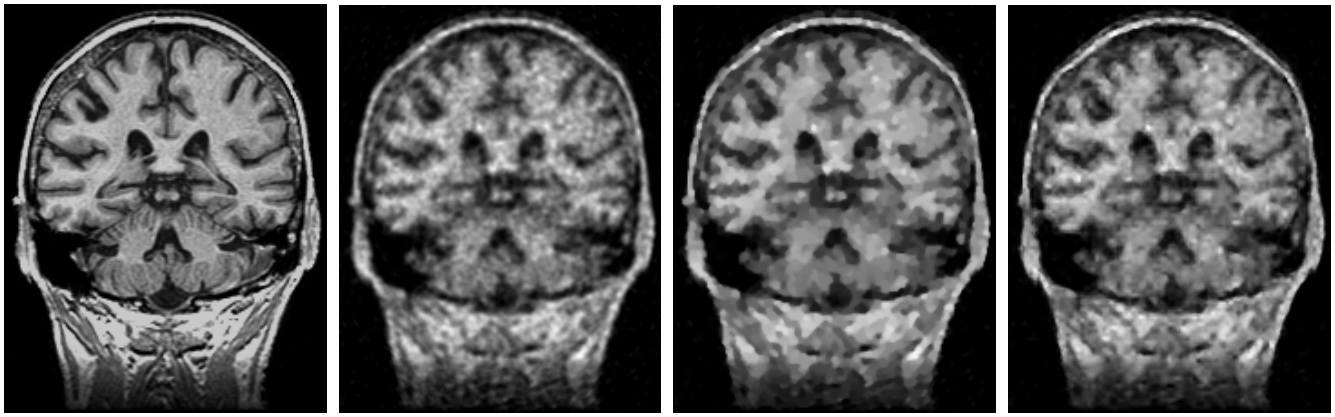
After performing the ICD update of x_i, we update $\varepsilon_s^{(n)}$ for all nodes in $\tilde{S}_i^{(n)}$.

For a general wavelet basis, the computational complexity associated with the optimization of the prior model is $O(N(\log N)^2)$ multiplications for a full update of x. For the special case of a Haar wavelet decomposition, this complexity becomes $O(N \log N)$.

4 RESULTS

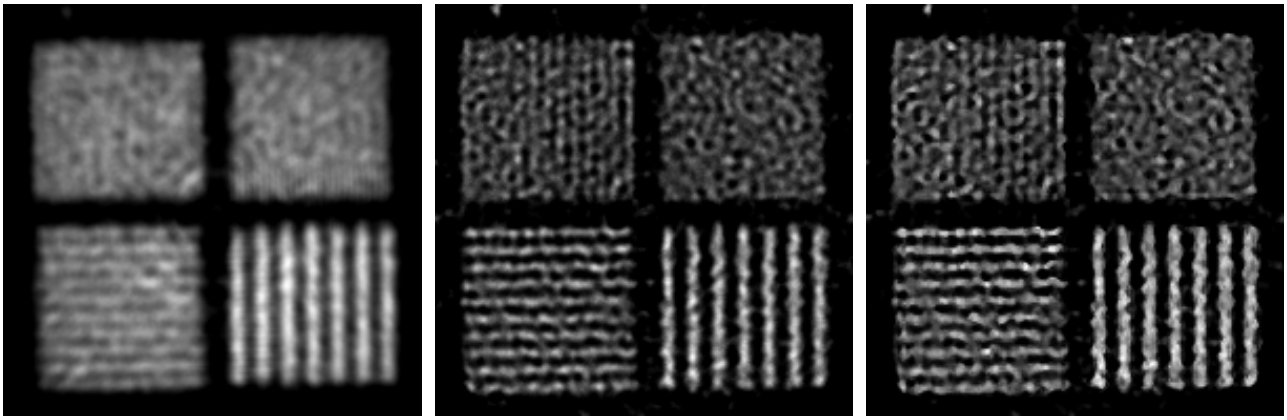
We present results for two data sets. The first set is simulated emission data generated by forward projecting the MRI data in Fig. 3(a). The second data set was collected using a bar phantom on the IndyPET scanner. The proposed algorithm was implemented using a 2-D Haar wavelet decomposition with $L = 5$ resolution levels. For all results, the algorithm was trained on a set of 40 MRI images. The MRI phantom in Fig. 3(a) was not included in the training set.

For the simulated data set, the projection data was calculated at 128 evenly spaced angles each with 256 parallel projections, assuming a field of view of 20 cm. The simulated projection beam had a triangular profile of width 2.3 mm. The data samples were formed by Poisson random variables with the appropriate means. The average number of counts per projection was 235. The reconstructions were computed at a resolution of 256×256 pixels. Figure 3 compares the reconstruction using the proposed method to standard Bayesian MAP reconstructions using a Gaussian MRF (GMRF) and a generalized Gaussian MRF (GGMRF)[7] prior model with shape parameter $p = 1.5$. For each method, the regularization parameter was adjusted to achieve minimum mean square error between reconstruction and ground truth. The proposed method produces sharper edge detail than the GMRF MAP algorithm. Alternatively, the GGMRF produces sharp edges



(a) Original Image (b) MAP, GMRF prior (c) MAP, GGMRF prior (d) Proposed algorithm

Figure 3: Reconstructions using simulated data.



(a) CBP (b) MAP, GMRF prior (c) Proposed algorithm

Figure 4: Reconstructions using bar phantom data collected on the IndyPET scanner.

but removes substantial detail, typical of spatially homogeneous non-Gaussian models.

Reconstructions of the data set collected on the IndyPET tomography scanner are shown in Fig. 4. The phantom consists of four quadrants, each containing an array of parallel acrylic bars with equal gaps between bars. The widths of the bars are 4.5, 3.0, 2.25, and 1.5 mm for the four quadrants. The gaps were filled with radioactive FDG, resulting in a square wave emission pattern. The total number of counts was 9.0 million. The reconstructions were computed at a resolution of 256×256 pixels for a field of view of 18 cm. In Fig. 4, the reconstruction using the proposed method is compared to convolution backprojection (CBP) using a Hanning reconstruction filter and to a Bayesian MAP reconstruction using a Gaussian MRF prior model. The proposed method and the GMRF MAP algorithm used a physical system model that is based on point source measurements of the scanner's point spread function in the sinogram domain. For both methods, the regularization parameters were adjusted to achieve the best visual reconstruction quality. The proposed method produces a reconstruction with sharper edges and more accurately preserves the square wave profile than either of the other reconstruction methods.

5 REFERENCES

[1] T. Frese, C. A. Bouman, and K. Sauer. Multiscale models for Bayesian inverse problems. In *Proceedings of SPIE Conference on Wavelet Applications in Signal and Image Processing VII*, volume 3813, pages 85–96, Denver, CO, July 19-23 1999.

[2] Khalid Daoudi, Austin B. Frakt, and Alan S. Willsky. Multiscale autoregressive models and wavelets. *IEEE Trans. on Information Theory*, 45(3):828–845, 1999.

[3] R. D. Nowak and R. G. Baraniuk. Wavelet-domain filtering for photon imaging systems. *IEEE Trans. on Image Processing*, 8(5):666–678, May 1999.

[4] N. C. Rouze, W. Winkle, and G. D. Hutchins. IndyPET - a high resolution, high sensitivity dedicated research scanner. In *Proc. of IEEE Nucl. Sci. Symp. and Med. Imaging Conf.*, Seattle, WA, October 1999.

[5] C. B. Atkins, C. A. Bouman, and J. P. Allebach. Tree-based resolution synthesis. In *Proc. of the Image Proc., Image Quality, Image Capture Systems Conference (PICS '99)*, pages 405–410, Savannah, GA, April 25-28 1999.

[6] C. A. Bouman and K. Sauer. A unified approach to statistical tomography using coordinate descent optimization.

IEEE Trans. on Image Processing, 5(3):480–492, March 1996.

[7] C. A. Bouman and K. Sauer. A generalized Gaussian image model for edge-preserving map estimation. *IEEE Trans. on Image Processing*, 2(3):296–310, July 1993.