

# A WAVELET-BASED METHOD FOR MULTISCALE TOMOGRAPHIC RECONSTRUCTION \*

M. Bhatia, W. C. Karl, and A. S. Willsky

Stochastic Systems Group  
Laboratory for Information and Decision Systems  
Massachusetts Institute of Technology  
Cambridge, Massachusetts 02139  
Telephone: (617) 253-3816 Telefax: (617) 258-8553  
Email: <mbhatia@mit.edu>

December 18, 1993

---

\*This work was supported by the Air Force Office of Scientific Research under grant F49620-92-J-0002, by the Office of Naval Research under grant N00014-91-J-1004, and by the US Army Research Office under grant DAAL03-92-G-0115.

## Abstract

We represent the standard ramp filter operator of the filtered back-projection (FBP) reconstruction in different bases composed of Haar and Daubechies compactly supported wavelets. The resulting multiscale representation of the ramp filter matrix operator is approximately diagonal. The accuracy of this diagonal approximation becomes better as wavelets with larger number of vanishing moments are used. This wavelet-based representation enables us to formulate a multiscale tomographic reconstruction technique wherein the object is reconstructed at multiple scales or resolutions. A complete reconstruction is obtained by combining the reconstructions at different scales. Our multiscale reconstruction technique has the same computational complexity as the FBP reconstruction method. It differs from other multiscale reconstruction techniques in that 1) the object is defined through a multiscale transformation of the projection domain, and 2) we explicitly account for noise in the projection data by calculating maximum a posteriori probability (MAP) multiscale reconstruction estimates based on a chosen fractal prior on the multiscale object coefficients. The computational complexity of this MAP solution is also the same as that of the FBP reconstruction. This is in contrast to commonly used methods of statistical regularization which result in computationally intensive optimization algorithms. The framework for multiscale reconstruction presented here can find application in object feature recognition directly from projection data, and regularization of imaging problems where the projection data are noisy.

**Key words: multiresolution reconstruction, wavelets, tomography, stochastic models.**

# 1 Introduction

In this work we present a multiresolution approach to the problem of reconstructing an image from tomographic projections. In general, a multiresolution framework for tomographic reconstruction may be natural or desirable for a variety of reasons. First, the data under consideration may be naturally acquired at multiple resolutions, e.g. if data from detectors of differing resolutions are used. In addition, the phenomenon may itself be naturally multiscale. For example, in the medical field self-similar or fractal models have been effectively used for the liver and lung [10–12]. Furthermore, it may be that, even if the data and phenomenon are not multiscale, our ultimate objectives are multiresolution in some way. For example, even though our data may be acquired at a fine level we may actually only care about aggregate or coarse scale quantities of the field. Such is often the case in ocean acoustic tomography or functional medical imaging. Conversely, if we are only interested in imaging high frequency details within the object (for example, boundaries), then we could directly obtain these features by extracting only the finer scale information in the data. Or, indeed, it may be that we want to use different resolutions in different areas – e.g. in nondestructive evaluation of aircraft we may want to look for general corrosion over an entire plane, but focus attention on certain suspect rivets to look for cracks. Using conventional techniques we would first have to reconstruct the entire field and then use post-processing to extract such features. A final compelling motivation for the use of multiresolution methods in estimation and reconstruction problems, is that they generally lead to extremely efficient algorithms, as in [13].

The conventional, and most commonly used, method for image reconstruction from tomographic projections is the Filtered Back-Projection (FBP) reconstruction technique [1]. In the standard FBP reconstruction as applied to a complete set of noiseless projections<sup>1</sup> the projection data at each angle are first filtered by a high-pass “ramp” filter and then back-projected. In this paper, we work in a different, multiscale transform space. The matrix representation of the resulting multiscale filtering operator is approximately diagonal. This enables us to formulate an efficient multiscale tomographic reconstruction technique that has the same computational complexity as that of the FBP reconstruction method. Perhaps more significantly, however, the different scale components of our proposed multiscale reconstruction method induce a corresponding multiscale representation of the underlying object and, in particular, provide estimates of (and thus information about) the field or object at a variety of resolutions at *no additional cost*. This provides a natural framework for explicitly assessing the *resolution-accuracy tradeoff* which is critical in the case of noisy or incomplete data.

Noisy imaging problems arise in a variety of contexts (e.g. low dose medical imaging, oceanography, and in several applications of nondestructive testing of materials) and in such cases standard techniques such as FBP often yield unacceptable results. These situations generally reflect the fact that more degrees of freedom are being sought than are really supported by the data and hence some form of regularization is required. Conventionally, this problem of reconstruction from noisy projection data is regularized by one of the following two techniques. First, the FBP ramp filter may be rolled off at high frequencies thus attenuating high frequency noise at the expense of not reconstructing the fine scale features in the object [7, 8]. This results in a fast, though ad hoc, method for regularization. The other common method for regularization is to solve for a maximum a posteriori probability (MAP) estimate of the object based on a 2-D (spatial) Markov random field (MRF) prior model [25, 26]. This results in a statistically regularized reconstruction which allows

---

<sup>1</sup>According to Llacer [3], “a complete data set could be described as sufficient number of line projections at a sufficient number of angular increments such that enough independent measurements are made to allow the image reconstruction of a complete bounded region.”

the inclusion of prior knowledge in a systematic way, but leads to optimization problems which are extremely computationally intensive. In contrast to these methods, we are able to extend our multiscale reconstruction technique in the case of noisy projections to obtain a multiscale MAP object estimate which, while retaining all of the advantages of statistically-based approaches, is obtained with the *same* computational complexity as the FBP reconstruction. We do this by realizing that for ill-posed problems the lower resolution (i.e. the coarser scale) reconstructions are often more reliable than their higher resolution counterparts and by capturing such intuition in prior statistical models constructed *directly in the multiscale domain*. Similar to the noise-free case, we also obtain these MAP estimates at multiple scales, essentially for free.

Finally, the FBP reconstruction method is not suitable for imaging problems where the projection data are incomplete (limited angle and/or truncated projections) [2,3]. These problems are encountered in many applications in medicine, non-destructive testing, oceanography and surveillance. Though the work presented here focuses on the case of complete data, our wavelet-based multiscale framework has the potential of overcoming this limitation and we briefly discuss such possibilities in the conclusion to the paper. We also refer the reader to a subsequent paper [20] where we consider an extension to the incomplete data case based on a similar multiscale framework.

Wavelets have been recently applied to tomography by other researchers as well. Peyrin et al [15] have shown that the back-projection of ramp-filtered and wavelet-transformed projection data corresponds to a 2-D wavelet decomposition of the original object. Their method differs from ours in several ways. First, the work in [15] does not deal with noise in the projection data. In contrast, our framework allows for the solution of statistically regularized problems at no additional cost when the projection data are noisy. Second, in [15], the object is represented in the original spatial domain by a *2-D wavelet basis*, the expansion coefficients of which are then obtained from the projection data. Instead of this, we represent the object in the *projection domain* by expanding the FBP basis functions (i.e. strips) in a *1-D wavelet basis*. This has the advantage that our multiscale basis representation of the object is closer to the measurement domain than the multiscale representation in [15]. One consequence is that our algorithms for multiscale reconstruction are no more complex than the FBP method. Another, is that our framework also allows for the simple and efficient solution of statistically regularized problems at no additional cost when the projection data are noisy. Sahiner and Yagle use the wavelet transform to perform spatially-varying filtering by reducing the noise energy in the reconstructed image over regions where high-resolution features are not present [16]. They also apply wavelet based reconstruction to the limited angle tomography problem by assuming approximate a priori knowledge about the edges in the object that lie parallel to the missing views [17]. Again as in [15], in [16,17] the object is represented in the original spatial domain by a *2-D wavelet basis* which is much different than our multiscale object representation. DeStefano and Olson [18], and Berenstein and Walnut [19] have also used wavelets for tomographic reconstruction problems, in particular to localize the Radon transform in even dimensions. Through this localization the radiation exposure can be reduced when a local region of the object is to be imaged. The work in [18] and [19] does not provide a framework for multiscale reconstruction, however, which is the central theme here. In addition, our reconstruction procedure also localizes the Radon transform, though we do not stress this particular application in this work.

The paper is organized as follows. Section 2 contains preliminaries. In Section 2.1 we describe the standard tomographic reconstruction problem and in Section 2.2 we describe the FBP reconstruction technique. We outline the theory of 1-D multiscale decomposition in Section 2.3. In Section 3, we develop the theory behind our wavelet-based multiscale reconstruction method starting from the FBP object representation. In Section 4 we build on this framework to provide a fast method for obtaining MAP regularized reconstructions from noisy data. The conclusions are

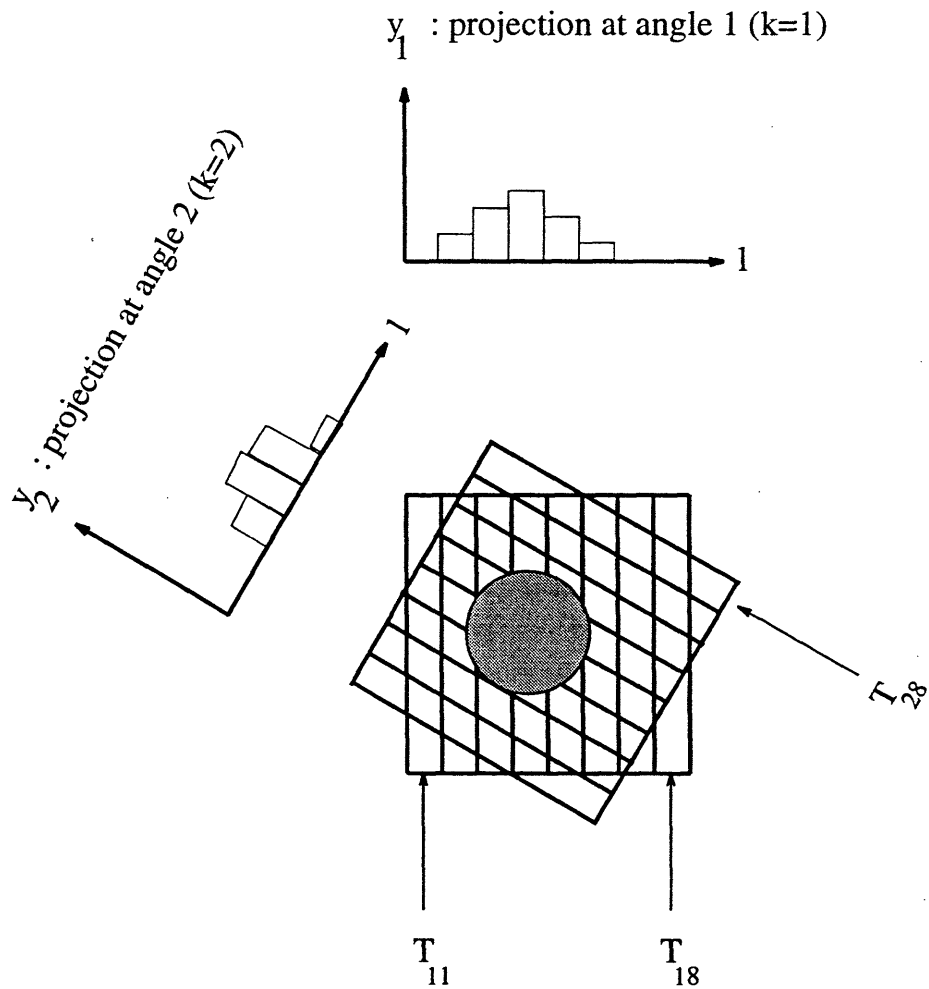


Figure 1: The projection measurements for an object,  $f$  (shaded), at two different angular positions ( $k = 1$  and  $k = 2$  respectively). The number of parallel strip integrals in each angular projection,  $N_s$ , is 8 in this case. Three basis functions,  $T_{11}, T_{18}, T_{28}$ , which are the indicator functions of the corresponding strips, are also shown.

presented in Section 5. Appendix A summarizes the mathematical notation used throughout this paper. Appendix B contains certain technical details.

## 2 Preliminaries

### 2.1 The Tomographic Reconstruction Problem

In tomography the goal is to reconstruct an object or a field,  $f$ , from line-integral projection data [1]. For a parallel-beam imaging geometry, the projection data consists of parallel, non-overlapping strip integrals through the object at various angles (refer to Figure 1). Each angular position corresponds to a specific source-detector orientation. Suppose we have  $N_\theta$  uniformly spaced angular positions between  $0^\circ$  and  $180^\circ$  and  $N_s$  parallel strip integrals at each angular position. Let us label the observation corresponding to projection  $\ell$  at angular position  $k$  by  $y_k(\ell)$ , where  $k = 1, \dots, N_\theta$  and  $\ell = 1, \dots, N_s$ . Furthermore, let  $T_{k\ell}$  be the indicator function of the strip integral corresponding to

this observation so that  $T_{k\ell}$  has value one within that strip and zero otherwise. Given this notation,

$$y_k(\ell) = \iint_{\Omega} f(u, v) T_{k\ell}(u, v) du dv \quad k = 1, \dots, N_\theta; \ell = 1, \dots, N_s \quad (1)$$

where  $(u, v)$  are the usual rectangular spatial coordinates and the integration is carried over a region of interest  $\Omega$ .

Due to practical considerations, we have to work with a discretized version of (1). By using standard discretization techniques (see for example [21]), the projection data at angle  $k$  is given by

$$y_k = T_k f \quad (2)$$

where  $T_k$  is an  $N_s \times N_s^2$  matrix representing  $\{T_{k\ell}(u, v); \ell = 1, \dots, N_s\}$  and  $f$  is an  $N_s^2 \times 1$  vector representing  $f(u, v)$  on an  $N_s \times N_s$  square pixel lattice, and  $y_k$  is the corresponding vector of measurements  $y_k(\ell)$ . Thus row  $\ell$  of  $T_k$  is the (discrete) representation of the strip function  $T_{k\ell}(u, v)$  and the inner product of  $f$  with this strip yields the data contained in the corresponding entry of  $y_k$ . The tomographic reconstruction problem then reduces to finding an estimate  $\hat{f}$  of the discretized object  $f$  given the projection data contained in the  $\{y_k; k = 1, \dots, N_\theta\}$ .

## 2.2 The Filtered Back-Projection Reconstruction Technique

The filtered back-projection (FBP) reconstruction technique is the most commonly used method for image reconstruction from tomographic data. It is based directly on the Radon inversion formula which is valid (i.e. yields exact reconstructions) only when a continuum of noise-free line integral projections from all angles are used [1]. In practice, as indicated in (1), we only have access to sampled projection data which are collected using strips of finite width. In this case, the quality of the FBP reconstruction is a function of the quality and fineness of the corresponding projection data used. We refer the reader to [23, 24] for details on sampling requirements for the Radon transform. In this work we assume that we sample finely enough to produce good reconstruction in the noiseless case.

In the FBP reconstruction, the object is expanded in a non-orthogonal basis that is closely related to the data acquisition process and the coefficients of this expansion are then found from local processing of the data in each angular projection. In particular, the estimated object is represented as a linear combination of the same functions  $T_{k\ell}(u, v)$  along which the projection data are collected. Similar to (2), a discretized version of this representation may be obtained as:

$$\hat{f} = \sum_{k=1}^{N_\theta} T_k^T x_k \quad (3)$$

where the  $N_s$  vector  $x_k$  contains the object coefficient set at angle  $k$ . Note that (3) can be interpreted as the back-projection operation [1] where the object coefficients  $x_k$  are back-projected along the basis functions  $T_k$  at each angle  $k$  and then the contributions from all  $N_\theta$  angles are added to get the overall reconstruction  $\hat{f}$ .

To complete the reconstruction the coefficients  $x_k$  must now be determined. The standard FBP method calculates them for each angle  $k$  according to the Radon inversion formula by filtering the projection data  $y_k$  at that particular angle with a ramp filter [1]. Thus, for a fixed angle  $k$ :

$$x_k = R y_k \quad (4)$$

where the matrix  $R$  captures this ramp-filtering operation. Thus (3) and (4) together represent the two operations used in the standard FBP reconstruction.

In Section 3 we apply a 1-D multiscale change of basis to the coefficients  $x_k$  and observations  $y_k$  using the wavelet transform [4,22]. One effect of this multiscale transformation will be to “compress” (sparsify) the filtering operator. Beyond simple compression of this operator, however, our method results in an associated multiscale transformation of the basis functions contained in  $T_k$ , and thus leads naturally to a framework for the reconstruction of objects at multiple resolutions, and hence the extraction of information at multiple resolutions. A key point in our multiscale reconstruction method is that, as opposed to what is done in other multiscale reconstruction techniques (for example, [15]), we do not directly expand the object estimate (i.e.  $\hat{f}$ ) in a spatial 2-D wavelet basis but rather we expand the FBP basis functions  $\{T_k\}$  in a *1-D wavelet basis* which then *induces* a corresponding 2-D multiscale object representation. The multiscale versions of the object expansion coefficients,  $\{x_k\}$ , “live” in the strip integral (i.e. the projection) domain rather than in the original object or spatial domain. Thus, as we have said, our multiscale basis representation of the object is closer to the measurement domain than the multiscale representations in previous work, resulting in very efficient algorithms.

### 2.3 1-D Wavelet Transform Based Multiscale Decomposition

Here we present a brief summary of the wavelet-based multiscale decomposition of 1-D functions. We do not intend this as a complete treatment of the topic and intentionally suppress many details. The interested reader is referred to any of the many papers devoted to this topic, e.g. the excellent paper [22]. A multiresolution dyadic decomposition of a discrete 1-dimensional signal  $x(n)$  of length  $2^N$  is a series of approximations  $x^{(m)}(n)$  of that signal at finer and finer resolutions (increasing  $m$ ) with dyadically increasing complexity or length and with the approximation at the finest scale equaling the signal itself, i.e.  $x^{(N)}(n) = x(n)$ . By considering the incremental detail added in going from the  $m$ -th scale approximation to the  $(m+1)$ -st we arrive at the *wavelet transform*. In particular, if  $x^{(m)}$  is the vector containing the sequence  $x^{(m)}(n)$  and  $\xi^{(m)}$  is the corresponding vector of detail added in proceeding to the next finer scale, then one can show [27] that the evolution of the approximations in scale satisfies an equation of the form:

$$x^{(m+1)} = L^T(m)x^{(m)} + H^T(m)\xi^{(m)} \quad (5)$$

where  $L(m)$  and  $H(m)$  are matrices (linear transformations) which depend on the particular wavelet chosen and are far from arbitrary and  $L^T(m)$  and  $H^T(m)$  denote their transposes (i.e. their adjoints). The operators  $L^T(m)$  and  $H^T(m)$  serve to interpolate the “low” and “high” frequency (i.e. approximation and detail) information, respectively, at one scale up to the next finer scale. The  $2^m$ -vector  $\xi^{(m)}$ , containing the information added in going from scale  $m$  to  $(m+1)$ , is composed of the *wavelet transform coefficients* at scale  $m$  and (5) is termed the *wavelet synthesis equation*.

Starting from a “coarsest” approximation  $x^{(0)}$  (usually taken to be the average value of the signal) then, it is possible to recursively and efficiently construct the different scale approximations through (5) by using the complete set of wavelet coefficient vectors  $\{\xi^{(m)}\}$ . This layered construction is shown graphically in Figure 2a, where our approximations are refined through the addition of finer and finer levels of detail as we go from right to left until the desired scale of approximation is achieved. In particular, the original signal  $x$  is obtained by adding all the interscale detail components  $\xi^{(m)}$  to the initial approximation  $x^{(0)}$ . For a given signal  $x$  the complete set of these elements uniquely captures the signal and thus corresponds to a simple change of basis. In addition, note from Figure 2 that the intermediate approximation  $x^{(m)}$  at scale  $m$  is generated using only

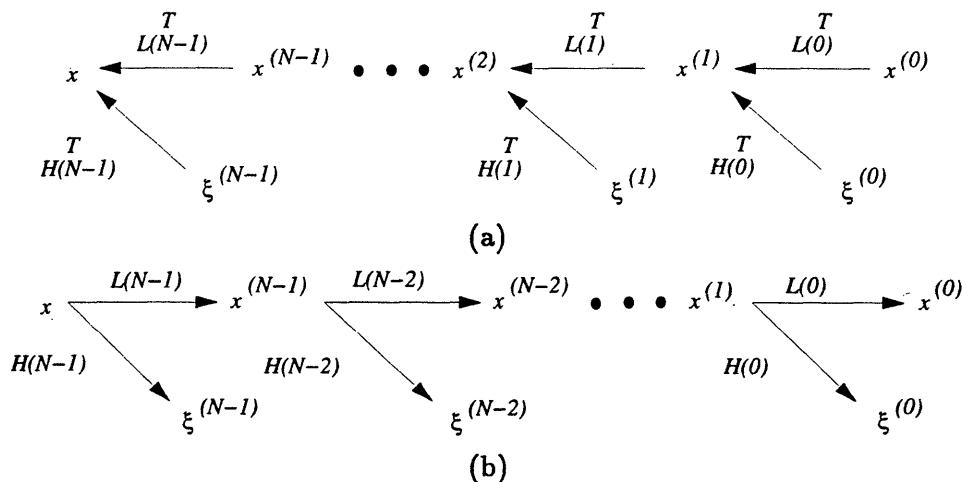


Figure 2: (a) Tree diagram for wavelet transform synthesis. We start from a coarsest approximation  $x^{(0)}$  on the right and progressively add finer levels of detail  $\xi^{(m)}$  as we proceed to the left, thus refining the original approximation to the signal. The original (finest scale) sequence is obtained as the final output on the left. (b) Tree diagram for wavelet transform analysis. Starting from a finest level signal  $x$  in the left we recursively peel off layers of detail  $\xi^{(m)}$  as we proceed to the right and the next coarser scale representation  $x^{(m)}$ .

the corresponding subset of the complete wavelet coefficient set (e.g. to obtain  $x^{(2)}$  we use only  $x^{(0)}$ ,  $\xi^{(0)}$ , and  $\xi^{(1)}$ ). The representation of this intermediate approximation at the original finest scale can be found by repeated interpolation of the information in  $x^{(m)}$  through the application of  $L^T(m')$ ,  $m' \geq m$ . This interpolation up to the finest scale corresponds to effectively assuming that additional, finer scale, detail components  $\xi^{(m')}$ ,  $m' \geq m$  are zero in our representation of the signal. It is such intermediate scale approximations and the detail necessary to go between them that give the wavelet transform its natural multiscale interpretation, and indeed we exploit such interpretations in Sections 3 and 4 to obtain induced multiscale *object* representations.

Beyond the recursive computation of the approximations, it is also possible to compute the components of the *decomposition* itself (i.e. the wavelet coefficients) recursively by exploiting the same multiscale structure. In particular, as shown in [27] the wavelet coefficient vectors  $\xi^{(m)}$  (and  $x^{(0)}$ ) can be obtained from the following recursion defining the *wavelet analysis equations*, which is illustrated in Figure 2b:

$$x^{(m)} = L(m) x^{(m+1)}, \quad \xi^{(m)} = H(m) x^{(m+1)} \quad (6)$$

where  $L(m)$  and  $H(m)$  are the same operators defined in connection with (5). The operators  $L(m)$  and  $H(m)$  correspond roughly to low and high pass filters followed by downsampling by a factor of 2, respectively. The figure shows how these wavelet coefficient vectors at each scale are obtained by “peeling off” successive layers of detail as we proceed from finer to coarser scales (left to right in the figure). This recursive structure yields algorithms for computation of the wavelet transform coefficients that are extremely efficient. For convenience in the development to follow, we will capture the overall operation which takes a vector  $x$  containing a discrete signal to the vector  $\xi$  containing all of its corresponding wavelet transform elements  $\{\xi^{(m)}\}$  and  $x^{(0)}$  by the matrix



operator  $W$  as follows:

$$W\mathbf{x} = \begin{bmatrix} \xi^{(N-1)} \\ \vdots \\ \xi^{(0)} \\ \mathbf{x}^{(0)} \end{bmatrix} \triangleq \xi \quad (7)$$

Since the transform is invertible and the wavelet basis functions are orthonormal, it follows that  $W^{-1}$  exists and further that  $W$  is a unitary matrix, i.e. that  $W^{-1} = W^T$ . From the above discussion, the matrix  $W$  captures the operation of the operators  $L(m)$  and  $H(m)$ , and thus depends on the underlying chosen wavelet. In our work in this paper, in addition to the Haar wavelet we will use wavelets from an especially popular family of these functions due to Daubechies [4], the separate elements of which are denoted  $D_n$ , where  $n$  is an indication of the support size of the corresponding filters contained in  $L(m)$  and  $H(m)$ . Finally, since our signals are of finite length, we need to deal with the edge effects which occur at the ends of the interval in the wavelet transform. While there are a variety of ways in which to do this, such as modifying the wavelet functions at the ends of the interval in order to provide an orthogonal decomposition over the interval [28], we have chosen here to use one of the most commonly used methods, namely that of cyclically wrapping the interval which induces a circulant structure in  $L(m)$  and  $H(m)$  [5,22]. While this does introduce some edge effects, these are of negligible importance for the objectives and issues we wish to emphasize and explore and for the applications considered here. Further, the methods we describe can be readily adapted to other approaches for dealing with edge effects as in [28] and the references contained therein.

As noted above, the intermediate approximations  $\mathbf{x}^{(m)}$  and their finest scale representation may be obtained by using only part of the full wavelet coefficient set during synthesis, effectively assuming the finer scale detail components are zero. For convenience in the discussion to follow we capture this partial zeroing operation in the matrix operator  $A(m)$ , which nulls the upper  $N - m$  subvectors of the overall wavelet vector  $\xi$  and thus retains only the information necessary to construct the *approximation*  $\mathbf{x}^{(m)}$  at scale  $m$ :

$$A(m) \triangleq \text{block diag} [0_{(2^N-2^m)}, I_{(2^m)}] \quad (8)$$

where  $0_p$  is a  $p \times p$  matrix of zeros and  $I_q$  is a  $q \times q$  identity matrix. Also it will prove convenient to define a similar matrix operator  $D(m)$ , that retains only the information in  $\xi$  pertaining to the *detail* component at scale  $m$  by zeroing all but the sub-vector corresponding to  $\xi^{(m)}$ :

$$D(m) \triangleq \text{block diag} [0_{(2^N-2^{m+1})}, I_{(2^m)}, 0_{(2^m)}] \quad (9)$$

Finally, with these definitions note that we have the following scale recursive relationship for the partially zeroed vectors, in the spirit of (5):

$$A^{(m+1)} \xi = A(m) \xi + D(m) \xi \quad (10)$$

### 3 The Multiscale Reconstruction Technique

In this section we derive our 1-D wavelet-based multiscale reconstruction technique. We start by applying a wavelet-derived multiscale change of basis  $W$  to the FBP object coefficients  $\mathbf{x}_k$ , which will induce a natural multiresolution object representation. We then show how the coefficients of our new multiscale representation can be computed directly from corresponding multiscale versions of

the data, in the same way that  $x_k$  is computed directly from  $y_k$  in the standard FBP method. Taken together these two components define a multiscale reconstruction algorithm, analogous in structure to the FBP method. An important point is that our approach does *not* start with a decomposition of the object in a 2-D wavelet basis and attempt to then find the resulting coefficients, but rather works directly in the projection domain. The multiscale nature of our object representation in the 2-D or spatial domain arises naturally from the original FBP definitions and our multiscale decomposition of the  $x_k$ , and thus we retain the simplicity and efficiency of this popular method.

### Multiscale Object Representation

We start by applying a multiscale change of basis, as defined by the matrix  $W$  in Section 2.3, to the original set of object coefficients  $x_k$  at each angle  $k$  to obtain an equivalent set of multiscale object coefficients as follows:

$$\xi_k = W x_k \quad (11)$$

Thus, for a given choice of wavelet defining  $W$ , the vector  $\xi_k$  contains the corresponding wavelet coefficients and coarsest level approximation (i.e. the average) associated with  $x_k$  and thus forms a multiresolution representation of this signal. More importantly, by reflecting this change of basis into the original FBP object representation (3), we naturally induce a corresponding *multiscale representation of the object* through the creation of a corresponding set of transformed multiscale basis functions. In particular, substituting (11) into (3) we obtain:

$$\hat{f} = \sum_{k=1}^{N_\theta} (T_k^T W^T) (W x_k) \triangleq \sum_{k=1}^{N_\theta} \mathcal{T}_k^T \xi_k \quad (12)$$

where  $\mathcal{T}_k = W T_k$ , is now a matrix representing the transformed, multiscale basis functions at angle  $k$ .

Before proceeding, let us consider these transformed bases functions contained in  $\mathcal{T}_k$  in more detail. Recall from Section 2.1 that the rows of  $T_k$  are composed of the (discretized) original strip basis functions at angle  $k$  along which the data were collected, c.f. (2). Similarly the rows of the transformed matrix  $\mathcal{T}_k$  will contain the corresponding (discretized) multiscale object basis functions at angle  $k$ . The wavelet transform operator matrix  $W$ , acting identically on each column of  $T_k$ , will thus form the new multiscale basis functions at that angle from linear combinations of the corresponding original strip functions, where these linear combinations correspond precisely to a *1-dimensional* wavelet transform perpendicular to the projection direction. This transformation of the basis functions is shown schematically in Figure 3 (which corresponds to the case of the rectangular, Haar wavelet). The original strip basis functions (rows of  $T_k$ ) are illustrated in the left half of the figure, while the corresponding collection of multiscale basis functions (rows of  $\mathcal{T}_k$ ) are shown in the right half. The heavy boundaries illustrate the support extent of the corresponding basis element while the “+” or “-” (together with shading) notionally indicate the sign of the function over this region. Note that the number of basis elements in the original (left half) and the multiscale (right half) framework are the same, as they must be since the multiscale framework involves an orthonormal change of basis. We may naturally group the multiscale 2-D spatial basis elements into a hierarchy of scale related components based on their support extent or spatial localization, as shown in the figure. The basis elements defining the  $m$ -th scale in such a group are obtained from the rows of  $\mathcal{T}_k$  corresponding to (i.e. scaled by) the associated wavelet coefficients  $\xi_k^{(m)}$  at that scale. We can see that the basis functions of these different scale components, though arising from a *1-dimensional* multiscale decomposition, naturally represent behavior of the *2-dimensional*

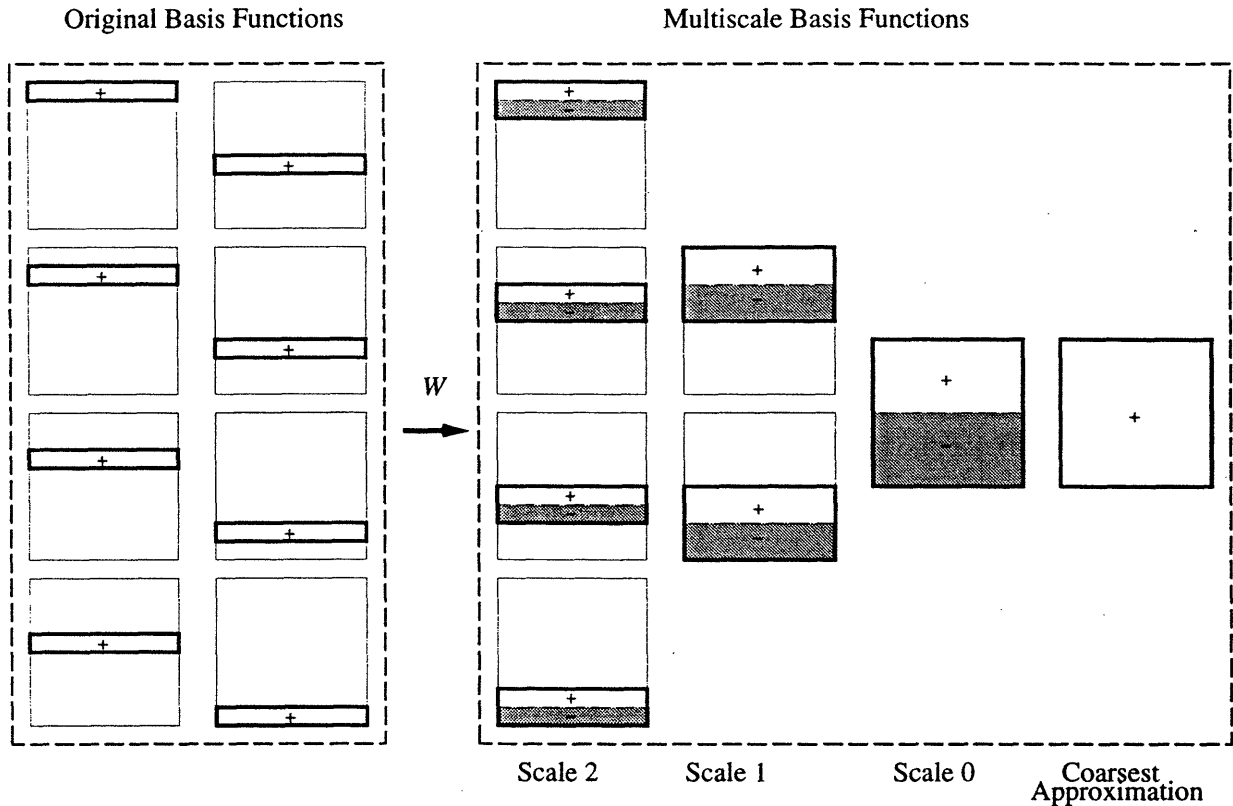


Figure 3: Example of relationship between original strip basis functions contained in  $T_k$  (shown in the left half of the figure) and transformed multiscale basis functions of  $T_k$  (shown in the right half of the figure) for a fixed angle  $k$  corresponding to the Haar wavelet. The multiscale basis functions may be naturally grouped into different scale components based on their spatial extent (or, equivalently, their relation to the coefficients in  $\xi_k$ ), as shown.

object at different resolutions, directly corresponding to the different scale components contained in the transformed vector  $\xi_k$ . In particular, in defining the overall object  $\hat{f}$ , the multiscale basis functions at scale  $m$  and angle  $k$  are weighted by the corresponding detail component  $\xi_k^{(m)}$ . The overall object is then represented by a superposition of such components at *all* angles  $k$ , as captured in the sum in (12).

So far we have simply transformed the representation of the original finest scale object estimate  $\hat{f}$ . But the preceding discussion together with the development in Section 2.3 suggests how to use our new multiscale decomposition  $\xi_k$  and corresponding basis functions  $T_k$  to obtain a multiscale decomposition of the object estimate in the original space. Such a multiresolution decomposition can be obtained through (12) by using a series of approximations to  $x_k$  at successively finer scales, thereby inducing a series of corresponding approximate representations of the object. In particular, we *define* the  $m$ -th scale approximation  $\hat{f}^{(m)}$  to  $\hat{f}$  as:

$$\hat{f}^{(m)} \triangleq \sum_{k=1}^{N_\theta} T_k^T (A(m) \xi_k) \quad (13)$$

where recall that the  $m$ -th scale approximation  $(A(m) \xi_k)$  is obtained by zeroing the finer scale components in the vector of 1-D wavelet transform coefficients of  $\xi_k$ , as discussed in Section 2.3.

Thus the approximation  $\hat{f}^{(m)}$  uses only the  $m$  coarsest scale components of the full vector  $\xi_k$ . Similarly, by  $\Delta \hat{f}^{(m)}$  we denote the additional detail required to go from the object approximation at scale  $m$  to that at scale  $(m + 1)$ , which is given by:

$$\Delta \hat{f}^{(m)} \triangleq \sum_{k=1}^{N_\theta} T_k^T (D(m) \xi_k) \quad (14)$$

where recall that the detail vector  $(D(m) \xi_k)$  is obtained by zeroing all but the corresponding level of detail  $\xi_k^{(m)}$  in  $\xi_k$ . Combining the object approximation and detail definitions (13) and (14) with the scale recursive relationship (10) we see that the *object itself* satisfies the following scale recursive relationship, whereby the object approximation at the next finer scale is obtained from the approximation at the current (coarser) scale through the addition of the incremental detail at this scale, just as for the 1-D case treated in Section 2.3:

$$\hat{f}^{(m+1)} = \hat{f}^{(m)} + \Delta \hat{f}^{(m)} \quad (15)$$

Note that our multiscale object representation given in (13) and corresponding scale recursive construction (15) is induced naturally by the structure of the individual 1-D wavelet-based multiscale decompositions at each angle  $k$  and is *not* simply a 2-D wavelet transform of the original object estimate  $\hat{f}$ . In other words, we are not simply relating the coefficients of a 2-D multiscale decomposition of  $\hat{f}$  based in the original object domain to those of a 1-D decomposition of the data in the projection domain, but rather we are allowing a multiscale projection domain decomposition to *induce* a corresponding, and thus naturally well matched, multiscale object representation. In particular, the  $m$ -th scale approximation of the object  $\hat{f}^{(m)}$  is created as a linear combination of the corresponding  $m$  coarsest multiscale basis functions (c.f. Figure 3) summed over all angles  $k$  (note that the coefficients finer than level  $m$  in  $(A(m) \xi_k)$  are zero and use the object definition (13)). As can be seen, our resulting object representation lives close to the projection domain in which data is gathered, with advantages in efficiency as we will see.

### Multiscale Coefficient Determination

We now have a natural multiscale object representation framework through (13), (14), and (15) that is similar in spirit to the FBP case (3). To complete the process and create multiscale object estimates from data we must find the multiscale object coefficients  $\xi_k$  (which contain all the information we need). Further we desire to find these object coefficients directly from corresponding multiscale tomographic observations. Aside from simply being an evocative notion (e.g. directly relating scale-specific data features to corresponding scale-specific object characteristics), such an approach should be more efficient, in that we would expect coarse scale object characteristics to be most strongly affected by coarse or aggregate data behavior and, conversely, fine scale object characteristics to depend most strongly on fine scale data behavior. Said another way, we would expect the relationship between such multiscale data and object elements to be nearly diagonal, and this is indeed the case.

To the above ends, we perform a wavelet-based multiscale change of basis to the data sequences  $y_k$ , similar to object oriented one in (11), to obtain an equivalent set of multiscale observations:

$$\eta_k \triangleq W y_k \quad (16)$$

where, recall,  $W$  is a matrix taking a discrete sequence to its wavelet transform. We may now easily obtain our desired direct relationship between the multiscale representation of the data at

angle  $k$  in  $\eta_k$  and the multiscale object coefficients  $\xi_k$  at the same angle by combining the two transformations (11) and (16) together with the original FBP relation (4) to obtain:

$$\xi_k = \mathcal{R} \eta_k \quad (17)$$

where  $\mathcal{R} = W^T R W$  is the multiscale data filter, corresponding to the ramp filter  $R$  of the usual FBP case. As we show through examples later, the operator  $R$  is compressed by the wavelet operator so that  $\mathcal{R}$  is nearly diagonal. Further, higher compression is achieved if Daubechies wavelets  $D_n$  with larger  $n$  are used. This observation is consistent with the observations of Beylkin et al [6], since  $R$  is a pseudo-differential operator.

### The Overall Multiscale Algorithm

We are now in a position to present our overall multiscale reconstruction method. By comparing the FBP equations (3) and (4) to the corresponding multiscale equations (12) and (17), respectively, we see that our complete multiscale reconstruction process for estimation of  $\hat{f}$  parallels that of the standard FBP reconstruction, in that identical and independent processing is performed on the multiscale data sets  $\eta_k$  at each angle to obtain the corresponding multiscale object coefficients  $\xi_k$  at that angle, which are then back-projected along corresponding multiscale basis functions  $T_k$  and combined to obtain the final object estimate. Thus our overall procedure, given next, is *no more complex than the standard FBP method*.

#### Algorithm 1 (Multiscale Reconstruction)

1. For a given choice of wavelet, form the multiscale filter matrix  $\mathcal{R} = W R W^T$  (the multiscale counterpart of the original ramp filter) to process the data at each angle.  $\mathcal{R}$  is nearly diagonal.
2. For each angle  $k$  perform the following:
  - (a) Find the multiscale observations  $\eta_k$  by taking the 1-D wavelet transform of the projection data at angle  $k$ ,  $\eta_k = W y_k$ .
  - (b) Calculate the multiscale object coefficient set  $\xi_k = \mathcal{R} \eta_k$  by filtering the multiscale observations.
  - (c) Back-project  $\xi_k$  along the corresponding multiscale basis functions  $T_k$ ,  $T_k^T \xi_k$ .
3. Combine the object contributions from the individual back-projections at each angle to obtain the overall estimate,  $\sum_k T_k^T \xi_k$ .

Beyond simply finding a finest scale object estimate as described in Algorithm 1, however, we also have a method to reconstruct the underlying object at *multiple resolutions* through (13), (14) and (15) and thus for easily obtaining information about the object at multiple scales. In particular, if an approximation  $\hat{f}^{(m)}$  at scale  $m$  is desired, then in Algorithm 1 we need only replace  $\xi_k$  by  $(A(m) \xi_k)$  in Step 2c and 3. In particular, this simply amounts to zeroing detail components in  $\xi_k$  which are finer than scale  $m$ . Further, if instead we want to reconstruct the detail  $\Delta \hat{f}^{(m)}$  added at a particular scale, we need only replace  $\xi_k$  by  $(D(m) \xi_k)$  in Step 2c and 3 of Algorithm 1. Similarly, this simply amounts to zeroing all but the desired scale of detail  $\xi_k^{(m)}$  in  $\xi_k$ . Note that such intermediate scale information about  $\hat{f}$  can even be efficiently found by calculating *only* those elements necessary for reconstructing the scale of interest – i.e. all of  $\xi_k$  is not required. For example, if all that is required is a coarse estimate of the object and not the full reconstruction, only the coarsest elements of  $\xi_k^{(m)}$  are required. Conversely if only fine scale features are to be reconstructed, then only the finest scale detail components of  $\xi_k^{(m)}$  are needed.

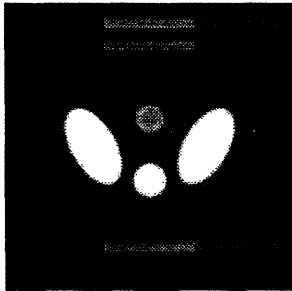


Figure 4: Phantom used for reconstruction experiments. The phantom is  $256 \times 256$  and projections are gathered at 256 equally spaced angles ( $N_\theta = 256$ ) with 256 strips per angle ( $N_s = 256$ ).

### Examples

We now show some examples of our multiscale reconstruction framework. Figure 4 shows the  $256 \times 256$  phantom used in the experiments of this section. Projection data were collected at 256 equally spaced angles ( $N_\theta = 256$ ) with 256 strips used for each projection ( $N_s = 256$ ). First we show a series of approximate reconstructions using the Daubechies  $D_3$  wavelet for the multiscale decomposition  $W$ . Figure 5 shows the various scale approximate object reconstructions  $\hat{f}^{(m)}$  for the entire range of scales  $m = 1, \dots, 8$ . The approximations get finer from left to right and top to bottom (so that the upper left frame is  $\hat{f}^{(1)}$  and the bottom middle frame corresponds to  $\hat{f}^{(8)}$ ). The bottom row, right shows the FBP reconstruction for comparison. Note in particular, that the finest scale approximation  $\hat{f}^{(8)}$  is identical to the FBP estimate  $\hat{f}$ . The intermediate scale estimates demonstrate how information is gathered at different scales. For example, in the scale 3 reconstruction  $\hat{f}^{(3)}$  (top right in the figure) though only 8 of the full 256 coefficient elements in the vectors  $\xi_k$  are being used, we can already distinguish separate objects. By scale 4 (middle row, left) we can start to identify the separate bright regions within the central larger object, while by scale 5 this information is well localized. Even at this comparatively fine scale we are still only using about 12% of the full object coefficient set.

In Figure 6 we show the corresponding detail components  $\Delta \hat{f}^{(m)}$  for the same phantom. Again, the additive detail becomes finer going from left to right and top to bottom in the figure. Notice that the fine scale, edge based, features of the phantom are clearly visible in the  $\Delta \hat{f}^{(4)}$  and  $\Delta \hat{f}^{(5)}$  reconstructions (center row, middle and right in the figure), showing that structural information can be obtained from these detail images alone. Recall that these images provide the added information needed in going from the object approximation at one scale to that at the next finer scale (as provided in Figure 5).

As we discussed earlier, the wavelet-based multiscale transformation of both the representation  $x_k$  and data  $y_k$  also serves to compress the ramp filter matrix  $R$  so that the corresponding multiscale filter matrix  $\mathcal{R}$  is nearly diagonal. As we argued earlier, this reflects the fact that coarse scale object characteristics are most strongly affected by coarse or aggregate data behavior and, conversely, fine scale object characteristics tend to depend most strongly on fine scale data behavior. One consequence is that a very good approximation to the exact reconstruction procedure of Algorithm 1 can be achieved by ignoring the off-diagonal terms of  $\mathcal{R}$  in (17). These off-diagonal terms capture both intra and inter-scale couplings. Further, this approximation to the exact reconstruction becomes better as Daubechies wavelets  $D_n$  with larger  $n$  are used. To illustrate this point, in Figure 7 we show complete (finest scale) reconstructions  $\hat{f}$  of the same phantom as before, based on the same projection data but using a diagonal approximation to  $\mathcal{R}$  in (17) and Algorithm 1 for

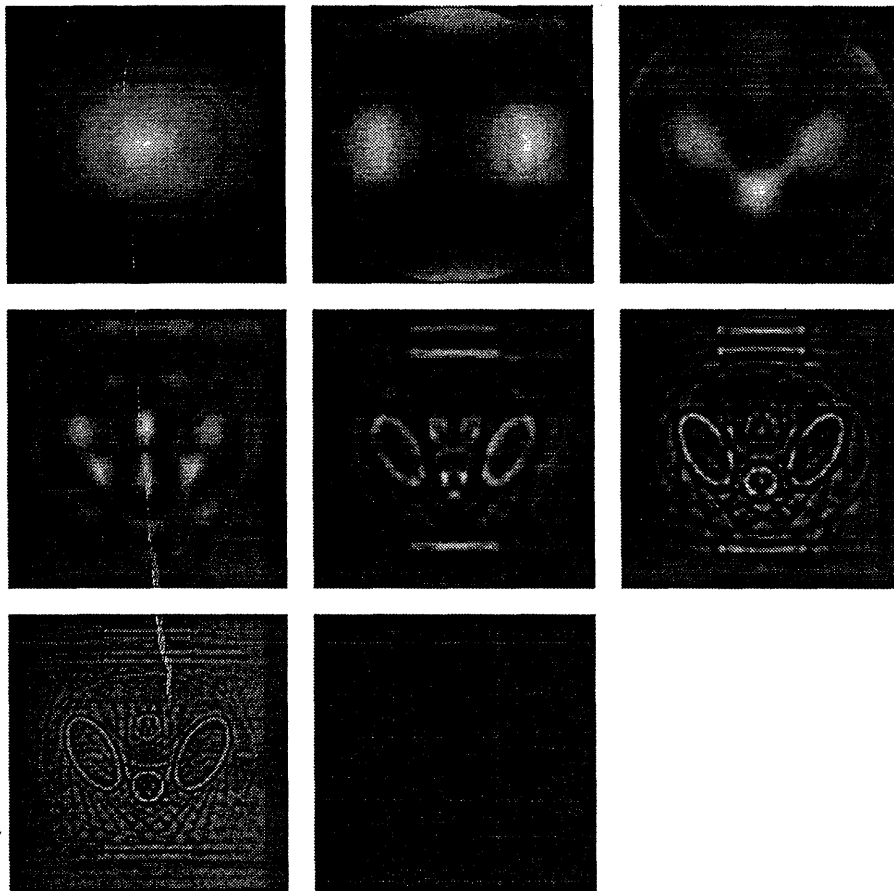


Figure 6: The detail added between successive scales in the reconstructions of Figure 5. First row, left:  $\Delta\hat{f}(0)$ . First row, middle:  $\Delta\hat{f}(1)$ . First row, right:  $\Delta\hat{f}(2)$ . Second row, left:  $\Delta\hat{f}(3)$ . Second row, middle:  $\Delta\hat{f}(4)$ . Second row, right:  $\Delta\hat{f}(5)$ . Third row, left:  $\Delta\hat{f}(6)$ . Third row, middle:  $\Delta\hat{f}(7)$ .

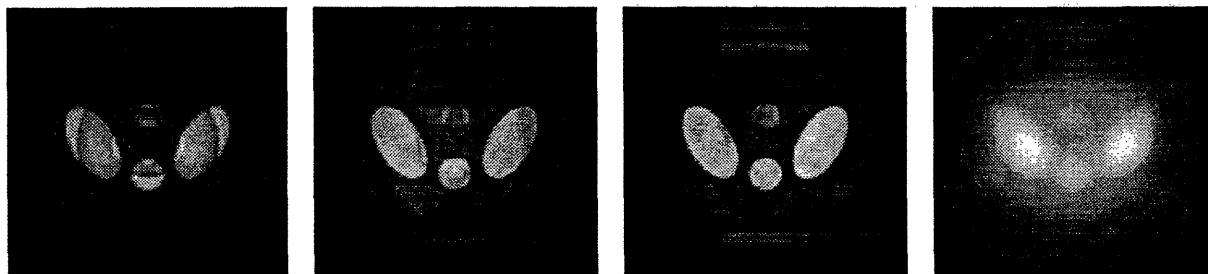


Figure 7: Complete finest scale multiscale reconstructions for phantom of Figure 4 for different approximate filtering operators. The left three frames show approximate multiscale reconstructions using only the diagonal elements of  $\mathcal{R}$  corresponding to different choices of the underlying wavelet: First column = Haar. Second column =  $D_3$ . Third column =  $D_8$ . For comparison, the right-most frame shows an equivalent approximate FBP reconstruction using only the diagonal elements of  $R$ , demonstrating the superiority of the multiscale based approximations.

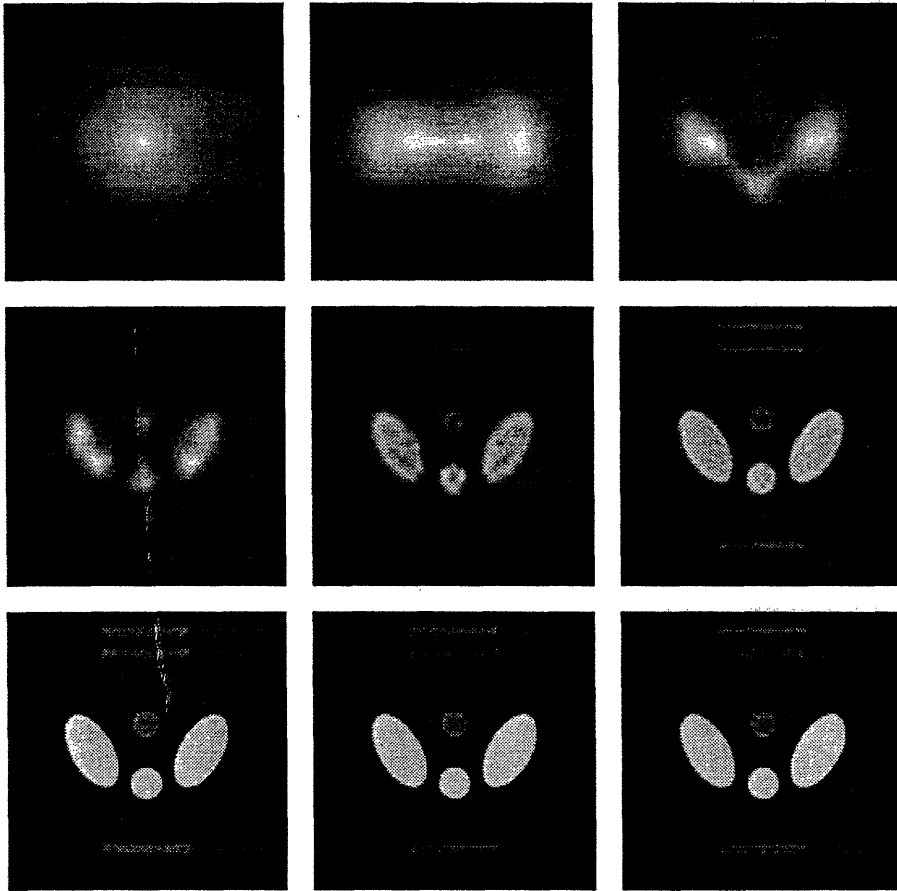


Figure 5: Approximation reconstructions of phantom of Figure 4 at various scales, using  $D_3$  wavelet. First row, left:  $\tilde{f}^{(1)}$ . First row, middle:  $\tilde{f}^{(2)}$ . First row, right:  $\tilde{f}^{(3)}$ . Second row, left:  $\tilde{f}^{(4)}$ . Second row, middle:  $\tilde{f}^{(5)}$ . Second row, right:  $\tilde{f}^{(6)}$ . Third row, left:  $\tilde{f}^{(7)}$ . Third row, middle:  $\tilde{f}^{(8)}$ . The third row, right shows the corresponding FBP reconstruction  $\hat{f}$  for comparison. The FBP reconstruction is the same as  $\tilde{f}^{(8)}$ , since this is the complete reconstruction.

a variety of choices of the wavelet defining  $W$ . For the reconstructions we use only the diagonal elements of  $\mathcal{R}$  (which account for 0.0031% of all the elements for this case) in the calculation of  $\xi$ , effectively setting all off-diagonal elements to zero. Reconstructions corresponding to Daubechies wavelets  $D_n$  with increasing  $n$  (in particular Haar or  $D_1$ ,  $D_3$ , and  $D_8$ ) are shown from left to right in the figure. It can be seen from the improvement in the reconstructions that the accuracy of the diagonal approximation becomes better as  $D_n$  wavelets with increasing  $n$  are used in the definition of  $W$ . In particular, the approximations can be seen to compare very favorably with the standard FBP reconstruction. For comparison we also show on the far right in Figure 7 a corresponding approximate FBP reconstruction obtained using a diagonal approximation to the original ramp-filter matrix  $R$  for reconstruction. It can be seen that a diagonal approximation in the multiscale domain results in far better reconstructions than a similar approximation in the original domain, indicating that the multiscale transformation of data and coefficients has served to decouple the resultant quantities.

In summary, we have formulated a 2-D multiscale object reconstruction method in terms of approximation and detail images. This method is derived from the classical FBP method and thus



is well matched to reconstruction from projection data. The associated 2-D multiresolution object representation is induced by a 1-D wavelet-based change of basis to the original FBP projection space object coefficients. While the resulting representations are similar in spirit to a direct 2-D multiresolution decomposition of the original object, in that approximations are produced at a series of scales along with the detail necessary to proceed from one such approximation to the next finer one, our approach does not correspond to such a direct orthonormal decomposition. As a result it is fundamentally different from previous multiscale-related work in tomography (for example, [15]). In these approaches such a direct 2-D expansion of the object (i.e. a 2-D wavelet transform) is used to directly define the approximation and detail images, the coefficients of which are then calculated from the projection data. In contrast, all of our multiscale quantities inherently “live” in the projection domain. As a result, our representation is closer to the measurement domain than previous multiscale representations, and in particular implies that our approach is no more computationally complex than FBP. To this point we have focused on noiseless reconstructions. Next, we build on our multiscale reconstruction method to obtain a fast method for computing regularized reconstructions from *noisy* projections.

## 4 Multiscale Regularized Reconstructions

In this section we consider the estimation of an object  $f$  from *noisy* projection observations. We extend our multiscale reconstruction method presented in Section 3 to obtain statistically regularized estimates which may be simply and efficiently computed, in particular, with no more effort than is required for the standard FBP reconstruction. This regularized solution is obtained by first solving for the Maximum A Posteriori Probability (MAP) estimate [29] of the multiscale object coefficients,  $\hat{\xi}_k$ , corresponding to a certain naturally derived multiscale prior model and then back-projecting these multiscale coefficient estimates along the corresponding multiscale basis functions as before.

The presence of noise in projection data often leads to reconstructions by standard methods, such as FBP, that are unacceptable and thus require some form of regularization. Traditionally, two broad approaches have been used in the generation of regularized object estimates from such noisy projection data. Perhaps the simplest approach has been to simply roll off the ramp filter used in the standard FBP reconstruction at high frequencies. This is called apodization [7] and several different windows are typically used for this purpose, for example Hanning, Hamming, Parzen, Butterworth etc. [8]. The assumption is that most the object energy occurs at low frequencies while the most disturbing noise-derived artifacts occur at high frequency. The high frequency roll-off thus attenuates these components at the expense of not reconstructing the fine scale features in the object. Since the overall procedure is essentially the same as the original FBP method, the result is a fast, though ad hoc, method for regularization. The other traditional approach to regularizing the noisy data problem is statistically based. This method starts with a statistical model for the noisy observations based on (2):

$$y_k = T_k f + n_k \quad (18)$$

where  $n_k$  is taken as an additive noise vector at angle  $k$ . This observation model is then coupled with a 2-D Markov random field (MRF) prior model [25, 26] for  $f$  to yield a direct MAP estimate of the object  $\hat{f}$ . While statistically based, thus allowing the systematic inclusion of prior information, the 2-D spatially-local MRF prior models used for the object generally lead to optimization problems that are *extremely* computationally complex. As a result, these methods have traditionally not found favor in practical applications.

In contrast to the above two techniques, we will develop a multiscale MAP object estimate that, while retaining all of the advantages of statistically-based approaches, is obtained with the same computational complexity as the FBP reconstruction. To accomplish this, we continue to work in the projection domain, as the FBP method does, and build our statistical models there, rather than in the original object domain. As in Section 3, we then allow the resulting projection domain coefficients to induce a 2-D object representation through the back-projection and summation operations. To this end we start with an observation equation relating the noisy data  $y_k$  to the *FBP object coefficients*  $x_k$ , rather than the corresponding 2-D object  $f$  as is done in (18). Such a relationship may be found in the FBP relationship (4), which in the presence of noise in the data becomes:

$$y_k = R^{-1}x_k + n_k, \quad n_k \sim \mathcal{N}(0, \Lambda_{n_k}) \quad (19)$$

where, recall  $R$  is the FBP ramp filter operator<sup>2</sup>, the notation  $z \sim \mathcal{N}(m, \Lambda)$  denotes a Gaussian distribution of mean  $m$  and covariance  $\Lambda$  and  $I_n$  denotes an  $n \times n$  identity matrix. In particular, we assume the  $\Lambda_{n_k} = \lambda_k I_{N_s}$ , i.e. that the noise is uncorrelated from strip to strip but may have different noise covariances at different angles, capturing the possibility that the data at different projections may be of differing quality (e.g. due to different sensors or imaging configurations). Further, we assume that the noise is uncorrelated from angle to angle, so that  $n_k$  is independent of  $n_j$ ,  $k \neq j$ . This model of independent noise in the projection domain is well justified for most tomographic applications.

As in Section 3, for purposes of estimation we desire a relationship between multiscale representations of the data, object coefficients, and noise. Working in the multiscale transform domain will again allow us to obtain induced multiresolution estimates of the object. Such a multiscale oriented relationship between the quantities of interest can be found by combining (19) with the multiresolution changes of bases (11) and (16) based on  $W$  (defined in Section 2.3) to obtain:

$$\eta_k = \mathcal{R}^{-1}\xi_k + \nu_k, \quad \nu_k \sim \mathcal{N}(0, \Lambda_{\nu_k}). \quad (20)$$

where  $\nu_k = Wn_k$  is the multiscale transformed noise vector at angle  $k$  with  $\Lambda_{\nu_k} = W\Lambda_{n_k}W^T = \lambda_k I_{N_s}$  as its corresponding covariance. This equation relates our observed noisy multiscale data  $\eta_k$  to our desired multiscale object coefficients  $\xi_k$  through the multiscale filtering operator  $\mathcal{R}$ . Note that the assumption of uncorrelated noise from angle to angle and strip to strip in the original projection domain results in uncorrelated noise from angle to angle and multiscale strip to multiscale strip in the multiscale domain, since  $W$  is an orthonormal transformation.

### The Multiscale Prior Model

To create a MAP estimate of the multiscale object coefficients  $\xi_k$ , we will combine the observation equation (20) with a prior statistical model for the desired unknown multiscale coefficient vectors  $\xi_k$ . Multiresolution *object* estimates and the detail between them can then be easily obtained by using the resulting MAP coefficient estimates  $\hat{\xi}_k$  at multiple scales in the multiscale object definitions (13) and (14), as was done previously in Section 3.

We base our prior model of the object coefficients *directly* in scale-space. Such scale-space based prior models are desirable for a number of reasons, e.g. they have been shown to lead to extremely efficient scale-recursive algorithms [9, 13] and they parsimoniously capture self-similar

<sup>2</sup>Note that (19) assumes that  $R^{-1}$  exists. For the case where  $R$  represents an ideal ramp filter this will indeed *not* be the case, as this operator nulls out the DC component of a signal. For filters used in practice, however, this inverse *does* exist and the expression given in (19), based on such a filter is well defined. Details may be found in Appendix B.

behavior, thus providing realistic models for a wide range of natural phenomenon. In particular, such self-similar models can be obtained by choosing the detail components  $\xi_k^{(m)}$  (i.e. the wavelet coefficients at each scale) as independent,  $\mathcal{N}(0, \sigma^2 2^{-\rho m})$  random variables [14]. The parameter  $\rho$  determines the nature, i.e. the texture, of the resulting self-similar process while  $\sigma^2$  controls the overall magnitude. This model says that the variance of the detail added in going from the approximation at scale  $m$  to the approximation at scale  $m + 1$  decreases geometrically with scale. If  $\rho = 0$  the resulting finest level representation (the elements of  $\mathbf{x}_k$ ) correspond to samples of white noise (i.e. are completely uncorrelated), while as  $\rho$  increases the components of  $\mathbf{x}_k$  show greater long range correlation. Such self-similar models are commonly and effectively used in many application areas such as modeling of natural terrain and other textures, biological signals, geophysical and economic time series, etc. [10–14].

In addition to defining the scale varying probabilistic structure of the detail components of  $\xi_k$ , we also need a probabilistic model for the element of  $\xi_k$  corresponding to the coarsest scale approximation of  $\mathbf{x}_k$ , i.e.  $\mathbf{x}_k^{(0)}$ . This term describes the DC or average behavior of  $\mathbf{x}_k$ , of which we expect to have little prior knowledge. As a result we choose this element as  $\mathcal{N}(0, \bar{\Lambda}_\xi)$ , where the (scalar) uncertainty  $\bar{\Lambda}_\xi$  is chosen sufficiently large to prevent a bias in our estimate of the average behavior of the coefficients, letting it be determined instead by the data.

In summary, we use a prior model for the components of the multiscale coefficient vectors  $\xi_k$  which is defined directly in scale-space and which corresponds to a self-similar, fractal-like prior model for the corresponding object coefficients  $\mathbf{x}_k$ . In particular, this model is given by  $\xi_k \sim \mathcal{N}(0, \Lambda_\xi)$  with  $\xi_k$  independent from angle to angle and where:

$$\begin{aligned} \Lambda_\xi &= \text{block diag} \left[ \Lambda_\xi^{(N-1)}, \dots, \Lambda_\xi^{(1)}, \Lambda_\xi^{(0)}, \bar{\Lambda}_\xi \right] \\ \Lambda_\xi^{(m)} &= \sigma^2 2^{-\rho m} I_{2^m} \end{aligned} \quad (21)$$

Again, this model not only assumes that the sets of multiscale object coefficients,  $\xi_k$ , are *independent from angle to angle* but also that these coefficients are *independent from scale to scale*, that they are *independent and identically distributed within a given scale*, and finally that their *variance decreases geometrically proceeding from coarse to fine scales*. Obviously other choices may be made for the statistics for the multiscale object coefficients, and we discuss some particularly interesting possibilities in the conclusions. The choice we have made in (21) while simple, is well adapted to many naturally occurring phenomenon. In addition, since the *observation noise* power is uniform across scales or frequencies, the geometrically decreasing variance of this prior model implies that the projection data will most strongly influence the reconstruction of coarse scale features and the prior model will most strongly influence the reconstruction of fine scale features. This reflects our belief that the fine scale behavior of the object (corresponding to high frequencies) is the most likely to be corrupted by noise. Finally, our choice of prior model in (21) results in efficient processing algorithms for the solution of the corresponding MAP estimate, in particular with no more complexity than the standard FBP reconstruction.

### The Multiscale MAP Estimate

We are now in a position to present our overall algorithm for computing a MAP [29] multiscale object estimate  $\hat{\xi}_k$ . Since the data at each angle  $\eta_k$  and the corresponding prior model for  $\xi_k$  are independent from angle to angle, the MAP estimates of the vectors  $\xi_k$  decouple. In particular, the estimate of  $\xi_k$  at each angle, based on the observations (20) and the prior model (21) is given by:

$$\hat{\xi}_k = \left[ \Lambda_\xi^{-1} + \mathcal{R}^{-T} \Lambda_{\nu_k}^{-1} \mathcal{R}^{-1} \right]^{-1} \mathcal{R}^{-T} \Lambda_{\nu_k}^{-1} \eta_k \triangleq \bar{\mathcal{R}} \eta_k \quad (22)$$

where the regularized multiscale filter operator  $\overline{\mathcal{R}}$  is defined in the obvious way. This regularized filtering matrix is exactly analogous to the unregularized filtering operator  $\mathcal{R}$  of (17) for the noise free case. In this regularized case, however,  $\overline{\mathcal{R}}$  now also depends on both the noise model  $\Lambda_{\nu_k}$  and the prior object model  $\Lambda_\xi$ . If the noise variance is low relative to the uncertainty in the prior model (so  $\Lambda_{\nu_k}^{-1}$  is large) then  $\overline{\mathcal{R}}$  will approach  $\mathcal{R}$  and the estimate will tend toward the standard unregularized one. Conversely, as the noise increases,  $\overline{\mathcal{R}}$  will depend to a greater extent on the prior model term  $\Lambda_\xi$  and the solution will be more regularized or smoothed.

Finally, as in the noise-less case, the resulting *object estimate*  $\hat{f}$  is then obtained by back-projecting the estimated multiscale object coefficients  $\hat{\xi}_k$  along the corresponding multiscale basis functions  $\mathcal{T}_k$  and combining the result. The overall structure of this regularized reconstruction parallels that of the original FBP method, and therefore is of the *same computational complexity as FBP*. In summary, our overall, efficient regularized multiscale estimation algorithm is given by the following procedure, which parallels our unregularized multiscale reconstruction algorithm:

### Algorithm 2 (Regularized Multiscale Reconstruction)

1. Find the regularized multiscale filter matrix  $\overline{\mathcal{R}}$  (the multiscale regularized counterpart of the original ramp filter) by doing the following:
  - (a) For a given choice of wavelet, form the unregularized multiscale filter matrix  $\mathcal{R} = W R W^T$  as before.
  - (b) Choose the model parameters  $\lambda_k$  specifying the variances of the observation noise processes and thus defining  $\Lambda_{\nu_k}$  c.f. (20).
  - (c) Choose the multiscale prior model parameters  $\sigma^2$ ,  $\rho$  and  $\overline{\Lambda}_\xi$  specifying the magnitude and texture of the model and the uncertainty in its average value, respectively, and generate the prior covariance matrix  $\Lambda_\xi$  through (21).
  - (d) Form  $\overline{\mathcal{R}} = \left[ \Lambda_\xi^{-1} + \mathcal{R}^{-T} \Lambda_{\nu_k}^{-1} \mathcal{R}^{-1} \right]^{-1} \mathcal{R}^{-T} \Lambda_{\nu_k}^{-1}$ .
2. For each angle  $k$  perform the following:
  - (a) Find the multiscale observations  $\eta_k$  by taking the 1-D wavelet transform of the projection data at angle  $k$ ,  $\eta_k = W y_k$ .
  - (b) Calculate the regularized multiscale object coefficient set  $\hat{\xi}_k = \overline{\mathcal{R}} \eta_k$  by filtering the multiscale observations.
  - (c) Back-project  $\hat{\xi}_k$  along the corresponding multiscale basis functions  $\mathcal{T}_k$ ,  $\mathcal{T}_k^T \hat{\xi}_k$ .
3. Combine the regularized object contributions from the individual back-projections at each angle to obtain the overall regularized object estimate,  $\sum_k \mathcal{T}_k^T \hat{\xi}_k$ .

As before, we may also easily obtain regularized reconstructions of the object at multiple resolutions by using (13) and (14) together with the MAP coefficient estimates  $\hat{\xi}_k$ . In particular, to obtain the approximation  $\hat{f}^{(m)}$  at scale  $m$  then we need only replace  $\hat{\xi}_k$  by  $(A(m) \hat{\xi}_k)$  (corresponding to simply zeroing some of the terms in  $\hat{\xi}_k$ ) in Step 2c and 3. Similarly, the corresponding object detail components  $\Delta \hat{f}^{(m)}$  at scale  $m$  may be obtained by using  $(D(m) \hat{\xi}_k)$  in place of  $\hat{\xi}_k$  in these steps.

While Algorithm 2 is already extremely efficient, in that 2-D multiscale regularized object estimates are generated with no more complexity than is needed for the standard FBP method, additional gains may be obtained by exploiting the ability of the wavelet transform operator  $W$

to compress the FBP filtering operator  $R$ . Recall, in particular, that the (unregularized) multiscale filtering matrix  $\mathcal{R} = WRW^T$  is nearly diagonal, with this approximation becoming better as Daubechies wavelets  $D_n$  with increasing  $n$  are used in the specification of  $W$ . Based on our assumptions, the matrices  $\Lambda_\xi$  and  $\Lambda_{\nu_k}$ , specifying the prior model and observation covariances respectively, are already diagonal. If in addition  $\mathcal{R}^{-1}$  were also a diagonal matrix, then from (22) we see that  $\bar{\mathcal{R}}$  itself would be diagonal, with the result that the “filtering” in Step 2b of Algorithm 2 would simply become point by point scaling of the data. To this end we will assume that the wavelet transform  $W$  truly diagonalizes  $R$  by effectively ignoring the small, off-diagonal elements in  $\mathcal{R}^{-1}$ . That is, we assume<sup>3</sup>:

$$\mathcal{R}^{-1} \approx \text{diag}(r_1, r_2, \dots, r_{N_s}) \quad (23)$$

where  $r_i$  are the diagonal elements of  $\mathcal{R}^{-1}$ . Now let us represent the diagonal prior model covariance matrix as  $\Lambda_\xi = \text{diag}[p_1, p_2, \dots, p_{N_s}]$ , and recall that  $\Lambda_{\nu_k} = \lambda_k I_{N_s}$ . Using these quantities together with our approximation to  $\mathcal{R}^{-1}$  in the specification of the estimate (22) yields an approximate expression for  $\hat{\xi}_k$ :

$$\hat{\xi}_k \approx \text{diag} \left( \frac{r_1}{r_1^2 + (\lambda_k/p_1)}, \frac{r_2}{r_2^2 + (\lambda_k/p_2)}, \dots, \frac{r_{N_s}}{r_{N_s}^2 + (\lambda_k/p_{N_s})} \right) \eta_k \triangleq \tilde{\mathcal{R}} \eta_k \quad (24)$$

where the *approximate* MAP filtering matrix  $\tilde{\mathcal{R}}$  is defined in the obvious way. Our experience is that when  $W$  is defined using Daubechies wavelets of order 3 or higher (i.e. using  $D_3, D_4, \dots$ ), the estimates obtained using  $\tilde{\mathcal{R}}$  in place of the exact regularized filter  $\bar{\mathcal{R}}$  in Algorithm 2 are visually indistinguishable from the exact estimates where  $\mathcal{R}^{-1}$  is not assumed to be diagonal. Indeed, it is actually this approximate filtering operator  $\tilde{\mathcal{R}}$  that we use to generate the example reconstructions we show next.

Before proceeding, however, let us examine our MAP regularized filtering operator  $\bar{\mathcal{R}}$  in more detail to understand how our multiscale MAP estimation procedure relates both to the standard FBP method and the ad hoc regularization obtained through apodization of the FBP filter. The MAP estimates  $\hat{\xi}_k$  induce corresponding estimates  $\hat{x}_k$  of the original object coefficients  $x_k$  through the change of basis (11) and, similarly,  $\eta_k$  and  $y_k$  are related through (16). Thus, the multiscale MAP estimation operation specified by (22) imposes a corresponding relationship between the original finest scale quantities  $\hat{x}_k$  and  $y_k$ , given by:

$$\hat{x}_k = (W^T \bar{\mathcal{R}} W) y_k \triangleq R_{\text{eff}} y_k \quad (25)$$

where the effective multiscale MAP regularized filtering matrix  $R_{\text{eff}}$  is defined in the obvious way. The effect of this MAP regularized filter can now be compared to the standard FBP or apodized ones. The behavior of the matrix operator  $R_{\text{eff}}$  can be most easily understood by examining its corresponding frequency domain characteristics. To this end, in Figure 8 we plot the magnitude of the Fourier transform of the central row of effective regularized matrix  $R_{\text{eff}}$  corresponding to a variety of choices of the model or regularization parameters  $\lambda_k$  (the noise variance) and  $\rho$  (the decay rate across scales of the added detail variance) for fixed  $\sigma^2 = 1$  (overall prior model amplitude) and  $\bar{\Lambda}_\xi = 1$  (prior model DC variance). We also plot, with heavy lines, the magnitude of the Fourier transform of the corresponding central row of the standard FBP ramp filter matrix  $R$  for comparison. From Figure 8, we can see that in the multiscale MAP framework regularization is

<sup>3</sup>One can imagine another level of approximation where we set the off-diagonal elements of  $\mathcal{R}$  itself to zero *prior* to inversion rather than those of  $\mathcal{R}^{-1}$ . This further approximation results in reconstructions which are visually very similar to what we obtain here.

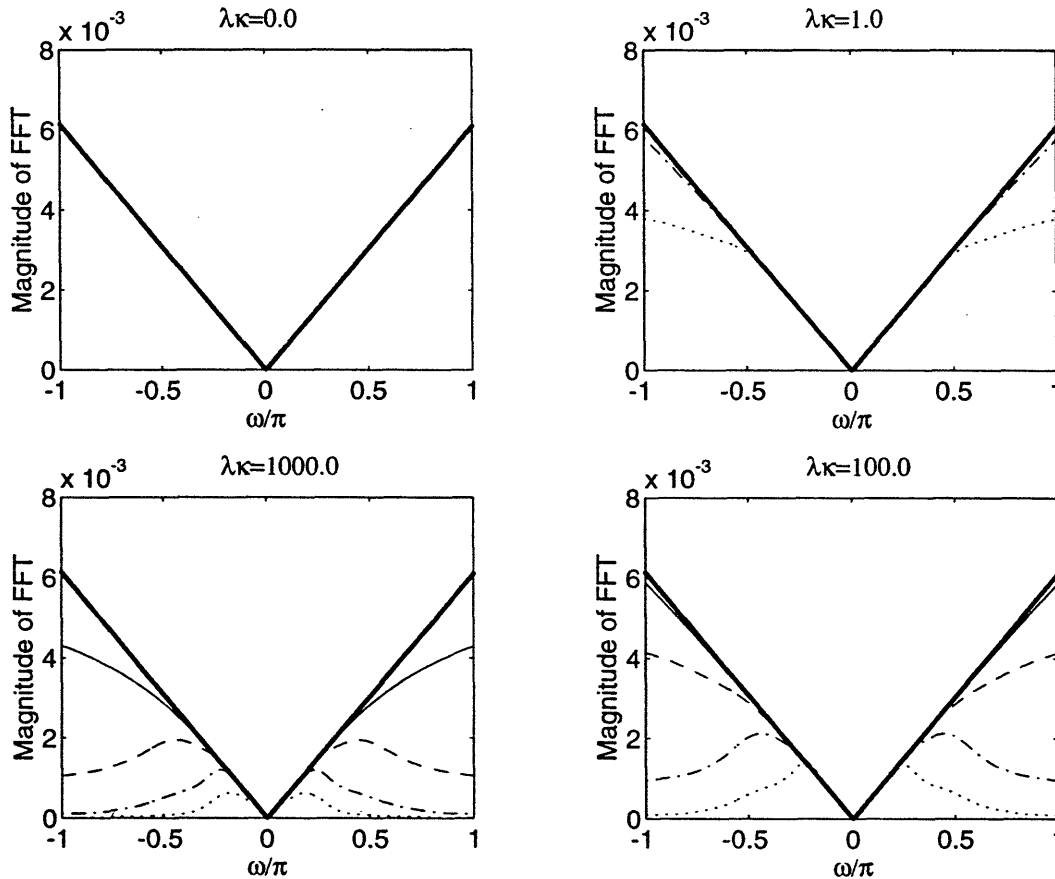


Figure 8: The Fourier transform of the central row of  $R_{\text{eff}}$  for different values of regularization parameters  $\rho$  and  $\lambda_k$ , illustrating the effect of the multiscale regularizing filter in the frequency domain. In each of the subplots, the V-shaped heavy line corresponds to the standard FBP ramp filter and the four curves from top to bottom correspond to  $\rho = 0.5$  (solid line), 1.0 (dashed line), 1.5 (dashdot line) and 2.0 (dotted line) respectively (in some subplots some of the lines overlap). In all cases we fixed  $\sigma^2 = 1$  (the overall prior model amplitude) and  $\bar{\Lambda}_\xi = 1$  (the prior model DC variance).

basically achieved by rolling off the ramp filter at high frequencies, the same principle as used in the ad hoc, apodization regularized FBP reconstructions. We also see that decreasing the observation noise variance  $\lambda_k$  for a fixed prior model structure  $\rho$ , or conversely, increasing the variance of the detail added in proceeding from coarse to fine scales in the prior model (i.e. decreasing  $\rho$ ) for a fixed observation noise variance  $\lambda_k$ , leads to decreased regularization as reflected in decreased high frequency attenuation. This behavior is reasonable, in that in the first case, the data becomes less noisy while in the second the uncertainty in the prior model becomes larger. In both these cases one would want to put more reliance on the data (i.e. less regularization).

In summary then, our multiscale based regularization approach, though derived from statistical considerations and possessing all the advantages of such methods (e.g. the ability to incorporate prior knowledge in a rational way, the ability to do performance analysis and understand the relative importance of various sources of uncertainty, etc.), obtains results at no greater (and in some cases with substantially less) computational complexity than standard unregularized or ad

hoc approaches. In addition, we obtain, *essentially for free*, estimates at multiple resolutions and thus the ability to extract information from data at multiple scales.

### Examples

Next we show some examples of reconstructions using our multiscale methods in the presence of noise. The same  $256 \times 256$  phantom shown in Figure 4 was used for all experiments. In each case projection data for the phantom were again generated at  $N_\theta = 256$  equally spaced angles with  $N_s = 256$  strips in each projection. These noise-free values were then corrupted through the addition of independent, zero-mean Gaussian noise to yield our observations. The variance  $\lambda_n$  of this additive noise depended on the experiment and was chosen to yield an equivalent signal-to-noise ratio (SNR) of the resulting observations, defined as:

$$\text{SNR (dB)} = 10 \log \frac{\sum_{k=1}^{N_\theta} \|T_k f\|^2}{\lambda_n N_\theta N_s} \quad (26)$$

where, recall,  $T_k f$  is the noise-free projection data at angle  $k$ . Finally, in all multiscale reconstructions we show here the Daubechies  $D_3$  wavelet was used in the definition of  $W$  for the reconstruction.

The first example, shown in Figure 9, demonstrates reconstruction from noisy data using the *unregularized* multiscale approach of Section 3. The variance  $\lambda_n$  of the added noise was chosen to yield a SNR of 5 dB. This figure shows the various scale approximate object reconstructions  $\hat{f}^{(m)}$  corresponding to the unregularized Algorithm 1 for the complete range of scales  $m = 1, \dots, 8$ . As before, the approximations become finer from left to right and top to bottom (so that the upper left frame is  $\hat{f}^{(1)}$  and the bottom middle frame corresponds to  $\hat{f}^{(8)}$ ). The bottom right frame shows the standard FBP reconstruction based on the noisy data. Since  $\hat{f}^{(8)}$  corresponds to the unregularized complete finest scale reconstruction it is also the same as the standard FBP reconstruction based on the noisy data for this case. The figure illustrates the resolution-accuracy tradeoff inherently captured in the multiscale framework and confirms the point that even in the unregularized case, information from noisy observations can be focused by stopping the reconstruction at a coarse scale, for example scale 5 (center row, middle in the figure). The finer scale detail contributions  $\Delta \hat{f}^{(m)}$ ,  $m \geq 5$  are evidently mainly noise which obscure the object features. In particular, in the finest scale reconstruction (i.e. the standard FBP reconstruction) the object is almost completely lost in the noise.

Next we show estimates generated by our multiscale MAP regularized estimation method discussed in this section. Figure 10 shows the various scale approximate object reconstructions  $\hat{f}^{(m)}$  corresponding to our multiscale MAP estimate of  $\hat{\xi}_k$  using noisy data with same SNR (i.e. SNR = 5 dB) as in Figure 9. The MAP estimate  $\hat{\xi}_k$  was generated using the extremely efficient approximate expression (24), which, for the Daubechies  $D_3$  wavelet we are using, was indistinguishable from the corresponding estimate based on the exact expression (22). Again the approximations become finer from left to right and top to bottom in the figure. For these reconstructions we chose the modeled observation noise variance as  $\lambda_k = 5.5 \times 10^5$ . For the statistical model parameters of the prior, the decay rate across scale of the added detail variance was chosen as  $\rho = 1.5$ , the overall magnitude of the prior was set to  $\sigma^2 = 11$ , and the variance of the prior model average value was  $\bar{\Lambda}_\xi = 1$ . The effect of the regularization can be readily seen in its ability to suppress noise in the finest scale reconstruction. For comparison, the standard FBP reconstruction for this case is given on the bottom row, right in Figure 10. In addition, the multiscale nature of the information focusing can be seen in the scale evolution of the estimates. In particular, there appears to be little difference between scale 5 and finer scale estimates in the figure, suggesting that little additional information

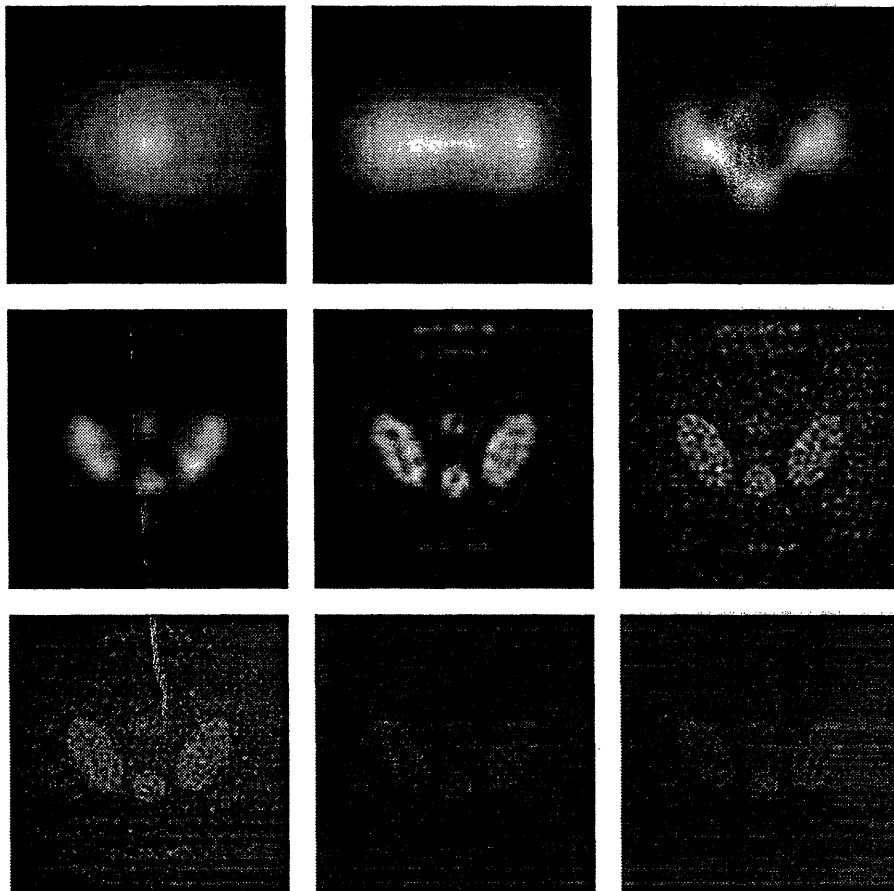


Figure 9: Reconstructions of phantom of Figure 4 from 5 dB SNR projection data based on unregularized Algorithm 1 using  $D_3$  wavelet. Reconstructions are shown at various scales demonstrating the smoothing effect that can be achieved. First row, left:  $\tilde{f}^{(1)}$ . First row, middle:  $\tilde{f}^{(2)}$ . First row, right:  $\tilde{f}^{(3)}$ . Second row, left:  $\tilde{f}^{(4)}$ . Second row, middle:  $\tilde{f}^{(5)}$ . Second row, right:  $\tilde{f}^{(6)}$ . Third row, left:  $\tilde{f}^{(7)}$ . Third row, middle:  $\tilde{f}^{(8)}$ . The standard FBP is shown in the third row, right for comparison. The FBP reconstruction is the same as  $\tilde{f}^{(8)}$ , since this is the complete unregularized reconstruction.

is being obtained in proceeding to such finer scales, that the additional degrees of freedom being added at such finer scales are not really being supported by the data, and thus that we should stop the reconstruction at this coarser scale. Further, estimates at scale 5 and coarser appear quite similar to the corresponding *unregularized* estimates in Figure 9, showing that these coarser scale estimates are dominated by the data and are not very dependent on the prior model at this point anyway.

Finally, in Figure 11, we show a series of finest scale multiscale MAP regularized reconstructions, corresponding to different choices of the prior model texture as determined by the parameter  $\rho$ . The same phantom as before is used, but we use observations with a SNR of  $-10$  dB (much worse than used above). The standard FBP reconstruction is shown for comparison in the far right image of the figure. The object is completely lost in the FBP reconstruction at this extreme level of noise. The MAP reconstructions are shown in the first three frames of the figure, with a smoother, more correlated prior model being used as we proceed from left to right. The specific



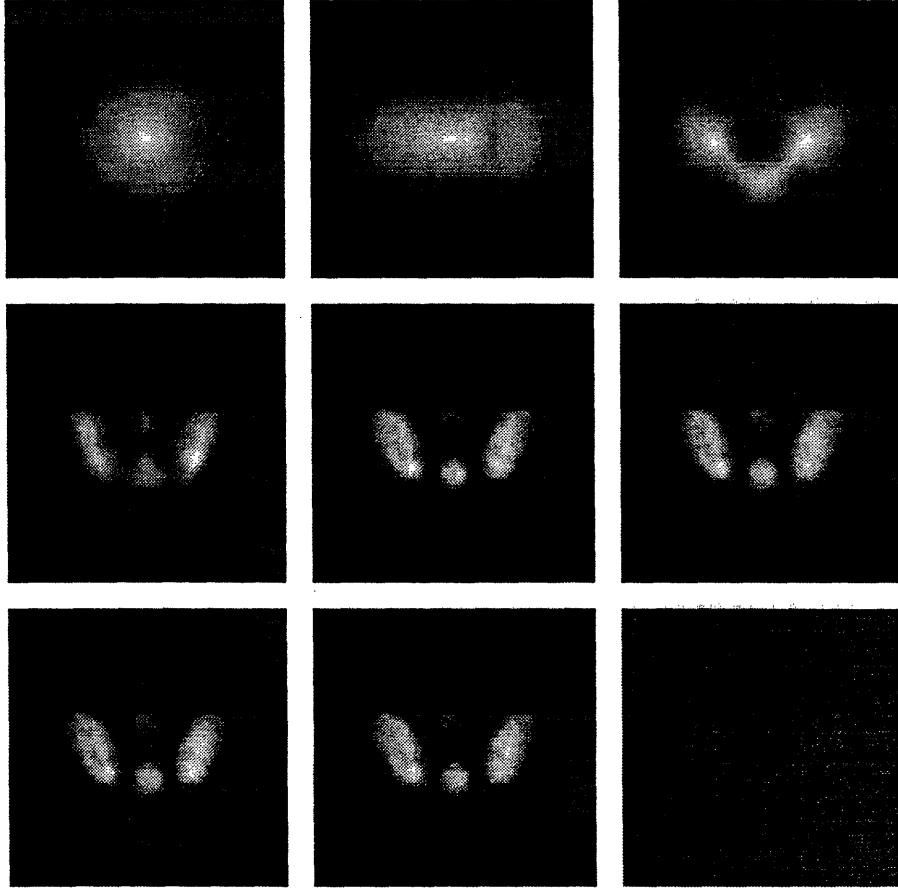


Figure 10: Multiscale MAP regularized reconstructions at various scales of phantom of Figure 4 from 5 dB SNR projection data using  $D_3$  wavelet. The values of the statistical model parameters used are  $\lambda_k = 5.5 \times 10^5$ ,  $\rho = 1.5$ ,  $\sigma^2 = 11$ ,  $\bar{\Lambda}_\xi = 1$ . First row, left:  $\hat{f}^{(1)}$ . First row, middle:  $\hat{f}^{(2)}$ . First row, right:  $\hat{f}^{(3)}$ . Second row, left:  $\hat{f}^{(4)}$ . Second row, middle:  $\hat{f}^{(5)}$ . Second row, right:  $\hat{f}^{(6)}$ . Third row, left:  $\hat{f}^{(7)}$ . Third row, middle:  $\hat{f}^{(8)}$ . For comparison, the standard FBP reconstruction for this case is given in the third row, right. The improved ability of the regularized reconstructions to extract information is demonstrated.

multiscale MAP model parameters were chosen as follows. The observation noise variance was chosen as  $\lambda_k = 1.7 \times 10^7$ . The overall prior model magnitude was set to  $\sigma^2 = 17$  while the prior model DC variance was set to  $\bar{\Lambda}_\xi = 1$ . The prior model texture parameter  $\rho$  took on the values  $\{0.5, 1.0, 1.5\}$ . The increased smoothness in the prior can be seen to be reflected in increased smoothness of the corresponding estimates. Note also the ability of the algorithm to pull out at least the global object features in the presence of this substantial amount of noise. Again, the more highly smoothed reconstructions (corresponding to higher values of  $\rho$ ) appear quite similar to the coarser level, unregularized reconstructions shown previously, showing that we are really accessing the coarse level information in the data.

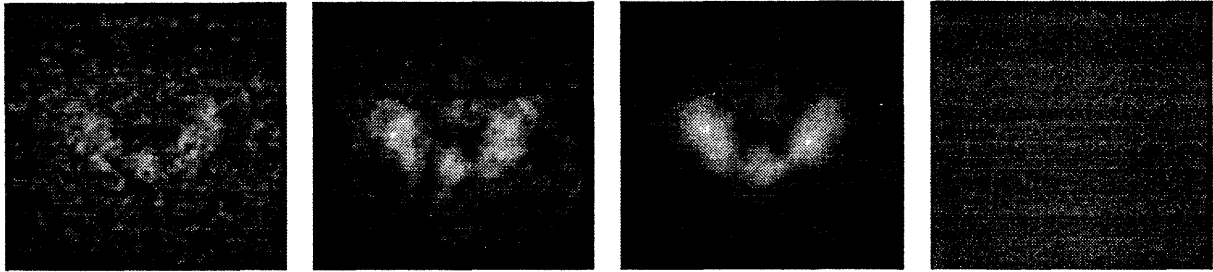


Figure 11: Multiscale MAP regularized reconstructions of the phantom of Figure 4 at the finest scale from  $-10$  dB SNR observations for different choices of prior model texture,  $\rho$ , with  $\lambda_k = 1.7 \times 10^7$ ,  $\sigma^2 = 17$ , and  $\bar{\Lambda}_\xi = 1$ , are shown in the first three frames: Frame 1:  $\rho = 0.5$ . Frame 2:  $\rho = 1.0$ . Frame 3:  $\rho = 1.5$ . For comparison the standard FBP reconstruction is shown in the last frame on the far right.

## 5 Conclusions

In this paper we have developed a wavelet-based multiscale tomographic reconstruction technique which is different from other multiscale techniques in the following respects. First, our 2-D multiscale object representation is naturally induced by expanding the FBP coefficients, and hence basis functions (i.e. strips), in a *1-D wavelet basis*. This is in contrast to other multiscale reconstruction techniques which *begin* with a 2-D object representation obtained from a full 2-D wavelet decomposition of the object space. These techniques must subsequently relate the inherently 1-D projection data to these fundamentally 2-D object coefficients. In contrast, the multiscale representation resulting from our approach, arising as it does from the projection strips themselves, is much closer to the measurement domain. The result is a highly efficient method to compute our multiscale object coefficients, in particular, no more complex than the widely used standard FBP operation. Yet, unlike the FBP method, our multiscale reconstructions also provide a framework for the extraction and presentation of information at multiple resolutions from data. Further, our resulting multiscale relationships between data and object allow extremely simple *approximations* to be made to our exact relationships with *virtually no loss in resulting image quality*, thus further improving the potential efficiency of our approach. Such approximations are not possible with the standard FBP method, as they result in severe artifacts.

In addition, based on this wavelet-based multiscale framework, we have proposed a statistically-based multiresolution MAP estimation algorithm. This method provides statistically regularized reconstructions from noisy data, and does so at multiple resolutions, *at no more effort than is required for the standard FBP method*. This approach, based on the construction of prior models directly in scale-space, allows for the inclusion of natural, self-similar prior models into the reconstruction process. In contrast, conventional statistically-based regularization methods, utilizing MRF-type prior models constructed directly in (finest scale) object space, lead to extremely complex and taxing optimization problems. The result has typically been that such statistically motivated methods have been shunned in practice in favor of fast, though ad hoc, approaches. Our results provide a bridge between these two extremes. Further, in providing estimates at multiple resolutions, our results provide tools for the assessment of the resolution versus accuracy tradeoff, wherein we expect coarser scale features of data to be more accurately determined than finer scale ones. Though we did not exploit this ability in the present paper, our formulation also allowed for the possibility of combining data from projections of fundamentally different quality, through

the specification of different noise variances  $\lambda_k$  at different angles. The resulting estimates do not correspond to a simple FBP or even rolled off FBP reconstruction, yet are easily obtain in our framework. Finally, as before, our multiscale MAP approach leads to algorithms which are amenable to an additional level of approximation, with resulting improved efficiency, again at virtually no loss in corresponding reconstruction quality.

Even though this paper concentrates on the complete-data case, our multiscale reconstruction method has the potential of overcoming this limitation and dealing with limited or missing data problems. First note that our ability to combine projections of different quality already does this to some extent, in that some projections can be down weighted. More important, however, is the structure of the prior model covariance. Specifically, while the prior model (21) assumes that the multiscale object coefficients,  $\xi_k$ , are independent from angle to angle, we would intuitively expect the coarse scale object coefficients at different projection angles to actually be highly correlated with each other, and further for this correlation to decrease at finer scales. Such a correlation structure across projection angles would help us estimate at least the coarse scale object coefficients to a good accuracy even if the projection data at certain angles are missing. The current angular independence of the coefficients  $\xi_k$  corresponds to an overall covariance structure for these variables (i.e. the vector of all  $\xi_k$ 's) which is block diagonal, and it is this block diagonality that is partially responsible for the extreme efficiency of our current technique. Using the proposed, more correlated prior covariance structure would correspond in this paradigm to the addition of off-diagonal terms. At first, such a proposal would seem to make things dramatically worse from a computational perspective, since we must now contend with what corresponds to the inversion of a full rather than a block-diagonal matrix. All is not lost, however, for at least two reasons, both related to the fact that we are building our prior models directly in scale space. First, the addition of only coarse scale correlations may be sufficient to regularize a given problem, with the result that only a few, low dimensional, off diagonal elements need be added to the prior model covariance (recall, at coarser scales there are far fewer model elements – e.g. at the coarsest scale there is only one per angle). These few additional coarse scale terms could then be aggregated into a slightly larger corresponding covariance block, returning us to the block diagonal case, but with one block slightly larger than the rest. More significantly perhaps, however, is that recent research has demonstrated that certain scale-based prior models (which correspond to tree structures), though corresponding to highly correlated fields, can lead to extremely efficient scale-recursive estimation algorithms [9,13]. We examine such variations of prior covariance structure, including the combination of projections of differing quality in a future paper.

## A Summary of Notation

Table 1 summarizes the notation used in this paper.

Notation	Explanation
$N_\theta$	Number of angular projections.
$N_s$	Number of strip integrals in each angular projection.
$k, \ell$	FBP quantities are indexed by $k, \ell$ . $k$ is the angle of projection, $k = 1, \dots, N_\theta$ . $\ell$ is the strip within the angular projection, $\ell = 1, \dots, N_s$ .
$f$	The discretized object defined on a $N_s \times N_s$ square grid.
$\hat{f}$	The reconstructed object.
$y_k$	Projection data set at angle $k$ , $k = 1, \dots, N_\theta$ . $y_k = [y_k(1) y_k(2) \dots y_k(N_s)]^T$ .
$x_k$	FBP object coefficient set at angle $k$ , $k = 1, \dots, N_\theta$ . $x_k = [x_k(1) x_k(2) \dots x_k(N_s)]^T$ .
$T_k$	The discrete strip projection operator at angle $k$ , $y_k = T_k f$ .
$T_k^T$	The back-projection operator at angle $k$ , $\hat{f} = \sum_k T_k^T x_k$ .
$R$	The matrix representing the FBP ramp filtering operation, $x_k = R y_k$ .
$W$	The matrix realization of the discrete 1-D wavelet transform operation.
$D_n$	Daubechies wavelet, where $n$ indicates the support size of the corresponding filter.
$m$	The multiscale quantities are indexed in scale by $m$ ; Scales become finer with increasing $m$ .
$\eta_k$	The 1-D wavelet transform of $y_k$ , $\eta_k = W y_k$ .
$\xi_k$	The 1-D wavelet transform of $x_k$ , $\xi_k = W x_k = [(\xi_k^{(N-1)})^T \dots (\xi_k^{(0)})^T (x_k^{(0)})^T]^T$
$\xi_k^{(m)}$	Detail vector of $x_k$ at scale $m$ containing the wavelet transform coefficients of $x_k$ at scale $m$ .
$x_k^{(0)}$	Coarsest scale approximation of $x_k$ .
$T_k$	The multiscale projection operator at angle $k$ , $\eta_k = T_k f$ , $T_k = W T_k$ .
$T_k^T$	The multiscale back-projection operator at angle $k$ , $\hat{f} = \sum_k T_k^T \xi_k$ .
$\hat{f}^{(m)}$	The approximate object reconstruction at scale $m$ .
$\Delta \hat{f}^{(m)}$	The object detail reconstruction at scale $m$ , $\hat{f}^{(m+1)} = \hat{f}^{(m)} + \Delta \hat{f}^{(m)}$
$\mathcal{R}$	The multiscale filter, $\xi_k = \mathcal{R} \eta_k$ , $\mathcal{R} = W R W^T$ .
$\rho, \sigma^2, \bar{\Lambda}_\xi$	The MAP prior model parameters; $\rho$ is the decay rate of prior variance across scales. $\sigma^2$ is the overall prior model amplitude. $\bar{\Lambda}_\xi$ is the prior model DC variance.
$n_k$	The noise vector at angle $k$ , $y_k = R^{-1} x_k + n_k$ ; $n_k \sim \mathcal{N}(0, \Lambda_{n_k})$ , $\Lambda_{n_k} = \lambda_k I_{N_s}$ .
$\nu_k$	The 1-D wavelet transform of $n_k$ , $\nu_k = W n_k$ , $\nu_k \sim \mathcal{N}(0, \Lambda_{\nu_k})$ , $\Lambda_{\nu_k} = \lambda_k I_{N_s}$ ; $\eta_k = \mathcal{R}^{-1} \xi_k + \nu_k$ .
$\bar{\mathcal{R}}$	The regularized multiscale filter.
$\tilde{\mathcal{R}}$	An approximation to above obtained by ignoring off-diagonal elements in $\mathcal{R}$ .

Table 1: Notation used in this paper.

## B Details about the formation of FBP ramp-filter matrix $R$

In this work we take the matrix  $R$  to represent the practical FBP filtering operator. In the ideal case, this FBP ramp filter operation is given by:

$$F_N^{-1} H_N F_N \quad (27)$$

where  $F_N$  is a  $N \times N$  matrix representing the 1-D Fourier transform operation on a sequence of length  $N$ , and  $H_N$  is a  $N \times N$  diagonal matrix containing ideal high-pass “ramp” filter coefficients for a length  $N$  sequence. The matrix  $H_N$  has a diagonal entry of 0 since it gives zero weight to the frequency cell centered around 0. Thus the ideal ramp filter coefficient matrix  $H_N$ , and hence the matrix (27), is not invertible. In practice, however, to avoid dishing (i.e. interperiod interference) and DC artifacts, a filtering operator  $R$  is used which is constructed according to [1]:

$$R = S F_{2N}^{-1} H_{2N} F_{2N} S^T \quad (28)$$

where  $F_{2N}$  is a  $2N \times 2N$  matrix representing the 1-D Fourier transform operation on a sequence of length  $2N$ ,  $H_{2N}$  is a  $2N \times 2N$  diagonal matrix containing the corresponding ideal ramp filter coefficients, and the  $N \times 2N$  zero-padding matrix  $S$  is given by

$$S = \begin{bmatrix} 0 & I_N & 0 \end{bmatrix}. \quad (29)$$

If we define  $H$  to be the equivalent  $N \times N$  practical ramp filter coefficient matrix such that:

$$R = F_N^{-1} H F_N = S F_{2N}^{-1} H_{2N} F_{2N} S^T \quad (30)$$

then  $H$  can be seen to be given by:

$$H = F_N S F_{2N}^{-1} H_{2N} F_{2N} S^T F_N^{-1}. \quad (31)$$

One can show that  $H$  is a diagonal matrix and the diagonal elements of  $H$  are almost identical to  $H_N$  except that  $H$  gives small positive weighting to the frequency cells centered around 0 (refer to [1] for a plot of the diagonal elements of  $H$ ). Thus  $H$  has no 0 diagonal entry, resulting in an invertible  $R = F_N^{-1} H F_N$ . The only issue that remains now is of the conditioning of such a  $R$ . The above procedure for computing  $R$  results in a relatively well-conditioned matrix, with the condition number of  $R$  ranging from 24 for  $N = 16$  to 389 for  $N = 256$ . Intermediate values of  $N$  result in condition numbers between 24 and 389. In the work of this paper we use this practical, and thus invertible, filtering operator given in (30) for all calculations.

## References

- [1] A. C. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging*, IEEE Press, 1988.
- [2] K. M. Hanson, “Bayesian and Related Methods in Image Reconstruction from Incomplete Data,” in *Image Recovery Theory and Application*, Edited by H. Stark, Academic Press, 1987.
- [3] J. Llacer, “Theory of Imaging With a Very Limited Number of Projections,” *IEEE Transactions on Nuclear Science*, Vol. NS-26, No. 1, February 1979, pp. 596-602.
- [4] I. Daubechies, *Ten Lectures on Wavelets*, S.I.A.M., 1992, pp. 167-213.

- [5] N. H. Getz, "A Perfectly Invertible, Fast, and Complete Wavelet Transform for Finite Length Sequences: The Discrete Periodic Wavelet Transform," *Proceedings of the SPIE Annual Conference on Mathematical Imaging: Wavelet Applications in Signal and Image Processing*, San Diego, July 1993.
- [6] G. Beylkin, R. Coifman, and V. Rokhlin, "Fast Wavelet Transforms and Numerical Algorithms I," *Communications on Pure and Applied Mathematics*, Vol. XLIV, 1991, pp. 141-183.
- [7] H. H. Barrett, "Image Reconstruction and the Solution of Inverse Problems in Medical Imaging," in *Medical Images: Formation, Handling and Evaluation, Proceedings of the NATO Advanced Study Institute on the Formation, Handling and Evaluation of Medical Images, Portugal, 1988*, Springer Verlag, 1992, pp. 21-22.
- [8] Donner Algorithms for Reconstruction Tomography, RECLBL Library Users Manual, Lawrence Berkeley Laboratory, University of California, 1977, pp. 35-42.
- [9] K. C. Chou, A. S. Willsky, and A. Benveniste, "Multiscale Recursive Estimation, Data Fusion, and Regularization," *To appear in IEEE Transactions on Automatic Control*, March 1994.
- [10] E. B. Cargill, H. H. Barrett, R. D. Fiete, M. Ker, D. D. Patton, and G. W. Seeley, "Fractal Physiology and Nuclear Medicine Scans," *SPIE Medical Imaging II*, Vo. 914, 1988, pp. 355-361.
- [11] B. J. West, and A. L. Goldberger, "Physiology in Fractal Dimensions," *American Scientist*, Vol. 75, 1987, pp. 354-365.
- [12] C-C. Chen, J. S. Daponte, and M. D. Fox, "Fractal Texture Analysis and Classification in Medical Imaging," *IEEE Transactions on Medical Imaging*, Vol. 8, No. 2, June 1989, pp. 133-142.
- [13] M. Luetgten, W. C. Karl, and A. S. Willsky, "Efficient Multiscale Regularization with Applications to the Computation of Optical Flow," *to appear in IEEE Transactions on Image Processing*, January 1994.
- [14] G. W. Wornell, and A. V. Oppenheim, "Estimation of Fractal Signals from Noisy Measurements using Wavelets," *IEEE Transactions on Signal Processing*, Vol. 40, No. 3, March 1992, pp. 611-623.
- [15] F. Peyrin, M. Zaim, and R. Goutte, "Multiscale Reconstruction of Tomographic Images," *Proceedings of the IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis*, October 1992, pp. 219-222.
- [16] B. Sahiner, and A. E. Yagle, "Image Reconstruction From Projections Under Wavelet Constraints," *to appear in IEEE Transactions on Signal Processing*, December 1993.
- [17] B. Sahiner, and A. E. Yagle, "Limited Angle Tomography Using the Wavelet Transform," Preprint, October 1993.
- [18] J. DeStefano, and T. Olson, "Wavelet Localization of the Radon Transform in Even Dimensions," *Proceedings of the IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis*, October 1992, pp. 137-140.

- [19] C. Berenstein, and D. Walnut, "Local Inversion of the Radon Transform in Even Dimensions Using Wavelets," Center For The Applications Of Mathematics, George Mason University, Reference: CAM-21/93, January 1993.
- [20] M. Bhatia, W. C. Karl, and A. S. Willsky, "Using Natural Wavelet Bases and Multiscale Stochastic Models for Tomographic Reconstruction," Laboratory for Information and Decision Systems, MIT, Technical Report LIDS-P-2196.
- [21] M. H. Buonocore, W. R. Brody, and A. Macovski, "A Natural Pixel Decomposition for Two-Dimensional Image Reconstruction," *IEEE Transactions on Biomedical Engineering*, Vol. BME-28, No. 2, 1981, pp. 69-78.
- [22] S. G. Mallat, "A Theory of Multiresolution Signal Decomposition: The Wavelet Representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 11, No. 7, 1989, pp. 674-693.
- [23] P. A. Rattey, and A. G. Lindgren, "Sampling the 2-D Radon Transform," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. ASSP-29, No. 5, 1981, pp. 994-1002.
- [24] A. M. Cormac, "Sampling the Radon Transform with Beams of Finite Width," *Phys. Med. Biol.*, Vol. 23, No. 6, 1978, pp. 1141-1148.
- [25] P. J. Green, "Bayesian Reconstructions from Emission Tomography Data Using a Modified EM Algorithm," *IEEE Transactions on Medical Imaging*, Vol. 9, No. 1, 1990, pp. 84-93.
- [26] K. Sauer, and C. Bouman, "Bayesian Estimation of Transmission Tomograms Using Segmentation Based Optimization," to appear in *IEEE Transactions on Nuclear Science*.
- [27] G. Strang, "Wavelets and Dilation Equations: A Brief Introduction," *SIAM Review*, Vol. 31, No. 4, 1989, pp. 614-627.
- [28] I. Daubechies, "Wavelet on the interval," *Progress in Wavelet Analysis and Applications*, Editors Y. Meyer and S. Roques, Frontieres, 1992, pp. 95-107.
- [29] H. L. Van Trees, *Detection, Estimation, and Modulation Theory*, John Wiley & Sons, 1968, pp. 54-63.