TOTAL VARIATION AND WAVELET REGULARIZATION METHODS IN EMISSION TOMOGRAPHY

Pavel Kisilev, Michael Zibulevsky, and Yehoshua Y. Zeevi

Abstract **– A classical technique for reconstruction of Emission Tomography (ET) images from measured projections is based on the maxim likelihood (ML) estimation achieved with the Expectation Maximization (EM) algorithm. We incorporate the wavelet transform (WT) and total variation (TV) based penalties into the ML framework, and compare performance of the EM algorithm and the recently proposed conjugate barrier (CB) algorithm. Using the WT- and TV-based penalties allows one to embed regularization procedures into the iterative process. In the case of the WT-based penalty, we impose a subset of wavelet coefficients with a desired resolution on the objective function. It appears that the CB algorithm outperforms substantially the EM algorithm in penalized reconstruction. Properties of the optimization algorithms along with WT- and TV-based regularization are demonstrated on image reconstructions of a synthetic brain phantom, and the quality of reconstruction is compared with standard methods.**

Keywords **medical imaging, positron emission tomography, expectation maximization, wavelets, total variation**

1. INTRODUCTION

ET is a medical imaging technique that enables one to quantify a distribution of radioactivity within the body, and, as such, it is useful in detection and identification of pathological tissue. In this technique, radioactive tracers, injected into the body of a patient, emit photons, which are detected in distinct detector pairs, or bins. By counting the number of photons detected in the various bins, one measures the projection of the tracer distribution at different angles. A classical technique for the reconstruction of 2D and 3D ET images from measured projections is based on the maximum likelihood (ML) framework [1]. Utilizing particular properties of the Poisson process leads to the Expectation Maximization (EM) algorithm for ET reconstruction [2]. This algorithm provides reliable reconstruction results with high resolution. Alternatively, the ML reconstruction can be performed by the recently proposed conjugate barrier (CB) algorithm, which has several advantages over the EM. The drawback of using both algorithms for minimization of the likelihood function is that they do not exploit the localized nature of images and their multiresolution structure. As a result, the reconstruction achieved by both algorithms suffers from lack of smoothness, and, at the same time, from lack of sharpness of edges.

Furthermore, in practice, it is desirable to carry out reconstruction on low statistics (i.e. noisy data). Under these circumstances, the maximum likelihood estimate at highest resolution contains high frequency noise even though the original image is known to be relatively smooth. Therefore, reconstruction of images from their projections requires some kind of regularization that usually represents a trade-off between accuracy and resolution. For example, image reconstruction using so-called blobs functions was discussed in [3] and [4]. It allows one to control the smoothness properties of the reconstructed image on the expense of resolution. This is achieved by choosing appropriately the value of the smoothness parameter of the blobs functions.

Ideally, a lower resolution reconstruction should be applied to regions with no edges. On the other hand, keeping higher resolution components preserves local features in the reconstructed image. This provides desired regularization, so that a trade-off between increasing resolution and noise suppression is achieved. Such a strategy requires some prior knowledge of edges' location. This knowledge can be obtained from another imaging procedure, e.g. the one based on the X-Ray computerized tomography (CT). In [10], the authors assumed that this prior information is available, and used this information to build a penalty template in the wavelet domain.

In practice, such prior knowledge is rarely available. It was shown recently that the TV method appears to be one of the most successful regularization approaches to ill-posed problems. For example, this method was applied for blind deconvolution of images [11]. In [12], the TV penalty was applied to image reconstruction from tomographic data. The TV penalty represents kind of a weak

prior about the object structure. In particular, it assumes that the underlying image contains edges, which is usually the case for medical images.

We utilize the wavelet transform (WT) and the Total Variation (TV) functional in our penalties, and show that they affect reconstruction in similar ways. The desired regularization is accomplished in a natural way; using either the WT- or the TV-based penalties allows one to embed regularisation procedures into the iterative process. This task is accomplished by either 1) penalizing for the lack of sparsity of the gradient of the reconstructed image, in the case of the TV-based penalty, or 2) imposing a new set of parameters on a subset of wavelet coefficients corresponding to desired resolutions, and penalizing for the lack of sparsity only this subset, in the case of the WT-based penalty.

It turns out that the penalized CB algorithm achieves the best trade-off between accuracy and resolution. In particular, it keeps improving the contrast while lowering the noise level with iterations. To our knowledge, such a result was not achieved with any of the existing ET reconstruction techniques. The WT-based penalty produces more natural reconstructed images at the earlier iterations, than the TV-based penalty , while the last provides the best trade-off between accuracy and resolution.

We first describe the basics of the positron ET (PET) image reconstruction technique based on the ML framework, and describe EM and CB algorithms. Next, we consider some computational issues related to the penalty calculations. Finally, we discuss the advantages of the proposed methods and illustrate their potential by experimental study.

2. ML RECONSTRUCTION OF PET IMAGES

EM algorithm

We first summarize some of the main results of L. Shepp and Y. Vardy [2] who pioneered the ML image reconstruction in PET by application of the EM algorithm.

Let the total number of photons detected in each bin be $y(b)$, $b = 1,\ldots,B$. Let the body be divided into voxels (or pixels), and the number of photons generated independently within each voxel be $n(v)$, $v = 1,\ldots,V$. Generation of photons in each voxel is described by the Poisson process, characterized by the expected value of photons $\lambda(v)$:

$$f(n|\lambda(v)) = P(n(v) = n|\lambda(v)) = e^{-\lambda(v)}\frac{\lambda(v)^n}{n!} \tag{1}$$

The values $\lambda(v)$ depend on the tracer distribution, which, in turn, depend on the tissue structure in the body. In addition, variables $y(b)$ are independent and Poissonian with expected values $\lambda(b)$:

$$f(y|\lambda(b)) = P(y(b) = y|\lambda(b)) = e^{-\lambda(v)}\frac{\lambda(b)^y}{y!} \tag{2}$$

Let $p(v,b)$ denote the probability of the event that a photon emitted from voxel v is detected in bin b, forming a matrix with VxB entries. The probability matrix values depend on various physical factors such as scanner geometry, detector efficiency, and the composition of the body being scanned. The issue of computing $p(v,b)$ will be discussed later.

From (2) the log-likelihood function for the measurements $y(b)$ is given by

$$L(\lambda) = \sum_v \lambda(v)p(v) - \sum_b y(b)\log\sum_v \lambda(v)p(v,b), \tag{3}$$

where $\boldsymbol{\lambda} = \{\lambda(v),\ v = 1,\ldots,V\}$ is the set of unknown parameters, and

$$p(v) = \sum p(v,b) \tag{4}$$

is the so-called sensitivity image (i.e. the probability that an emission from v is detected).

To solve (3), the EM algorithm was applied to the PET reconstruction problem leading to the following formula

$$\lambda^{(k+1)}(v) = \frac{\lambda^{(k)}(v)}{p(v)}\sum_{b=1}^{B}\frac{y(b)p(v,b)}{\sum_{v'=1}^{V}\lambda^{(k)}(v')p(v',b)} \qquad \forall v \tag{5}$$

for iteratively approximating a maximizer of $L(\boldsymbol{\lambda})$.

Conjugate Barrier (CB) algorithm

The CB algorithm, recently proposed by Ben-Tal and Nemirovski [5], belongs to the general class of the Gradient Descent algorithms.

Let the function h be defined as:

$$h(\lambda) = \begin{cases} \| \lambda \|_p, & \lambda \in \Delta \\ +\infty, & \lambda \notin \Delta, \end{cases}$$

where $\|\cdot\|_p$ denotes the l_p norm, and Δ is the domain of valid values of λ. In our setting, wherein λ are the values of normalized image intensity

$$\Delta \equiv \{\lambda \in R^n \mid \lambda \geq 0, \sum_1^V \lambda(v) = \rho\},$$

where ρ is the total number of detector counts.

Let h^* be the so-called conjugate function of h:

$$h^*(\xi) = \sup_{\lambda \in \Delta}[\xi^T \lambda - h(\lambda)], \tag{6}$$

where ξ is called the conjugate image, so that $\xi, \lambda \in R^V$. The initial value ξ^0 can be initialized arbitrary (we use a matrix of ones).

The following two iterative steps summarize the CB algorithm:

$$\begin{aligned} &\text{step 1:} && \lambda^{n+1} = \nabla h_*(\xi^{n+1}) \\ &\text{step 2:} && \xi^{n+1} = \xi^n - \gamma^n \nabla L(\lambda^n), \\ &\text{where} && \gamma^n = \frac{\gamma^0}{\sqrt{n}} \|\nabla L(\lambda^n)\|_\infty \end{aligned} \tag{7}$$

where $\nabla L(\lambda^n)$ is the gradient of the Log-likelihood function at the n-th iteration, γ^n is a positive step size at the n-th iteration, γ^0 is a small positive constant and values of λ^0 are initialized with the value $1/V$.

The value of the conjugate function h^* at a given point ξ is found according to (6) by solving the following optimization problem

$$\min_{x \in \Delta}[\|\lambda\|_p - \xi^T \lambda \,|],$$

so that $h^*(\xi)$ is given as the optimal value of this problem, and $\nabla h^*(\xi)$ as the optimal solution. (For more details see [5] and [6]).

The CB algorithm has several advantages over the EM [7]: 1) its ordered sets version always converges 2) the rate of convergence is known and independent of the dimension of the problem, 3)

the error bound is of order $O\left(\sqrt{\log V / K}\right)$, where K is the number of iterations. In addition, its computational cost is comparable to the EM.

3. TOTAL VARIATION AND WAVELET PENALTIES

As was mentioned before, natural PET data are usually very noisy due to a short acquisition time and various scatter effects. Exact minimization of the log-likelihood function of such noisy data leads to a very noisy reconstructed image. In such cases, various penalty functions reflecting smoothness of the noise-free image or other prior information are used in order to improve quality of reconstruction. In this case, penalized log-likelihood function takes the form

$$\begin{aligned} L_p(\lambda) &= L(\lambda) + \mu \mathrm{H}(\lambda) \\ &= \sum_v \lambda(v) p(v) - \sum_b y(b) \log \sum_v \lambda(v) p(v,b) + \mu \mathrm{H}(\lambda), \end{aligned} \tag{8}$$

where $\mathbf{H}(\lambda)$ is a penalty function and μ is its weight parameter, which can be chosen based on the estimated signal to noise ratio (see, for example [18]). (In this paper, we choose it experimentally).
The gradient of the penalized log-likelihood function is:

$$\nabla L_p(\lambda(v)) = p(v) - \sum_{b=1}^{B} \frac{y(b) p(v,b)}{\sum_{v'=1}^{V} \lambda^{(k)}(v') p(v',b)} + \mu \nabla \mathrm{H}(\lambda) \tag{9}$$

In the case of the CB algorithm, the update of the conjugate image in the second step in (7) is performed according to the gradient of the penalized log-likelihood function (above).
The corresponding iterative formula for the penalized EM algorithm is

$$\lambda^{(k+1)}(v) = \frac{\lambda^{(k)}(v)}{p(v) + \mu \nabla \mathrm{H}(\lambda)} \sum_{b=1}^{B} \frac{y(b) p(v,b)}{\sum_{v'=1}^{V} \lambda^{(k)}(v') p(v',b)}, \quad \forall v. \tag{10}$$

Several penalty functions were proposed in the literature (see for example [8], [9]). These are usually applied in the original image domain. In this work we propose a new method that utilizes penalty function, defined in the wavelet domain. Our penalty function is intimately related to the Total Variation method.

Total variation (TV) penalty

In our study, we use the following formula for the TV penalty:

$$\mathrm{T}(\lambda) = \sum \left\| \nabla \lambda(v) \right\|_2 . \tag{11}$$

Since the expression for the TV penalty (above) is non-differentiable in locations where $\left\| \nabla \lambda(v) \right\|_2 = 0$, we replace it with the approximation of the norm of gradient for a 2D image $\lambda(v)$:

$$\mathrm{H}_{TV}(\lambda(v)) = \sqrt{\lambda_x^2(v) + \lambda_y^2(v) + \eta},$$

where $\lambda_x(v)$ and $\lambda_y(v)$ are derivatives in directions x and y, wherein the index v in a 2D case is a pair (i,j), i.e. pixel's coordinates, and η is the parameter which controls the smoothness of the penalty. The above approximation is crucial for smooth optimization methods[1].

In order to minimize penalized log-likelihood function we need a gradient of the TV penalty term.

We use a simple approximation of the derivatives λ_x and λ_y:

$$\begin{aligned} \lambda_x(i,j) &= \lambda(i+1,j) - \lambda(i,j) \\ \lambda_y(i,j) &= \lambda(i,j+1) - \lambda(i,j) \end{aligned} \tag{12}$$

The partial derivatives of the penalty function with respect to λ_x and λ_y are:

$$\begin{aligned} \frac{\partial \mathrm{H}_{TV}}{\partial \lambda_x(i,j)} &= \frac{1}{\sqrt{\lambda_x^2(i,j) + \lambda_y^2(i,j) + \eta}} \lambda_x(i,j) \\ \frac{\partial \mathrm{H}_{TV}}{\partial \lambda_y(i,j)} &= \frac{1}{\sqrt{\lambda_x^2(i,j) + \lambda_y^2(i,j) + \eta}} \lambda_y(i,j) \end{aligned} \tag{13}$$

By carefully analysing the contributions of the neighbouring pixels with coordinates (i,j), $(i+1,j)$ and $(i,j+1)$ from (12) into the term $\nabla_{\lambda(i,j)} \mathrm{H}_{TV}$, we arrive at the following formula:

$$\nabla_{\lambda(i,j)} \mathrm{H}_{TV} = -\left(\frac{\partial \mathrm{H}_{TV}}{\partial \lambda_x(i,j)} - \frac{\partial \mathrm{H}_{TV}}{\partial \lambda_x(i-1,j)} \right) - \left(\frac{\partial \mathrm{H}_{TV}}{\partial \lambda_y(i,j)} - \frac{\partial \mathrm{H}_{TV}}{\partial \lambda_y(i,j-1)} \right). \tag{14}$$

Wavelet-based penalty

Let $\{\varphi_m\}$, $m = 1,\ldots,M$, be an orthonormal wavelet basis of an M-dimensional vector space, and let the discrete image $\lambda(v)$, belonging to this vector space, be represented by this basis as follows:

$$\lambda(v) = \sum_{m=1} a_m \varphi_m(v) \tag{15}$$

where

$$a_m = \sum_{v=1}^{V} \lambda(v) \varphi_m(v), \tag{16}$$

Or, in matrix form

$$\lambda = \Phi \mathbf{a},$$
$$\mathbf{a} = \Phi^T \lambda.$$

[1] The CB algorithm is designed for non-smooth optimization, and it can work even for η=0.

(For more detailed discussion of the wavelet transform, see the Appendix).

Suppose that the wavelet coefficients of the image are known *apriori* to be sparse, i.e. only a small (unknown) part of them significantly differs from zero. Under this assumption, the sum of the absolute values of the coefficients represents a natural penalty, which forces the coefficients of the reconstructed image to become sparse. Such a penalty is very popular in denoising and compression problems (see for example [13]).

Suppose further that only a subset of the coefficients is known to be sparse. For example, for 2D images, the detail coefficients usually are sparse, while approximation ones are quite 'dense'. In this case, it is reasonable to penalize only detail coefficients. Moreover, sometimes it is reasonable to penalize only a subset of the detail coefficients at high resolutions. Note, that the Wavelet Packet (WP) decomposition can be used instead of the WT, since the former provide a much richer decomposition structure. In this case, the choice of the subset of coefficients that is most appropriate to penalize, can be based on some apriori information (see, for example, [14], in the context of the blind source separation problem).

Let the matrix $\boldsymbol{a}$ of the wavelet transform coefficients of the image be partitioned as $\boldsymbol{a} = [\boldsymbol{b}\ \boldsymbol{c}]$, where all the coefficients in $\boldsymbol{c}$ are known *apriori* to be close to zero. This assumption represents sort of prior knowledge about the emission means λ, e.g. knowledge about a particular organ's metabolism, which defines in turn smoothness, presence of edges and other characteristics of the image λ. Therefore, such a parametric model based on the WT can be easily incorporated into the Bayesian reconstruction approach.

In this case, the ML reconstruction problem is formulated as:

$$\begin{cases} \max\limits_{\lambda}\{L(\lambda)\} \\ s.t.:\ (1)\,\lambda \geq 0 \\ \qquad (2)\,\boldsymbol{c}(\lambda) = 0 \end{cases}$$

In practice, we replace the above formulation with

$$\begin{cases} \min\limits_{\lambda}\{-L(\lambda) + \mu\left\|\boldsymbol{c}(\lambda)\right\|_{p}\} \\ \text{s.t.:}\ \lambda \geq 0 \end{cases}$$

In [10], the authors used the l_2 norm that is more appropriate in cases wherein some prior knowledge of the edges' location is available, and it is possible to build template-based penalties. In this work we do not require any such prior knowledge, but suppose that the image λ has a sparse representation in wavelet dictionary. In this case, the l_1 norm of the wavelet coefficients is a more appropriate choice than the l_2 norm, since it forces small coefficients to become zero, and does not penalize large coefficients as strong as the quadratic penalty does.

Since edges 'live' at high resolutions, in our experiments we choose subset $\boldsymbol{c}$ to be constructed from the detail coefficients of the WT at the highest resolutions. Thereafter, our penalty is defined as

$$H_{WT} = \sum_{s \in S} |c_s|, \tag{17}$$

where S is the subset of the coefficients at the desired resolutions so that $c = \{c_s\}_{s \in S}$. Again, we approximate the modulus function by $|c| = \sqrt{c^2 + \zeta}$ with regularization parameter ζ.

Taking into account that $c = \Phi_S \lambda$, where matrix Φ_S is constructed from the wavelet basis vectors indexed on S, the gradient of the penalty can be written as

$$\nabla_\lambda H_{WT} = \Phi_S^* \tilde{c},$$

where Φ_S^* is the adjoint of Φ_S and vector $\tilde{c} = \left[\frac{c_s}{\sqrt{c_s^2 + \zeta}} \right]_{s \in S}$.

On the relation of the TV and the wavelet penalties.

In this section, we explain why a wavelet-based penalty can be used as a penalty for the sparseness of edges of a 2D function, and, as so, can mimic the TV-based penalty.

Suppose that the intensity of the original image has jumps, or, in other words, sharp edges. It is known that the Haar wavelet basis provides good representation of such images. Let us take a look at the Haar wavelet coefficients at different resolutions. Wavelet detail coefficients in the x-direction and at the finest resolution j are obtained by calculating the scalar product of the image with the translated versions of the wavelet:

$$\varphi(j,x,y) = \begin{cases} 0.5, & x = 0,\ y = [0,1] \\ -0.5, & x = 1,\ y = [0,1] \\ 0, & otherwise \end{cases}$$

A shifted version of this wavelet is shown in Figure 1, upper left plot.

Similarly, the y-direction detail coefficients as well as the diagonal detail coefficients are obtained by calculating the scalar products with the corresponding wavelets (Figure 1, upper middle and right plots). This operation produces essentially a finite *differentiation* of the image in the corresponding direction. Therefore, at the finest resolution, the number of the corresponding detail coefficients which differ from zero correspond roughly to the number of jumps of the image in x,y and diagonal directions. For natural images, edges appear only in a few locations, therefore there should be only a few non-zero coefficients at the finer resolutions. Otherwise, if we observe many non-zero coefficients, it indicates the presence of noise.

Proceeding to the next, coarser resolution level j-1, the corresponding wavelet is

$$\varphi(j-1,x,y) = \begin{cases} 0.25, & x = [0,1],\ y = [0,3] \\ -0.25, & x = [1,2],\ y = [0,3] \\ 0, & otherwise \end{cases}$$

This wavelet and its two complementary brothers are shown in the lower row of Figure 1.

At this resolution, the number of non-zero detail coefficients still corresponds to the number of jumps, but the total number of coefficients at this level is 4 times smaller, and so is the sparsity. If we further proceed to coarser resolutions, we will encounter levels where the support of wavelet functions is comparable to the 'flat' parts of the image, i.e. parts between jumps. In this case, most of the coefficients are expected to be non-zero, and, therefore, sparsity will fade away. This is in particular noticeable for the approximation (i.e. low-pass) coefficients. That is why there is no reason to include wavelet coefficients at the coarser resolutions into the penalty.

In the case of the Haar wavelet basis, calculation of scalar products of the image with basis functions at the highest resolution is essentially equivalent to the calculation of gradients of the image. Therefore, the TV-based penalty can be considered as kind a WT penalty, in a particular case wherein the Haar coefficients at the highest resolution are penalized. Both, the sum of absolute values of corresponding subset of wavelet coefficients and the sum of absolute values of gradients, represent a measure of sparseness. (Note, that in this work we were in particular interested in improved reconstruction of edges of a 2D function. If our 'feature of interest' was a texture (instead of edges), we would penalize a different subset of coefficients).

4. EXPERIMENTAL RESULTS

We carried out tests with the Shepp-Logan phantom; a model used in tomography for evaluating properties of reconstruction algorithms. The phantom was discretized into a 128x128 image. We slightly modified it by adding a hot spot (Figure 2), which we used for calculation of the contrast and noise suppression properties of reconstruction algorithms. Projection data were simulated as follows: we applied the radon transform to the phantom, using 60 angular and 185 radial samples of projections. These projection data were used as a mean rate of a Poisson process. Random samples of projection data were generated according to the above Poisson process, arriving at overall $1.2e^5$ detector counts.

Criteria for comparison

There are two parameters which are useful in the contrast calculation: coefficient of variation (CV) and contrast recovery (CR). The CV is defined as the ratio of the standard deviation to the mean-value of the image over some region of interest (ROI). CR for cold lesions in a hot background is calculated as

$$CR_{cold} = 1 - C / H,$$

where C and H are means taken over cold and hot ROI, respectively. For hot lesions in a cold background we compute:

$$CR_{hot} = \frac{H / C - 1}{H_{true} / C_{true} - 1},$$

where H_{true} / C_{true} is the real ratio of hot lesion to the background in the phantom.
We use the above parameters in order to determine which algorithm provides the best trade-off between contrast and noise control.

Discussion

To illustrate the contrast improvement and noise suppression properties of plain, TV- and WT-based penalized versions of EM and CB algorithms (we will refer to them as EMTV, EMWT, CBTV, CBWT, respectively), we compare their plots of CV versus CR with those characteristics of plain EM and CB algorithms (Figure 3). In this figure, for each optimization method there is a curve whose each point corresponds to the CV versus CR pair at a particular iteration. Each algorithm was iterated 50 times, with the penalty parameter $\mu = 0.005$. Generally speaking, when comparing two such curves, the lower curve achieves a better contrast-to-noise-intensity trade-off. (Note, that the curves are monotonic, besides the first few iterations, and the rightmost point on a curve corresponds to the last iteration). It is clear that, when comparing plain versions of the algorithms, the CB has no advantages over the EM. Moreover, the EM algorithm achieves the same contrast as the CB at the earlier iteration. In contrast, when penalties are applied, the CB outperforms the EM by achieving a much better contrast-to-noise-intensity trade-off (compare, for example, Figure 9 and Figure 10), although the CB versions achieve the same contrast as the corresponding EM versions at a later iteration. TV-based penalized versions of EM and CB algorithms outperform slightly the WT-based ones.
The remarkable result is that, in the case of the CBTV algorithm, the curve is monotonically decreasing. This means that both, noise suppression and contrast, are improved with iterations. This phenomenon is depicted also in Figure 6, which shows that the CR curve is increasing for each algorithms, while the CV curve is decreasing only in the case of the CBVT algorithm.
In Figure 4 we compare the contrast-to-noise-intensity trade-off in the case of TV- and WT-based penalties, obtained with various values of penalty parameter. Clearly, a better contrast-to-noise-suppression trade-off is achieved when a larger penalty is used, although this slows down slightly the contrast increment.
In order to verify the performance of the algorithms at late stages of optimization process we run 1000 iterations (Figure 5). The CBTV algorithm provides the best and consistent improvement of the contrast to noise intensity trade-off.
Examples of images reconstructed by various combinations of algorithms and penalties are depicted in Figure 7 - Figure 13. It is easy to see, that, although the TV-based penalty provides a better contrast to noise trade-off than the WT-based one, the image reconstructed after 50 iterations looks more natural and artefacts-free for the WT-based penalty (compare for example Figure 10 and Figure 12). On the other hand, after 1000 iterations the CBWT algorithm reveals artefacts and contrast-to-noise deterioration (see curves on Figure 5, and compare Figure 13 and Figure 14). This result can be

explained by the fact that the WT-based penalty misses some of the edges because of the downsampling operation performed in the dyadic WT. As a result, the smallest shift at the finest resolution is 2 pixels, and, therefore, those missed locations were not penalized and prevailed at the later stages of optimization. Note also, that the reconstruction obtained by the CBTV suffers slightly from the stepped artefacts around diagonal edges. The reason is that diagonal edges are penalized about $\sqrt{2}$ times stronger than horizontal and vertical ones, as a result of using the rectangular grid (using, for example, the hexagonal grid could reduce this factor). This phenomenon can be addressed in more details, but this topic is out of scope of this paper.

5. CONLUSIONS

Numerical results and comparisons, concerning convergence properties and the quality of reconstruction of the proposed WT- and TV-based penalized algorithms versus plain algorithms indicate that using either WT or TV penalty significantly improves the contrast-to-noise trade-off. Penalties are in particular useful when they are applied to the CB algorithm, which outperforms the EM for all kinds of penalties and parameters.

The combination of the CB algorithm with the TV penalty achieves the best contrast to noise trade-off, and, most importantly, the CBTV algorithm improves the contrast *and* suppresses noise at the same time, monotonically with increasing number of iterations.

At the earlier iterations (up to about 70), the CBWT algorithm produce artefacts-free reconstructed images which look more natural than the ones obtained with the CBTV. Nevertheless, the performance of the CBWT algorithm deteriorates after some 100 iterations, as a result of a downsampling procedure (see above). Our current research is concentrated on improving the performance of the WT-based algorithm, which, we feel, has more flexibility than the TV-based one.

6. REFERENCES

[1] A. Dempster, M. Laird, and D. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm. *Jour. Royal Stat. Soc., 39, no 1, 1-38.*

[2] L. Shepp, Y. Vardi, (1982). Maximum likelihood reconstruction for emission tomography. *IEEE Med. Imaging, 1, 113-122.*

[3] R.M. Lewitt (1992). Alternatives to voxels for image representation in iterative reconstruction algorithms. *Phys. Med. Biol., Vol. 37, No. 3, 705-716.*

[4] F. Jacobs, et al. (1999) A comparative study of 2D reconstruction algorithms using pixels and optimized blobs applied to Fourier rebinned 3D data. *International meeting on fully 3D reconstruction in radiology and nuclear medicine, 43-46.*

[5] Ben-Tal, A. and Nemirovski, A. (1999). "The conjugate barrier method for non-smooth convex optimization", TR #5/99, Oct. 1999, Minerva Optimization Center, Technion.

[6] Margalit, T, et al. (1999). "The ordered sets mirror descent and conjugate barrier optimization algorithms adapted to the 3D PET reconstruction problem", TR #7/99, Nov. 1999, Minerva Optimization Center, Technion.

[7] Nemirovsky, A. Personal communications.

[8] J. A. Fessler, N. H. Clinthorne, W. L. Rogers. "Regularized emission image reconstruction using imperfect side information". Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf., vol. 3, pp. 1991-1995, 1991.

[9] D. F. Yu, J. A. Fessler. "Three-dimensional non-local edge-preserving regularization for PET transmission reconstruction". Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf., 2000. To appear. (oral)

[10] P. Kisilev, M. Jacobson, Y.Y. Zeevi. Utilizing Wavelet Transform for ML Reconstruction in Positron Emission Tomography. CCIT Report #522, Jan. 2001.

[11] T. F. Chan, C.K. Wong. Total Variation Blind Deconvolution, *IEEE Transactions on Image Processing, to appear.*

[12] E. Jonsson, S. Huang and T. Chan, Total Variation Regularization in Positron Emission Tomography, TR 9848, November 1998.

[13] S. Chen, D. Donoho, and M. Saunders. Atomic decomposition by basis pursuit. 1996. http:/www-stat.stanford.edu/~donoho/Reports/.

[14] P.Kisilev, M. Zibulevsky, and Y.Y. Zeevi, B. Pearlmutter. Multiresolution Framework for Blind Source Separation. CCIT Report #317, June 2000.

[15] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, 1998.

[16] D. Weitzer, D. Stanhill, and Y.Y. Zeevi, Non-separable two-dimensional multiwavelet transform for image coding and compression, Proc. SPIE, 3309 (1997), pp. 944-954.

[17] R. Coifman, Y. Meyer, and M. Wickerhauser, *Wavelet analysis and signal processing*, in Wavelets and their applications, Jones and Barlett. Ruskai B. et al. Eds., Boston, 1992.

[18] H.W. Engl, M. Hanke, and A. Neubauer, *Regularization of inverse problems,* Kluwer, Dorbrecht 1996.

Acknowledgement

This research has been supported by the Ollendorff Minerva Center and by the Technion V.P.R. Fund for the promotion of sponsored research – The Goldberg Fund (050-011).

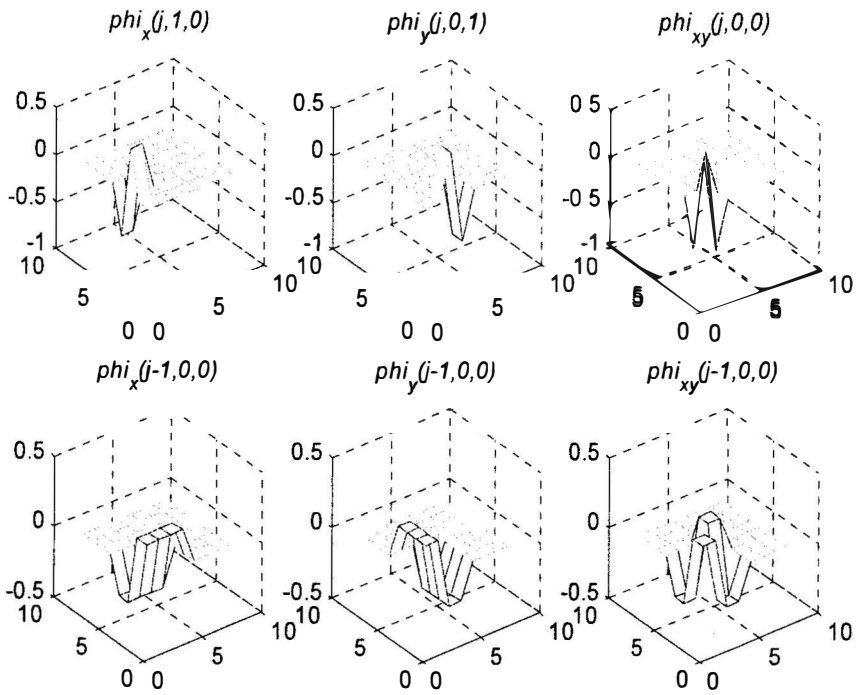

Figure 1. Haar wavelet basis functions at two consecutive levels of resolution

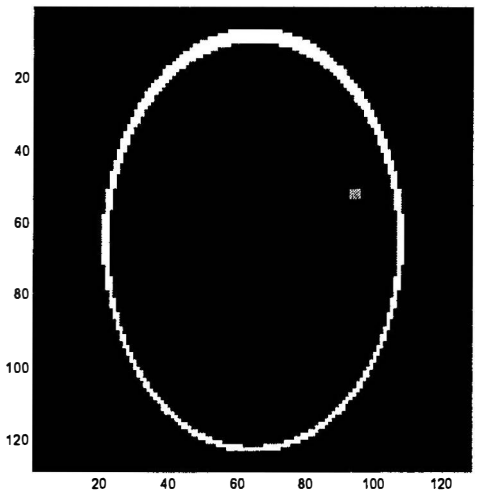

Figure 2. Modified Shepp-Logan phantom

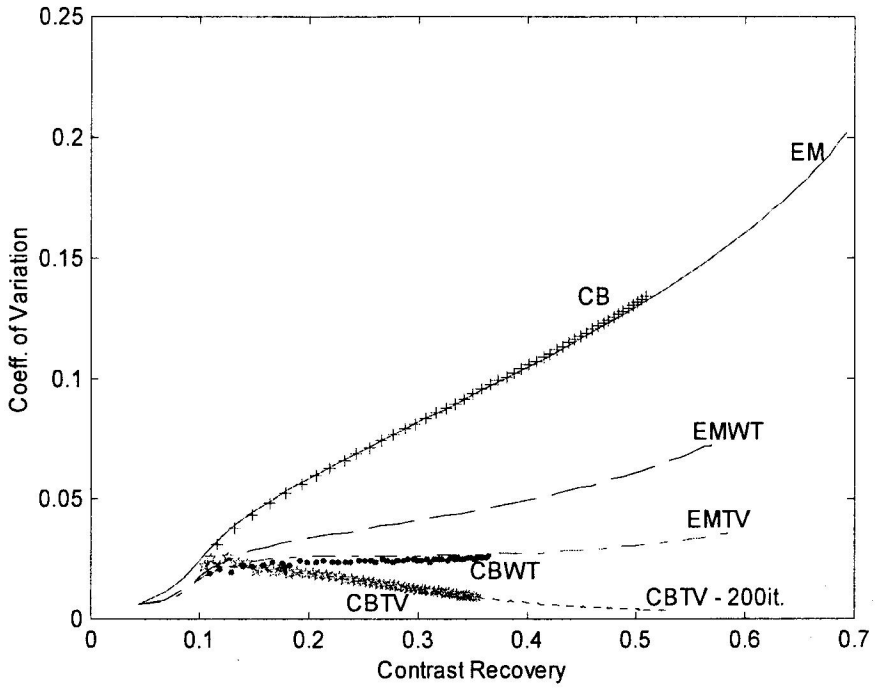

Figure 3. Coefficient of Variation versus Contrast Recovery; 50 iterations.

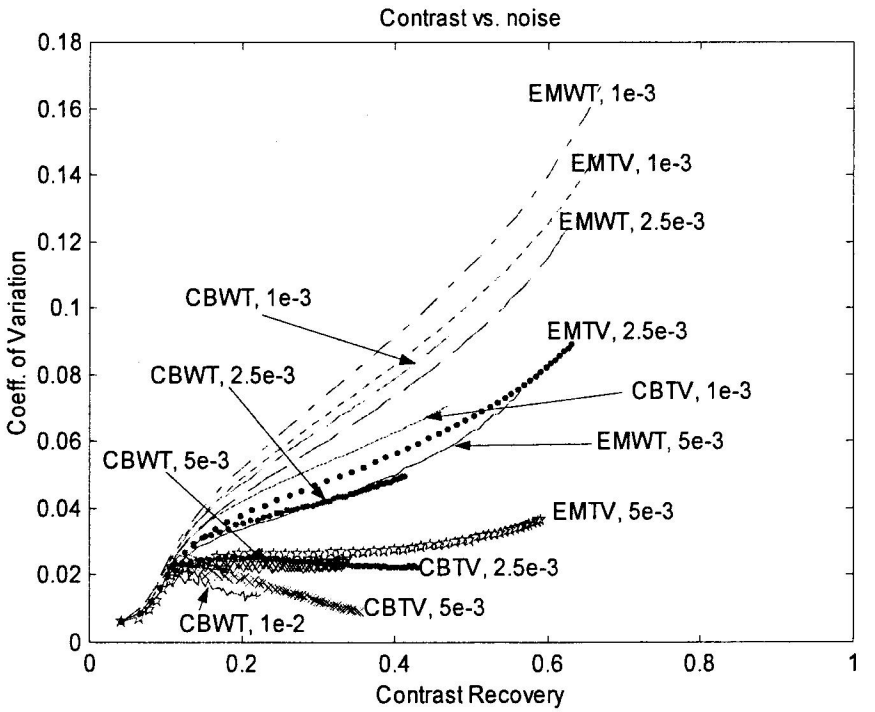

Figure 4. Contrast-to-noise-intensity trade-off for TV and wavelet penalties with various values of penalty parameter μ; 50 itérations.

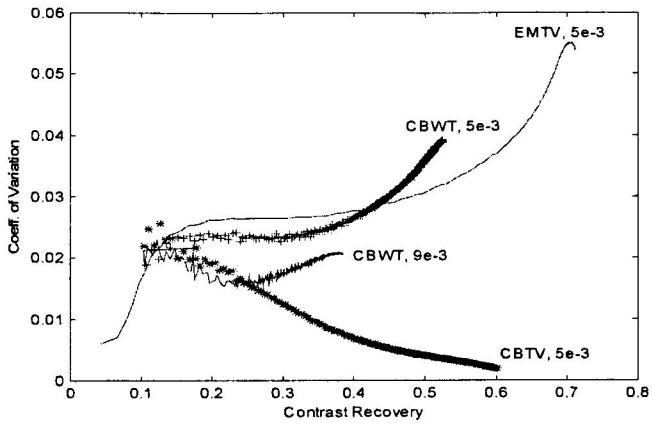

Figure 5. Coefficient of Variation versus Contrast Recovery; 1000 iterations.

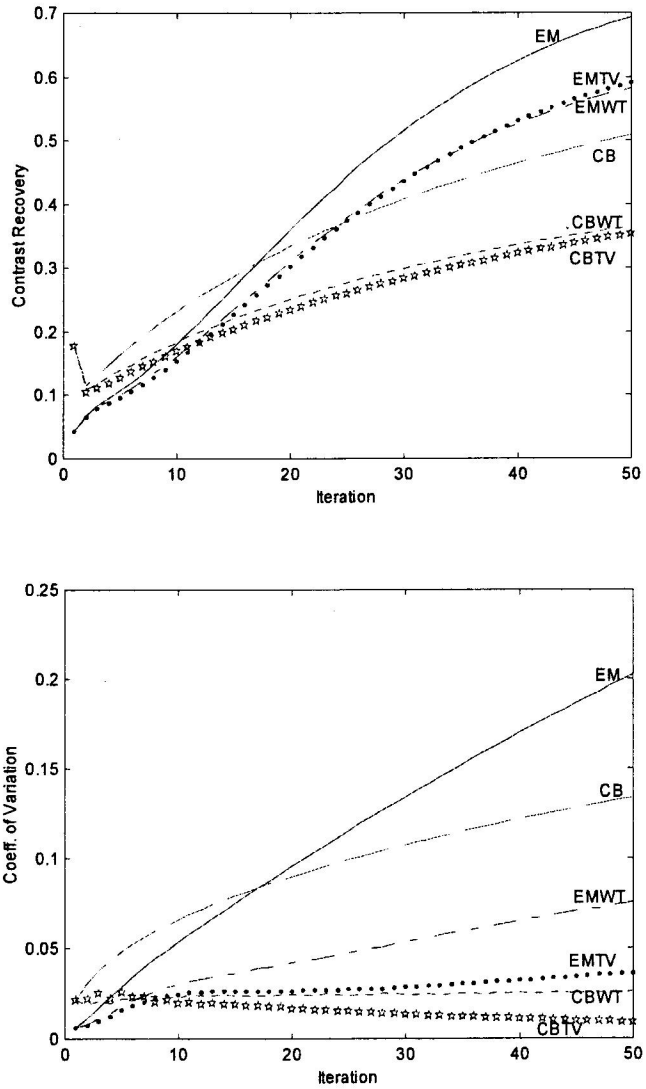

Figure 6. Contrast Recovery (upper) and Coefficient of Variation (lower) vs. iteration

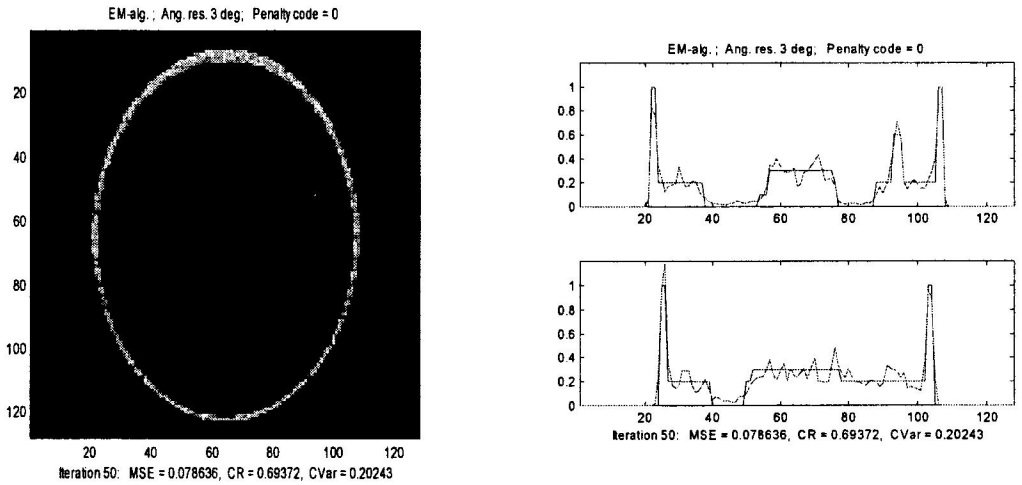

Figure 7. Reconstruction with the EM algorithm: image (left) and its two slices (right)

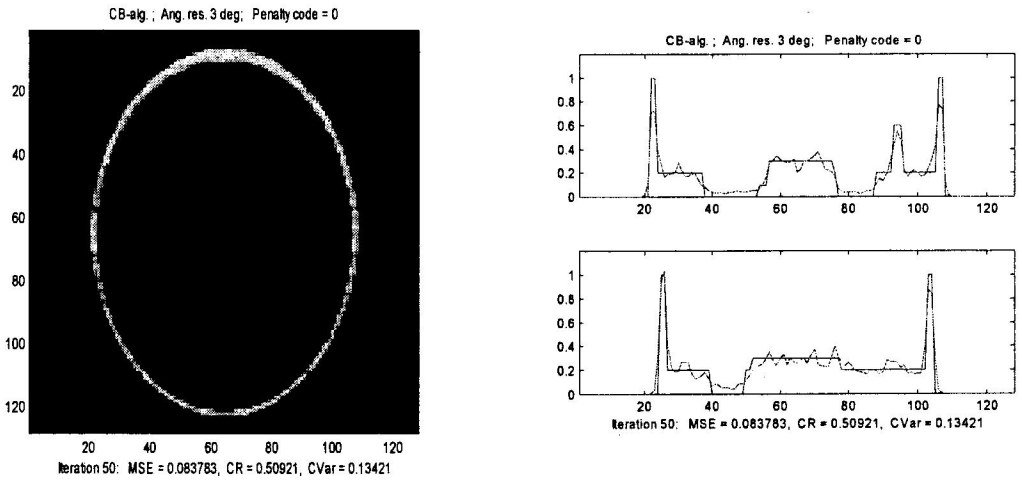

Figure 8. Reconstruction with the CB algorithm: image (left) and its two slices (right)

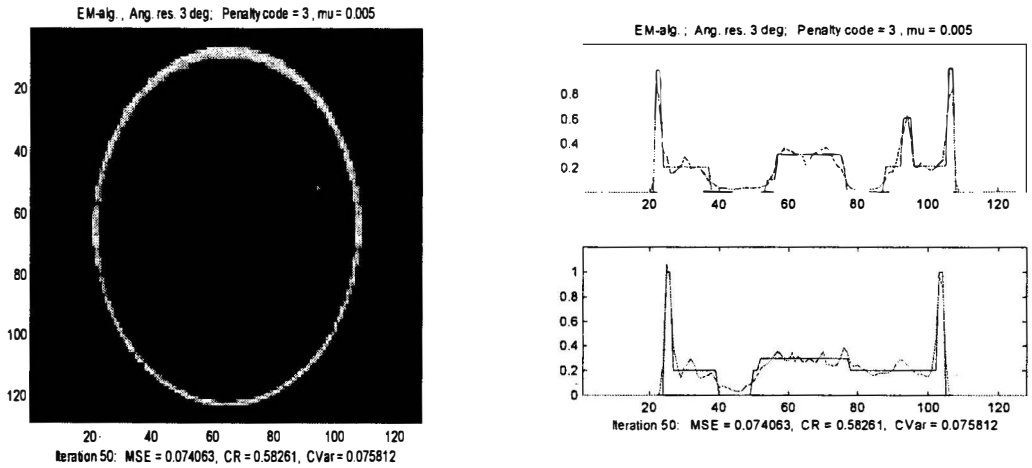

Figure 9. Reconstruction with the EMWT algorithm: image (left) and its two slices (right)

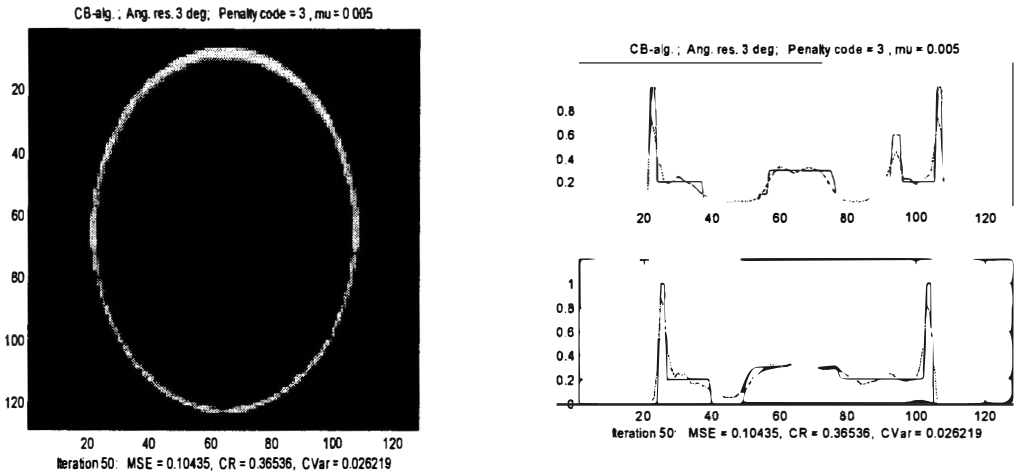

Figure 10. Reconstruction with the CBWT algorithm: image (left) and its two slices (right)

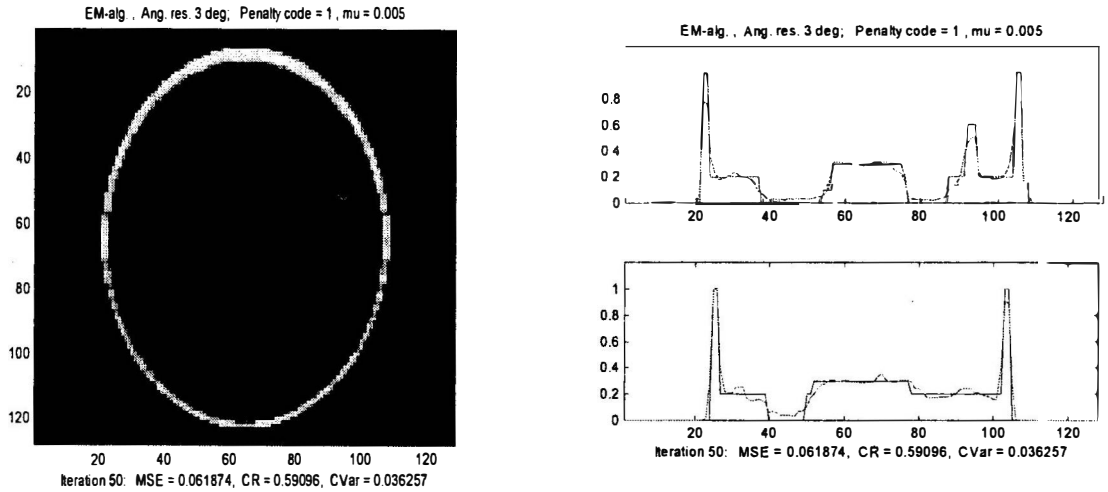

Figure 11. Reconstruction with the EMTV algorithm: image (left) and its two slices (right)

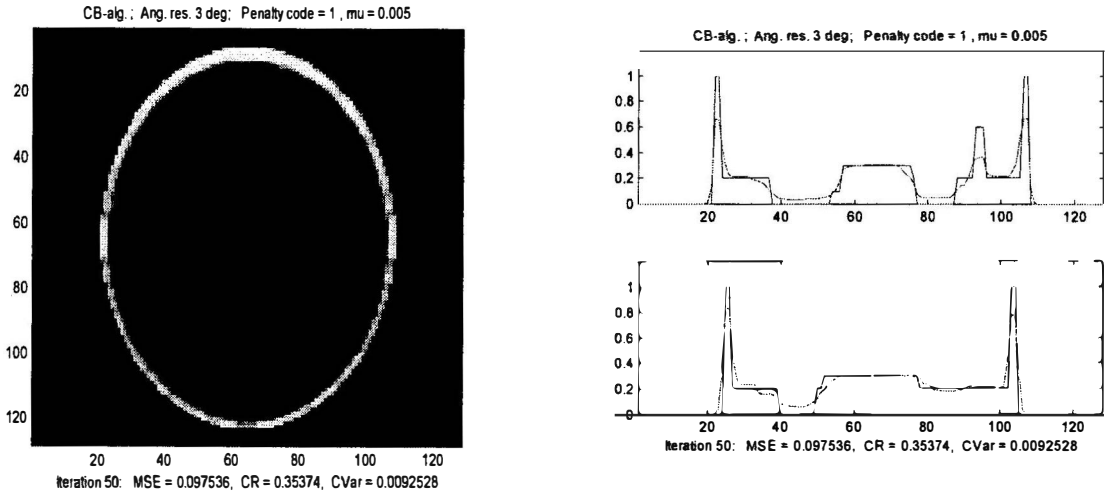

Figure 12. Reconstruction with the CBTV algorithm: image (left) and its two slices (right)

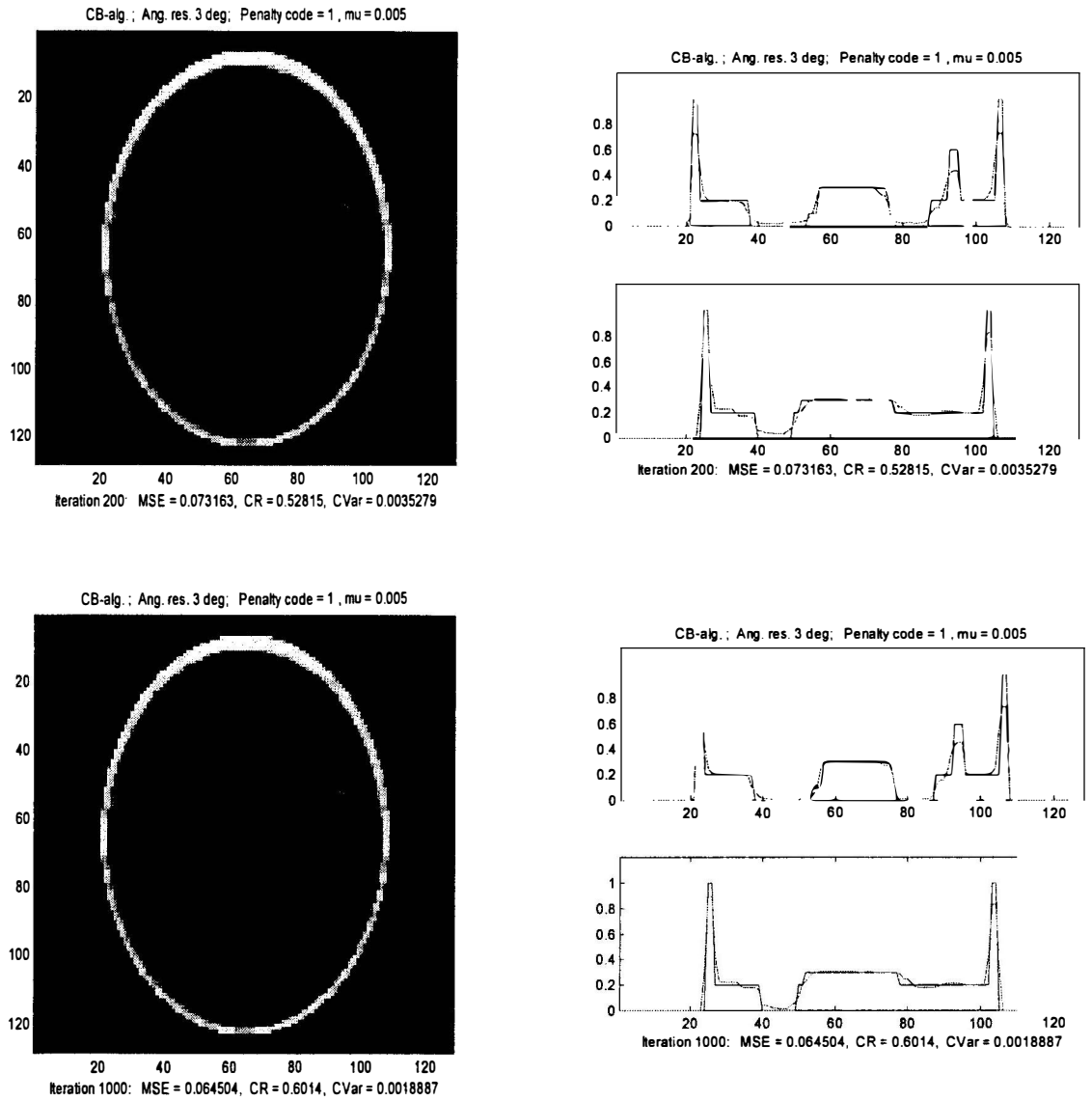

Figure 13. Reconstruction with the CBTV algorithm: after 200 iterations (upper row), after 1000 iterations (lower row)

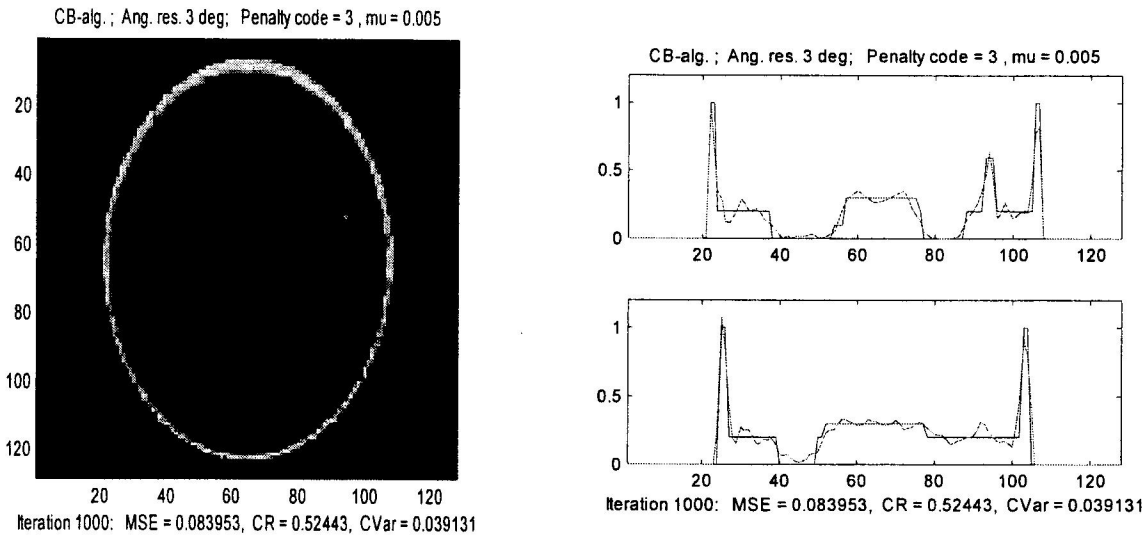

Figure 14. Reconstruction with the CBWT algorithm after 1000 iterations

Appendix
Wavelet based multiresolution decomposition and wavelet packets

The following brief outline of the pyramidal algorithm for multiresolution analysis is based on the formalism of Mallat [16]. The algorithm is given here for one-dimensional functions and can be defined in a similar way for the dyadic, separable two dimensions, in which case, one obtains a 4-band pyramid or tree (see below). (One can also obtain a 2-band pyramid by using a quinc sampling scheme and two-dimensional nonseparable wavelets [17]).

The multiresolution framework is based on two sets of subspaces of $L^2(R)$ - V_j and $W_j, j \in Z$, which satisfy the following conditions:

1) $...V_{-1} \supset V_0 \supset V_1...$
2) $W_j \cap W_k = \emptyset; \forall j \neq k \in Z$
3) $V_j \cap W_j = \emptyset$
4) $V_j \cup W_j = V_{j+1}; \forall j \in Z$

In addition, V_j is dense in $L^2(R)$ for $j \rightarrow \infty$. Index j is a scale parameter, and the scale becomes finer (i.e., the resolution increases) with increasing j.

The projection of a function $f(t)$ onto subspace V_j is an approximation of $f(t)$ at scale j, and the projection of the function $f(t)$ onto subspace W_j provides the detail of $f(t)$ at scale j. The detail function contains the information that is lost in the transition from $A_j f(t)$ to $A_{j+1} f(t)$.

Starting from a single function $\phi(t)$, the father wavelet, it is possible to generate an orthonormal basis of the scaling functions $\phi_{jk}(t) = 2^{j/2}\phi(2^j t - k), j,k \in Z$, spanning V_j. Similarly, starting from a single function $\psi(t)$, the mother wavelet, one can obtain an orthonormal basis of wavelet functions $\psi_{jk}(t) = 2^{j/2}\psi(2^j t - k)$, spanning W_j, where k is a shift parameter.

Thus the approximation and detail of the function at scale j are represented by

$$A_j f(t) = \sum_{k \in Z} a_j(k)\phi_{jk}(t), \tag{1A}$$

$$D_j f(t) = \sum_{k \in Z} d_j(k)\psi_{jk}(t), \tag{2A}$$

where

$$a_j(k) = < f(t), \phi_{jk}(t) >, \tag{3A}$$

$$d_j(k) = < f(t), \psi_{jk}((t) > . \tag{4A}$$

Since the subspaces V_j and W_j are orthonormal, a unique inverse transform exists.

A multi-level analysis representation of such a sampled function in terms of its details and approximation functions up to some reference level J is given by:

$$f(t) = A_J f(t) + \sum_{j \geq J} D_j f(t). \tag{5A}$$

Further, let the sequences h and g, describing a lowpass and a highpass filters, respectively,

satisfy the following relationships, the so-called refinement equations:

$$\phi(t) = \sum_{k \in Z} \sqrt{2}\, h(k)\phi(2t-k), \tag{6A}$$

$$\psi(t) = \sum_{k \in Z} \sqrt{2}\, g(k)\phi(2t-k). \tag{7A}$$

The approximation and detail coefficients at scale j, a_j and d_j, respectively, can be recursively calculated from the approximation coefficients a_{j+1} at a finer scale, by convolving it with filters $h(k)$ and $g(k)$, followed by a dyadic down-sampling, i.e.

$$a_j(k) = h * a_{j+1}(2k), \tag{8A}$$

$$d_j(k) = g * a_{j+1}(2k), \tag{9A}$$

where * denotes a convolution operation. The next step in a multi-level decomposition splits the approximation coefficients a_j using the same operations, producing recursively the coefficients a_{j-1} and d_{j-1}.

Let s be a vector containing L samples of a continuous function $f(t)$, and let L be a power of 2. In this case, we can consider these samples as approximation coefficients of $f(t)$ at the finest (available) scale $J = \log_2 L$, and a multi-level wavelet analysis of a sampled signal s can be rewritten in the matrix form:

$$c = Ws, \tag{10A}$$

where a new features vector in the wavelet domain has the following coefficients structure: $c = [a_0, d_0, d_1, ..., d_{J-1}]^T$, and W is a matrix of an orthonormal wavelet basis, describing the multiresolution decomposition given in (8A) and (9A). The corresponding tree structure of coefficients for the case of a 3-level decomposition is depicted in Figure 1A.

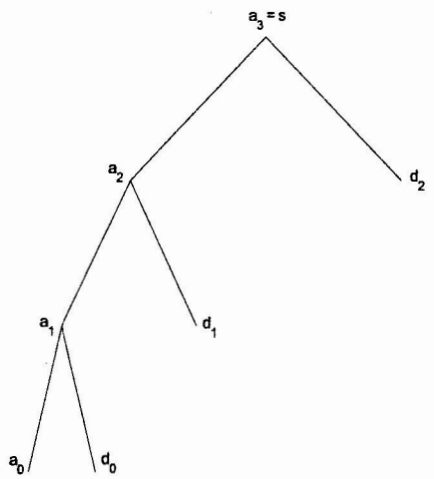

Figure 1A. Multiresolution wavelet

The complement operations of those given in (8A) and (9A), i.e. the calculation of a finer scale

approximation coefficients a_{j+1} from a coarser scale approximation and detail coefficients a_j and d_j, is performed by a dyadic upsampling followed by the corresponding convolutions with filters $h(k)$ and $g(k)$

$$a_{j+1}(k) = \sum_{m \in Z} h(2m-k)a_j(k) + \sum_{m \in Z} g(2m-k)d_j(k). \tag{11A}$$

A corresponding multi-level wavelet synthesis of a sampled signal s from its wavelet coefficients can be written in the matrix form as:

$$s = W^T c. \tag{12A}$$

Let us define the following sequence of functions $\varphi_n(t)$, $n \in N$ so that

$$\varphi_{2n}(t) = \sum_{k \in Z} \sqrt{2}\, h(k)\varphi_n(2t-k), \tag{13A}$$

$$\varphi_{2n+1}(t) = \sum_{k \in Z} \sqrt{2}\, g(k)\varphi_n(2t-k), \tag{14A}$$

where $\varphi_0(t) = \phi(t)$, and $\varphi_1(t) = \psi(t)$. Thus, starting from the family of generating functions $\varphi_n(t)$, $n \in N$, and constructing from them orthogonal wavelet waveforms, we arrive at the triple-indexed family of functions:

$$\varphi_{jnk}(t) = 2^{j/2}\varphi_n(2^j t - k),\ j,k \in Z,\ n \in N. \tag{15A}$$

As before, j,k are the scale and shift parameters, respectively, and n is an explicit frequency parameter, related to the number of oscillations of a particular generating function $\varphi_n(t)$. The set of functions $\varphi_{jn}(t)$ forms a (j,n) wavelet packet. This set of functions can be split into two parts at a coarser scale: $\varphi_{j-1,2n}(t)$ and $\varphi_{j-1,2n+1}(t)$. It follows that these two form an orthonormal basis of the space which is spanned by $\varphi_{jn}(t)$. Thus, we arrive at a family of wavelet packet functions on a binary tree (Figure 2A). The nodes of this tree are numbered by two indices: the depth of the level $j = 0, 1, .., J$, and the number of node $n = 0, 1, 2, 3, ..., 2^j - 1$ at the specified level.

Using such a family of functions allows one to analyze given signals not only with a scale-oriented decomposition but also with reference to frequency subbands. There are 2^L wavelet bases in such a library (L is the number of samples in the signal), and, therefore, 2^L ways to represent a given signal. Naturally, the library contains the wavelet basis. There are several criteria for choosing the best basis, i.e. the best tree, for a given signal (see, [18] for a survey).

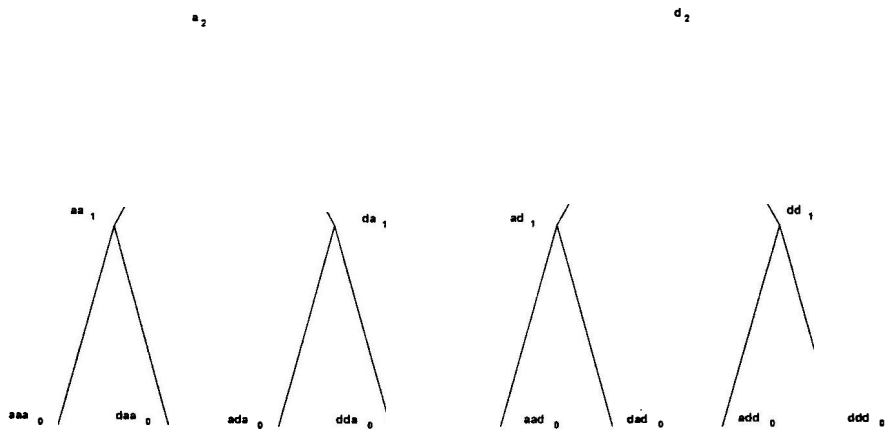

Figure 2A. Wavelet packets decomposition tree

The decomposition coefficients $c_{jnk} = \langle s, \varphi_{jnk} \rangle$ are also split into $(j\ n)$ sets corresponding to the nodes of the tree. There is a fast way to compute them, using banks of *conjugate mirror filters*, as is implemented in the fast wavelet transform [16]. As is clear from the above discussion, a multiresolution wavelet decomposition of a function consists of splitings of its approximations at successively coarser resolutions into two parts. In the wavelet packet decomposition both the approximation and detail signals split at each level into two parts. This results in a much richer decomposition structure, and the corresponding binary tree of coefficients in this case is complete (Figure 2A).

The *Center for Communication and Information Technologies* (CCIT) is managed by the Department of Electrical Engineering.

Technical Report is listed also as
EE PUB #1270, January 2001.