A Bayesian MAP-EM Algorithm for PET Image
Reconstruction Using Wavelet Transform

Jian Zhou, Jean-Louis Coatrieux, *Fellow, IEEE,* Alexandre Bousse, Huazhong Shu, *Member, IEEE,* and
Limin Luo, *Senior Member, IEEE*

A Bayesian MAP-EM Algorithm for PET Image Reconstruction Using Wavelet Transform

Jian Zhou, Jean-Louis Coatrieux, *Fellow, IEEE,* Alexandre Bousse, Huazhong Shu, *Member, IEEE,* and Limin Luo, *Senior Member, IEEE*

Abstract—In this paper, we present a PET reconstruction method using the wavelet-based maximum a posteriori (MAP) expectation-maximization (EM) algorithm. The proposed method, namely WV-MAP-EM, shows several advantages over conventional methods. It provides an adaptive way for hyperparameter determination. Since the wavelet transform allows the use of fast algorithms, WV-MAP-EM also does not increase the order of computational complexity. The spatial noise behavior (bias/variance and resolution) of the proposed MAP estimator is analyzed. Quantitative comparisons to MAP methods with Markov Random Field (MRF) prior models point out that our alternative method, wavelet-base method, offers competitive performance in PET image reconstruction.

Index Terms—maximum a posteriori, wavelet transform, expectation-maximization, image reconstruction, positron emission tomography.

I. Introduction

STATISTICAL methods have attracted much interest in the positron emission tomography (PET) reconstruction. The Bayesian *maximum a posteriori* (MAP) algorithm is one of the techniques suggested to improve the quality of reconstruction (e.g., [1]–[3]). Conventional MAP algorithms control the noise behavior by introducing the so-called image *a priori* information. Such prior plays the role of a smoothness constraint that penalizes the roughness of image estimate and then reduces the noise level.

Markov Random Field (MRF) approaches have been widely used for reconstruction, where image features, such as lines and edges, can be well modeled by means of local neighborhood structures. Although the MRF-based MAP algorithms have proven to be a powerful tool when dealing with the reconstruction problem, it is also computationally intensive. This results from the correlation model of MRF that makes difficult the parameter estimation. In addition, MAP models depend on

the hyperparameters that describe the *a priori* distribution. In MRF, the optimal choices for these parameters are quite model and object dependent and their setting in concrete examples may be nontrivial [4]. To overcome this problem, Fessler and Roger [5] proposed approximate expressions to predict the spatial behavior of MAP estimators for a desired spatial resolution, and left the hyperparameter setting to the users. Similar works can be found in [6] and [7], using a MAP with a spatial domain quadratic prior. For nonquadratic penalties or even more complicated statistical models, approximate expressions may bring inaccuracy.

The present paper considers a wavelet-based alternative. Wavelets as time-scale and time-frequency analysis tools have been widely used in signal processing, but their application in tomographic reconstruction is recent [8]–[12], and still growing [13]–[18]. Our method departs from previous works by: 1) the determination of image wavelet coefficients from the observed projection data, and 2) the image reconstruction via the wavelet inversion. The first step is accomplished by using a Bayesian MAP algorithm with an *a priori* constraint directly imposed on the distribution of random wavelet coefficients. It does not transform the measured PET data, and thus the original statistics of observations can be well preserved when performing optimal coefficient restoration. Moreover, since wavelet coefficients are often modeled with decorrelation property, this, compared to the MRF model, makes our method better suited for parameter optimization.

Here, the expectation-maximization (EM) algorithm allows deriving an efficient iterative approach to find the required wavelet coefficients. EM provides an elegant way for parameter estimation by assuming the existence of values for additional but *hidden* (or *missing*) parameters [19]. By using the EM algorithm, our wavelet-based reconstruction method can be implemented quickly and efficiently. Moreover, we observe that the hyperparameters describing the wavelet coefficient prior model can also be approached in terms of the *threshold estimator*. Recently, a similar heuristic solution has been successfully applied to the study of wavelet-based image deblurring [20][21].

This paper is organized as follows. In Section 2, we first define the wavelet-based PET reconstruction problem together with the EM optimization. Section 3 compares experimentally the designed algorithm with other conventional image domain MRF-based MAP estimations. Finally, conclusions and directions of future work are drawn.

II. METHOD

A. The wavelet-based measurement model

Without loss of generality, we consider in this paper the problem of two-dimensional PET image reconstruction. Let us suppose that the underlying radioactivity distribution function can be discretized into M pixels array which are ordered lexicographically in vector $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_M]^T$ where the superscript 'T' indicates the transpose operation. Considering a PET imaging system with N measurements collected in vector $\boldsymbol{y} = [y_1, \ldots, y_N]^T$, we suppose that the ensemble average of the measurement $\boldsymbol{y}$ is related to $\boldsymbol{\lambda}$ by

$$\mathrm{E}[\boldsymbol{y}] = \boldsymbol{H}\boldsymbol{\lambda} \tag{1}$$

where $\boldsymbol{H}$ is the $N \times M$ system matrix as well as the forward projection operator that characterizes the response of the imaging system.

Since we are dealing with digital images, we limit our studies on the orthogonal dyadic discrete wavelet transform (DWT). The DWT of a two-dimensional image $\boldsymbol{\lambda}$ is defined by

$$\boldsymbol{c} = \boldsymbol{W}_{l_0}\boldsymbol{\lambda} \tag{2}$$

where $\boldsymbol{W}_{l_0}$ represents the DWT matrix with a maximum decomposition level l_0, and $\boldsymbol{c}$ is the vector collecting the wavelet coefficients. In practice, the construction of the operator $\boldsymbol{W}_{l_0}$ depends on the selection of wavelet bases. For example, in the discrete case, one can use a pair of quadratic mirror filters to approximate the required wavelet operators [22]. The wavelet inversion operator is denoted by $\boldsymbol{W}_{l_0}^{-1}$ or simply $\boldsymbol{W}_{l_0}^T$ (due to the orthogonality). If the coefficients $\boldsymbol{c}$ are known exactly, then we are able to reconstruct $\boldsymbol{\lambda}$ perfectly by the wavelet inversion:

$$\boldsymbol{\lambda} = \boldsymbol{W}_{l_0}^T\boldsymbol{c}. \tag{3}$$

It has been noted that fast algorithms [23] are available for the efficient calculation of DWT and its inversion. By substituting (3) into (1), the imaging model can be expressed by

$$\mathrm{E}[\boldsymbol{y}] = \boldsymbol{H}\boldsymbol{W}_{l_0}^T\boldsymbol{c}. \tag{4}$$

B. The MAP estimation for wavelet coefficients

Instead of solving the unknown radioactivity $\boldsymbol{\lambda}$, we are interested in recovering its wavelet coefficients $\boldsymbol{c}$ in advance. As thus, we seek the Bayesian MAP estimation for coefficients $\boldsymbol{c}$. This is achieved by maximizing the *posterior distribution* $f(\boldsymbol{c}|\boldsymbol{y})$, i.e.,

$$\widehat{\boldsymbol{c}}_{\mathrm{MAP}} = \arg\max_{\boldsymbol{c}} f(\boldsymbol{c}|\boldsymbol{y}). \tag{5}$$

Using the Bayes formula, we obtain

$$f(\boldsymbol{c}|\boldsymbol{y}) \propto f(\boldsymbol{y}|\boldsymbol{c})f(\boldsymbol{c}) \tag{6}$$

where $f(\boldsymbol{c})$ is the *prior distribution* of wavelet coefficients, and $f(\boldsymbol{y}|\boldsymbol{c})$ is the *likelihood distribution* which is characterized by the imaging model (4). In this paper, the measurement data $\boldsymbol{y}$ are assumed to be Gaussian distributed with unknown parameter $\boldsymbol{c}$. Then the likelihood distribution is:

$$f(\boldsymbol{y}|\boldsymbol{c}) \propto \\ \exp\left\{-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{H}\boldsymbol{W}_{l_0}^T\boldsymbol{c})^T\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{y} - \boldsymbol{H}\boldsymbol{W}_{l_0}^T\boldsymbol{c})\right\} \tag{7}$$

where $\boldsymbol{\Sigma}_0$ denotes the data covariance matrix. While it is more appropriate to use a measure of Poisson distribution, the Gaussian distribution has been shown to be approximately Poisson with means greater than five (see, e.g., [24]). On the other hand, in modern PET systems that precorrect the data, the Poisson assumption no longer holds. The exact likelihood function describing these data is complicated and difficult to work with. Therefore, it has been suggested that a minimization of a weighted least squares (WLS) functional can be used [25], which is also equivalent to consider a Gaussian distribution like (7) in place of the complicated distribution of raw data. However, it should be noted that the accuracy of the Gaussian approximation decreases when the number of random prompts is low.

The specification of the covariance matrix $\boldsymbol{\Sigma}_0$ is important when the Gaussian approximation is used. As it has been demonstrated in [25], it is sufficient to use a diagonal matrix $\boldsymbol{\Sigma}_0 = \mathrm{diag}\left\{\sigma_1^2, \ldots, \sigma_N^2\right\}$ to capture the second-order statistics of the measured data. Here, σ_n^2, $n = 1, \ldots, N$, is the variance for nth measurement data y_n that can be estimated beforehand using several useful 'data plug-in' techniques [6][25][26].

As for the *a priori* coefficient distribution, it has been well reported in the literature (e.g., [27]–[32]) that image wavelet coefficients are characterized with sparsity (i.e., the heavy-tailed histogram), hence, any distribution possessing the sparse property is legitimate to model the prior distribution of wavelet representation. The generalized Gaussian distribution is well suited for fitting the mode and the heavy tail behavior of wavelet coefficients distributions. The employed prior distribution in this paper is the *Laplacian* distribution:

$$f(\boldsymbol{c}) \propto \exp\left\{-\beta\|\boldsymbol{c}\|_1\right\} \tag{8}$$

where $\|\cdot\|_1$ denotes the ℓ^1-norm and β is the adjustable hyperparameter. It belongs to the family of generalized Gaussian distribution and has proven to be useful for the wavelet coefficients sparsity description [31][32].

C. EM optimization

According to the terminology accepted in the EM algorithm literature [1],[19],[33],[34],[35], the measurement vector $\boldsymbol{y}$ is called the incomplete data vector. Since the mean of $\boldsymbol{y}$ is linear in $\boldsymbol{c}$, the natural complete-data for the EM algorithm solving (5) is also linear in $\boldsymbol{c}$ [36] (or see the section IV of [37]). Here, we consider a complete-data $\boldsymbol{x}$ which satisfies

$$\begin{cases} \boldsymbol{x} & \sim & \mathcal{N}(\boldsymbol{W}_{l_0}^T\boldsymbol{c}, \boldsymbol{\Sigma}_1), \\ \boldsymbol{y} & \sim & \mathcal{N}(\boldsymbol{H}\boldsymbol{x}, \boldsymbol{\Sigma}_2). \end{cases} \tag{9}$$

where $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{C})$ denotes the normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{C}$. The admissible condition for this $\boldsymbol{x}$ is that the two normal distributions are independent and consistent with (7), i.e., $\boldsymbol{H}\boldsymbol{\Sigma}_1\boldsymbol{H}^T + \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_0$, as required.

The EM algorithm is an iterative procedure which produces a sequence $\widehat{\boldsymbol{c}}^0$, $\widehat{\boldsymbol{c}}^1$, $\widehat{\boldsymbol{c}}^2$,..., of estimations maximizing (5). It gets $\widehat{\boldsymbol{c}}^{k+1}$ from $\widehat{\boldsymbol{c}}^k$, by first considering the conditional expectation of $\boldsymbol{c}$ over the complete data set (the E-step), i.e.,

$$Q(\boldsymbol{c}; \widehat{\boldsymbol{c}}^k) = \mathrm{E}[\log f(\boldsymbol{c}|\boldsymbol{x})|\boldsymbol{y}, \widehat{\boldsymbol{c}}^k] \tag{10}$$

and then choosing $\widehat{c}^{k+1}$ as c which maximizes this conditional expectation (the M-step).

Applying the Bayesian formula to $f(c|x)$, we obtain:

$$
\begin{aligned}
Q(c; \widehat{c}^k) &= \log f(\overline{x}^k | c) + \log f(c) \\
&\quad + \text{terms independent of } c
\end{aligned}
\tag{11}
$$

where

$$
\overline{x}^k = \mathrm{E}[x | y; \widehat{c}^k].
\tag{12}
$$

By standard properties of joint Gaussian distributions, it can further be shown that

$$
\overline{x}^k = W_{l_0}^T \widehat{c}^k + \Sigma_1 H^T \Sigma_0^{-1}(y - H W_{l_0}^T \widehat{c}^k).
\tag{13}
$$

Equation (11) is recognized as the *log-posterior distribution* of wavelet coefficient c using the conditional mean $\overline{x}^k$ as the direct measurement data. Therefore, $\widehat{c}^{k+1}$ is also a MAP solution. When combined with the prior distribution (8), it is not difficult to derive that

$$
\begin{aligned}
\widehat{c}^{k+1} = \Big(&\mathrm{sign}\left\{ W_{l_0} \overline{x}^k \right\} \cdot \\
&\max\left\{ \left| W_{l_0} \overline{x}^k \right| - \beta W_{l_0} \Sigma_1 W_{l_0}^T \mathbf{1}, 0 \right\} \Big)
\end{aligned}
\tag{14}
$$

where $\mathbf{1}$ denotes the unity vector. It is obvious that the update in (14) is consistent with the well-known *soft-thresholding* formula [41] with the coefficient threshold $\beta W_{l_0} \Sigma_1 W_{l_0}^T \mathbf{1}$.

D. Selection of appropriate Σ_1 and β values

Under the given admissible condition, there are possibly many choices for Σ_1 and Σ_2. One simple but very convenient choice for Σ_1 is in the form

$$
\Sigma_1 = \delta^2 I
\tag{15}
$$

where δ^2 is the assumed noise variance in $\overline{x}_k$, and I the identity matrix. For example, from the above admissible condition, we have $\Sigma_2 = \Sigma_0 - H \Sigma_1 H^T = \Sigma_0 - \delta^2 H H^T$. Since Σ_2 is known to be semi-positive definite, we are able to determine a δ that fulfills

$$
\det |\Sigma_0 - \delta^2 H H^T| \geq 0.
\tag{16}
$$

With some elementary algebra, we can deduce:

$$
\delta^2 \leq 1/\xi_{\max}
\tag{17}
$$

where $\xi_{\max}$ is the largest eigenvalue of $\Sigma_0^{-1} H H^T$. Apparently, the parameter δ^2 should be adaptively selected by the system properties and the statistical characteristics of measured PET data. On the other hand, substituting (15) into (13) yields

$$
\overline{x}^k = W_{l_0}^T \widehat{c}^k + \delta^2 H^T \Sigma_0^{-1}(y - H W_{l_0}^T \widehat{c}^k).
\tag{18}
$$

This shows that the parameter δ^2 likely plays the role of controlling the speed of iteration. Therefore, a sound choice considers to select the upper bound of δ^2, i.e., $\delta^2 = 1/\xi_{\max}$.

The simplest method for finding the largest eigenvalue of a matrix is the Power method, also called the vector iteration method. The power iteration algorithm starts with a nonzero vector v_0, which may be a random vector. The method is described by the iteration

$$
v_{k+1} = \frac{\Sigma_0^{-1} H H^T v_k}{\| \Sigma_0^{-1} H H^T v_k \|_2}, \quad k = 0, 1, \dots.
\tag{19}
$$

and the sequence $(\xi_{\max})_k$, defined by $\frac{v_k \Sigma_0^{-1} H H^T v_k}{v_k^T v_k}$, converges to the largest eigenvalue.[1] The computational cost of this algorithm is relatively low. In our experiments, we observed that it takes us only a few iterations to estimate the converged largest eigenvalue. For a detailed discussion on the power algorithm, see [38][39] for example.

Like many other MAP algorithms, the hyperparameter β adjusts the strength of the prior constraint. It can be seen from (14) that large β would eliminate a lot of coefficients and thus, would lead to oversmoothed reconstructions. An appropriate β has to be selected in order to reduce noise effects while keeping important coefficients undestroyed. One often used approach in EM considers the adaptive hyperparameter that maximizes the marginal distribution $m(\overline{x}^k; \beta) = \int f(\overline{x}^k | c) f(c; \beta) dc$ (see [33] and [40] for example). However, in our case, this is numerically intractable. Early works on approximate expressions have been proposed [5], [6] and [7] to calculate the local impulse response and covariance of MAP estimators. They help choosing β for a desired spatial resolution. However, since our prior model is clearly nonquadratic, a direct use of these approximations may lead to inaccuracies.

Here, the parameter β is approached in an heuristic way. As pointed out by (11), the M-step is equivalent to seek a MAP estimation for the following measurement model: $\overline{x}_k = W_{l_0}^T c + n$ with a sparsity constraint in wavelet coefficients c. Notice that the noise n is Gaussian with zero mean and covariance Σ_1. This is similar to a wavelet-based linear inverse problem that is widely addressed in the area of image denoising. Eq. (14) has shown that the MAP estimate is equivalent to a soft-thresholding rule. Using (15), we can show that the threshold is $\beta \delta^2$. Now our motivation for a choice of β is to consider a threshold estimator. Using again (15), i.e., assuming that the noise in $\overline{x}_k$ is white with variance δ^2, we obtain the *universal* threshold estimator [41]: $\sqrt{2\delta^2 \log M}/\sqrt{M}$. In comparison to other estimators, this estimator introduces less unknown parameters and has the lowest computational cost since it depends only on the image size when δ is set in advance. The threshold is independent of data $\overline{x}^k$, which results in coefficient estimation asymptotically optimal in the minimax sense [42] (note that this does not mean that the image reconstruction is near optimal). In addition, since it does not change with iteration, so does the parameter β (see at below). Usually, this is preferred for image reconstruction since an iteration-adaptive β could lead to more complicated MAP estimators (than the one with a fixed *a priori* β) whose noise behavior may be difficult to predict. By letting

---

[1]The largest eigenvalue of $\Sigma_0^{-1} H H^T$ is equal to the one of $H^T \Sigma_0^{-1} H$ which is known as the Fisher information matrix. Thus, another way to find out the largest eigenvalue is to use an approximated block Toeplitz Fisher information matrix [5] and then compute it with the Fast Fourier Transform as suggested by [6].

$\beta\delta^2 = \sqrt{2\delta^2 \log M}/\sqrt{M}$, we have $\beta = \sqrt{2 \log M}/(\delta\sqrt{M})$. Since δ^2 is chosen to be $1/\xi_{\max}$, we arrive at

$$\beta = \sqrt{2\xi_{\max} \log M}/\sqrt{M}. \tag{20}$$

Since $\xi_{\max}$ depends on data, our β is data dependent. We experimentally found that such β is indeed a good choice. Of course, by considering other threshold estimators, other β values can be found out.

E. Compensation for nonnegativity constraints

The development of wavelet-based MAP-EM method does not account for the nonnegativity constraint typically used in emission tomography reconstruction. This would result in large errors in covariance estimates for low-activity regions [6][43]. An improvement can be made by considering a positive hidden image x. For example, we can assume the *truncated* Gaussian random variable x in model (9). i.e., by neglecting negative values in the hidden image x. A similar idea was early suggested in [6] to compensate the nonnegativity constraint of the conventional MAP reconstruction. However, since the noise is non-Gaussian, we found that it is really complicated to derive an explicit coefficient update formulation.

To reduce this effort, it seems reasonable for us to assume that the conditional mean of the hidden image holds the same nonnegativity property. Under this assumption, we can easily modify the mean $\overline{x}^k$ as follows:

$$\overline{x}_+^k \equiv \max\{\overline{x}^k, 0\} \tag{21}$$

which clips the negative values in vector $\overline{x}^k$. Although this *ad hoc* modification can not ensure the rigorous nonnegativity in reconstruction and perhaps impacts the monotonic convergence property of the EM algorithm, as we show later, it does improve the accuracy over a method that ignores the nonnegativity constraint.

F. Algorithm implementation and computational complexity

We summarize our wavelet-based reconstruction algorithm (namely WV-MAP-EM) as follows:

1) *Setup Σ_0 using a data plugin technique, calculate δ^2 using $1/\xi_{\max}$, and compute β according to (20). Choose an initial coefficient vector $\widehat{c}^0$;*
2) *Update coefficients according to (13), (14), and (21);*
3) *Repeat step 2) until some specific stopping conditions are verified;*
4) *Obtain image estimate $\widehat{\lambda}$ using the wavelet inversion transform (3).*

The key step of our algorithm is 2) which is illustrated in Fig. 1. It can be seen that a full updating mainly requires four operations: W_{l_0}, $W_{l_0}^T$, H and H^T. Here, we have neglected the operation in the soft-thresholding (15) because it can be performed without the need of multiplication. Either W_{l_0} or $W_{l_0}^T$ in the worst case with levels $(\log_2 M)/2$ (M is supposed to be a power of 2) incurs the complexity of $\mathcal{O}(LM \log M)$ [23] where L is the length of FIR filters (typically small), while the forward and backward projections
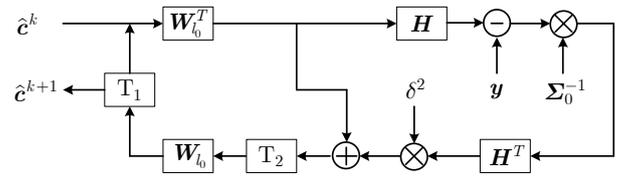

Fig. 1. The complete programming framework for the updating of wavelet coefficients. T_1 represents the coefficient thresholding operation (14) while T_2 is the max operation (21) compensating the nonnegativity constraint.
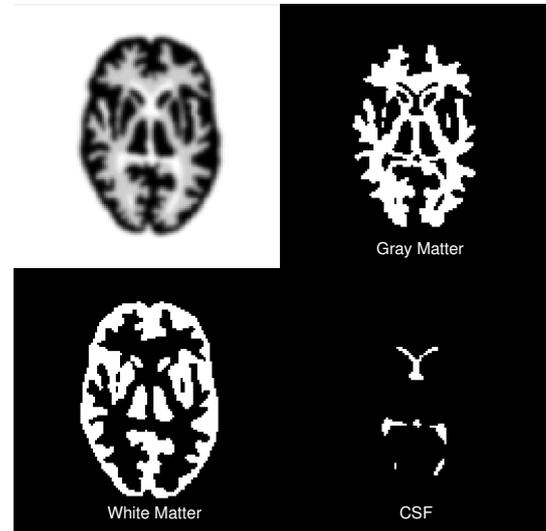

Fig. 2. The Hoffman brain phantom and the corresponding ROI regions used in our experimental studies.

require a complexity less than $\mathcal{O}(MN)$. Since $N \gg L \log M$ holds in most cases, it gives an overall approximated complexity of $\mathcal{O}(MN)$. In other words, although the proposed algorithm does involve additional efforts to compute the wavelet transforms, it does not increase the order of complexity.

III. EXPERIMENTS

A. Data description

The Hoffman brain phantom (Fig. 2) was used to evaluate the performance of the proposed method. The image was a 128×128 array of squared pixels with 4.7mm size. We simulated a PET scan with 192 radial bins and 192 angles spaced uniformly over $180°$. The used ray spacing is 3.3mm. We used pseudo-random log-normal variants with a standard deviation of 0.3 to simulate detector pairs with non-uniform detector efficiencies and scaled so that the total number of Poisson random prompts was 1.8×10^6. The factor corresponding to a uniform field of 5% random coincidences was also considered in our experiments. Another 40 replicate noisy sinograms were generated in a similar way, which were used to evaluate the noise property of the wavelet-based MAP estimate.

B. Reconstruction

In the WV-MAP-EM method, we used Qi and Leahy's "modified plug-in" technique [6] to compute the variance estimations of the measured PET data. The Daubechies' wavelets
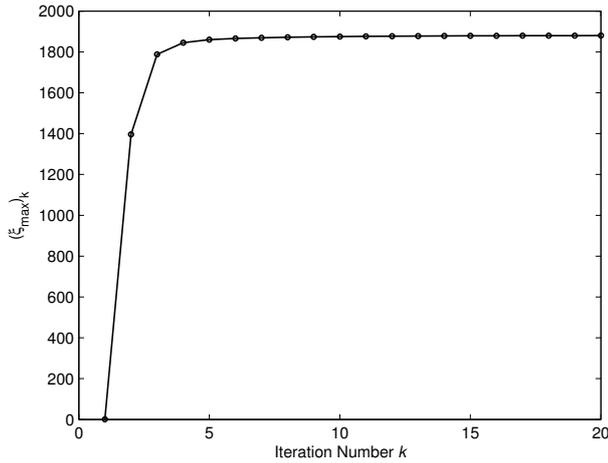
Fig. 3.   Vector iteration method solving the largest eigenvalue of $\boldsymbol{\Sigma}_0^{-1}\boldsymbol{H}\boldsymbol{H}^T$.

having four vanishing moments were adapted to compute the forward and inverse wavelet transform. The discrete wavelet transform was performed up to level 3. The largest eigenvalue was approximated by using the vector iteration method. The initial vector $\boldsymbol{v}_0$ was a unit vector. We stopped the iteration by checking: $|(\xi_{\max})_{k+1} - (\xi_{\max})_k|/(\xi_{\max})_k < 10^{-6}$. Fig. 3 shows a sequence of $(\xi_{\max})_k$ produced by this method. We see that only 20 iterations are enough for convergence. Besides the proposed method, we also investigated the conventional MRF-based MAP methods for the purpose of comparisons. Consider the often-used MRF based on the second-order neighborhood system, the prior distribution can be modeled by:

$$f(\boldsymbol{\lambda}) \propto \exp\left\{ -\gamma \sum_{m=1}^{M} \left[\sum_{j \in \eta_m^1} V(\lambda_m - \lambda_j) + \frac{1}{\sqrt{2}} \sum_{j \in \eta_m^2} V(\lambda_m - \lambda_j) \right] \right\} \quad (22)$$

where $V(\cdot)$ is called the potential function, γ is the corresponding hyperparameter, η_m^1 denotes the horizontal and vertical neighborhood of pixel m and η_m^2 denotes the diagonal neighborhood of pixel m. Two typical potential functions, $V(x) = |x|$ and $V(x) = x^2$, were studied in this paper that can be considered as nice approximations of the *total-variational* (TV) and *Tikhonov* (or the quadratic prior) smoothness constraints. For each MRF model, we used Green's one-step late (OSL) MAP-EM algorithm [3] to yield the required MAP solutions. The resulting methods were named thereafter as TV-MAP-EM and TIK-MAP-EM respectively. The OSL algorithm is popular due to its simplicity, but it can diverge. However we observed from our empirical results that it always converges if the hyperparameter γ is not too large (for instance smaller than 6)[2]. Since TV is not differentiable, to apply OSL, we considered its subgradient as suggested in [45] with

---

[2]The De Pierro's method (block sequential regularized expectation maximization (BSREM) algorithm [44] with 16 equal size blocks) has also been implemented. The results show very little MSE difference between the OSL and BSREM when γ is not large. It is however superior in convergence speed.

$d|x|/dx = \text{sign}(x)$. The optimal γ values were obtained by minimizing the mean-squared-error (MSE) criterion (defined in the next subsection). We applied a greedy searching strategy in which 1000 different γ values uniformly spaced in the interval $[10^{-4}, 10^1]$ were used. For each γ, we carried out either TV-MAP-EM or TIK-MAP-EM, and the lowest MSE after each reconstruction was recorded as the quantity evaluating γ. We safely selected the global optimal hyperparameter which was the one generating the minimum of 1000 lowest MSE values. Note that this procedure was quite time consuming.

Both TV-MAP-EM and TIK-MAP-EM were initialized with the same filtered-backprojection (FBP) reconstruction where a Ram-Lak filter (multiplied by a Hamming window) with the cut-off frequency of 0.8 (normalized) was used. For a fair comparison, the initial wavelet coefficients in our WV-MAP-EM method were also calculated using the wavelet transform of FBP. All MAP methods were stopped after 200 iterations.

To investigate which type of wavelets can better fit our objective, three other wavelets: Haar wavelet, coiflets and symlets were introduced (see [10] for a review). In order to further explore the benefit of the wavelet-based MAP method, we also attempted the use of *undecimated* DWT (UDWT) as an alternatives to DWT. The UDWT is known as the linear and shift invariant discrete wavelet transform, which simply eliminates the down-sampling process in the implementation of a DWT. It has been shown of relevance in many signal processing applications [46] [47]. Hence, UDWT in this paper was expected to yield higher visual quality reconstructions. Note UDWT is no longer orthogonal, and thus the actual coefficient updating process may differ from (14). However, as it was suggested in [20], we can still use the same updating framework (see Fig. 1). The corresponding results were obtained by performing 200 iterations together with the four different types of wavelet.

C. Evaluation

To quantify the quality of reconstructions, we define the *percent* mean-squared-error as follows:

$$\% \text{MSE} = \|\widehat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}^{\text{true}}\|_2^2 / \|\boldsymbol{\lambda}^{\text{true}}\|_2^2 \times 100\% \quad (23)$$

where $\boldsymbol{\lambda}^{\text{true}}$ represents the true phantom data and $\|\cdot\|_2$ denotes the ℓ^2-norm. The vertical profiles of reconstructed images were used to show the edge-preserving capacity of the various wavelet-based MAP methods.

A region of interest (ROI) analysis was also carried out on gray matter, white matter and CSF. The bias and variance over an ROI are defined as

$$\text{BIAS} = (\overline{\mu}_{\text{ROI}} - \mu_{\text{ROI}}^{\text{true}})/\mu_{\text{ROI}}^{\text{true}}, \quad (24)$$

$$\text{VAR} = \frac{1}{|\text{ROI}|} \sum_{j \in \text{ROI}} (\lambda_j - \overline{\mu}_{\text{ROI}})^2 \quad (25)$$

where $\mu_{\text{ROI}}^{\text{true}}$ denotes the true value of region intensity; $\overline{\mu}_{\text{ROI}}$ is the average intensity of the given ROI.

In order to evaluate the noise properties of the wavelet-based MAP estimates, we calculated the standard deviation
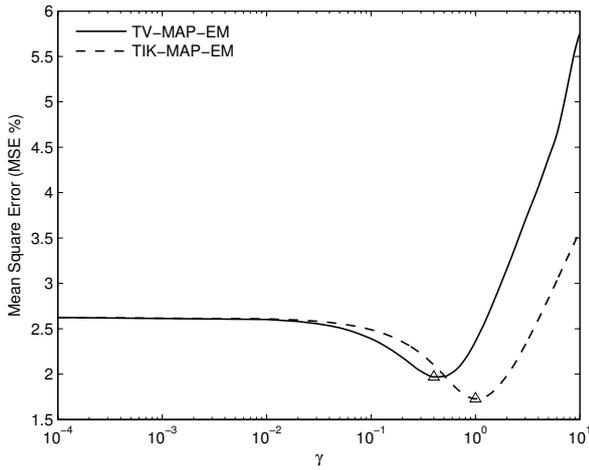
Fig. 4.   The lowest MSE change curves generated by the greedy enumeration strategy. The global minimum MSE values were marked with triangles. The corresponding γ values were treated as the optimal hyperparameters for the MRF prior models.

image using the following formula:

$$\boldsymbol{\lambda}^{\text{std}} = \sqrt{\frac{1}{39}\sum_{i=1}^{40}\|\widehat{\boldsymbol{\lambda}}^{(i)} - \overline{\widehat{\boldsymbol{\lambda}}}\|_2^2} \qquad (26)$$

where $\widehat{\boldsymbol{\lambda}}^{(i)}$ is the MAP estimate for the ith noisy sinogram and $\overline{\widehat{\boldsymbol{\lambda}}}$ is the average reconstruction which is given by

$$\overline{\widehat{\boldsymbol{\lambda}}} = \frac{1}{40}\sum_{i=1}^{40}\widehat{\boldsymbol{\lambda}}^{(i)}. \qquad (27)$$

The average standard deviation (ASTD) over each pixel was computed by

$$\text{ASTD} = \frac{1}{|\Omega|}\sum_{j\in\Omega}\lambda_j^{\text{std}} \qquad (28)$$

where Ω denotes the object region.

Three local impulses (with magnitude 100 on a particular pixel) sampling the gray matter, white matter and CSF were also added to the replicate data sets. Reconstruction of the impulses was obtained by subtracting the average reconstruction of the impulse-free data from the average reconstruction of the data containing the added impulses. The equivalent Gaussian kernel with the proper full width at half maximum (FWHM) was used to evaluate the resolution property.

In addition to the computational complexity, the CPU time costs of different MAP methods were recorded throughout the whole iterative procedure. The average costs for every 5 iterations were plotted in curves.

IV. RESULTS

Fig. 4 plots the lowest MSE changes according to γ. As we can see, in this particular experiment, the optimal γ for TV-MAP-EM is obtained from $10^{-0.6} \approx 0.25$ while for TIK-MAP-EM it is close to $10^{0.0} = 1.0$. It also can be viewed that TIK-MAP-EM leads to smaller MSE than TV-MAP-EM. This
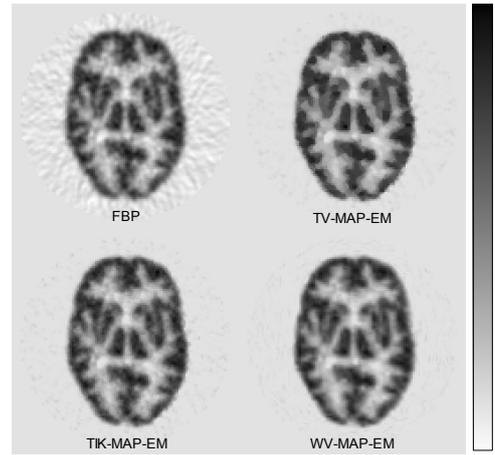

Fig. 5.   Reconstructions of FBP, TV-MAP-EM ($\gamma = 0.25$), TIK-MAP-EM ($\gamma = 1.0$) and WV-MAP-EM ($\beta = 1.497$). Here, β is adaptively selected by the noise variance $\delta^2 = 5.289 \times 10^{-4}$. Note that all images are scaled to the same gray levels.
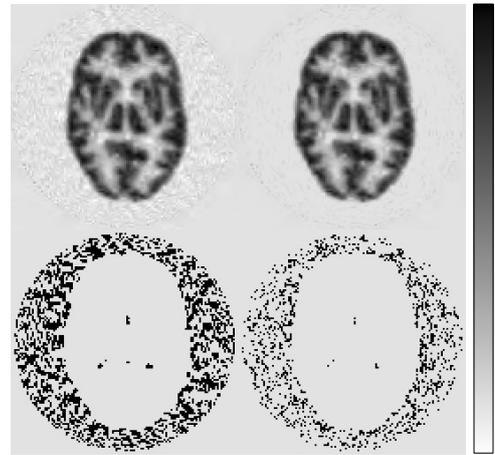

Fig. 6.   Reconstructions of WV-MAP-EM with (top-right) and without (top-left) the compensation of nonnegativity constraint. The black spots in images of the second row indicate negative values of the corresponding reconstructions of the first row. Notice the significant decrease of negative values when using our nonnegativity constrained WV-MAP-EM.

may indicate that the Tikhonov prior model can be superior to TV in terms of mean squared error.

Fig. 5 shows reconstructions using different methods. The FBP reconstruction clearly suffers from the "spike" noise artifacts that, however, can be successfully suppressed using MAP methods (note that the MAP reconstructions exhibit very different noise characteristics). It is clear from Fig. 5 that the TV-MAP-EM method can provide a piecewise smoothed image with distinct regions and sharp boundaries. This comes from the strong edge-preserving property of TV prior model in which the high amplitude would be interpreted as an edge and thus can be well preserved. In spite of this advantage, TV-MAP-EM produces an overall homogenous image in contrast to a real tomographic image. Both TIK-MAP-EM and WV-MAP-EM provide smoothed images that differ noticeably from the reconstruction obtained with TV-MAP-EM. The WV-MAP-EM method offers slightly better visual quality.

Fig. 6 shows the reconstructions of WV-MAP-EM with

TABLE I
ROI BIAS AND VARIANCE ANALYSIS (BIAS $\pm$ VAR). THE VALUES IN
BRACKET SHOW THE BIAS AND VARIANCE OF RECONSTRUCTION
WITHOUT THE NONNEGATIVITY CONSTRAINT.

	gray matter	white matter	CSF
TV-MAP-EM	0.25 ± 3.96	0.05 ± 5.94	0.41 ± 2.49
TIK-MAP-EM	0.56 ± 4.02	-0.16 ± 4.91	0.32 ± 2.29
WV-MAP-EM	0.43 ± 3.89 (0.39 ± 9.76)	-0.16 ± 4.56 (-0.16 ± 4.76)	0.18 ± 2.55 (-0.23 ± 3.76)

TABLE II
MSE(%) AND ASTD (IN BRACKETS) COMPARISONS FOR DIFFERENT
MAP METHODS.

	TV-MAP-EM		TIK-MAP-EM	
	1.969		1.788	
	(0.1572)		(0.1368)	
	WV-MAP-EM			
	Haar	Daubechies	Coiflet	Symlet
DWT	1.928	1.829	1.783	1.795
	(0.1520)	(0.1394)	(0.1385)	(0.1406)
UDWT	2.604	1.723	1.506	1.716
	(0.1075)	(0.1071)	(0.1154)	(0.1030)

and without the compensation of nonnegativity constraint. The improvement in accuracy can be clearly found out by checking the negative values of the second row. Similar conclusions can be drawn from the ROI analysis as it has been shown in Tab. I. Notice the improvement in the bias and variance performance after using the nonnegativity compensation.

Images in the first row of Fig. 7 are produced by WV-MAP-EM combined with DWT. It can be seen that the reconstruction with Haar wavelet shows piecewise smooth effects like those produced by TV-MAP-EM, while reconstructions using the other three wavelets are too similar to be visually distinguished. The second row of Fig. 7 illustrates the WV-MAP-EM reconstructions using UDWT. Clearly the images generated by this method present much higher visual quality (by comparison with those in the first row). The Haar wavelet-based WV-MAP-EM also depicts a piecewise smoothing comparable to TV-MAP-EM (as shown on Fig. 5). When the other wavelets (Daubechies, coiflet and symlet) are used, the WV-MAP-EM also provides quite visually similar images but their qualities are surprisingly higher than those in the first row. This is consistent with the theoretical studies on UDWT reported in the signal processing literature [46][47].

The MSE's corresponding to different MAP-EM methods are gathered in Tab. II. WV-MAP-EM with DWT is close to TIK-MAP-EM, while it is superior to other MAP-EM methods when UDWT is used (except for the Haar wavelet). Moreover, the coiflet seems more suitable for PET reconstruction since in this case WV-MAP-EM shows the smallest MSE.

Fig. 8 compares the vertical profile of the reference phantom data with the same profiles throughout the reconstructions as given in Fig. 7. The agreement between the reference and the MAP reconstruction is shown by a good overlap of the symbols over the reference lines. These results suggest that WV-MAP-EM also provide an edge-preserving image
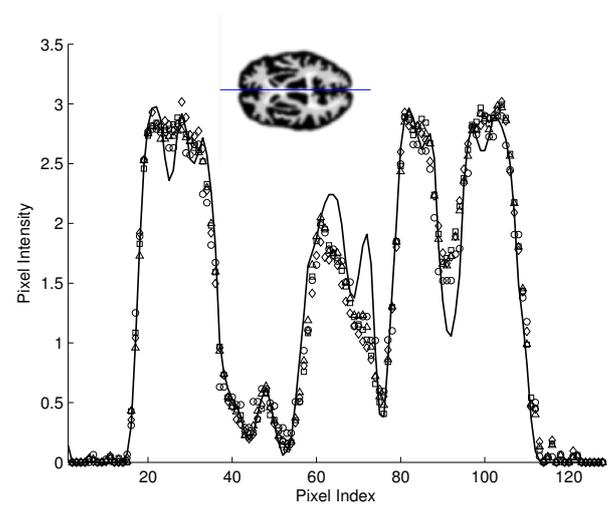
(a)
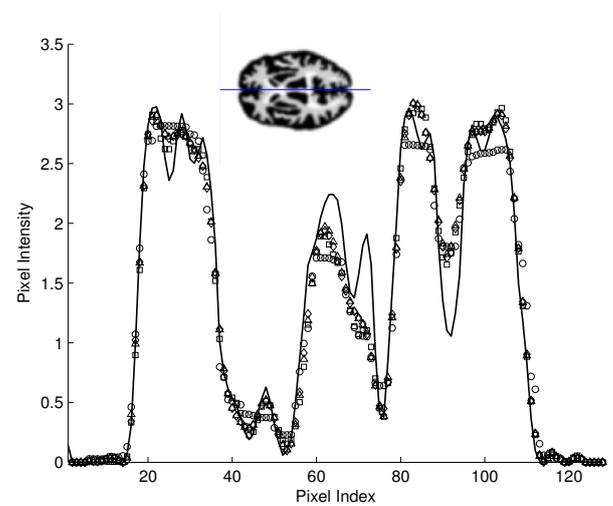
(b)

Fig. 8. Vertical profiles through WV-MAP-EM reconstructions (shown in Fig. 7) with wavelets of Haar (circles), Daubechies (diamond), Coiflet (squares) and Symlet (triangles) compared to the reference image (solid line). (a): DWT; (b): UDWT.

reconstruction.

Fig. 9 illustrates the noise properties of the WV-MAP-EM reconstruction with different types of wavelet. The images show the standard deviation at each pixel computed from the 40 replicate reconstructions. Using the Haar wavelet transform, the standard deviation of the noise appears spatially correlated. This becomes more noticeable when the undecimated Haar wavelet transform was used. Conversely, the standard deviation becomes uniform when other wavelets are used (an interesting noise texture can also be observed when using DWT-based WV-MAP-EM). To sum up, DWT (the top row of Fig. 9) presents slightly higher deviation values than UDWT (the bottom row of Fig. 9). Taking the symlet based reconstructions (the third column of Fig. 9) as an example, the ASTD with DWT was on average 25% greater than with UDWT (see Tab. II).

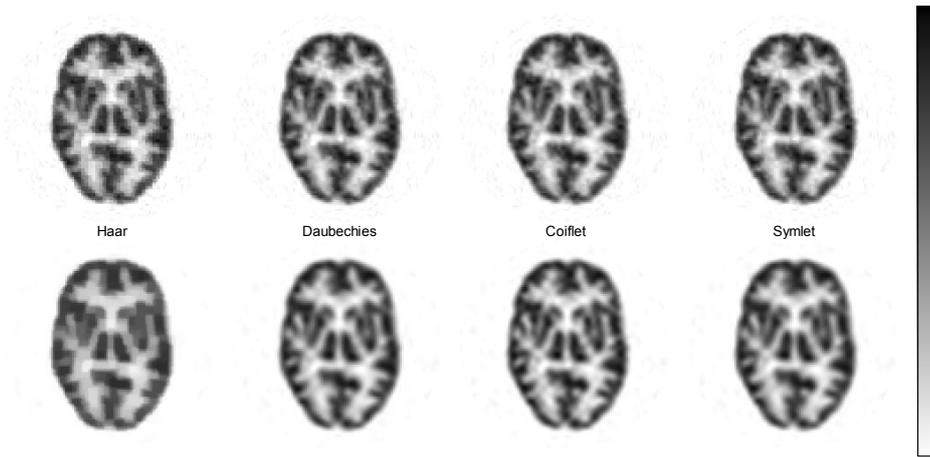Fig. 10 shows horizontal profiles crossing each impulse

Fig. 7.   WV-MAP-EM reconstructions using DWT (the top row) and UDWT (the bottom row) with $\beta = 1.497$. All wavelets have four moments equal to zero except the Haar wavelet. Note that all images are scaled to the same gray levels.
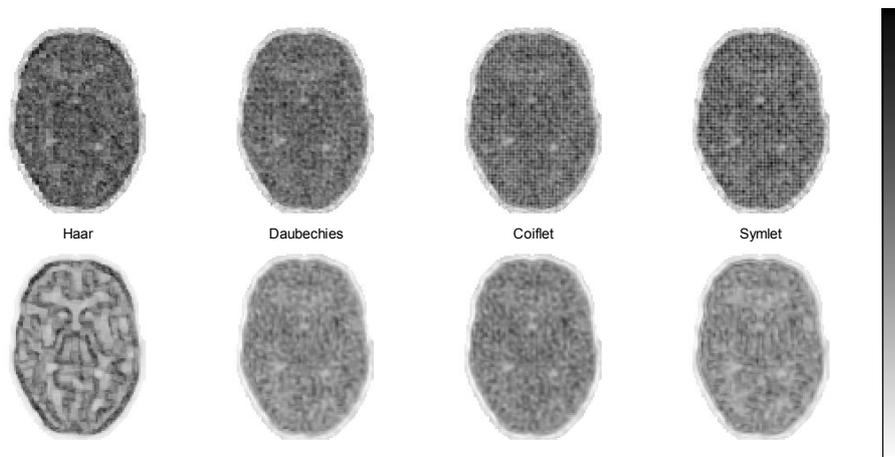

Fig. 9.   Standard deviation images computed from reconstruction of 40 replicate noise projection data. Top row: WV-MAP-EM reconstruction with DWT. Bottom row: WV-MAP-EM reconstruction with UDWT. The object mask (not shown) was used to display the valid region only. Note that all images are scaled to the same gray levels.

TABLE III
THE APPROXIMATED FWHM VALUES (UNIT IN PIXEL) AT DIFFERENT LOCATIONS IN WV-MAP-EM RECONSTRUCTIONS WITH DWT AND UDWT.

	gray matter	white matter	CSF
DWT	1.98	2.07	2.03
UDWT	2.15	2.19	2.21

TABLE IV
THE AVERAGE CPU TIME COST (UNIT IN SECOND).

TV-MAP-EM	TIK-MAP-EM	WV-MAP-EM (DWT)	WV-MAP-EM (UDWT)
0.662	0.659	0.594	0.668

drawn in the image of Fig. 10. Three profiles have been concatenated and the approximated FWHM values are listed in Tab. III. As we can see, a slight loss of resolution has been observed using UDWT-based WV-MAP-EM when compared with the DWT-based method. Note also that the resolution of the wavelet-based reconstruction was not perfectly uniform: the amount of smoothing was inversely proportional to the distance to the center.

Fig. 11 shows the average CPU costs. These results were obtained by averaging the CPU time costs for every 5 itera-

tions. The overall average CPU costs (i.e., the average value for the total 200 iterations) are listed in Tab. IV. As we can see, two image-domain MAP-EM methods, TV-MAP-EM and TIK-MAP-EM, have similar computational costs. WV-MAP-EM using DWT costs less CPU time than the other MAP-EM methods. These are inherited from fast algorithms for wavelet transform and inversion. The UDWT-based WV-MAP-EM requires more computational load than the DWT-based one. This is because the elimination of the down-sampling process increases at the same time the computational complexity of UDWT. However, the actual time consumption of the UDWT-based WV-MAP-EM is only slightly larger than those needed by the other two image-domain MAP-EM methods. Note that all computations are done with a desktop PC (Intel processor,
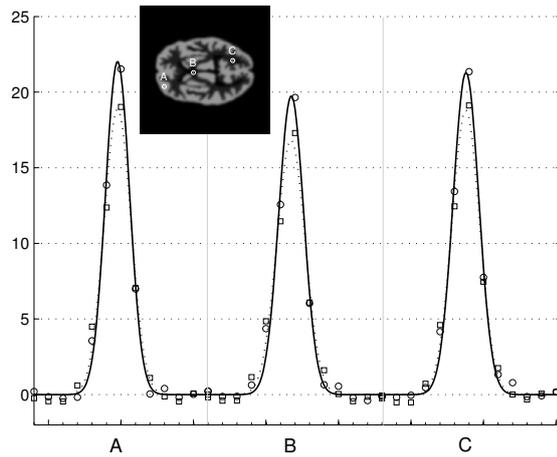
Fig. 10.   Concatenated horizontal profiles of three small impulses obtained from WV-MAP-EM reconstructions with DWT (circles) and UDWT (squares). The solid and dot lines are the approximated Gaussian kernels. The center of circles added to the image indicate the locations of the impulses. A, B and C represent the white matter, the CSF, and the gray matter respectively.
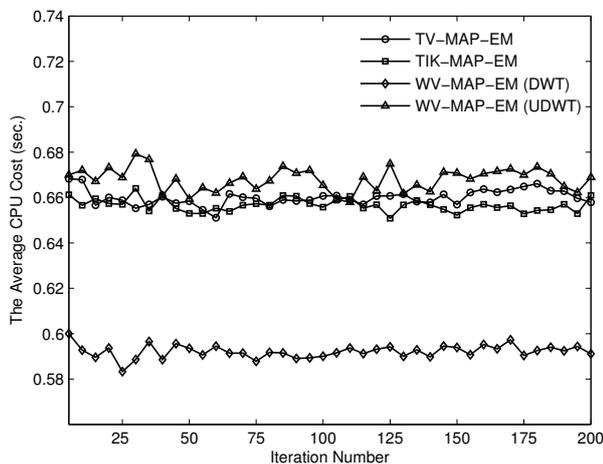

Fig. 11.   The average CPU time costs as the function of iteration. Here, the Daubechies wavelets were used in WV-EM-MAP. Each point represents an average value after every 5 iterations.

2.4-GHz CPU and 1GB RAM) using Matlab 7.1.

V. CONCLUSION AND DISCUSSION

Statistical iterative methods are used to improve the noise properties of PET reconstructions. In order to control the noise level inherent to tomographic reconstruction from noisy projections, MAP is introduced. However, the conventional MAP using MRF prior model has proven difficult to set the intrinsic hyperparameters. Although in this paper this was overcome using the greedy searching method, the computational load remains high and can not be used in clinical routine.

In the present work, we have proposed a wavelet-based MAP-EM reconstruction algorithm. Rather than reconstructing the unknown radioactivity distribution image, we were interested in recovering its wavelet coefficients beforehand. As it

has been pointed out, the resulting WV-MAP-EM method has strong relations with the image domain MAP methods but it indeed offers better performance by at least the following two aspects: 1) The hyperparameters describing the wavelet domain prior model are relatively easy to deal with, and moreover they can be adaptively selected from the system properties and the noise characteristics of measurement data; 2) The implementation of WV-MAP-EM shows computational advantages due to the availability of fast algorithms (all algorithms were implemented without any acceleration technique). Of course, this latter point is partially due to the choice of Laplacian prior distribution which led to a closed-form coefficient updating formula. Finally, the use of UDWT provided even better visual quality but with a slightly higher computation and lower resolution.

Further work is needed to combine other choices of wavelet threshold estimator, to discuss further the optimality of wavelet-based MAP reconstruction method. Fast EM optimization is also highly desired to speed up the convergence rate. Approximations describing either the covariance or the resolution property of our specific wavelet-based MAP algorithm are worth deriving. Our ongoing works are devoted to these topics, and improvements are expected to meet further clinical demands.

VI. ACKNOWLEDGMENT

REFERENCES

[1] E. Levitan and G. T. Herman, "A maximum a posteriori probability expectation maximization algorithm for image reconstruction in emission tomography," *IEEE Trans. Med. Imag.*, vol. MI-6, pp. 185–192, Sept. 1987.

[2] T. Hebert and R. Leahy, "A generalized EM algorithm for 3D Bayesian reconstruction from Poisson data using Gibbs priors," *IEEE Trans. Med. Imag.*, vol. 8, pp. 194–202, 1989.

[3] P. J. Green, "Bayesian reconstructions from emission tomography data using a modified EM algorithm," *IEEE Trans. Med. Imag.*, vol. 9, pp. 84–93, 1990.

[4] G. Winkler, Image Analysis, Random Fields and Dynamic Monte Carlo Metods. New York: Springer-Verlag, 1995.

[5] J. Fessler and W. Rogers, "Spatial resolution properties of penalized-likelihood image reconstruction: space-invariant tomographs," *IEEE Image Process.*, vol. 5, pp. 1346–1358, 1996.

[6] J. Qi and R. Leahy, "Resolution and noise properties of MAP reconstruction for fully 3D PET," *IEEE Trans. Med. Imag.*, vol. 19, pp. 493–506, 2000.

[7] J. W. Stayman and J. A. Fessler, "Regularization for uniform spatial resolution properties in penalized-likelihood image reconstruction," *IEEE Trans. Med. Imag.*, vol. 19, pp. 601–615, 2000.

[8] A. H. Delaney and Y. Bresler, "Multiresolution tomographic reconstruction using wavelets," *IEEE Trans. Image Process.*, vol. 4, pp. 799–813, June 1995.

[9] M. Bhatia, W. C. Karl, and A. S. Willsky, "A wavelet-based method for multiscale tomographic reconstruction," *IEEE Trans. Med. Imag.*, vol. 15, pp. 92–101, Feb. 1996.

[10] M. Unser and A. Aldroubi, "A review of wavelets in biomedical applications," *Proc. IEEE*, vol. 84, pp. 626–638, Apr. 1996.

[11] A. Raheja and A. P. Dhawan, "Wavelet based multiresolution expectation maximization image reconstruction algorithm for positron emission tomography," *Computerized Med. Imag. Graphics*, vol. 24, pp. 359–376, 2000.

[12] R. D. Nowak and E. D. Kolaczyk, "A statistical multiscale framework for Poisson inverse problem," *IEEE Trans. Inform. Theory*, vol. 46, pp. 1811–1825, Aug. 2000.

[13] N. Y. Lee and B. J. Lucier, "Wavelet methods for inverting the Radon transform with noisy data," *IEEE Trans. Image Process.*, vol. 10, no. 1, pp. 79–94, 2001.

[14] Y. Choi, J. Y. Koo, and N. Y. Lee, "Image reconstruction using the wavelet transform for positron emission tomography," *IEEE Trans. Med. Imag.*, vol. 20, pp. 1188–1193, Nov. 2001.

[15] T. Frese, C. A. Bouman, and K. Sauer, "Adaptive wavelet graph model for Bayesian tomographic reconstruction," *IEEE Trans. Image Process.*, vol. 11, pp. 756–770, July 2002.

[16] C. De Mol and M. Defrise, "A note on wavelet-based inversion methods, in: Inverse Problems, Image Analysis and Medical Imaging," M. Z. Nashed and O. Scherzer eds, Series "Contemporary Mathematics", vol. 313, pp. 85–96, American Mathematical Society, 2002.

[17] J. Kalifa, A. Laine, and P. D. Esser, "Regularization in tomographic reconstruction using thresholding estimators," *IEEE Trans. Med. Imag.*, vol. 22, no.3, pp. 351–359, 2003.

[18] N. Y. Lee and Y. Choi, "A modified OSEM algorithm for PET reconstruction using wavelet processing," *Computer Methods and Programs in Biomedicine*, vol. 80, pp. 236–245, 2005.

[19] G. McLachlan, and T. Krishnan, The EM Algorithm and Extensions. New York: Wiley, 1996.

[20] M. A. T. Figueiredo, and R. D. Nowak, "An EM algorithm for wavelet-based image restoration," *IEEE Trans. Image Process.*, vol. 12, pp. 906–916, Aug. 2003.

[21] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Comm. Pure Appl. Math*, vol. 57, pp.1413–1541, 2004.

[22] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, pp. 674–693, 1989.

[23] S. G. Mallat, A Wavelet Tour of Signal Processing. second ed. San Diego: Academic Press, 1999.

[24] K. Huang, Introduction to Statistical Physics. London: Taylor & Francis, 2001.

[25] J. A. Fessler, "Penalized weighted least squares image reconstrcution for positron emission tomography," *IEEE Trans. Med. Imag.*, vol. 13, pp. 290–300, June, 1994.

[26] Q. Li, E. Asma, J. Qi, *et al*, "Accurate estimate of the fisher information matrix for the PET image reconstruction problem," *IEEE Trans. Med. Imag.*, vol. 23, pp. 1057–1064, 2004.

[27] E. P. Simoncelli and E. H. Adelson, "Noise removal via Bayesian wavelet coring," *In Third Internat. Conf. on Imaging Processing, IEEE Signal Processing Society*, Lausanne, Switzerland, 1999, pp. 279–382.

[28] A. Achim, A. Bezerianos, and P. Tsakalides, "Novel Bayesian multiscale method for speckle removal in medical ultrasound images," *IEEE Trans. Med. Imag.*, vol. 20, pp. 772–783, 2001.

[29] S. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Trans. Image Process.*, vol. 9, pp. 1522–1531, 2000.

[30] L. Boubchir and J. M. Fadili, "A closed-form nonparametric Bayesian estimator in the wavelet domain of images using an approximate α-stable prior," *Pattern Recognition Letters*, vol. 27, pp. 1370–1382, 2006.

[31] P. Moulin and J. Liu, "Analysis of multiresolution image denoising schemes using generalized-gaussian and complexity priors," *IEEE Trans. Inform. Theory.*, vol. 38, pp. 719–747, 1998.

[32] M. Belge, M. E. Kilmer, and E. L. Miller, "Wavelet domain image restoration with adaptive edge-preserving regularization," *IEEE Trans. Image Processing*, vol. 9, pp. 597–608, 2000.

[33] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Stat. Soc.*, vol. 29, pp. 1–38, 1977.

[34] L. A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *IEEE Trans. Med. Imag.*, vol. MI-1, pp. 113–122, 1982.

[35] K. Lange and R. Carson, "EM reconstruction algorithms for emission and transmission tomography," *J. Comput. Assist. Tomogr.*, vol. 8, pp. 306–316, 1984.

[36] M. Feder and E. Weinstein, "Parameter estimation of superimposed signals using the EM algorithm," *IEEE Trans. Acoust. Speech, Signal Process.*, vol. 36, no. 4, pp. 477–489, Apr. 1988.

[37] J. A. Fessler and A. O. Hero, "Space-alternating generalized expectation-maximization algorithm," *IEEE Trans. Signal. Process.*, vol. 42, no. 10, pp. 2664–2677, 1994.

[38] J. Demmel, Applied Numerical Linear Algebra. SIAM, Philadelphia, 1997.

[39] G. Golub and C. Van Loan. Matrix Computations. The Johns Hopkins University Press, Baltimore, third edition, 1996.

[40] J. O. Berger, Statistical Decision Theory and Bayesian Analysis. New York: Springer-Verlag, 1985.

[41] D. L. Donoho, "Denoising by soft-thresholding," *IEEE Trans. Inform. Theroy.*, vol. 41, pp. 613–627, 1995.

[42] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–455, 1994.

[43] J. Qi and R. M. Leahy, "A theoretical study of the contrast recovery and variance of MAP reconstructions from PET data," *IEEE Trans. Med. Imag.*, vol. 18, pp. 293–305, Apr. 1999.

[44] A. R. De Pierro and M. E. B. Yamagishi, "Fast EM-like methods for maximum 'a posteriori' estimates in emission tomography," *IEEE Trans. Med. Imag.,* vol. 20, pp. 280–288, 2001.

[45] C. Bouman and K. Sauer, "A generalized Gaussian image model for edge-preserving MAP estimation," *IEEE Trans. Image Process.,* vol. 2, pp. 296–310, 1993.

[46] R. R. Coifman and D. L. Donoho, "Translation-invariant de-noising," *in Wavelet and Statistics, Lecture Notes in Statistics*, pp. 125–150, Berlin: Springer Verlag, 1995.

[47] M. Lang, H. Guo, J. Odegard, C. S. Burrus, and R. Wells, "Noise reduction using an undecimated discrete wavelet transform," *IEEE Signal Process. Lett.*, vol. 3, pp. 10–12, Jan. 1995.