COMPUTED TOMOGRAPHY RECONSTRUCTION BASED ON A HIERARCHICAL MODEL AND VARIATIONAL BAYESIAN METHOD

Li Wang[⋆] *Ali Mohammad-Djafari*[⋆†] *Nicolas Gac*[⋆] *Mircea Dumitru*[⋆]

[⋆] Laboratoire des Signaux et Systèmes, CentraleSupelec, France
[†]CNRS, France

ABSTRACT

In order to improve the quality of X-ray Computed Tomography (CT) reconstruction for Non Destructive Testing (NDT), we propose a hierarchical prior modeling with a Bayesian approach. In this paper we present a new hierarchical structure for the inverse problem of CT by using a multivariate Student-t prior which enforces sparsity and preserves edges. This model can be adapted to the piecewise continuous image reconstruction problems. We demonstrate the feasibility of this method by comparing with some other state of the art methods. In this paper, we show simulation results in 2D where the image is the middle slice of the Shepp-Logan object but the algorithms are adapted to the big data size problem, which is one of the principal difficulties in the 3D CT reconstruction problem.

Index Terms— Computed Tomography (CT), Non Destructive Testing (NDT), Hierarchical Model, Bayesian Method, Student-t prior models

1. INTRODUCTION.

The concept of an inverse problem is now a familiar one to most scientists and engineers, particularly in the field of signal and image processing. In a nutshell, it involves the estimation of an unknown quantity from another observable quantity which is linked to it through a mathematical relationship known as the forward model. In this paper, we are concerned with CT. The domains of CT application range from medical imaging to NDT. X-ray computed tomography (X-ray CT) is nowadays a common technology in NDT. In X-ray tomography, the reconstructed 3D image represents the attenuation function of the volume being examined. The 3D X-ray CT image reconstruction, which is recently a popular topic in inverse problem, is usually implemented in GPU to reduce the computation time [1].

To model and solve an inverse problem, firstly a mathematical model is considered. In this paper, Cartesian coordinates is used. Different coordinates can also be applied according to the character of model. [2] has presented a 3D CT reconstruction using polar coordinates, and [3] has introduced the method using hexagonal grids. Various methods have

been used for doing the inversion in the forward model. The most widely used is MAP estimation method [4], the Markov Chain Monte Carlo (MCMC) method [5], Discrete Algebraic Reconstruction Technique (DART) reconstruction method [6] and the super-resolution method [7]. In this article we adapt a Bayesian approach. The Bayesian approach can be done via a hierarchical model, which assigns priors for the image and the noise. In this Bayes framework, by using the priors and the likelihood, we obtain the posterior distrubution, from which we estimate the unknowns of the model, together with their corresponding variances, via two techniques: the Posterior Mean (PM) estimation via Variational Bayesian Approach (VBA) and the Joint Maximum A Posterior (JMAP).

One of the estimation that we used is the PM via VBA. More details of VBA are given in [4,8–10]. For the Bayesian-based methods, by assuming the prior distributions and likelihood distribution, the posterior distribution can be obtained. Different assumptions for the prior distribution will encode different prior knowledgement. Some prior models are discussed in [11–15].

In the mathematical model, $\boldsymbol{f} \in \mathbb{R}^N$ represents all the pixels or voxels of the object and $\boldsymbol{g} \in \mathbb{R}^M$ represents all the projection data. The projection system is denoted by the linear matrix $\boldsymbol{H} \in \mathbb{R}^{M \times N}$, which depends on the geometry of the acquisition system. The forward model is expressed as:

$$\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\epsilon}, \tag{1}$$

with $\boldsymbol{\epsilon} \in \mathbb{R}^M$ representing the noise.

The inverse problem consists of using the real projection data to infer the original object $\boldsymbol{f}$. Filtered Back Projection (FBP): $\widehat{\boldsymbol{f}} = \left(\boldsymbol{H}^t \boldsymbol{H}\right)^{-1} \boldsymbol{H}^t \boldsymbol{g}$ is a common inversion methods.

Generally, the problem is ill-posed. To solve this problem, the regularization theory and Bayesian inversion are widely used. The regularization approach reconstructs $\boldsymbol{f}$ by minimizing the criterion:

$$\widehat{\boldsymbol{f}} = \arg \min_f \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2 + \lambda \mathcal{R}\left(\boldsymbol{f}\right) \tag{2}$$

where λ is a regularization parameter and $\mathcal{R}\left(\boldsymbol{f}\right)$ indicates the criterion penalty, for example $\mathcal{R}\left(\boldsymbol{f}\right) = \|\boldsymbol{f}\|^2$, $\mathcal{R}\left(\boldsymbol{f}\right) = |\boldsymbol{f}|$. Our Bayesian method uses a hierarchical model with proba-

bility distribution functions (pdf) for the prior on $\boldsymbol{f}$, the likelihood function from the data, and relevant parameters.

The rest of this paper is organized as follows: in Section 2, the Bayesian method is presented, Section 3 provides the new hierarchical model. Different reconstruction algorithms are introduced in Section 4. We present the experimental results in Section 5, and conclusions are drawn in Section 6.

2. BAYESIAN APPROACH BASIS.

Using the forward model, Equation (1), and assigning the likelihood $p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{H})$ and the prior law $p(\boldsymbol{f})$, the Bayesian Rule provides the posterior law:

$$p(\boldsymbol{f}|\boldsymbol{g}, \boldsymbol{H}) = \frac{p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{H})p(\boldsymbol{f})}{p(\boldsymbol{g}|\boldsymbol{H})} \qquad (3)$$

where

$$p(\boldsymbol{g}|\boldsymbol{H}) = \int p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{H})p(\boldsymbol{f})d\boldsymbol{f} \qquad (4)$$

We proceed to infer the unknown $\boldsymbol{f}$ by using the MAP and PM algorithms.

3. THE HIERARCHICAL MODEL.

Firstly, we assign a pdf for ϵ, the noise in Equation (1). We consider a non-stationary noise, so we propose a zero-mean Normal distribution, with unkown variance v_ϵ. In order to estimate the variance, we assign it a pdf. We choose an Inverse Gamma (IG) with parameters α and β defining expected mean and uncertainty.

$$\begin{cases} p(\boldsymbol{\epsilon}|v_\epsilon) = \mathcal{N}(\boldsymbol{\epsilon}|0, v_\epsilon \boldsymbol{I}) \\ p(v_\epsilon|\alpha_{\epsilon_0}, \beta_{\epsilon_0}) = \mathcal{IG}(v_\epsilon|\alpha_{\epsilon_0}, \beta_{\epsilon_0}) \end{cases} \qquad (5)$$

The image $\boldsymbol{f}$ can be expressed as:

$$\boldsymbol{f} = \boldsymbol{D}\boldsymbol{z}, \qquad (6)$$

where $\boldsymbol{z}$ is the sparse representation of $\boldsymbol{f}$ and $\boldsymbol{D}$ is the corresponding sparse transformation, for example, the wavelet transformation. Student-t probability density function is assigned to $\boldsymbol{z}$ in order to enforces the sparsity and thus preserves the edges of image. Many other distribution could also be used to enforce sparsity. Those can be divided in 3 categories: Generalized Gaussian (GG), mixture of Gaussian or heavy-tailed distribution. Student-t is one of the heavy-tailed distributions, with the convenient that it can be expressed as the marginal of a Normal-IG distribution.

In this paper, we estimate the edges $\boldsymbol{z}$ first, with the image $\boldsymbol{f}$ following.

The Student-t distribution can be expressed as the marginal of a Gaussian-IG distribution:

$$\mathcal{S}t(x|\alpha, \beta) = \int_0^\infty \mathcal{N}(x|0, v)\mathcal{IG}(v|\alpha, \beta) \, dv \qquad (7)$$

Assuming that $\boldsymbol{z}$ is separable, i.e. $p(\boldsymbol{z}) = \prod_j^N p(z_j)$, we can easily represent the prior distribution for all the elements of the vector $\boldsymbol{z}$:

$$\begin{aligned} p(z_j|\alpha_{z_{j0}}, \beta_{z_{j0}}) &= \mathcal{S}t\left(z_j|\alpha_{z_{j0}}, \beta_{z_{j0}}\right) \\ &= \int_0^\infty \mathcal{N}(z_j|0, v_{z_j})\mathcal{IG}(v_{z_j}|\alpha_{z_{j0}}, \beta_{z_{j0}})dv_{z_j} \end{aligned} \qquad (8)$$
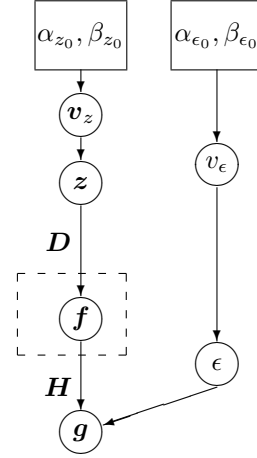

Fig. 1: The Hierarchical Model.

Equation 5, together with the Student-t prior model proposed to $\boldsymbol{f}$ which is expressed as in Equation 8, gives the proposed hierarchical model. A graphical representation of this hierarchical model is shown in Figure 1, and the corresponding equations in Equation 9.

$$\begin{cases} p(\boldsymbol{g}|\boldsymbol{z}, v_\epsilon) = \mathcal{N}(\boldsymbol{g}|\boldsymbol{H}\boldsymbol{D}\boldsymbol{z}, v_\epsilon \boldsymbol{I}) \\ p(\boldsymbol{z}|\boldsymbol{v}_z) = \mathcal{N}(\boldsymbol{z}|0, \boldsymbol{V}_z), \ \boldsymbol{V}_z = \text{diag}\left[\boldsymbol{v}_z\right], \boldsymbol{v}_z = [v_{z_1}, \cdots, v_{z_N}] \\ p(\boldsymbol{v}_z|\alpha_{z_0}, \beta_{z_0}) = \prod_j^N \mathcal{IG}(v_{z_j}|\alpha_{z_0}, \beta_{z_0}) \\ p(v_\epsilon|\alpha_{\epsilon_0}, \beta_{\epsilon_0}) = \mathcal{IG}(v_\epsilon|\alpha_{\epsilon_0}, \beta_{\epsilon_0}) \end{cases}$$
$$(9)$$

The posterior distribution is obtained from Equation 9 via Bayes rule:

$$\begin{aligned} p(\boldsymbol{z}, \boldsymbol{v}_z, v_\epsilon|\boldsymbol{g}) &\propto p(\boldsymbol{g}|\boldsymbol{z}, v_\epsilon)p(\boldsymbol{z}|\boldsymbol{v}_z) \\ &\quad p(\boldsymbol{v}_z|\alpha_{z_0}, \beta_{z_0})p(v_\epsilon|\alpha_{\epsilon_0}, \beta_{\epsilon_0}). \end{aligned} \qquad (10)$$

From here we can apply various algorithms to solve the reconstruction problem. In this paper, we will discuss the JMAP method and the PM via VBA method.

4. RECONSTRUCTION ALGORITHMS.

The main Bayesian computational tools are: Maximum A Posteriori (MAP) method [4], Markov Chain Monte Carlo (MCMC) method [5], Expectation-Maximization (EM) method [16] and the Variational Bayesian (VB) method.

In this paper, we will mainly consider the MAP method and the VBA methods.

4.1. The Joint MAP algorithm.

In JMAP method, we estimate $\boldsymbol{f}$ and the other parameters involved in the model, i.e. the variances v_ϵ and $\boldsymbol{v}_z$, iteratively by maximizing the posterior law:

$$(\widehat{\boldsymbol{z}}, \widehat{\boldsymbol{v}}_z, \widehat{v}_\epsilon) = \arg \max_{(\boldsymbol{z}, \boldsymbol{v}_z, v_\epsilon)} \{ p(\boldsymbol{z}, \boldsymbol{v}_z, v_\epsilon | \boldsymbol{g}) \} \quad (11)$$

As we have mentioned above, the size of data in our research is so huge that we can not calculate directly the inverse of matrix in the steps of updating $\boldsymbol{z}$. Instead, we use the gradient descent method to avoid the costly computation. The updating rules are summarized below:

$$\begin{cases} \textbf{iter}: \widehat{\boldsymbol{z}}^{(k+1)} = \widehat{\boldsymbol{z}}^{(k)} - \widehat{\gamma}_z^{(k)} \nabla \mathcal{J}(\widehat{\boldsymbol{z}}^{(k)}) \\ \widehat{v}_{z_j} = \frac{\beta_{z_{j0}} + \frac{1}{2} \widehat{z}_j^2}{\alpha_{z_{j0}} + \frac{3}{2}} \\ \widehat{v}_\epsilon = \frac{\beta_{\epsilon_0} + \frac{1}{2} \|\boldsymbol{g} - \boldsymbol{HD}\widehat{\boldsymbol{z}}\|^2}{\alpha_{\epsilon_0} + \frac{M}{2} + 1} \end{cases} \quad (12)$$

where

$$\begin{cases} \mathcal{J}(\boldsymbol{z}) = \frac{1}{2} \widehat{v}_\epsilon^{-1} \|\boldsymbol{g} - \boldsymbol{HD}\boldsymbol{z}\|^2 + \frac{1}{2} \left\| \widehat{\boldsymbol{Y}} \boldsymbol{z} \right\|^2, \ \ \widehat{\boldsymbol{Y}} = \widehat{\boldsymbol{V}}_z^{-\frac{1}{2}} \\ \widehat{\gamma}_z^{(k)} = \frac{\left\| \nabla \mathcal{J}(\widehat{\boldsymbol{z}}^{(k)}) \right\|^2}{\widehat{v}_\epsilon^{-1} \left\| \boldsymbol{HD} \nabla \mathcal{J}(\widehat{\boldsymbol{z}}^{(k)}) \right\|^2 + \left\| \widehat{\boldsymbol{Y}} \nabla \mathcal{J}(\widehat{\boldsymbol{z}}^{(k)}) \right\|^2} \end{cases}$$

and $\nabla \mathcal{J}(\cdot)$ is the gradient of $\mathcal{J}(\cdot)$.

4.2. The VBA algorithm.

We compute the PM via VBA method. Firstly, the posterior distribution $p(\boldsymbol{z}, \boldsymbol{v}_z, v_\epsilon | \boldsymbol{g})$ is approached by a separable distribution $q(\boldsymbol{z}, \boldsymbol{v}_z, v_\epsilon)$:

$$p(\boldsymbol{z}, \boldsymbol{v}_z, v_\epsilon | \boldsymbol{g}) \approx q(\boldsymbol{z}, \boldsymbol{v}_z, v_\epsilon) = q_1(\boldsymbol{z}) q_2(\boldsymbol{v}_z) q_3(v_\epsilon). \quad (13)$$

This approaximation is done by minimizing the Kullback-Leibler divergence, defined as:

$$\begin{aligned} &\mathcal{KL}\left(q(\boldsymbol{z}, \boldsymbol{v}_z, v_\epsilon) : p(\boldsymbol{z}, \boldsymbol{v}_z, v_\epsilon | \boldsymbol{g}) \right) \\ &= \int q(\boldsymbol{z}, \boldsymbol{v}_z, v_\epsilon) \ln \frac{q(\boldsymbol{z}, \boldsymbol{v}_z, v_\epsilon)}{p(\boldsymbol{z}, \boldsymbol{v}_z, v_\epsilon | \boldsymbol{g})} d\boldsymbol{z} d\boldsymbol{v}_z dv_\epsilon \end{aligned} \quad (14)$$

The details of this minimization are discussed in [17]. Thanks to the conjugacy property of the prior model, we obtain the conjugate distributions for $\boldsymbol{z}$ and the parameters.

$$q(z_j) = \mathcal{N}(z_j | \widetilde{m}_{z_j}, \widetilde{\sigma}_{z_j}^2) \text{ where}$$

$$\begin{cases} \widetilde{m}_{z_j} = \frac{\langle \widetilde{v}_\epsilon^{-1} \rangle \widetilde{b}_j}{\langle \widetilde{v}_\epsilon^{-1} \rangle h_j + \langle \widetilde{v}_{z_j}^{-1} \rangle} \\ \widetilde{\sigma}_{z_j}^2 = \frac{1}{\langle \widetilde{v}_\epsilon^{-1} \rangle h_j + \langle \widetilde{v}_{z_j}^{-1} \rangle} \end{cases}$$

$$q(v_{z_j}) = \mathcal{IG}(v_{z_j} | \widetilde{\alpha}_{z_j}, \widetilde{\beta}_{z_j}) \text{ where}$$

$$\begin{cases} \widetilde{\alpha}_{z_j} = \alpha_{z_0} + \frac{1}{2} \\ \widetilde{\beta}_{z_j} = \beta_{z_0} + \frac{1}{2} \left(\widetilde{m}_{z_j}^2 + \widetilde{\sigma}_{z_j}^2 \right) \end{cases}$$

$$q(v_\epsilon) = \mathcal{IG}(v_\epsilon | \widetilde{\alpha}_\epsilon, \widetilde{\beta}_\epsilon) \text{ where}$$

$$\begin{cases} \widetilde{\alpha}_\epsilon = \alpha_{\epsilon_0} + \frac{M}{2} \\ \widetilde{\beta}_\epsilon = \beta_{\epsilon_0} + \frac{1}{2} \left(\|\boldsymbol{g} - \boldsymbol{HD}\widetilde{\boldsymbol{m}}_z\|^2 + \sum_j h_j \widetilde{\sigma}_{z_j}^2 \right) \end{cases} \quad (15)$$

where

$$\begin{cases} h_j = \left[\boldsymbol{D}^t \boldsymbol{H}^t \boldsymbol{HD} \right]_{jj} \\ \widetilde{b}_j = \left[\boldsymbol{D}^t \boldsymbol{H}^t \left(\boldsymbol{g} - \boldsymbol{HD}\widetilde{\boldsymbol{m}}_z \right) \right]_j + h_j \widetilde{m}_{z_j} \\ \langle \widetilde{v}_\epsilon^{-1} \rangle = \frac{\widetilde{\alpha}_\epsilon}{\widetilde{\beta}_\epsilon}, \ \left\langle \widetilde{v}_{z_j}^{-1} \right\rangle = \frac{\widetilde{\alpha}_{z_j}}{\widetilde{\beta}_{z_j}} \end{cases}$$

5. NUMERICAL EXPERIMENTATION.

Several experiments in 2D image analyse have been carried out in order to demonstrate the properties of the proposed model. The original image is the middle slice of Shepp-Logan of size 128×128. The sparse operator $\boldsymbol{D}$ is the Haar wavelet transform. The projection matrix $\boldsymbol{H}$ is only used in the geometry projection $\boldsymbol{Hf}$ and back-projection $\boldsymbol{H}^t \boldsymbol{g}$ operators. The reconstructed images presented in Figure 2 show that the proposed hierarchical model can lead to results with more details and preserve better the edges of image.

The relative errors of reconstruction $\delta_f = \left\| \boldsymbol{f} - \widehat{\boldsymbol{f}} \right\|^2 / \|\boldsymbol{f}\|^2$ using different algorithms are shown in Figure 3. Figure 3(a) corresponds to a projection without noise while Figure 3(b) corresponds to a projection with 20dB noise. We compared the JMAP and VBA method using the proposed model with Least Square (LS) method, quadratic regularization (QR) method with fixed λ parameter and Gauss Markov Potts (GMP) methods. With the proposed model, both the JMAP and VBA algorthms works better than the non-hierarchical models, and their performance are as good as the GMP method which also can be considered as a hierarchical one. The JMAP method with the three-level hierarchical model has a better convergence property than the other methods.

Figure 4 illustrate the influence of the noise and number of projections in our application. Figure 4(a) shows that the error of reconstruction is in inverse proportion to the noise value. Figure 4(b) indicates that by using 180 projections and 90
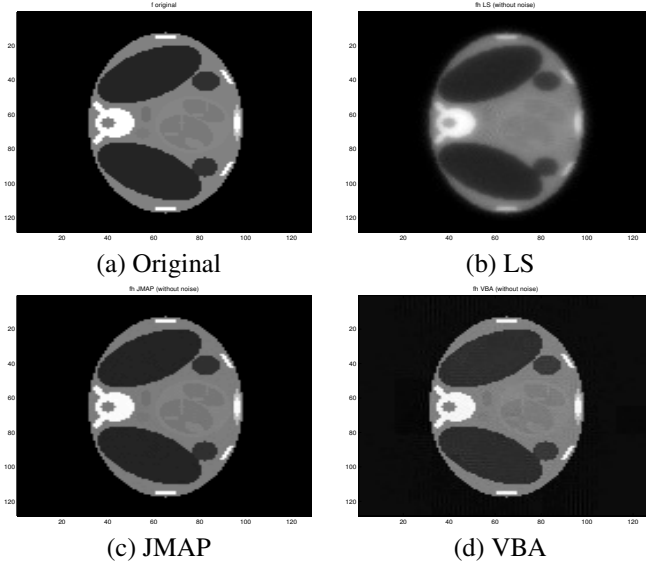
(a) Original   (b) LS

(c) JMAP   (d) VBA

Fig. 2: Images reconstructed from unnoised projection using different methods, with 180 projections and 30 iterations.

(a)

(b)

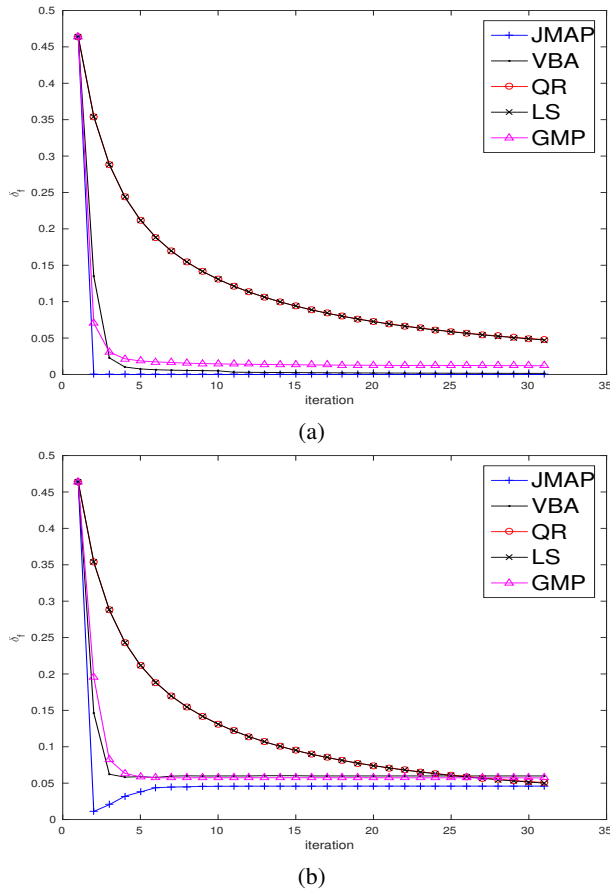**Fig. 3**: Comparaison of the convergency of different methods using (a) a projection without noise and (b) a projection with SNR=20dB.
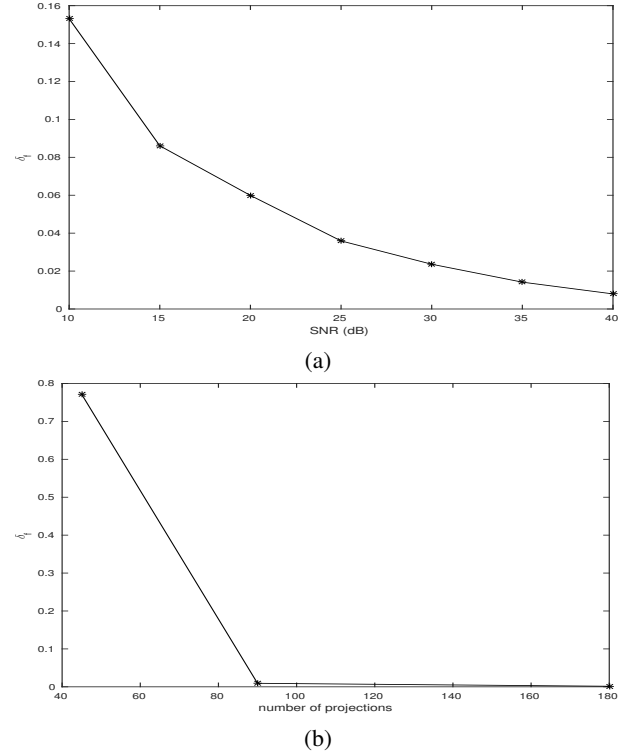

(a)

(b)

Fig. 4: Performance of the different methods as a function of (a) SNR with a great number of projections and (b) different number of projections. All the results are obtained with 30 iterations.

projections, the results won't change a lot, but 45 projections is not enough for a big data CT reconstruction problem.

6. CONCLUSION

We presented a three-level hierarchical model for the piece-wise continuous tomography reconstruction problem. We used a heavy-tailed Student-t distribution to model the sparse representation of image and the errors in order to achieve the sparsity property and model the non-stationality of the noise. Experimental results demonstrate the advantages of using the proposed three-level hierarchical model and a Student-t distribution for the prior model. Future work is to apply this method to the 3D CT problem where the GPU processor is used to reduce the computation time.

7. REFERENCES

[1] Nicolas Gac, Alexandre Vabre, Ali Mohammad-Djafari, et al., "Multi gpu parallelization of 3d bayesian ct algorithm and its application on real foam reonconstruction with incomplete data set," *Proceedings FVR 2011*, pp. 35–38, 2011.

[2] Christian Thibaudeau, Jean-Daniel Leroux, Réjean Fontaine, and Roger Lecomte, "Fully 3d iterative ct reconstruction using polar coordinates," *Medical physics*, vol. 40, no. 11, pp. 111904, 2013.

[3] Michael Knaup, Sven Steckmann, Olivier Bockenbach, and Marc Kachelrieß, "Ct image reconstruction using hexagonal grids," in *Nuclear Science Symposium Conference Record, 2007. NSS'07. IEEE*. IEEE, 2007, vol. 4, pp. 3074–3076.

[4] Ali Mohammad-Djafari, "Gauss-markov-potts priors for images in computer tomography resulting to joint optimal reconstruction and segmentation," *International Journal of Tomography & Simulation*, vol. 11, no. W09, pp. 76–92, 2009.

[5] Dani Gamerman and Hedibert F Lopes, *Markov chain Monte Carlo: stochastic simulation for Bayesian inference*, CRC Press, 2006.

[6] Kees Joost Batenburg and Jan Sijbers, "Dart: a practical reconstruction algorithm for discrete tomography," *Image Processing, IEEE Transactions on*, vol. 20, no. 9, pp. 2542–2553, 2011.

[7] W. Van Aarle, Kees Joost Batenburg, G Van Gompel, Elke Van de Casteele, and Jan Sijbers, "Super-resolution for computed tomography based on discrete tomography," *IEEE Transactions on Image Processing*, vol. 23, pp. 1181 – 1193, 03/2014 2014.

[8] Michael I Jordan, Zoubin Ghahramani, Tommi S Jaakkola, and Lawrence K Saul, "An introduction to variational methods for graphical models," *Machine learning*, vol. 37, no. 2, pp. 183–233, 1999.

[9] Dimitris G Tzikas, CL Likas, and Nikolaos P Galatsanos, "The variational approximation for bayesian inference," *Signal Processing Magazine, IEEE*, vol. 25, no. 6, pp. 131–146, 2008.

[10] Dimitris Tzikas, Aristidis Likas, and Nikolaos Galatsanos, "Variational bayesian blind image deconvolution with student-t priors," in *Image Processing, 2007. ICIP 2007. IEEE International Conference on*. IEEE, 2007, vol. 1, pp. I–109.

[11] Xiaojing Gu, Henry Leung, and Xingsheng Gu, "Bayesian sparse estimation using double lomax priors," *Mathematical Problems in Engineering*, vol. 2013, 2013.

[12] Michael Ting, Raviv Raich, and Alfred O Hero, "Sparse image reconstruction for molecular imaging," *Image Processing, IEEE Transactions on*, vol. 18, no. 6, pp. 1215–1227, 2009.

[13] Yang Chen, Jianhua Ma, Qianjin Feng, Limin Luo, Pengcheng Shi, and Wufan Chen, "Nonlocal prior bayesian tomographic reconstruction," *Journal of Mathematical Imaging and Vision*, vol. 30, no. 2, pp. 133–146, 2008.

[14] Ali Mohammad-Djafari and Lionel Robillard, "Hierarchical markovian models for 3d computed tomography in non destructive testing applications," *EUSIPCO, Florence, Italy*, 2006.

[15] Rafael Molina, Antonio López, José Manuel Martín, and Aggelos K Katsaggelos, "Variational posterior distribution approximation in bayesian emission tomography reconstruction using a gamma mixture prior.," in *VISAPP (Special Sessions)*, 2007, pp. 165–176.

[16] Václav Šmídl and Anthony Quinn, *The variational Bayes method in signal processing*, Springer, 2006.

[17] Ali Mohammad-Djafari, "Bayesian inference with hierarchical prior models for inverse problems in imaging systems," in *Systems, Signal Processing and their Applications (WoSSPA), 2013 8th International Workshop on*. IEEE, 2013, pp. 7–18.