Bayesian X-ray Computed Tomography using a Three-level Hierarchical Prior Model

Li Wang, Ali Mohammad-Djafari, Nicolas Gac

Bayesian X-ray Computed Tomography using a Three-level Hierarchical Prior Model

Li Wang[1,2,a)], Ali Mohammad-Djafari[1,2,3,b)] and Nicolas Gac[1,2,c)]

[1]*Laboratoire des Signaux et Systmes, CentraleSuplec, France*
[2]*Universit Paris Sud, France*
[3]*CNRS, France*

[a)]Corresponding author: li.wang@lss.supelec.fr
[b)]djafari@lss.supelec.fr
[c)]nicolas.gac@lss.supelec.fr

Abstract. In recent decades X-ray Computed Tomography (CT) image reconstruction has been largely developed in both medical and industrial domain. In this paper, we propose using the Bayesian inference approach with a new hierarchical prior model. In the proposed model, a generalised Student-t distribution is used to enforce the Haar transformation of images to be sparse. Comparisons with some state of the art methods are presented. It is shown that by using the proposed model, the sparsity of the sparse representation of images is enforced, so that edges of images are preserved. Simulation results are also provided to demonstrate the effectiveness of the new hierarchical model for reconstruction with fewer projections.
Keywords: Computed Tomography (CT), Bayesian Method, Hierarchical Model, Sparsity, Student-t distribution, Inverse Problem, Variational Bayesian Approach (VBA), Joint Maximum A Posterior (JMAP), Infinite Gaussian Scaled Mixture (IGSM)

INTRODUCTION

Computed Tomography (CT) has been a widely used technique in recent decades. Research on X-ray CT has been developing fast in both medical and industrial applications. By 1990, the famous Filtered Back Projection (FBP) had been developed. In the X-ray CT field, research started to move into the fully 3D domain, as medical and industrial scanners turned to 3D reconstruction methods and iterative reconstruction schemes. .
In X ray CT, the intensity of X ray is attenuated when passing through the object, and the parameters to be reconstructed is the linear attenuation coefficient inside the object under the test. The relation between the measured quantity g and the unknown distribution of the attenuation coefficient $f(x, y)$ is given by:

$$g = -\ln \frac{\boldsymbol{I}}{\boldsymbol{I}_0} = \int f(x, y) \mathrm{d}l \tag{1}$$

where (x, y) is the coordinate position of the object, $\boldsymbol{I}_0$ the initial X ray intensity, $\boldsymbol{I}$ the detected X ray intensity after passing through the object and dl the infinite signal length on the line joining the emitter and the detector.

Considering that different tissue has different attenuation coefficient values, for example the metal and air material in a component of industry applications, one can therefore distinguish the structure information through the attenuated data. Often, in order to distinguish the details of the object, a high resolution image is needed. It is therefore important to pay attention to the computational aspects, for example using the GPU processor as presented in [2]. A more troublesome case is when we consider reconstructing a dynamically changing object, for example a beating heart, or a component on a moving conveyor. This lead to the problem of reconstruction with less measuring time, hence less number of projections.
The Radon Transform (RT), with details introduced in [1], is one of the most commonly used forward modelling

approach that treats the X ray CT projections, with the expression:

$$g(r,\phi) = \mathcal{R}f(x,y) = \int f(x,y)\,\delta(x\cos\phi + y\sin\phi - r)\,\mathrm{d}x\,\mathrm{d}y \tag{2}$$

where x, y define the position of pixel value in the image, while r, ϕ define the perpendicular length from center point and the angle of X ray under consideration. Based on this transformation and its analytical inversion, there have been many analytical reconstruction methods. We may mention here one of the main methods called Filtered Back Projection (FBP) which can be summarized as $\widehat{f} = \mathcal{B}\boldsymbol{F}^{-1}\,|\Omega|\,\boldsymbol{F}g$ where $\boldsymbol{F}$ and $\boldsymbol{F}^{-1}$ are direct and inverse Fourier transform, $|\Omega|$ is a modulated filter and $\mathcal{B}$ is the Back-Projection operator. In order to solve the reconstruction problem algebraically, the discretization of the model gives the projection rule for each ray i through the pixel j:

$$g[i] = \sum_j H[i,j]f[j] \tag{3}$$

By adding up all the projection rays and all the direction of projections, the synthetic forward operator is available, with the object represented by a vector $\boldsymbol{f} \in \mathbb{R}^{N\times 1}$, projections represented by vector $\boldsymbol{g} \in \mathbb{R}^{M\times 1}$ and the linear projection system represented by matrix $\boldsymbol{H} \in \mathbb{R}^{M\times N}$. Accounting for additive noise of the linear system $\boldsymbol{\epsilon} \in \mathbb{R}^{M\times 1}$, the forward model is:

$$\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\epsilon} \tag{4}$$

According to this model, two basic analytical reconstruction methods find their discretized version: Back-Projection (BP) $\widehat{\boldsymbol{f}} = \boldsymbol{H}'\boldsymbol{g}$ and the Filtered Back-Projection (FBP) $\widehat{\boldsymbol{f}} = \boldsymbol{H}\left(\boldsymbol{H}\boldsymbol{H}'\right)^{-1}\boldsymbol{g}$ with analytical and algebraical details presented in [3]. The BP gives a rough blurred reconstruction image, while FBP is an efficient way of reconstruction when there are a great number of projections uniformly distributed around the object. In the cases with strong noise or with insufficient projections, it is no longer robust.

Deterministic Regularization based methods

Many image processing problems are ill-posed. It is necessary to regularize the solution by imposing an a priori constraint. Mathematically, this constraint is often expressed through a regularization function:

$$\mathcal{J}(\boldsymbol{f}) = \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2 + \lambda\mathcal{R}(\boldsymbol{f}) \tag{5}$$

where λ is called regularization parameter and it controls the tradeoff between data-model adequation and the prior knowledge expressed through the regularization term $\mathcal{R}(\boldsymbol{f})$. Different choices on the regularization terms are for example:

- L2 (Ridge regression): $\mathcal{R}(\boldsymbol{f}) = \|\boldsymbol{f}\|_2^2$
- L2 (Quadratic Regularization): $\mathcal{R}(\boldsymbol{f}) = \|\boldsymbol{D}\boldsymbol{f}\|_2^2$ where $\boldsymbol{D}$ is a linear operator.
- L1 (Lasso): $\mathcal{R}(\boldsymbol{f}) = \|\boldsymbol{f}\|_1$
- Total Variation: $\mathcal{R}(\boldsymbol{f}) = \|\boldsymbol{G}\boldsymbol{f}\|_1$ where $\boldsymbol{G}$ represents the gradient operator.
- Total Variation for images: $\mathcal{R}(\boldsymbol{f}) = \|\boldsymbol{G}_x\boldsymbol{f}\|_1 + \left\|\boldsymbol{G}_y\boldsymbol{f}\right\|_1$ where $\boldsymbol{G}_x$ and $\boldsymbol{G}_y$ represent correspondingly the horizontal gradient and vertical gradient.

From another point of view, regularization can also impose constraints on the coefficients $\boldsymbol{z}$ of the signal $\boldsymbol{f}$ in a proper basis or frame; i.e., $\boldsymbol{f} = \boldsymbol{D}\boldsymbol{z}$, known as synthesis prior presented in [4]:

$$\mathcal{J}(\boldsymbol{z}) = \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{D}\boldsymbol{z}\|^2 + \lambda\mathcal{R}(\boldsymbol{z}) \quad \text{and} \quad \widehat{\boldsymbol{f}} = \boldsymbol{D}\widehat{\boldsymbol{z}} \tag{6}$$

where $\boldsymbol{D}$ is the synthesis operator of a decomposition of $\boldsymbol{f}$ on a dictionary (e.g., gradient, wavelets, etc).

Classical regularization techniques for variational image restoration include constrained Least Square, Tikhonov-Miller method. Since the invention of Total Variation (TV) based regularization, variational image deblurring and its

extensions have received much more attention. The other class of competing schemes in the past decade were sparsity-based regularization. The class of l_1 regularized optimization problems has received much attention because of the presence of an l_1 regularization term, optimization problems are still very difficult to solve. Various of methods have been studied for example the Newton's method presented in [5] and the Split Bregman method presented in [6]. By using these kinds of classical optimization methods, we can get reconstruction results by setting suitable regularization parameters.
Choosing suitable regularization parameter values is necessary and challenging, and very often costly. The ways of chosing this parameter are for example Cross Validation [7] and L-Curve [8] regularization. By adding different regularization terms, l_2 or l_1, different constraints are considered. l_2 norm criterion gives globally smooth results, and enforce a roughness penalty on the solution. l_1 regularization, on the other hand, enforces the sparsity and therefore preserves edges of image during reconstruction.

Sparsity enforcing prior models

Normally the sparse property of a statistical variable is enforced by using three kinds of distributions:

- The Generalized Gaussian distributions: $p(x) = \frac{\beta}{2\alpha\Gamma(1/\beta)} \exp\left\{-\left(\frac{|x-\mu|}{\alpha}\right)^{\beta}\right\}$
- The Gaussian Mixture distributions: $p(x) = \sum_j^S w_j \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu_j}{\sigma_j}\right)^2\right\}$
- The heavy tailed distributions, of which the tails are not exponentially bounded.

The standard Student-t distribution is a heavy tailed one, with the expression:

$$\mathcal{S}t(x|\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}} \tag{7}$$

From the definition of its variance we easily figure out that there is a limit of the variance of Student-t distribution: $\mathrm{Var}[f] = \frac{\nu}{\nu-2} = 1 + \frac{2}{\nu-2} > 1$. This limit lead to the consequence that this heavy-tailed distribution can't have a small variance, therefore the sparsity couldn't be intensively enforced. In this paper we use a generalization of Student-t distribution (Stg) which can be obtained by using the Normal-Inverse Gamma marginalization property:

$$\mathcal{S}t_g(f|\alpha,\beta) = \int \mathcal{N}(f|0,v)\mathcal{IG}(v|\alpha,\beta)\,\mathrm{d}v = \frac{\Gamma(\alpha+1/2)}{\sqrt{2\pi\beta}\Gamma(\alpha)}\left(1+\frac{f^2}{2\beta}\right)^{-\left(\alpha+\frac{1}{2}\right)} \tag{8}$$

This generalization of Student-t distribution adds a supplementary parameter to the standard one, and hence is more able to control the sparsity rate of prior distribution.

Linear sparse transformation

In many applications, in particular in Non Destructive Testing (NDT), the objects are piece-wise homogeneous. Noticing that in the homogeneous zone, the gradient is zero, and the non-zero values only exist at the edges of the image. This gives a very convenient property to use the sparsity constraints. A piece-wise continuous image can then be represented by different type of sparse transforms. We use matrix $\boldsymbol{D}$ to represent the chosen transformation operator which gives a linear relationship between image and the sparse transform coefficients $\boldsymbol{z}$ with the relationship $\boldsymbol{f} = \boldsymbol{D}\boldsymbol{z}$. In this paper, the multilevel Haar Transformation, which is a form of the Discrete Wavelet Transform (DWT), is used, with benefits that it is an orthonormal operator with the property $\boldsymbol{D}'\boldsymbol{D} = \boldsymbol{I}$ which simplifies the mathematical computations and above all, the inverse operator $\boldsymbol{D}^{-1}$ and the transpose operator $\boldsymbol{D}'$ are interchangeable. The multilevel Haar Transformation with level l of an image $\boldsymbol{f} \in \mathbb{R}^{m\times n}$ is defined as:

$$\text{for } k = 1 \text{ to } l, \quad \boldsymbol{f}_{\mathcal{H}}^{(k)} = \mathcal{H}aar\left(\boldsymbol{f}_{\mathcal{H}}^{(k-1)}\left(1:\frac{m}{2^{k-1}}, 1:\frac{n}{2^{k-1}}\right)\right), \quad \boldsymbol{f}_{\mathcal{H}}^{(0)} = \boldsymbol{f} \text{ and } \boldsymbol{z} = \boldsymbol{f}_{\mathcal{H}}^{(l)} \tag{9}$$

Fig.(1) shows the original Shepp-Logan image, its gradient and its corresponding 5-level Haar transformation coefficients. The gradient of the image represents precisely the contours, and similarly for the high frequency coefficients

in the MH transform. By enforcing the sparsity of either the gradient or the high frequency coefficients of the MH transform, the edges of the image could be preserved. The benefit of the MH transform is that it is inversible and orthonormal.

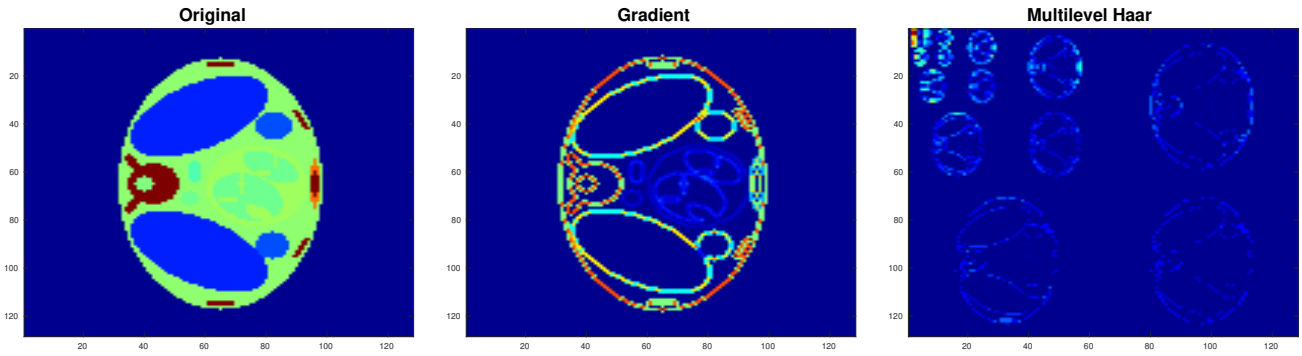


FIGURE 1: Relation of original image, its gradient and its multilevel Haar transform coefficients.

While using this linear dictionry representation, an additive white noise in always considered:

$$\boldsymbol{f} = \boldsymbol{D}\boldsymbol{z} + \boldsymbol{\xi} \tag{10}$$

The Proposed Unsupervised Hierarchical Haar Transform based Model (HHBM)

The Bayesian methods gives possibilities to do the reconstructions in an unsupervised way, meaning estimating variables as well as parameters. By using the Bayesian method, the prior informations can be combined with the data, and most importantly, it provides a convenient setting for a wide range of models adapting to different prior informations. In this paper, we propose a new hierarchical a priori model for the Bayesian inference in which the sparsity of coefficients of sparse transformation is enforced by using a heavy tailed prior distribution. In this method, both the image and the transformation coefficients are estimated alternatively.
To consider the estimation problem from a Bayesian point of view, we first define a likelihood based on the given data generation model and then we introduce sparsity to our estimate by assigning a suitable heavy-tailed prior distribution over the parameter vector $\boldsymbol{z}$. Generally, the noise in a linear model is supposed to be a independent and identically distributed (iid) white noise, thus belonging to an i.i.d. Gaussian noise with variance $\boldsymbol{v}_\epsilon$:

$$p(\boldsymbol{\epsilon}|\boldsymbol{v}_\epsilon) = \mathcal{N}(\boldsymbol{\epsilon}|\boldsymbol{0}, \boldsymbol{V}_\epsilon) \quad \text{where } \boldsymbol{V}_\epsilon = \text{diag}\,[\boldsymbol{v}_\epsilon] \tag{11}$$

According to the forward model shown in Eq.(4), the distribution of the additive noise gives exactly the distribution of the discrepancy of $\boldsymbol{g}$ and $\boldsymbol{H}\boldsymbol{f}$ and hence the likelihood is:

$$p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{v}_\epsilon) = \mathcal{N}(\boldsymbol{g}|\boldsymbol{H}\boldsymbol{f}, \boldsymbol{V}_\epsilon) \tag{12}$$

For the representation of the prior statistical model of image, the multilevel Haar Transformation is used as shown in Eq.(10), where the prior distribution is defined as a Normal distribution while the noise belonging to a Gaussian distribution $p(\boldsymbol{\xi}|\boldsymbol{v}_\xi) = \mathcal{N}(\boldsymbol{\xi}|0, \boldsymbol{V}_\xi)$:

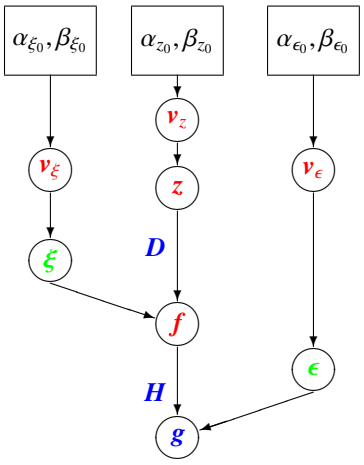
$$p(\boldsymbol{f}|\boldsymbol{z}, \boldsymbol{v}_\xi) = \mathcal{N}(\boldsymbol{f}|\boldsymbol{D}\boldsymbol{z}, \boldsymbol{V}_\xi) \quad \text{where } \boldsymbol{V}_\xi = \text{diag}\left[\boldsymbol{v}_\xi\right] \tag{13}$$

The Multilevel Haar Transformation of the image is considered as a sparse representation and a Generalized Student-t (Stg) distribution is used to enforce the sparsity property. The Stg distribution for coefficients $\boldsymbol{z}$: $p(\boldsymbol{z}) = \mathcal{S}t_g(\boldsymbol{z}|\alpha_{z_0}, \beta_{z_0})$ can be defined by a hierarchical structure as below:

$$p(\boldsymbol{z}|0, \boldsymbol{v}_z) = \mathcal{N}(\boldsymbol{z}|0, \boldsymbol{v}_z) \tag{14}$$

$$p(\boldsymbol{v}_z|\alpha_{z_0}, \beta_{z_0}) = \mathcal{IG}(\boldsymbol{v}_z|\alpha_{z_0}, \beta_{z_0}) \tag{15}$$

where $\alpha_{z_0}, \beta_{z_0} > 0$ are the two hyper-parameters of the hierarchical structured sparsity enforcing Generalised Student-t distribution. In particular, by initializing different α_{z_0} and β_{z_0} value, the scale of sparsity rate of distribution can be controlled.
Lastly, considering the positive property of the variances of noise, as well as that major part of the variance are near zero, the Inverse Gamma (IG) distribution is used to define variables $\boldsymbol{v}_\epsilon$ and $\boldsymbol{v}_\xi$. In this case, the additive noises are therefore belonging to the Generalized Student-t distribution.
A directed acyclic graph (DAG) of the proposed hierarchical Bayesian Haar Transform based model is shown in Fig.(2). The corresponding statistical structure model is on the right.



$$p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{v}_\epsilon) = \mathcal{N}(\boldsymbol{g}|\boldsymbol{H}\boldsymbol{f}, \boldsymbol{V}_\epsilon) \text{ where } \boldsymbol{V}_\epsilon = \text{diag}[\boldsymbol{v}_\epsilon]$$
$$p(\boldsymbol{f}|\boldsymbol{z}, \boldsymbol{v}_\xi) = \mathcal{N}(\boldsymbol{f}|\boldsymbol{D}\boldsymbol{z}, \boldsymbol{V}_\xi)$$
$$p(\boldsymbol{z}|\boldsymbol{v}_z) = \mathcal{N}(\boldsymbol{z}|0, \boldsymbol{V}_z) \text{ where } \boldsymbol{V}_z = \text{diag}[\boldsymbol{v}_z]$$
$$p(\boldsymbol{v}_z|\alpha_{z_0}, \beta_{z_0}) = \prod_j^N \mathcal{IG}(v_{z_j}|\alpha_{z_0}, \beta_{z_0})$$
$$p(\boldsymbol{v}_\epsilon|\alpha_{\epsilon_0}, \beta_{\epsilon_0}) = \prod_i^M \mathcal{IG}(v_{\epsilon_i}|\alpha_{\epsilon_0}, \beta_{\epsilon_0})$$
$$p(\boldsymbol{v}_\xi|\alpha_{\xi_0}, \beta_{\xi_0}) = \prod_j^N \mathcal{IG}(v_{\xi_j}|\alpha_{\xi_0}, \beta_{\xi_0})$$

FIGURE 2: Directed acyclic graph of proposed model.

The posterior model and JMAP estimation

So far we have presented the hierarchical structured prior treating the variables and parameters. To proceed with Bayesian inference, the computation of the joint posterior distribution over the variables and parameters is required. Using Bayes' law, this distribution is expressed as shown in Eq.(16):

$$p(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{v}_z, \boldsymbol{v}_\epsilon, \boldsymbol{v}_\xi|\boldsymbol{g}) \propto p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{v}_\epsilon)\, p(\boldsymbol{f}|\boldsymbol{z}, \boldsymbol{v}_\xi)\, p(\boldsymbol{z}|\boldsymbol{v}_z)\, p(\boldsymbol{v}_z|\alpha_{z_0}, \beta_{z_0})\, p(\boldsymbol{v}_\epsilon|\alpha_{\epsilon_0}, \beta_{\epsilon_0})\, p(\boldsymbol{v}_\xi|\alpha_{\xi_0}, \beta_{\xi_0}) \tag{16}$$

Where $\boldsymbol{z}$ represents the wavelet coefficients of the object $\boldsymbol{f}$, $\boldsymbol{D}$ the corresponding transform dictionary, $\boldsymbol{v}_z$ variance of $\boldsymbol{z}$, $\boldsymbol{v}_\epsilon$ variance of noise $\boldsymbol{\epsilon}$ and v_ξ variance of noise $\boldsymbol{\xi}$.

In this paper, the Joint Maximum A Posterior (JMAP) estimator is applied to estimate the variables and parameters. By using JMAP optimization, all the variables and parameters are optimized alternatively by maximizing the posterior distribution. The optimizing rule is expressed in Eq.(17):

$$[\widehat{\boldsymbol{f}}, \widehat{\boldsymbol{z}}, \widehat{\boldsymbol{v}}_z, \widehat{\boldsymbol{v}}_\epsilon, \widehat{\boldsymbol{v}}_\xi] = \arg \max_{\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{v}_z, \boldsymbol{v}_\epsilon, \boldsymbol{v}_\xi} p(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{v}_z, \boldsymbol{v}_\epsilon, \boldsymbol{v}_\xi|\boldsymbol{g}) \tag{17}$$

By estimating alternatively, the updating rule of all the variables and parameters are listed in Eq.(18)-Eq.(22):

$$\textbf{iter}: \widehat{\boldsymbol{f}}^{(k+1)} = \widehat{\boldsymbol{f}}^{(k)} - \widehat{\gamma}_f^{(k)} \nabla \mathcal{J}(\widehat{\boldsymbol{f}}^{(k)}) \tag{18}$$
$$\textbf{iter}: \widehat{\boldsymbol{z}}^{(k+1)} = \widehat{\boldsymbol{z}}^{(k)} - \widehat{\gamma}_z^{(k)} \nabla \mathcal{J}(\widehat{\boldsymbol{z}}^{(k)}) \tag{19}$$
$$\widehat{v}_{z_j} = \frac{\beta_{z_0} + \frac{1}{2}\widehat{z}_j^2}{\alpha_{z_0} + 3/2} \tag{20}$$
$$\widehat{v}_{\epsilon_i} = \frac{\beta_{\epsilon_0} + \frac{1}{2}\left(g_i - \left[\boldsymbol{H}\widehat{\boldsymbol{f}}\right]_i\right)^2}{\alpha_{\epsilon_0} + 3/2} \tag{21}$$

$$\widehat{v}_{\xi_j} = \frac{\beta_{\xi_0} + \frac{1}{2}\left(\widehat{f}_j - [\boldsymbol{D}\widehat{\boldsymbol{z}}]_j\right)^2}{\alpha_{\xi_0} + 3/2} \tag{22}$$

where

$$\mathcal{J}(\boldsymbol{f}) = \frac{1}{2}(\boldsymbol{g} - \boldsymbol{H}\boldsymbol{D}\boldsymbol{z})' \boldsymbol{V}_\epsilon^{-1}(\boldsymbol{g} - \boldsymbol{H}\boldsymbol{D}\boldsymbol{z}) + \frac{1}{2}(\boldsymbol{f} - \boldsymbol{D}\boldsymbol{z})' \boldsymbol{V}_\xi^{-1}(\boldsymbol{f} - \boldsymbol{D}\boldsymbol{z}) \tag{23}$$

$$\mathcal{J}(\boldsymbol{z}) = \frac{1}{2}(\boldsymbol{f} - \boldsymbol{D}\boldsymbol{z})' \boldsymbol{V}_\xi^{-1}(\boldsymbol{f} - \boldsymbol{D}\boldsymbol{z}) + \frac{1}{2}\boldsymbol{z}'\boldsymbol{V}_z^{-1}\boldsymbol{z} \tag{24}$$

$$\widehat{\gamma}_f^{(k)} = \frac{\left\|\nabla\mathcal{J}(\widehat{\boldsymbol{f}}^{(k)})\right\|^2}{\left\|\widehat{\boldsymbol{Y}}_\epsilon \boldsymbol{H}\nabla\mathcal{J}(\widehat{\boldsymbol{f}}^{(k)})\right\|^2 + \left\|\widehat{\boldsymbol{Y}}_\xi \nabla\mathcal{J}(\widehat{\boldsymbol{f}}^{(k)})\right\|^2} \quad \text{where } \widehat{\boldsymbol{Y}}_\epsilon = \widehat{\boldsymbol{V}}_\epsilon^{-\frac{1}{2}} \text{ and } \widehat{\boldsymbol{Y}}_\xi = \widehat{\boldsymbol{V}}_\xi^{-\frac{1}{2}} \tag{25}$$

$$\widehat{\gamma}_z^{(k)} = \frac{\left\|\nabla\mathcal{J}(\widehat{\boldsymbol{z}}^{(k)})\right\|^2}{\left\|\widehat{\boldsymbol{Y}}_\xi \boldsymbol{D}\nabla\mathcal{J}(\widehat{\boldsymbol{z}}^{(k)})\right\|^2 + \left\|\widehat{\boldsymbol{Y}}_z \nabla\mathcal{J}(\widehat{\boldsymbol{z}}^{(k)})\right\|^2} \quad \text{where } \widehat{\boldsymbol{Y}}_z = \widehat{\boldsymbol{V}}_z^{-\frac{1}{2}} \tag{26}$$

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

Computational aspects and GPU implementation

For the updating of $\boldsymbol{f}$ and $\boldsymbol{z}$, considering the big data size constraints and the impossibility of calculating the inverse of projection matrix $\boldsymbol{H}$, the gradient descent optimization algorithm is used, in which the parameters γ_f and γ_z in Eq.(25) and Eq.(26) are the corresponding descent step length. In this paper, considering orthonormal property, $\boldsymbol{D}' = \boldsymbol{D}^{-1}$, the matrix $\boldsymbol{D}^t$ corresponds to the multilevel Haar transformation operator, and $\boldsymbol{D}$ corresponds to the inverse multilevel Haar transformation. The covariance matrix $\boldsymbol{V}_z$, $\boldsymbol{V}_\epsilon$ and $\boldsymbol{V}_\xi$ are diagonal matrix, so that the inverse of them are simply inverse of all the diagonal elements. By using this method, all the calculations can be applied to a 3D CT reconstruction case. The projection, back-projection and Haar transformation operators could also be accelerated by using GPU processor.

Simulation Results

In the simulations, the middle slice image of the Shepp Logan object is used as the original image. The simulation image has size 128×128 and it consists of several homogeneous zones, each of which corresponds to a material. Thus it has a piece-wise continuous property. Parallel projections are applied in angles uniformly distributed from 0 to 180 degree. The reconstruction performance is measured in terms of the nomalized mean square error (NMSE), or the relative error δ_f, which is defined as:

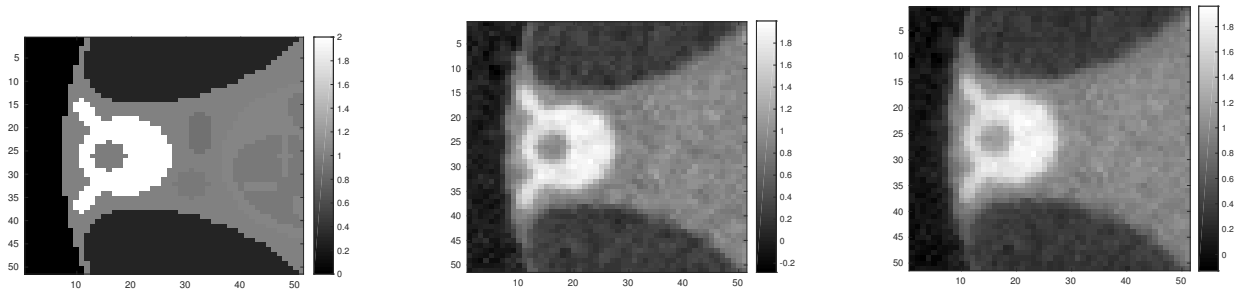
$$\delta_f = \text{NMSE} = \left\|\widehat{\boldsymbol{f}} - \boldsymbol{f}\right\|^2 / \|\boldsymbol{f}\|^2 \tag{27}$$

In Fig.(3), zoon of the reconstructed images by using different methods are listed, with the NMSE of reconstructed images shown below the corresponding image.

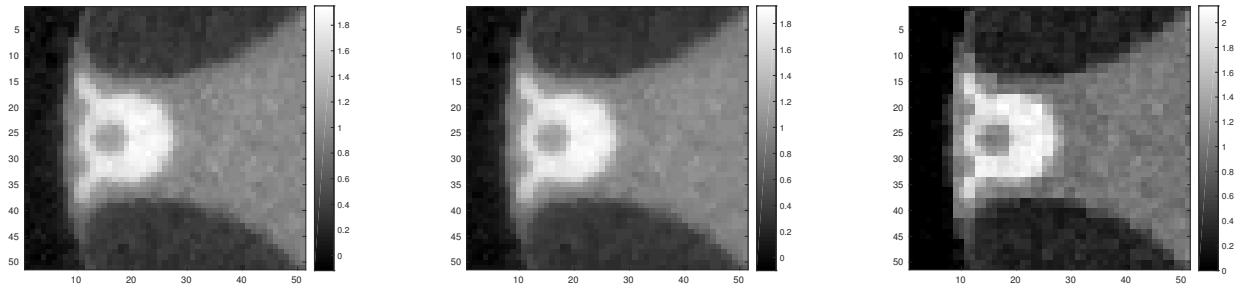
By using the Haar Transform base method, while enforcing the sparsity of the sparse transformation coefficients, the edges of the image are better preserved.

In Fig.(4) the profiles of the reconstructed images are compared. The HHBM method outperforms the other method on protecting edges.

The convergence of the relative error of reconstruction of different methods are shown in Fig.(5). Fig.(5)(a) compares different methos by using 128 projections, and Fig.(5)(b) compares the results by using 64 projections. Results proved that the HHBM method stay robust while using insufficient data.



Original FBP (NMSE = 0.3202) QR (NMSE = 0.0560)



Lasso (NMSE = 0.0541) TV (NMSE = 0.0528) HHBM (NMSE = 0.0380)

FIGURE 3: Zoon of reconstructed images with different methods: Filtered Back-Projection (FBP), Quadratic Regularization (QR), Lasso, Total Variation (TV) and the proposed HHBM, using 128 projection data and SNR=20dB.

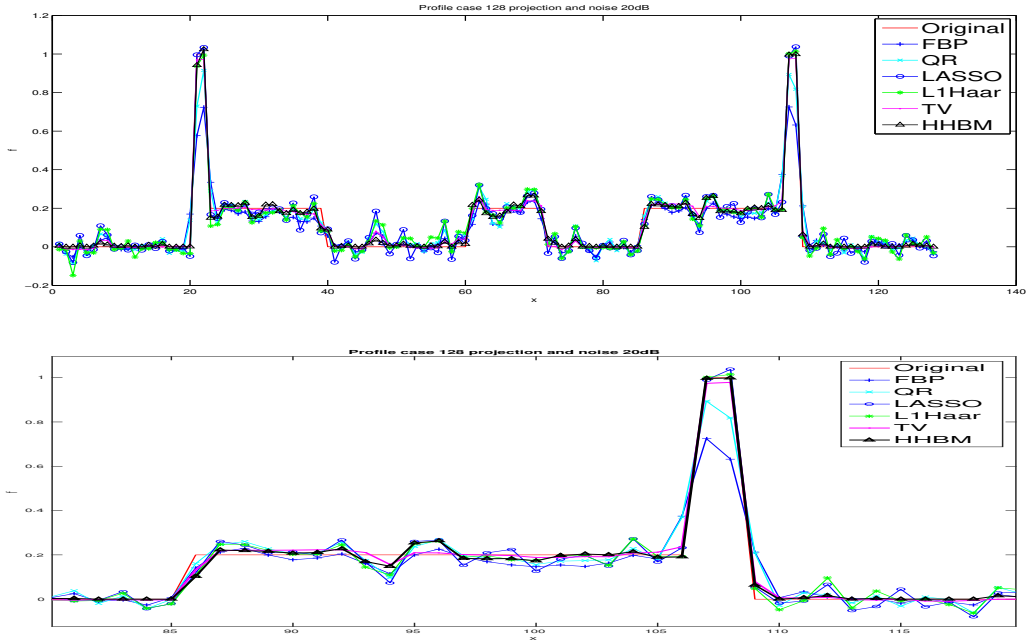


FIGURE 4: Comparison of a) profiles of reconstructed images and b) zone of the profiles. Results obtained within 128 projections and SNR=20dB.

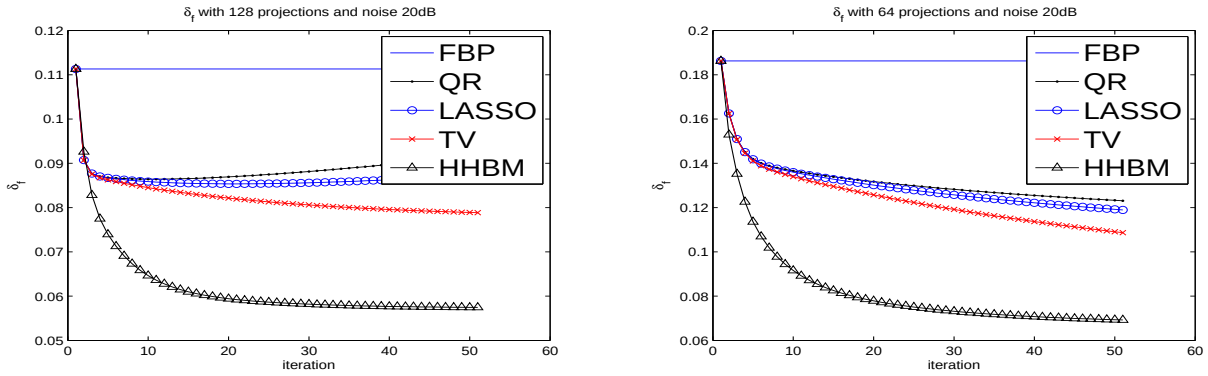


FIGURE 5: Relative error of reconstruction with 20dB noised data and a) 128 projections and b) 64 projections.

Conclusions and Perspectives

In this paper, we have introduced a new hierarchical Haar Transformation based model for the reconstruction of piece-wise continuous X ray Computed Tomography images. Reconstruction using this model enforces the sparsity of the sparse transformation coefficients and hence preserves the edges of reconstructed images. We show that by using our proposed method, the reconstruction of piece-wise continuous image will be obtained with better preserved edges and less projections.

Currently, we are working on the extension of the method to 3D applications, and using a Dual-Tree Discrete Wavelet Transformation as the sparse coefficients. Improved results with respect to noise are expected, due to the robust property of the DT-DWT. Further, we are planing more extended real data tests, including the huge data size industrial Non Destructive Testing one.

REFERENCES

[1] Rolf Clackdoyle and Michel Defrise. Tomographic reconstruction in the 21st century. *IEEE Signal Processing Magazine*, 27(4):60–80, 2010.

[2] 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*, pages 35–38, 2011.

[3] Ali Mohamad-Djafari. *Inverse problems in vision and 3D tomography*. John Wiley & Sons, 2013.

[4] Michael Elad, Peyman Milanfar, and Ron Rubinstein. Analysis versus synthesis in signal priors. *Inverse problems*, 23(3):947, 2007.

[5] Tony F Chan, Gene H Golub, and Pep Mulet. A nonlinear primal-dual method for total variation-based image restoration. *SIAM journal on scientific computing*, 20(6):1964–1977, 1999.

[6] Tom Goldstein and Stanley Osher. The split bregman method for l1-regularized problems. *SIAM journal on imaging sciences*, 2(2):323–343, 2009.

[7] Ron Kohavi et al. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Ijcai*, volume 14, pages 1137–1145, 1995.

[8] Per Christian Hansen and Dianne Prost O'Leary. The use of the l-curve in the regularization of discrete ill-posed problems. *SIAM Journal on Scientific Computing*, 14(6):1487–1503, 1993.