Implicit Gibbs Prior Models for Tomographic Reconstruction

Pengchong Jin
School of ECE, Purdue University
West Lafayette, IN, 47907
Email: jin36@purdue.edu

Eri Haneda
School of ECE, Purdue University
West Lafayette, IN, 47907
Email: haneda@ge.com

Charles A. Bouman
School of ECE, Purdue University
West Lafayette, IN, 47907
Email: bouman@purdue.edu

***Abstract*—Bayesian model-based inversion has been applied to many applications, such as tomographic reconstructions. However, one limitation of these methods is that prior models are quite simple; so they are not capable of being trained to statistically represent subtle detail in images.**

In this paper, we demonstrate how novel prior modeling methods based on implicit Gibbs distributions can be used in MAP tomographic reconstruction to improve reconstructed image quality. The concept of the implicit Gibbs distribution is to model the image using the conditional distribution of each pixel given its neighbors and to construct a local approximation of the true Gibbs energy from the conditional distribution. Since the conditional distribution can be trained on a specific dataset, it is possible to obtain more precise and expressive models of images which capture unique structures. In practice, this results in a spatially adaptive MRF model, but it also provides a framework that assures convergence. We present results comparing the proposed method with both state-of-the-art MRF prior models and K-SVD dictionary-based methods for tomographic reconstruction of images. Simulation results indicate that the proposed method can achieve higher resolution recovery.

I. INTRODUCTION

Recently, model-based iterative reconstruction (MBIR) has been shown to be effective in the reconstruction of X-ray CT data [1]–[3]. These algorithms have the advantage that they can incorporate more accurate models of both forward acquisition processes and the objects being reconstructed. More specifically, they fall into the general framework of Bayesian model-based inversion, and the reconstruction is computed as the maximum a posterior (MAP) estimate of the unknown image x from the indirect measurement y given by

$$\hat{x} = \arg\max_{x \geq 0} \{\log p(y|x) + \log p(x)\} \tag{1}$$

where $p(y|x)$ is the likelihood function corresponding to the forward projection model and $p(x)$ is the prior distribution of image x used for regularization.

For imaging problems, the most commonly-used prior is the Markov random field (MRF) with the following explicit form

$$p(x) = \frac{1}{z} \exp\left\{ - \sum_{\{s,r\} \in \mathcal{C}} b_{s,r} \rho(x_s - x_r) \right\} \tag{2}$$

where ρ is the non-negative, symmetric and absolutely non-decreasing potential function, $\mathcal{C}$ is the set of all pairwise cliques in the neighborhood system and $b_{s,r}$ are the weights for neighboring pixel pairs. While this simple model has been widely used [4], the limitation of this model is that the distribution is not very expressive and thus it is not well adapted to local image structures because of the pre-designed and fixed potential functions.

It is also worth noticing that recent research in image denoising indicates that dictionary-based regularization techniques have the potential to substantially improve results as compared to classical MAP image restoration methods [5]. Some research has been extended to inverse problems such as MRI [6] and image compression [7]. Nevertheless, these methods are generally difficult to be adapted to the classical Bayesian inverse framework that is widely used in problems such as tomographic reconstructions.

In this paper, we demonstrate how novel prior modeling methods based on implicit Gibbs distributions [8] can be applied to MAP tomographic reconstruction. The concept of the implicit Gibbs distribution is to model the image using the conditional distribution of each pixel given its neighbors. It is then possible to compute a local approximation of the true Gibbs distribution from the conditional distribution. Since the conditional distribution of the MRF can be learned from specific training data, it is possible to obtain more precise and expressive models of images which capture unique characteristics. In practice, this results in spatially adaptive MRF models, but it also provides a framework that assures convergence.

We present simulation results comparing our new method with both state-of-art MRF prior models and K-SVD dictionary-based methods for tomographic reconstruction of images. Results indicate that the proposed method has the potential to produce reconstructions with higher resolution and finer detail.

II. STATISTICAL MODEL FOR TOMOGRAPHIC RECONSTRUCTION

Let $x \in \mathbb{R}^M$ be the image vector and $y \in \mathbb{R}^N$ be the tomographic projection measurement. In the Bayesian statistical framework, both x and y are considered as random, and the reconstruction is computed as the maximum a posterior

This research was supported by ALERT DHS center Northeastern University. Pengchong Jin and Charles A. Bouman is currently with Purdue University. Eri Haneda is currently with GE Global Research.

(MAP) estimate given by [9]

$$\hat{x} = \arg\min_{x\geq 0}\left\{\frac{1}{2}(y-Ax)^T D(y-Ax) - \log p(x)\right\} \quad (3)$$

where A represents the forward projection matrix and D is a diagonal weighting matrix whose elements are inversely proportional to the variance of the corresponding projection measurements. $p(x)$ is the prior distribution of the image x, which is used to regularize the reconstructed image. The log prior term of commonly-used prior models in the form of (2) corresponds to the weighted sum of potential functions over all pairwise cliques.

III. Implicit Gibbs Prior Distribution

A. MAP Estimation with Implicit Gibbs Prior

To develop our new prior model, we consider a general Gibbs distribution given by

$$p(x) = \frac{1}{z}\exp\{-\mu(x)\} \quad (4)$$

where $\mu(x)$ is the Gibbs energy function, and z is the partition function. By Hammersley-Clifford theorem [10], x will satisfy the Markov property given by

$$p(x_s|x_r, r\neq s) = p(x_s|x_{\partial s}) \quad (5)$$

where x_s is an element of x and $x_{\partial s}$ are the neighbors of x_s.

Our approach is to first model the local conditional distribution $p(x_s|x_{\partial s})$, and then to derive an approximation of the Gibbs distribution. To be more specific, given a local conditional distribution model $p(x_s|x_{\partial s})$, we will construct a surrogate energy function $u(x;x')$ using $p(x_s|x_{\partial s})$ such that the surrogate energy function $u(x;x')$ will satisfy the following two conditions.

$$\begin{aligned} \mu(x') &= u(x';x') \quad (6)\\ \mu(x) &\leq u(x;x') \quad (7)\end{aligned}$$

Therefore, $u(x;x')$ is an upper bound of $\mu(x)$ and they have the same values at x'. Then iterative minimization of the surrogate cost will also generate a decreasing sequence of the original MAP cost, which insures convergence. The overall algorithm is summarized in Figure 1 where the weighting factor λ is inserted to control the regularization.

Estimate the model parameters of $p(x_s|x_{\partial s})$
Initialize x'
repeat
  Update surrogate energy function $u(x;x')$
  $x' \leftarrow \arg\min_{x\geq 0}\{\frac{1}{2}(y-Ax)^T D(y-Ax)+\frac{1}{\lambda^2}u(x;x')\}$
until x' has converged

Fig. 1. MAP estimation with implicit Gibbs prior

B. Design of Surrogate Energy Function

We formulate the surrogate energy function to be a quadratic function expanded at x' as follows

$$u(x;x') = \frac{1}{2}(x-x')^T B(x-x') + d^T(x-x') + c \quad (8)$$

where B is a symmetric matrix, d is a vector and c is a scalar independent of x. Notice that we could set $c=0$ without changing the solution of the optimization.

As shown in our previous paper [8], given the conditional distribution $p_s(x_s|x_{\partial s})$, the parameters B and d can be calculated as follows

$$\begin{aligned} d_s &= -\frac{\partial}{\partial x_s}\log p_s(x_s|x_{\partial s})|_{x=x'} \quad (9)\\ B &= \tilde{H} + \alpha\text{diag}\{\tilde{H}\} \quad (10)\\ \tilde{H} &= \frac{H+H^T}{2} \quad (11)\\ H_{s,r} &= -\frac{\partial^2}{\partial x_s \partial x_r}\log p_s(x_s|x_{\partial s})|_{x=x'} \quad (12)\end{aligned}$$

The main idea behind these equations is that the gradient of the surrogate energy and the true energy are matched at x', and the second derivative matrix B is chosen to make the surrogate energy upper bound the true energy at x'. Notice that d and H are the gradient and Hessian of the true energy function $\mu(x)$ at x'.

C. Conditional Distribution Model

We model the homogeneous conditional distribution, $p(x_s|x_{\partial s})$, using a Gaussian mixture similar to [11]. Mathematically, the conditional distribution $p(x_s|x_{\partial s})$ will be a weighted sum of Gaussian distributions as follows

$$p(x_s|x_{\partial s}) = \sum_k p(x_s|x_{\partial s},k)p(k|x_{\partial s}) = \sum_k \gamma_k \mathcal{N}(x_s|\mu_k,\sigma_k^2)\ , \quad (13)$$

where μ_k and σ_k^2 are the conditional means and variances of each Gaussian component, and γ_k are the mixing weights. Moreover, each conditional mean is a weighted sum of its neighboring pixels given by

$$\mu_k = A_k x_{\partial s} + \beta_k\ . \quad (14)$$

Following the assumptions in [11], we further assume that

$$p(k|x_s) = p(k|\phi) \quad (15)$$

where ϕ is a feature vector extracted from the local neighborhood. The distribution of the feature vector, ϕ, is also assumed to be conditionally Gaussian given the class k, so that

$$p(\phi,k) = \pi_k \mathcal{N}(z|\bar{\phi}_k,\sigma_\phi^2 I^{-1})\ . \quad (16)$$

For model simplicity, we constrain the mixture components to share the same diagonal covariance matrix.

To summarize the overall conditional distribution model, we have two sets of parameters, one is $(\{\pi_k,\bar{\phi}_k\}_k,\sigma_\phi^2)$ for the feature vector ϕ, and the other is $(\{\gamma_k,A_k,\beta_k,\sigma_k^2\}_k)$ for the conditional distribution of the pixel. The standard Expectation-Maximization (EM) algorithm can be used to train the model parameters.

Fig. 2. Simulated phantom with artifacts and noise moderately removed, uniform flat region created, and high resolution targets inserted. The white box in the image indicates the region where the noise standard deviation is evaluated. Display window is [0 1600] HU.

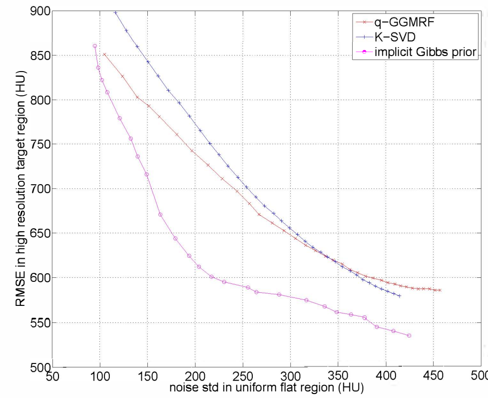

Fig. 3. Bias-variance trade-off curve. Horizontal axis is the noise standard deviation evaluated over the window indicated in Figure 2. Vertical axis is the RMSE in the right target high resolution region in Figure 2.

IV. Experimental Results

In this section, the reconstruction performance of different algorithms is presented. The simulated phantom, shown in Figure 2, is chosen from one slice of the 3D MBIR reconstruction of a CT scan of a bag [3]. The field of view is 512 mm $\times 512$ mm. Reconstruction artifacts are moderately cleaned up. Moreover, we made one region of the simulated phantom uniformly flat with a density of 1000 HU and also inserted high resolution target with a density of 2000 HU in order to quantitatively measure the noise and resolution performance. The simulated scanner has a 2D parallel-beam projection geometry with 363 detectors of width 2 mm. We generated the sinogram using 90 equal-angle views over 180 degree. The photon count in air calibration is $\lambda_0 = 500$. So each projection measurement y_i is approximately a Gaussian with variance $\frac{1}{\lambda_0 e^{-A_{i,*}x}}$ where $A_{i,*}$ is the i-th row of the matrix A.

For the implicit Gibbs prior, the conditional distribution is trained on a number of 5×5 patches and the number of mixture components are 32. For the K-SVD prior, the size of the patch is $n = 7 \times 7$, the size of the over-complete dictionary is $K = 256$, and the sparsity is $L = 3$ (see [5]). The parameters of both implicit Gibbs prior and K-SVD are trained offline on a training image dataset that does not include the testing phantom. For the q-GGMRF prior, we use $p = 2$, $q = 1.2$ and $c = 10$ HU (see [1]).

Figure 3 shows the RMSE in target high resolution region versus the noise standard deviation in the uniform flat region. Different points on the curve correspond to different amounts of regularization and the curve characterizes the bias-variance trade-off of each method. As shown in the figure, given a fixed noise variance, the implicit Gibbs prior gives the best resolution recovery in high resolution target region.

In Figure 5, we show the reconstructed images with different methods at the comparable noise level. It can be seen that

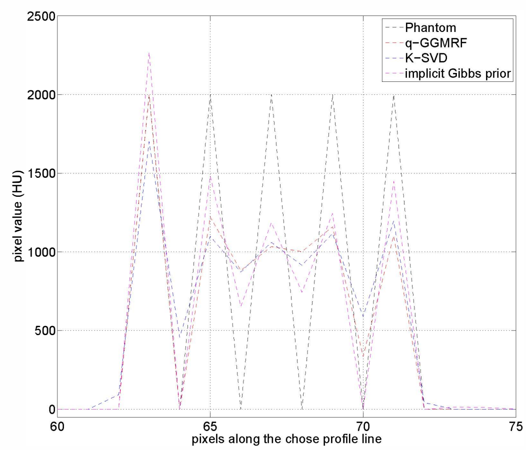

Fig. 4. CT values of pixels along a line through the target bars.

the fine details in the high resolution target region are best recovered in the reconstruction using the implicit Gibbs prior. Figure 4 shows the pixel profile along a line through the target bars. This plot indicates that the implicit Gibbs prior produces shaper edges. These results illustrate some potential benefits of the proposed method.

V. Conclusions

In this paper, we have introduced a novel prior model for Bayesian tomographic reconstruction. Instead of designing a particular potential functional for regularization, we approach the MAP estimation problem by constructing an upper bound of the Gibbs energy function from the local conditional distribution and iteratively solving the optimization, which also ensures convergence. We model the local conditional distribution using Gaussian mixture and train the parameters using EM algorithm. Simulations indicate that the proposed method has the potential to recover higher resolution details

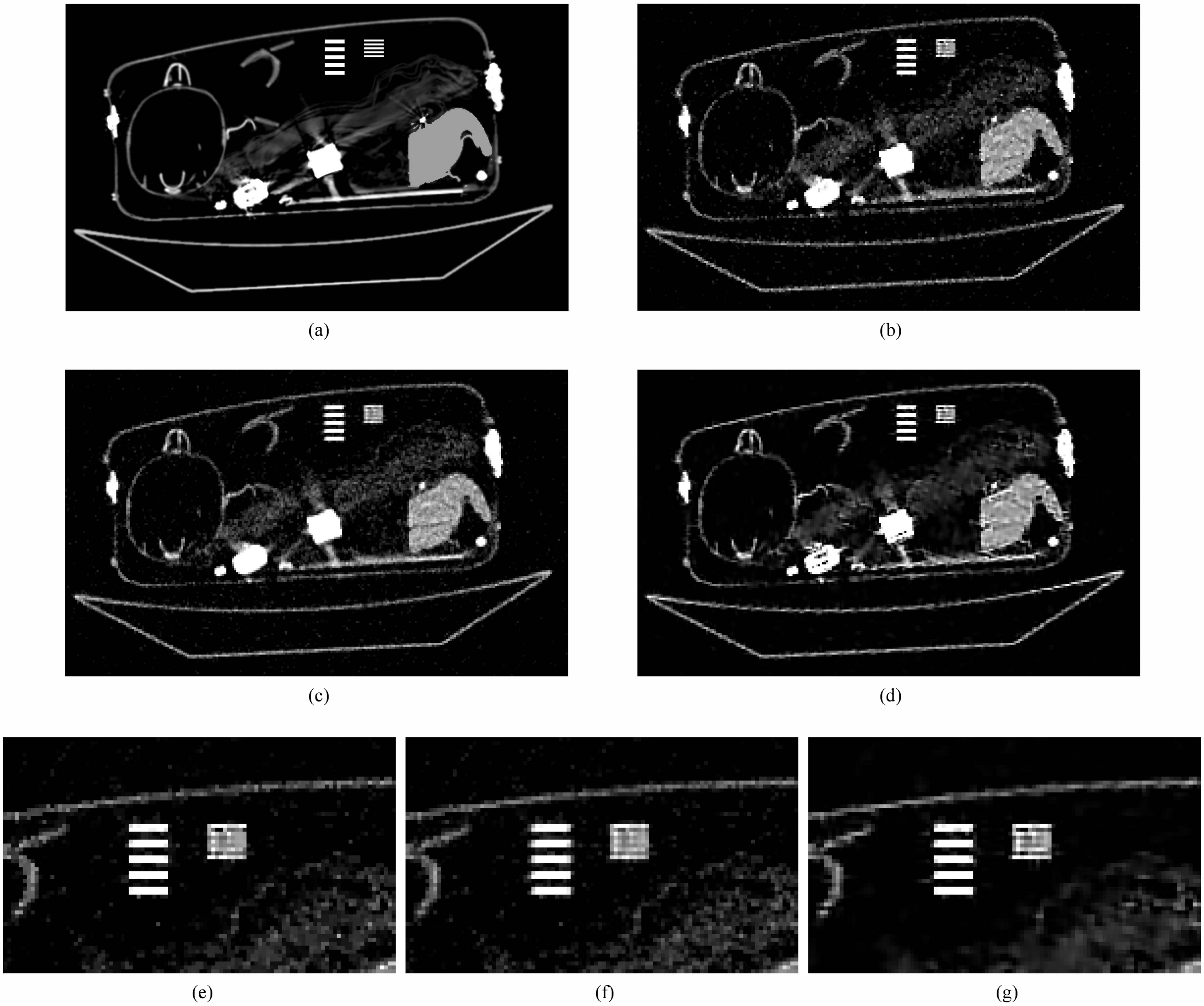

Fig. 5. Comparison of q-GGMRF, K-SVD and implicit Gibbs prior performance on the simulated phantom with comparable noise level. (a) phantom, (b) q-GGMRF, (c) K-SVD, (d) implicit Gibbs prior. (e) q-GGMRF zoomed-in, (f) K-SVD zoomed-in, (g) implicit Gibbs prior zoomed-in. Display window is [0 1600] HU.

at a particular noise level as compare to several other methods.

REFERENCES

[1] J.-B. Thibault, K. Sauer, C. Bouman, and J. Hsieh, "A three-dimensional statistical approach to improved image quality for multi-slice helical CT," *Med. Phys.*, vol. 34, no. 11, pp. 4526–4544, 2007.

[2] R. Zhang, J.-B. Thibault, C. Bouman, K. Sauer, and J. Hiseh, "A model-based iterative algorithm for dual-energy X-ray CT reconstruction," in *Proc. 2nd Intl. Mtg. on image formation in X-ray CT*, 2012, pp. 439–443.

[3] P. Jin, E. Haneda, C. A. Bouman, and K. D. Sauer, "A model-based 3D multi-slice helical CT reconstruction algorithm for transportation security application," in *Proc. 2nd Intl. Mtg. on image formation in X-ray CT*, 2012, pp. 297–300.

[4] 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.

[5] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, pp. 3736–3745, Dec. 2006.

[6] S. Ravishankar and Y. Bresler, "MR image reconstruction from highly undersampled k-space data by dictionary learning," *IEEE Trans. Med. Imag.*, vol. 30, pp. 1028–1041, May 2011.

[7] Y. Guo, D. Depalov, P. Bauer, B. Bradburn, J. Allebach, and C. Bouman, "Conditional entropy estimation-based dictionary design in binary document image compression with JBIG2," in *Proc. SPIE*, 2013.

[8] E. Haneda and C. Bouman, "Implicit piors for model-based inversion," in *Proc. IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Mar. 2012.

[9] K. Sauer and C. Bouman, "A local update strategy for iterative reconstruction from projections," *IEEE Trans. Signal Process.*, vol. 41, pp. 534–548, Feb. 1993.

[10] J. Besag, "Spatial interaction and the statistical analysis of lattice systems," *J. Royal Stat. Soc.*, vol. 36, no. 2, pp. 192–236, 1974.

[11] C. Atkins, C. Bouman, and J. Allebach, "Optimal image scaling using pixel classification," in *Proc. IEEE Intl. Conf. on Image Processing (ICIP)*, vol. 3, 2001, pp. 864–867.