A Gaussian Mixture MRF for Model-Based Iterative Reconstruction With Applications to Low-Dose X-Ray CT

Ruoqiao Zhang, Dong Hye Ye, *Member, IEEE*, Debashish Pal, *Member, IEEE*,
Jean-Baptiste Thibault, *Member, IEEE*, Ken D. Sauer, *Member, IEEE*, and Charles A. Bouman, *Fellow, IEEE*

Abstract—**Markov random fields (MRFs) have been widely used as prior models in various inverse problems such as tomographic reconstruction. While MRFs provide a simple and often effective way to model the spatial dependencies in images, they suffer from the fact that parameter estimation is difficult. In practice, this means that MRFs typically have very simple structure that cannot completely capture the subtle characteristics of complex images. In this paper, we present a novel Gaussian mixture Markov random field model (GM-MRF) that can be used as a very expressive prior model for inverse problems such as denoising and reconstruction. The GM-MRF forms a global image model by merging together individual Gaussian-mixture models (GMMs) for image patches. In addition, we present a novel analytical framework for computing MAP estimates using the GM-MRF prior model through the construction of surrogate functions that result in a sequence of quadratic optimizations. We also introduce a simple but effective method to adjust the GM-MRF so as to control the sharpness in low- and high-contrast regions of the reconstruction separately. We demonstrate the value of the model with experiments including image denoising and low-dose CT reconstruction.**

Index Terms—**Gaussian mixture model (GMM), image model, Markov random field (MRF), model-based iterative reconstruction (MBIR), patch-based method, prior modeling.**

I. INTRODUCTION

IN recent years, model-based iterative reconstruction (MBIR) has emerged as a very powerful approach to reconstructing images from sparse or noisy data in applications ranging from medical, to scientific, to non-destructive imaging [1]–[8]. The power of these methods is due to the synergy that results from modeling both the sensor (i.e., forward model) and the image being reconstructed (i.e., prior model). In medical applications, for example, MBIR has been demonstrated to substantially improve image quality by both reducing noise and improving resolution [9], [10].

While the MBIR forward model is typically based on the physics of the sensor, accurate prior modeling of real images remains a very challenging problem. Perhaps the most commonly used prior model is a very simple Markov random field (MRF) with only very local dependencies and a small number of parameters [3], [11]–[14]. Alternatively, total-variation (TV) regularization approaches can also be viewed as simple MRF priors [15]–[19]. Besides direct reconstruction, simple MRF priors have also been employed in applications such as sinogram smoothing and restoration [20]–[22]. While these models have been very useful, their simple form does not allow for accurate or expressive modeling of real images.

Alternatively, a number of approaches have been proposed for modeling the non-Gaussian distribution of pixels in image reconstruction. Hsiao *et al.* [23], [24] used a gamma mixture model in emission/transmission tomography; Wang *et al.* [25] used a Gaussian scale mixture to model the distribution of scattering data in muon tomography; and Mehranian *et al.* [26] used a Gaussian mixture model for PET/MRI reconstruction. In [25] and [26], the authors also derived a surrogate function for the log of a mixture of Gaussians, which they used in optimization. However, none of these mixture models accounted for the spatial correlation of the neighboring pixels in the reconstructed image, and therefore additional spatial models might be required [26].

More recently, methods such as K-SVD [27] and non-local means (NLM) [28] have been proposed which can be adapted as prior models in MBIR reconstruction. Though effective in denoising applications, other patch-based methods such as BM3D [29] are not directly suited for application in model-based reconstruction problems, and integration of such methods into iterative reconstruction is still an active topic of research [30]–[32]. While K-SVD can be adapted as a prior model [33], [34], it does not explicitly model the multivariate distribution of the image. This can lead to drawbacks in applications. For example, the K-SVD algorithm is designed to be invariant to scaling or average gray level of image patches. In applications such as CT reconstruction, this is a severe limitation since regions of different densities generally correspond to different tissues (e.g., bone and soft tissue) with distinctly different characteristics. While direct incorporation of the NLM filtering into iterative

reconstruction framework is still under study [30], [35], [36], there have been efforts to adapt this filter as the regularization in tomographic reconstruction [37]–[40]. Though producing promising results, these methods do not explicitly model the statistics of the image, and therefore are not consistent image models. A variety of research also adapted the ideas of dictionary learning to the problem of prior modeling in CT reconstruction [41]–[43].

Another approach to prior modeling is to allow different patches of the images to have different distributions. This approach has been used by both Zoran *et al.* [44] and Yu *et al.* [45] to construct non-homogeneous models of images as the composition of patches, each with a distinct Gaussian distribution. The Gaussian distribution of each patch is selected from a discrete set of possible distributions (i.e., distinct mean and covariance). The reconstruction is then computed by jointly estimating both the image and a discrete class for each patch in the image. This approach can be very powerful for modeling the different spatially varying characteristics in real images. However, the approaches suffer from the need to make hard classifications of each patch. These hard classifications can lead to artifacts when patch distributions overlap, as is typically the case when a large number of classes are used. Recently, patch models using Gaussian mixtures have also been applied in various applications [46]–[50].

In this paper, we introduce the Gaussian-mixture MRF (GM-MRF) image prior along with an associated method for computing the MAP estimate using exact surrogate functions. (see [51], [52] for early conference versions of our method). The GM-MRF model is constructed by seaming together patches that are modeled with a single Gaussian mixture (GM) distribution. The advantages of this approach are that:

1) The GM-MRF prior provides a theoretically consistent and very expressive model of the multivariate distribution of the image;
2) The GM-MRF parameters can be easily and accurately estimated by fitting a GM distribution to patch training data using standard methods such as the EM algorithm [53];
3) MAP optimization can be efficiently computed by alternating soft classification of image patches with MAP reconstruction using quadratic regularization.

To create a consistent global image model, we seam together the GM patch models by using the geometric mean of their probability densities. This approach, similar to the product-of-experts technique [54] employed in deep-learning, produces a single consistent probability density for the entire image. Moreover, we also show that the resulting GM-MRF model is a Markov random field (MRF) as its name implies.

Of course, an accurate image model is of little value if computation of the MAP estimate is difficult. Fortunately, we show that the GM-MRF prior has an exact quadratic surrogate function for its log likelihood. This surrogate function allows for tractable minimization of the MAP function using a majorization-minimization approach [55]. The resulting MAP optimization algorithm has the form of alternating minimization. The two alternating steps are soft classification for patches

followed by MAP optimization using quadratic regularization (i.e., a non-homogeneous Gaussian prior). Moreover, our approach to MAP optimization with the GM-MRF prior avoids the need for hard classification of individual patches. In practice, this means that patch-based GM-MRF models with a large numbers of overlapping mixture components can be used without adverse modeling effects. This allows for the use of very expressive models that capture fine details of image behavior. Note that Wang *et al.* [25] and Mehranian *et al.* [26] used similar technique to construct quadratic surrogate functions for the log of mixture of Gaussians. However, in this paper we formulate and prove a lemma that is more general and can be applied to a wider range of distributions as compared to previous methods.

It has been reported that MAP estimation with non-Gaussian priors can lead to contrast-dependent noise and resolution properties in reconstructed images [56]–[58]. Typically, the estimated image has higher noise and resolution at higher-contrast edges, and lower noise and resolution at lower-contrast edges. In practice, this non-homogeneity may result in undesirable image quality in specific applications.

In order to address this issue of non-homogeneity for the GM-MRF prior, we introduce a simple method for adjusting the GM components of the GMMRF prior, so as to control the sharpness in low and high contrast regions of the reconstruction separately.

Our experimental results indicate that GM-MRF method results in improved image quality and reduced RMS error in simple denoising problems as compared to simple MRF and K-SVD priors. We also show multi-slice helical scan tomographic reconstructions from both phantom and clinical data to demonstrate that the GM-MRF prior produces visually superior images as compared to filtered back-projection (FBP) and MBIR using the traditional q-GGMRF prior [3], especially under the condition of low-dose acquisition.

II. GAUSSIAN MIXTURE MARKOV RANDOM FIELD

MBIR algorithms work by computing the maximum *a posteriori* (MAP) estimate of the unknown image, x, given the measured data, y, by

$$\hat{x} \leftarrow \arg\min_{x \in \Omega} \left\{ -\log p(y|x) - \log p(x) \right\} . \quad (1)$$

In this framework, $p(y|x)$ is the conditional probability of y given x, which comprises the forward model of the measurement process. The density $p(x)$ is the prior model for x, which will be discussed in detail in this section.

Let $x \in \Re^N$ be an image with pixels $s \in S$, where S is the set of all pixels in x with $|S| = N$. Let $P_s \in \mathcal{Z}^{L \times N}$ be a patch operator that extracts a patch from the image, where the patch is centered at pixel s and contains L pixels. More precisely, P_s is a rank L matrix that has a value of 1 at locations belonging to the patch and 0 otherwise. Furthermore, we assume that each patch, $P_s x$, can be modeled as having a multivariate GM distribution with K components,

$$g(P_s x) = \sum_{k=1}^{K} \frac{\pi_k |R_k|^{-\frac{1}{2}}}{(2\pi)^{\frac{L}{2}}} \exp\left\{ -\frac{1}{2} \|P_s x - \mu_k\|_{R_k^{-1}}^2 \right\} , \quad (2)$$
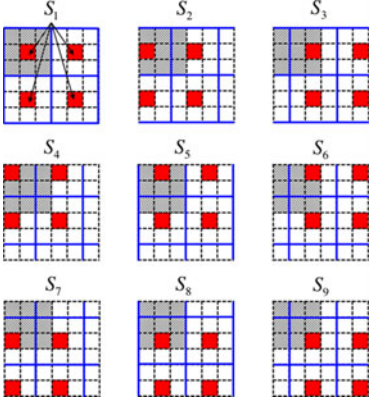
Fig. 1. 2-D illustration of the tiling method. Each blue grid represents one of nine distinct tilings with 3×3 patches on a 6×6 grid, i.e., $L = 9$, with the center pixel of each patch marked in red. Toroidal boundary condition is considered in this illustration. Note that there are exactly 9 distinct phase shifts of the tiling, each of which is determined by the center pixel of the first patch in the upper-left corner, which corresponds to a distinct pixel location in the shadowed patch.

where parameters π_k, μ_k, R_k represent the mixture probability, mean, and covariance, respectively, of the kth mixture component.

Then let $\{S_m\}, m \in \{1, \cdots, L\}$, be a partition of the set of all pixels into L sets, each of which tiles the image space. In other words, $\{P_s x\}_{s \in S_m}$ forms a set of non-overlapping patches, which contain all pixels in x. A simple 2-D example of this is when each $P_s x$ is a square $r \times r$ patch, and S_m is the set of pixels at each rth row and column. Then the set of patches, $\{P_s x\}_{s \in S_m}$, tiles the plane.

Importantly, there are exactly L distinct tilings of the image space where L is the number of pixels in a patch. In order to see why this is true, consider the 2-D example in Fig. 1. (Note that this tiling method can be easily extended to n-D space with $n \geq 3$ by using n-D patches.) Notice that each distinct tiling of the space is determined by the position of the center pixel for the first (e.g., upper left hand) patch since the positioning of the first patch determines the phase shift of the tiling. With this in mind, there are exactly L distinct phase shifts corresponding to the L pixels in a single patch. Using this notation, we model the distribution of each tiling as the product of distributions of all its patches, as

$$p_m(x) = \prod_{s \in S_m} g(P_s x) . \qquad (3)$$

In this case, $p_m(x)$ has the desired distribution for each patch. However, the discrete tiling of the space introduces artificial boundaries between patches. To remove the boundary artifacts, we use an approach similar to the product-of-experts approach in [54] and take the geometric average of the probability densities for all L tilings of the image space to obtain the resulting distribution

$$p(x) = \frac{1}{z} \left(\prod_{m=1}^{L} p_m(x) \right)^{\frac{1}{L}} = \frac{1}{z} \left(\prod_{s \in S} g(P_s x) \right)^{\frac{1}{L}} , \qquad (4)$$

where z is a normalizing factor introduced to assure that $p(x)$ is a proper distribution after the geometric average is computed.

Let $V(P_s x) = -\log\{g(P_s x)\}$. Then we formulate a GM-MRF model directly from (4) as

$$p(x) = \frac{1}{z} \exp\{-u(x)\} , \qquad (5)$$

with the energy function

$$u(x) = \frac{1}{L} \sum_{s \in S} V(P_s x) , \qquad (6)$$

and the potential function

$$V(P_s x) = -\log \left\{ \sum_{k=1}^{K} \frac{\pi_k |R_k|^{-\frac{1}{2}}}{(2\pi)^{\frac{L}{2}}} \exp \left\{ -\frac{\|P_s x - \mu_k\|_{R_k^{-1}}^2}{2} \right\} \right\} . \qquad (7)$$

Notice that $p(x)$ is a Gibbs distribution by (5). Therefore, by the renowned Hammersley-Clifford theorem [11], $p(x)$ is also an MRF.

III. MAP ESTIMATION WITH GM-MRF PRIOR

For typical model-based inversion problems, the log-likelihood function may be modeled under the Gaussian assumption as

$$-\log p(y|x) = \frac{1}{2} \|y - Ax\|_D^2 , \qquad (8)$$

where $A \in \Re^{M \times N}$ is the projection matrix with M measurements and N unknowns. The weighting D is a diagonal matrix with each diagonal element inversely proportional to the variance of the corresponding measurement.

A. Surrogate Prior

By substituting (5) and (8) into (1), we can calculate the MAP estimate with the GM-MRF prior as

$$\hat{x} \leftarrow \arg \min_{x \in \Omega} \left\{ \frac{1}{2} \|y - Ax\|_D^2 + u(x) \right\} . \qquad (9)$$

However, the function $u(x)$ is not well suited for direct optimization due to the mixture of logarithmic and exponential functions. Therefore, we will use a majorization-minimization approach, in which we replace the function $u(x)$ with a quadratic upper-bounding surrogate function.

More precisely, the objective of the majorization-minimization method is to find a surrogate function $u(x; x')$ that satisfies the following two conditions

$$u(x'; x') = u(x') \qquad (10)$$

$$u(x; x') \geq u(x). \qquad (11)$$

Intuitively, these conditions state that the surrogate function upper bounds $u(x)$ and that the two functions are equal when $x = x'$. Importantly, these conditions also imply that any reduction of $u(x; x')$ also must reduce $u(x)$.

In order to construct a surrogate function for our problem, we introduce the following lemma that is proved in Appendix. The lemma provides a surrogate function for a general class of
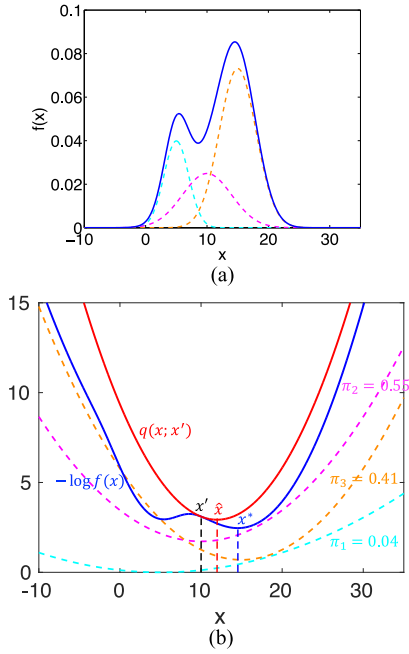
Fig. 2. Figure illustrates the lemma with a 1-D GM distribution. (a) GM distribution, (b) surrogate function. The quadratic function $q(x; x')$ is a surrogate function for the negative log of the GM distribution $f(x)$ at point x'. The surrogate function is a weighted sum of the quadratic exponents of the exponential functions in the GM distribution. The weights π_1, π_2, π_3 give the posterior probabilities of the point x' belonging to different GM components.

functions formed by the log of a sum of exponential functions. Since the potential function of (7) has this form, we can use this lemma to construct a surrogate function for our MAP estimation problem. Fig. 2 illustrates the usage of the lemma for a particular case of GM distribution. We note that the lemma generalizes the approaches of [25], [26], since it works for any exponential mixture rather than just mixtures of Gaussians.

Lemma (surrogate functions for logs of exponential mixtures):
Let $f : \Re^N \to \Re$ be a function of the form

$$f(x) = \sum_k w_k \exp\{-v_k(x)\} , \quad (12)$$

where $w_k \in \Re^+$, $\sum_k w_k > 0$, and $v_k : \Re^N \to \Re$. Furthermore $\forall (x, x') \in \Re^N \times \Re^N$ define the function

$$q(x; x') \triangleq -\log f(x') + \sum_k \tilde{\pi}_k (v_k(x) - v_k(x')) , \quad (13)$$

where $\tilde{\pi}_k = \frac{w_k \exp\{-v_k(x')\}}{\sum_l w_l \exp\{-v_l(x')\}}$. Then $q(x; x')$ is a surrogate function for $-\log f(x)$, and $\forall (x, x') \in \Re^N \times \Re^N$,

$$q(x'; x') = -\log f(x') \quad (14)$$

$$q(x; x') \geq -\log f(x) \quad (15)$$

Proof: see Appendix.

Since the function $u(x)$ specified by (6) and (7) has the same form as assumed by the lemma, we can use this lemma to find

a surrogate function with the following form

$$u(x; x') = \frac{1}{2L} \sum_{s \in S} \sum_{k=1}^{K} \tilde{w}_{s,k} \|P_s x - \mu_k\|_{R_k^{-1}}^2 + c(x') , \quad (16)$$

where x' is the current state of the image, $c(x')$ only depends on the current state, and the weights $\tilde{w}_{s,k}$ are given by

$$\tilde{w}_{s,k} = \frac{\pi_k |R_k|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} \|P_s x' - \mu_k\|_{R_k^{-1}}^2 \right\}}{\sum_{l=1}^{K} \pi_l |R_l|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} \|P_s x' - \mu_l\|_{R_l^{-1}}^2 \right\}} . \quad (17)$$

Note that the weights $\tilde{w}_{s,k}$ are only functions of the current image x'. Therefore, the optimization in (9) can be implemented as a sequence of optimizations as

$$\text{repeat}\left\{ \quad \hat{x} \leftarrow \arg \min_x \left\{ \frac{1}{2}\|y - Ax\|_D^2 + u(x; x') \right\} \quad (18) \right.$$

$$\left. x' \leftarrow \hat{x} \quad \right\} ,$$

with $u(x; x')$ being a quadratic prior that adapts to the current image at each iteration.

Importantly, the weights in (17) represent a soft classification of the current patch into GM components. This differs from existing approaches in which each patch is classified to be from a single component of the mixture [44], [45]. The previous methods [44], [45] performed discrete optimizations over the GM components to select one single component for each image patch. Thus, there is no fixed prior used to form a single consistent MAP estimate of the image. Instead, the prior model is iteratively adapted through the choice of a discrete class for each patch.

We believe that this discrete classification approach may cause inaccurate results or artifacts since it requires hard classification of patches into classes even when the class membership of a patch is ambiguous. In contrast, our proposed method uses a single consistent and spatially homogeneous non-Gaussian prior model of the entire image. A local minimum of the true posterior distribution is then computed by minimizing a series of convex surrogate functions.

Fig. 2(b) illustrates a simple 1-D example that demonstrates the benefit of the proposed method over previous methods when the distributions of mixture components heavily overlap. More precisely, the negative log of the underlying distribution, $-\log f(x)$, has a global minimum at $x^* = 14.54$. For a current point $x' = 10$, the previous method will select model 2 (magenta) since it produces the greatest posterior probability ($\pi_2 = 0.55$) among the three models. However, optimizing the resultant quadratic function (magenta) will give a stationary point at $x = 10$, which deviates from the global minimum x^* and cannot be improved even with multiple iterations. In contrast, the proposed method constructs a quadratic surrogate function (red), $q(x; x')$, of which the optimization will give $\hat{x} = 11.97$, which moves closer to the global minimum x^* as compared to the current point x'. Thus, by repeatedly constructing surrogate functions with a new $\hat{x}$ estimated from the previous iteration,
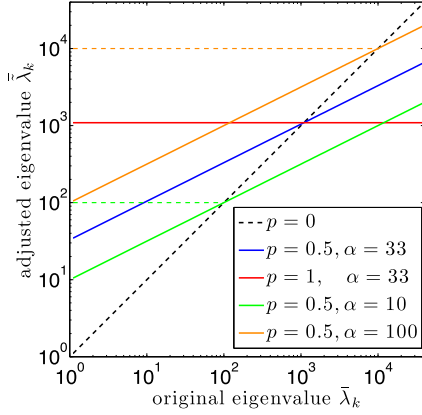
Fig. 3. The covariance scaling defined in (29) and (30) with various values of p and α on a log scale. The black dotted line shows the case when no scaling is present, i.e., $p = 0$. When $0 < p \leq 1$, the average eigenvalues $\bar{\bar{\lambda}}_k$ are "compressed" toward α^2, where eigenvalues further away from α^2 lead to greater change and a larger p results in greater compression. For a fixed value of p, increasing α increases the covariance of each GM component.

it will asymptotically reach the global minimum x^* for this particular example.

B. Optimization

We use the iterative coordinate descent (ICD) algorithm [12] to solve this quadratic minimization problem in (18). The ICD algorithm sequentially updates each of the pixels by solving a 1-D optimization problem, as

$$\hat{x}_j \leftarrow \arg\min_{x_j} \left\{ \frac{1}{2} \|y - Ax' + A_{*j}(x'_j - x_j)\|_D^2 + u(x_j; x') \right\}, \tag{19}$$

with the surrogate prior for x_j, as

$$u(x_j; x') = \frac{1}{2L} \sum_{r \in \mathcal{S}_j} \sum_{k=1}^{K} \tilde{w}_{r,k} \|P_r x - \mu_k\|_{R_k^{-1}}^2 + c(x'), \tag{20}$$

where the weights $\tilde{w}_{r,k}$ are given by (17) and S_j represents a set of center pixels whose patches contain pixel j.

By rearranging the terms, we can explicitly write (19) as a quadratic function of x_j, as

$$\hat{x}_j \leftarrow \arg\min_{x_j} \left\{ (\theta_1 + \varphi_1) x_j + \frac{\theta_2 + \varphi_2}{2} (x_j - x'_j)^2 + c(x') \right\}, \tag{21}$$

where $c(x')$ is constant to x_j and $\theta_1, \theta_2, \varphi_1, \varphi_2$ are given by

$$\theta_1 = A_{*j}^t D(Ax' - y), \tag{22}$$

$$\theta_2 = A_{*j}^t D A_{*j}, \tag{23}$$

$$\varphi_1 = \frac{1}{L} \sum_{r \in S_j} \sum_k \tilde{w}_{r,k} (P_r \delta_j)^t R_k^{-1} (P_r x' - \mu_k), \tag{24}$$

$$\varphi_2 = \frac{1}{L} \sum_{r \in S_j} \sum_k \tilde{w}_{r,k} (P_r \delta_j)^t R_k^{-1} (P_r \delta_j), \tag{25}$$

where the calculation of the projection matrix A follows the same procedure in [3]. The function $\delta_j \in \Re^{|S|}$ is a Kronecker delta function, which is a vector with a value of 1 at entry j and

with 0 elsewhere. Therefore, $P_r \delta_j$ is simply an operator that extracts a particular column from a matrix corresponding to the location of the pixel j within the patch operator P_r.

Solving (21) by rooting the gradient, we then have

$$\hat{x}_j \leftarrow x'_j - \frac{\theta_1 + \varphi_1}{\theta_2 + \varphi_2}. \tag{26}$$

IV. COVARIANCE CONTROL FOR GM-MRF

We will see that the GM-MRF distribution can be used to form a very accurate model of images. However, in applications such as CT reconstruction, the MAP estimate may not be visually appealing even with an accurate forward and prior model. This is because the MAP estimate with non-Gaussian priors tends to produce a reconstruction that is under-regularized (i.e., too sharp) in high-contrast regions and over-regularized (i.e., too smooth) in low-contrast regions [56]–[58]. While this variation in spatial resolution may produce a lower mean squared error (MSE), in particular applications it may not be visually appealing.

In order to address this problem of spatial variation in sharpness, in this section we introduce a simple parameterization for systematically controlling the covariance of each GM component of the GM-MRF model. In the experimental results section, we will then demonstrate that this simple parameterization can be used to effectively tune the visual quality of the MAP reconstruction. In real applications such as medical CT reconstruction, this covariance adjustment can be used to effectively fine-tune the rendering of specific tissue types, such as soft tissue, lung, and bone, which may have different desired characteristics.

We start by introducing regularization parameters, σ_x and $\{\sigma_k\}_{k=1}^K$, into the distribution given by

$$u_\sigma(x) = -\frac{1}{L\sigma_x^2} \sum_{s \in S} \log\{g_\sigma(P_s x)\}, \tag{27}$$

with the patch GM distribution

$$g_\sigma(P_s x) = \sum_{k=1}^{K} \frac{\pi_k |R_k/\sigma_k^2|^{-\frac{1}{2}}}{(2\pi)^{\frac{L}{2}}} \exp\left\{ -\frac{\sigma_k^2 \|P_s x - \mu_k\|_{R_k^{-1}}^2}{2} \right\}. \tag{28}$$

Notice that σ_x controls the overall level of regularization and that the K values of σ_k control the regularization of each individual component of the GMM. When the value of σ_x is increased, the overall reconstruction is made less regularized (i.e., sharper) and when the value of σ_k is increased, the individual GM component is made more regularized.

Now for a typical GM-MRF model there may be many components, so this would require the choice of many values of σ_k. Therefore, we introduce a simple method to specify these K parameters using the following equation

$$\sigma_k = \left(\bar{\lambda}_k/\alpha^2\right)^{p/2}, \tag{29}$$

where p and α are two user-selectable parameters such that $0 \leq p \leq 1$, $\alpha > 0$, and $\bar{\lambda}_k = |R_k|^{\frac{1}{L}}$ is the geometric average of the eigenvalues of R_k. Define $\tilde{R}_k = R_k/\sigma_k^2$ as the covariance matrix after scaling. Then its corresponding average eigenvalue
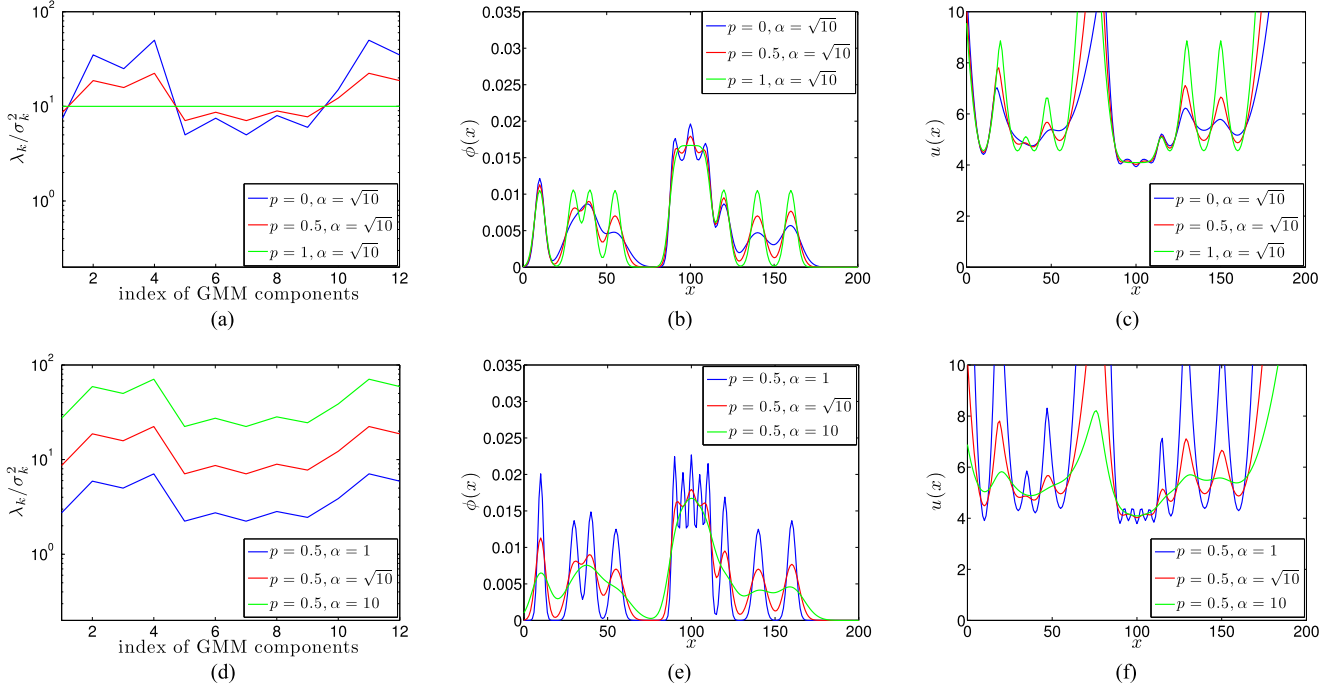
Fig. 4. 1-D illustration of the covariance scaling in (29). The 1-D energy function is given by $u(x) = -\log(\phi(x))$ with the GM distribution $\phi(x) = \sum_k \frac{1}{\sqrt{2\pi \lambda_k/\sigma_k^2}} \exp\{-\frac{1}{2\lambda_k/\sigma_k^2}(x-\mu_k)^2\}$, with λ_k the original variance and the scaling $\sigma_k = (\lambda_k/\alpha^2)^{p/2}$. (a) varying p with α fixed, (b) the resulting distribution, and (c) the resulting energy function; (d) varying α with p fixed, (e) the resulting distribution, and (f) the resulting energy function.

is given by

$$\bar{\bar{\lambda}}_k = \alpha^{2p} \bar{\lambda}_k^{1-p} \ . \tag{30}$$

Fig. 3 illustrates this scaling with various values of p and α.

In this model, the parameters p and α collectively compress the dynamic range of the average eigenvalues $\bar{\lambda}_k$ of all GM co-variance matrices. That is, for those GM components with large average eigenvalues of covariance, which typically correspond to high-contrast or structural regions, applying the scaling in (29) decreases the eigenvalues, which leads to increased regularization. Conversely, for those GM components with small average eigenvalues of covariance, which are typically associated with low-contrast or homogeneous regions, applying the scaling increases the eigenvalues and subsequently results in reduced regularization.

More specifically, p is the compression rate with a larger value resulting in greater compression of the dynamic range, and α defines a stationary point during the compression, i.e.

$$\bar{\bar{\lambda}}_k = \bar{\lambda}_k \ , \quad \text{if } \bar{\lambda}_k = \alpha^2 \ . \tag{31}$$

When $0 < p < 1$, the average eigenvalues $\bar{\bar{\lambda}}_k$ are "compressed" toward α^2, with eigenvalues further away from α^2 leading to greater change. When $p = 1$, all GM components have the same average eigenvalue α^2, while they maintain the original eigenvalues when $p = 0$. Fig. 4(a)–(c) illustrate the change in the distribution and the energy function as p varies.

In addition, the parameter α controls the "smoothness" of the GM distribution. With p fixed, increasing the value of α leads to a smoother distribution of (28), which potentially reduces the degree of non-convexity of the energy function in (27). Moreover, an increased α also reduces the overall regularization.

TABLE I
PARAMETER SELECTION FOR GM-MRF MODEL

Parameter	Selection	Effect
σ_x	≈ 1	MAP estimate
	$\gg 1$	large value $\rightarrow$ weak overall regularization
	$\ll 1$	small value $\rightarrow$ heavy overall regularization
p	0	unmodified GM-MRF
	0.5	regularization strength increases for GM components with large average eigenvalues; reduces for those with small average eigenvalues
	1	same regularization strength for all GM components
α	33 (HU)	large value $\rightarrow$ smooth prior distribution $\rightarrow$ weak regularization small value $\rightarrow$ peaky prior distribution $\rightarrow$ heavy regularization

Fig. 4(d)–(f) illustrate the change in the distribution and the energy function as α varies.

Table I presents the selection of the regularization parameters and the corresponding effect. Note that the parameter α is related to the reconstruction noise and therefore has the same unit as the reconstruction. For instance, in X-ray CT reconstruction, the parameter α is in Hounsfield Unit (HU).

V. METHODS

In this section, we present the datasets used for training and testing in our experiments. We also provide description for the training procedure in detail. In addition, we describe the different methods that used for comparison.

TABLE II
PARTITION OF THE TRAINING DATA

Group index, i	1	2	3	4	5	6
Mean (HU)	$[-1000\ -850)$	$[-850\ -200)$	$[-200\ 200)$	$[-200\ 200)$	$[-200\ 200)$	≥ 200
Standard deviation (HU)			$[0\ 25)$	$[25\ 80)$	≥ 80	
Number of patch samples	5×10^3	1×10^5	5×10^4	1×10^5	1×10^5	1×10^5
Number of GM components, K_i	1	15	5	15	15	15
Mixture weight, π_i	0.05	0.17	0.40	0.25	0.04	0.09

Each image patch was classified into one of the six groups based on its mean and standard deviation. The number of GM components for each group was empirically chosen. The mixture weights are determined by the proportions of corresponding groups in the whole training data and will be used when combining different GMMs to form a single model.
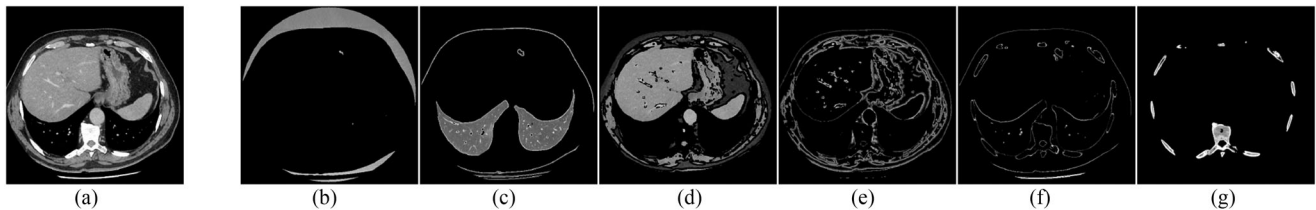

Fig. 5. Partition of the training data. (a): a 2-D axial slice from the 3-D image volume where training patches were extracted; (b)-(g): partition of the data based on the criteria in Table II, with (b) group 1, (c) group 2, (d) group 3, (e) group 4, (f) group 5, (g) group 6. Display window: (a) $[-160, 240]$ HU, (b) $[-1250, -750]$ HU, (c) $[-1400, 200]$ HU, (d)-(f) $[-210, 290]$ HU, (g) $[-300, 700]$ HU. Notice that different groups roughly capture different materials or tissue types in the image, as group 1 for air, group 2 for lung tissue, group 3 for smooth soft tissue, group 4 for low-contrast soft-tissue edge, group 5 for high-contrast edge, and group 6 for bone.

A. Training

We trained the GMM patch distribution, $g(P_s x)$ in (2), on clinical CT images using the standard EM algorithm with the software in [53]. Training data consisted of 2-D or 3-D overlapping patches extracted from the 3-D reconstruction of a normal-dose scan, acquired with a GE Discovery CT750 HD scanner in 64×0.625 mm helical mode with 100 kVp, 500 mA, 0.8 s/rotation, pitch 0.984:1, and reconstructed in 360 mm field-of-view (FOV). We have supplied typical images of the training dataset in the supplementary material.

Instead of training one GMM using all the patches, we partitioned the patches into different groups and then trained one GMM from each of the groups. In this way, we were able to collect sufficiently many samples from underrepresented groups to obtain accurate parameter estimates, while simultaneously limiting the data size for other groups to retain training efficiency. For the ith group, we trained the parameters, $\{\pi_{i,k}, \mu_{i,k}, R_{i,k}\}_k^{K_i}$, for one GMM, $g_i(P_s x)$, with K_i components. Then we merged all GMMs trained from different groups into a single GMM by weighted summation

$$g(P_s x) = \sum_{i=1}^{I} \pi_i g_i(P_s x), \qquad (32)$$

where the mixture weights π_i were determined by the natural proportions of corresponding groups in the whole training data.

More specifically, we partitioned the patches into six groups based on the mean and standard deviation as listed in Table II, where the partition thresholds were empirically determined to roughly reflect typical tissue types in a medical CT image. Fig. 5 illustrates different groups on a 2-D image slice. As shown in Fig. 5, different groups roughly capture different materials or tissue types in the image, as group 1 for air, group 2 for lung tissue, group 3 for smooth soft tissue, group 4 for low-contrast soft-tissue edge, group 5 for high-contrast edge, and group 6 for bone. With this partition, we were able to collect adequate patches for individual groups separately, especially for the underrepresented ones as group 4, 5, and 6. During the separate training process, we empirically fixed the number of GMM components, K_i, in the EM algorithm for each group. Table II also presents the mixture weights π_i for different GMMs, which were determined by the natural proportions of corresponding groups in the whole training data.

We trained three 2-D GM-MRF models consisting of 66 GM components with different patch sizes, i.e., 3×3, 5×5, and 7×7 patches, for an image denoising experiment. We also trained a number of 3-D GM-MRF models with different parameter settings to study the influence of model parameters in phantom studies. More precisely, we fixed the number of GM components to 66 and then trained different GM-MRFs with $3 \times 3 \times 3$ patches, $5 \times 5 \times 3$ patches, and $7 \times 7 \times 3$ patches, respectively. Similarly, we fixed the patch size to $5 \times 5 \times 3$ and then trained several GM-MRF models with various number of GM components, namely 6, 15, 31, 66, and 131 components. Details of the training procedures are provided in the supplementary material. Note that the covariance scaling introduced in Sec IV is applied to the trained model and hence does not require additional training.

The blue plot in Fig. 6 illustrates the square-rooted geometrically-averaged eigenvalues, $\bar{\lambda}_k$, of trained GM covariance matrices for the $5 \times 5 \times 3$ GM-MRF with 66 components. The numbers within the figure correspond to the indices of training groups in Table II. Within each group, the GM components are sorted from the most probable to the least probable based on
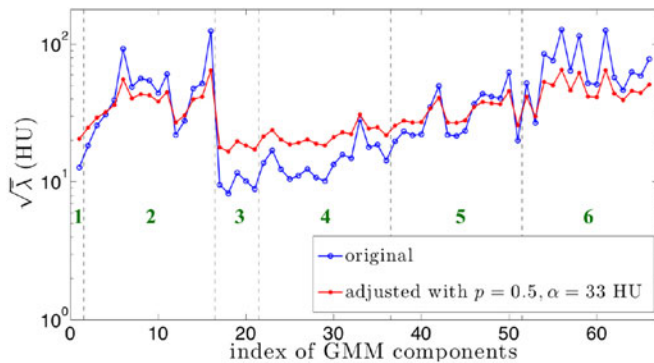
Fig. 6.   The covariances of the originally trained GMM and the adjusted GMM with $p = 0.5$ and $\alpha = 33$ HU in (29), for a $5 \times 5 \times 3$ patch case. More precisely, the figure plots the square root of geometrically-averaged eigenvalues $\bar{\lambda}_k$ of the GMM covariances R_k, as $\bar{\lambda}_k = |R_k|^{\frac{1}{L}}$ with $L = 75$. Numbers within the figure correspond to the indices of the training groups. Within each group, the GMM components are sorted from the most probable to the least probable. The adjusted model increases the eigenvalues for group 1, 3, 4, and decreases the eigenvalues for most of group 2, 5, and 6.

the trained mixture probabilities, $\pi_{i,k}$. Fig. 6 shows that different groups present different amounts of regularization strength. Note that there is a large variation in average eigenvalues for different groups, which leads to highly varying regularization strength for different image contents. For example, group 3 has much smaller average eigenvalues than groups 2 and 6, which indicates that during the reconstruction, patches dominated by group 3, typically the smooth soft-tissue patches, will be regularized more heavily than patches dominated by group 2 and 6, typically lung and bone patches respectively, and therefore will contain less noise in the reconstructed image.

As introduced in Section IV, we will apply the simple parameterization of (29) to the trained GM covariances to tune the visual quality. The red plot in Fig. 6 illustrates this adjusted model with $p = 0.5$ and $\alpha = 33$ HU. By adjusting the model parameters, we increase the eigenvalues of group 3 and 4, which will consequently reduce the regularization for smooth and low-contrast soft-tissue contents, while we decrease the eigenvalues of group 2, 5, and 6, which will lead to stronger regularization for lung, high-contrast edge, and bone.

B. Testing

We tested the proposed GM-MRF as a prior for model-based inversion problems. The testing data included different scans on the same and different patients as compared to the training data. All the data were acquired using GE Discovery CT750 HD scanners with 64×0.625 mm helical mode. For testing purposes, the same patient whose normal-dose scan was used for training, was scanned on the same imaging site with the same CT scanner, but using a different protocol with pitch 1.375:1 and a much lower current of 40 mA. Different patients for testing, including a GE Performance Phantom (GEPP), were scanned on various imaging sites with a variety of acquisition protocols. More precisely, the GEPP datasets were acquired with 120 kVp, 1 s/rotation, pitch 0.516:1, with four different magnitudes of tube current as 290 mA, 145 mA, 75 mA, and 40 mA, and were reconstructed

in 135 mm FOV. A normal-dose scan of a new patient was acquired with 120 kVp, 200 mA, 0.5 s/rotation, pitch 0.984:1, and reconstructed in 320 mm FOV. This particular dataset was used for both denoising and reconstruction experiments. The supplementary material includes reconstruction results of a low-dose scan acquired with another new patient.

1) 2-D Image Denoising: We first tested the GM-MRF model in a 2-D image denoising experiment. The ground-truth image in Fig. 7(a) was obtained from the reconstructed images of the aforementioned normal-dose scan of a new patient whose data was not used for training. Then, we added Gaussian white noise to the ground truth to generate the noisy image in Fig. 7(b). Different denoising methods were then applied to the noisy image.

We experimented with a few GM-MRFs with various sizes of patch models, that is, 2-D GM-MRF with 3×3, 5×5, and 7×7 patch models, to study the impact of patch sizes in the proposed GM-MRF model. In addition, for each patch size, we experimented with the original model obtained directly from training and the adjusted model with $p = 0.5, \alpha = 33$ HU in (29), to study the effect of covariance scaling.

We will compare our GM-MRF methods with a number of widely used methods, including the q-GGMRF method [3], K-SVD method [27], BM3D method [29], and NLM method [28]. The q-GGMRF method was implemented with 3×3 neighborhood with parameters $p = 2, q = 1.2, c = 10$ HU. The K-SVD method was performed by using the software in [59] with 7×7 patch size and 512 dictionary entries. The BM3D method was performed by using the software in [60] with 8×8 patch. The NLM method was performed by using the software in [61] with 9×9 patch and 21×21 search window. We adjusted the regularization strength for all methods to achieve the lowest root-mean-square error (RMSE) between the recovered image and the ground truth.

2) 3-D CT Reconstruction: We tested the GM-MRF method in a number of 3-D CT reconstruction experiments. We first conducted phantom studies using different GM-MRF models with various parameter settings to study the impact of change in model parameters on the final reconstructed images. Then, we picked only one GM-MRF model and then applied this model to all subsequent experiments with clinical datasets. Thus, no additional training was conducted for the test data.

We conducted both qualitative and quantitative comparisons by using GEPP reconstructions. The GEPP contains a plexiglas insert with cyclic water bars and a 50 μm diameter tungsten wire placed in water. We measured the mean and standard deviation within fixed ROIs in flat regions to assess the reconstruction accuracy and noise. In addition, we measured the modulation transfer function (MTF) using the wire to assess the in-plane resolution and contrast. We reported the 10% MTF since it reflects the visual resolution of the image, with higher value indicating finer texture, which is a desirable image quality especially for a low-dose condition. We compared the visual quality for clinical reconstructions.

We studied the influence of different model parameters in the GM-MRF model using GEPP reconstructions, including the number of GM components (K), the patch size (L), and param-
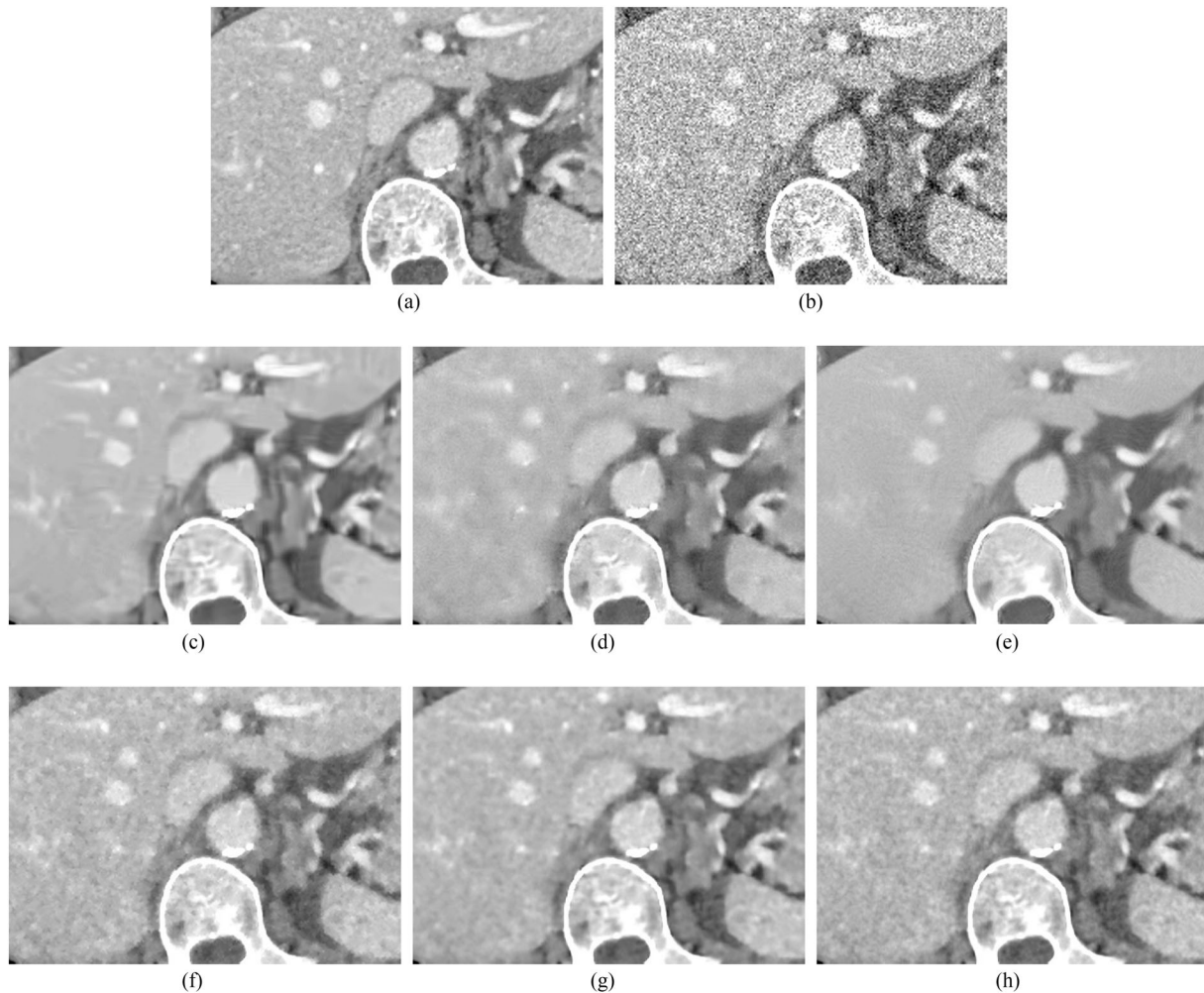
Fig. 7. Denoising results with different methods. (a) ground truth, (b) noisy (RMSE: 39.88 HU), (c) BM3D (RMSE: 13.35 HU), (d) K-SVD (RMSE: 14.57 HU), (e) NLM (RMSE: 14.82 HU), (f) q-GGMRF (RMSE: 15.96 HU), (g) original 5×5 GM-MRF (RMSE: 13.78 HU), (h) adjusted 5×5 GM-MRF (RMSE: 14.33 HU). Individual image is zoomed to a region containing soft tissue, contrast, and bone for display purposes. The RMSE value between each reconstructed image and the ground truth is reported. Display window: $[-100\ 200]$ HU. GM-MRF methods achieve lower RMSEs and better visual quality than q-GGMRF, K-SVD, and NLM methods. Though having a slightly higher RMSE, GM-MRF with original model preserves real textures in soft tissue without creating severe artifacts, while BM3D tends to over-smooth the soft tissue and introducing artificial, ripple-like structures. In addition, though compromising the RMSE than the original model, GM-MRF with the adjusted model produces better visual quality, especially for the soft-tissue texture.

eters p and α for covariance scaling. For fair comparison, we matched the noise level between different models by adjusting the global regularization parameter, σ_x, in (27). That is, for a given dose level, we matched the standard deviations within a selected flat region between different models such that the absolute difference of the two is within 1 HU.

In addition to evaluating the behavior of different parameters within the GM-MRF prior, we will also compare the MBIR using GM-MRF prior with two widely used reconstruction methods: FBP using a standard kernel and MBIR using q-GGMRF prior [3] with reduced regularization. Note that we intentionally reduced the regularization for MBIR with the q-GGMRF prior so as to obtain higher resolution, which lead to much higher noise as well. We matched the noise level between the q-GGMRF and GM-MRF methods. However, for low-dose datasets, it is challenging to match the noise between those two methods due to the excessive speckle noise produced by using an under-regularized q-GGMRF prior. Thus, we will instead

demonstrate that the GM-MRF prior achieves higher resolution with even less noise than the q-GGMRF prior. Besides direct comparison, we also use FBP reconstruction as an illustration of the current dose level.

VI. RESULTS

A. 2-D Image Denoising

Fig. 7 presents the denoising result with different methods. It shows that the mean-square-error (MSE) achieved by the 5×5 GM-MRF method with original covariances is slightly higher than the BM3D method, but significantly lower than the q-GGMRF, K-SVD, and NLM methods. Qualitatively, the GM-MRF method with the original model produces sharper edges and less speckle noise than the q-GGMRF method, and preserves more fine structures and detail than K-SVD and NLM methods. The BM3D method seems to produce more enhanced fine structures than the GM-MRF method due to its strong structure-
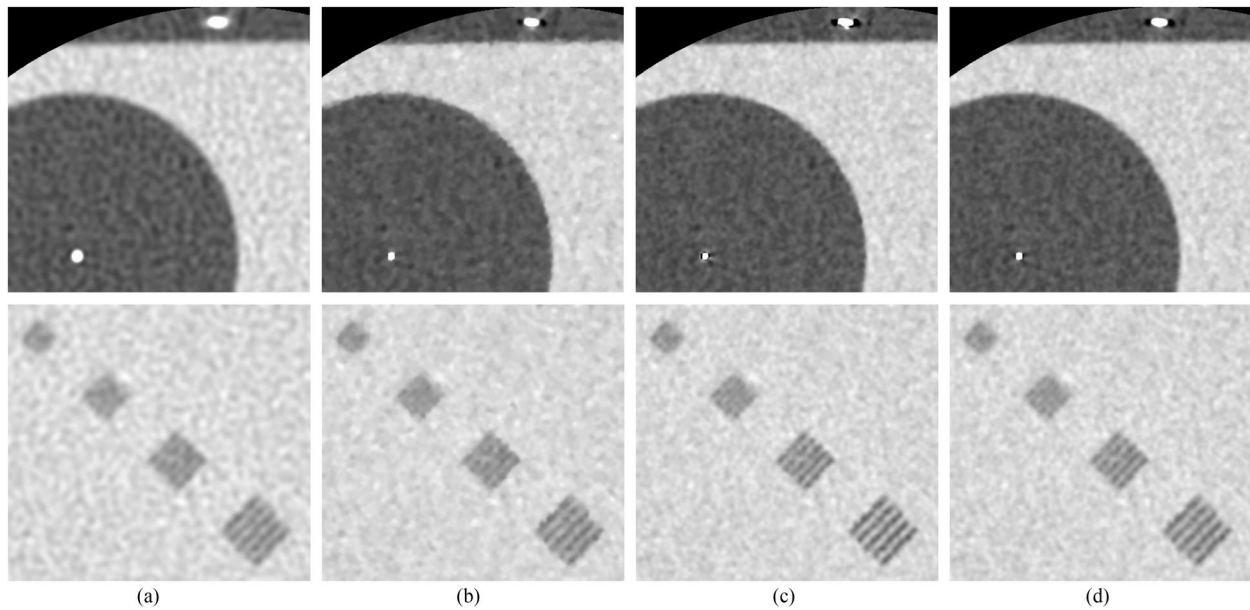
Fig. 8. GEPP reconstruction with data collected under 290 mA. Individual image is zoomed to small FOVs for display purposes. From left to right, the columns represent (a) FBP, (b) MBIR with q-GGMRF with reduced regularization, (c) MBIR with original $5 \times 5 \times 3$ GM-MRF, and (d) MBIR with adjusted $5 \times 5 \times 3$ GM-MRF with $p = 0.5, \alpha = 33$ HU. The top row shows the wire section and the bottom row shows the resolution bars with the display window as $[-85 \ 165]$ HU. MBIR with GM-MRF priors generate images with sharper high-contrast objects and better texture than MBIR with q-GGMRF prior at a comparable noise level. Moreover, for GM-MRF priors, the adjusted model produces better rendering of the high-contrast objects than the original model.
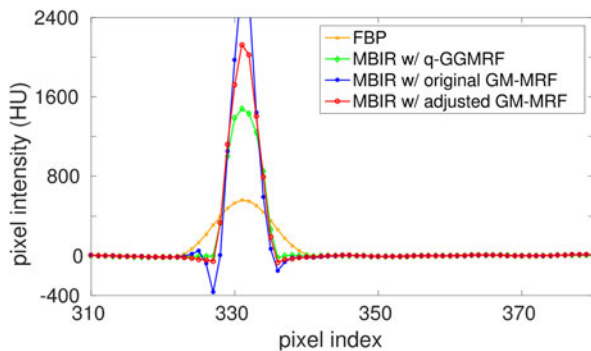

Fig. 9. Profile plot through the center of the tungsten wire in Fig. 8. As compared to the q-GGMRF result (green), there is an undershoot-like artifact near the wire in the MBIR result with the original GM-MRF (blue), which is substantially reduced by using the adjusted GM-MRF model (red). Also notice that increasing the regularization for high-contrast regions tends to reduce the contrast (reduced peak from blue to red); however, the resultant contrast still remains more substantial than the q-GGMRF result.

preserving behavior. However, it also tends to over-smooth the soft tissue region while creating some artificial, ripple-like structures and texture. Alternatively, the GM-MRF method is able to preserve some real texture in soft tissue without inducing severe artifacts, which can be important in some medical applications.

The supplementary material include the 2-D denoising results by using GM-MRF methods with different sizes of patch models. It is observed that the RMSE value decreases as the patch size increases in the GM-MRF models, since the model becomes more expressive as patch size increases. Moreover, denoising results produced by using the GM-MRF with larger patches appear visually more natural than results associated with the GM-MRF with smaller patches.

Interestingly, though compromising the MSE, the GM-MRF methods with adjusted models produce images with better visual quality than those with original models. The better visual quality is achieved with improved soft-tissue texture and better rendering of high-contrast structures. This is because the MAP estimate tends to over-regularize the low-contrast regions and under-regularize the high-contrast regions in the image. Therefore, by adjusting the regularization strength in different contrast regions, we may achieve desirable visual quality but with the compromise in the MSE.

B. 3-D Phantom Reconstruction

Fig. 8 shows the GEPP reconstruction under normal X-ray dosage, with zoomed-in images for the tungsten wire and cyclic bars. It shows that MBIR with the traditional q-GGMRF prior produces sharper images with less noise than FBP, as indicated by smoother homogeneous regions, a smaller reconstructed wire, and more enhanced cycling bars. As a further improvement, MBIR with the $5 \times 5 \times 3$ GM-MRF priors produce even sharper image than MBIR with the q-GGMRF prior at a comparable noise level. The GM-MRF priors also improve the texture in smooth regions over the q-GGMRF method by reducing the speckle noise and grainy texture. For the GM-MRF priors, the original model shows a sharper tungsten wire as compared to the adjusted model, since the adjusted model increases regularization for high-contrast edge (group 5) and bone (group 6), as shown in Fig. 6. However, the limited regularization for high-contrast edge and bone in the original model also leads to noisy rendering of high-attenuation objects, such as the non-circular tungsten wire and irregularly shaped small metal insertion.

Fig. 9 presents a profile line through the tungsten wire. As compared to the q-GGMRF result, there is clearly an under-
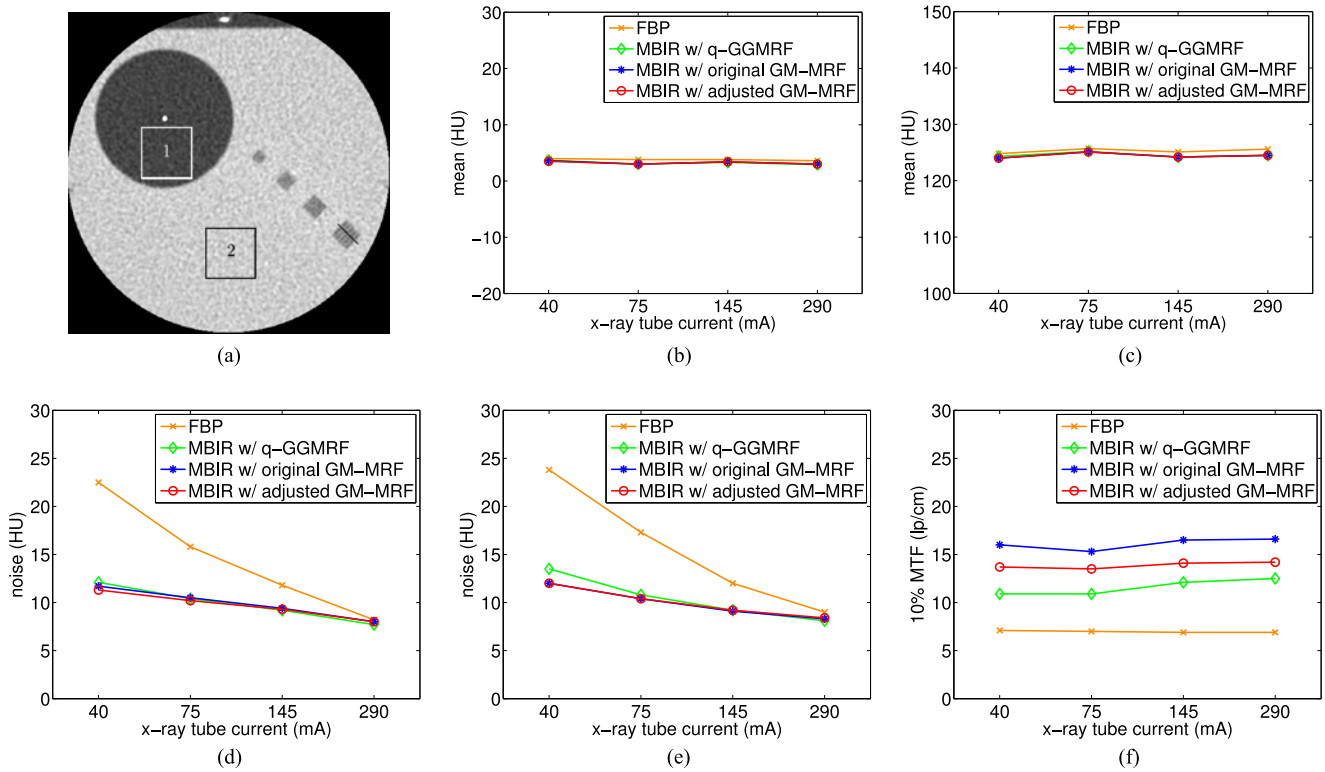
Fig. 10.   Quantitative measurements for GEPP reconstructions. (a) measurements, (b) mean within ROI 1, (c) mean within ROI 2, (d) noise within ROI 1, (e) noise within ROI 2, (f) 10% MTF. Four different magnitudes of X-ray tube current were used in data acquisition to achieve different X-ray dose levels. The mean values along with noise were measured within two different ROIs in (a). The MTF values were measured at the tungsten wire. Figure demonstrates that MBIR with $5 \times 5 \times 3$ GM-MRF priors improve the in-plane resolution in (f) while producing comparable or even less noise than FBP and MBIR with the q-GGMRF prior in (d) and (e), without affecting the reconstruction accuracy in (b) and (c).
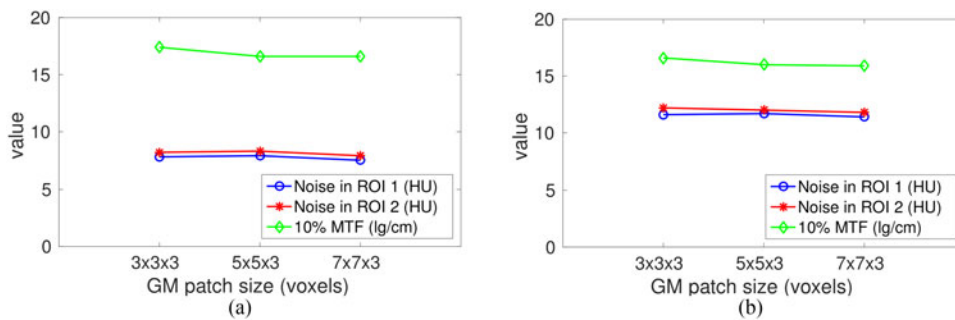

Fig. 11.   Quantitative measurements for GEPP reconstructions with different patch sizes with (a) 290 mA data and (b) 40 mA data. With matched noise level in homogeneous regions, trained GM-MRF models with various patch sizes can achieve similar high-contrast resolution.

shoot near the wire in the original GM-MRF result, which, however, is substantially reduced by using the adjusted GM-MRF. We believe that this type of artifact is caused by the under-regularization of high-contrast regions in the original GM-MRF model, and can be mitigated by increasing the regularization in those regions using the proposed covariance scaling method. To further improve the result, one may continue increasing the regularization for high-contrast regions by using the proposed systematic approach, or simply change the regularization for mixture components associated with high-contrast regions. Additionally, notice that increasing the regularization for high-contrast regions tends to reduce the contrast; however, the resultant contrast still remains more substantial than the q-GGMRF result.

The visual comparison is further verified by quantitative measurements in Fig. 10, which presents the measurements of reconstruction accuracy, noise, and resolution, of GEPP reconstructions at different dose levels. It shows that MBIR with $5 \times 5 \times 3$ GM-MRF priors improve the resolution (in Fig. 10(f)) while producing comparable or even less noise than FBP and MBIR with the q-GGMRF prior (in Fig. 10(d) and (e)), without affecting the reconstruction accuracy (in Fig. 10(b) and (c)).

Fig. 11 presents the quantitative measurements for GEPP reconstructions produced by various GM-MRF models. It shows that, with matched noise level in homogeneous regions, trained GM-MRF models with different patch sizes can achieve similar high-contrast resolution.

290 mA

40 mA

(a)   (b)   (c)   (d)   (f)
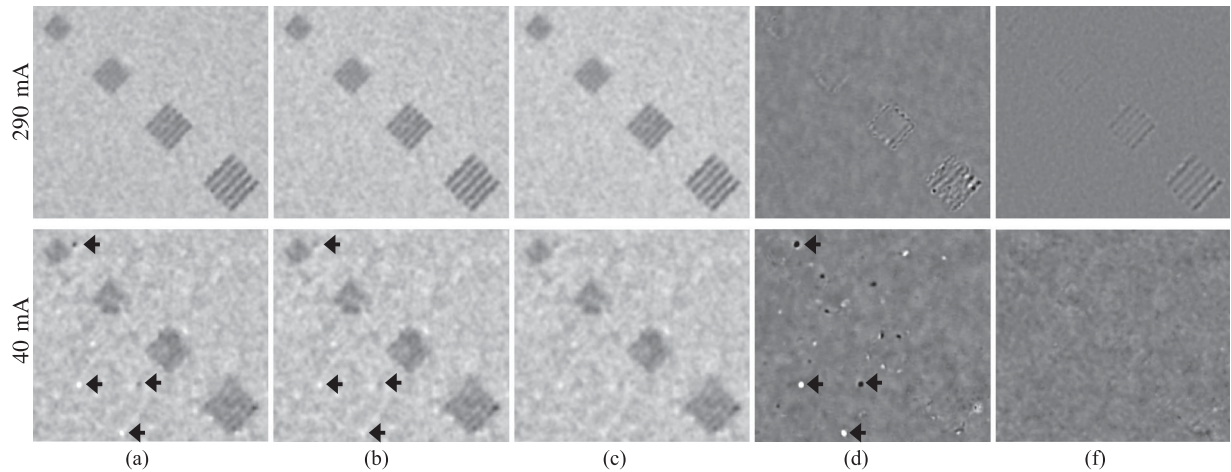
Fig. 12.   GEPP reconstructions with 66-component GM-MRF model with different sizes of patch models. (a) $3 \times 3 \times 3$ patch, (b) $5 \times 5 \times 3$ patch, (c) $7 \times 7 \times 3$ patch, (d) difference: (a)-(b), (e) difference: (b)-(c). Individual image is zoomed to a small FOV containing cyclic bars for display purposes. From top to bottom: GEPP reconstructions with 290 mA data and 40 mA data, respectively. Display window: for GEPP images: $[-110\ 190]$ HU; for difference images: $[-15\ 15]$ HU. The $3 \times 3 \times 3$ result with 40 mA data contains many noisy spots, as indicated by the arrows, which are suppressed substantially in the result with larger patch models. It is also observed that the $7 \times 7 \times 3$ result with 290 mA data has slightly blurred cyclic bars as compared to results with smaller patch models.

(a)   (b)   (c)
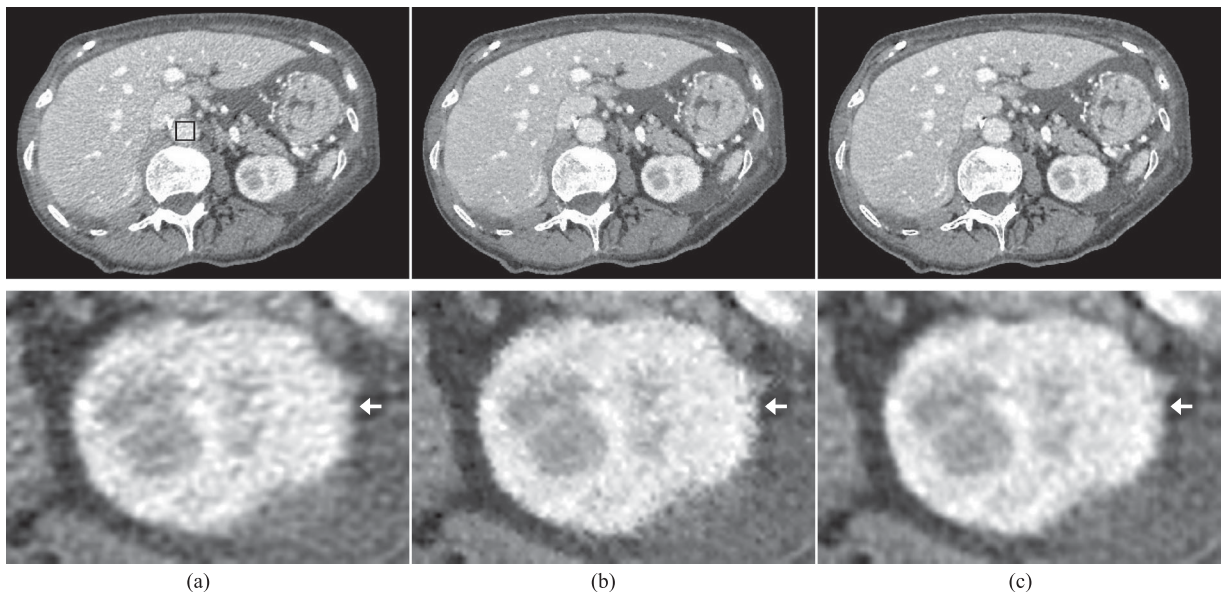
Fig. 13.   An abdominal axial slice of the normal-dose clinical reconstruction. From left to right, the columns represent (a) FBP (noise: 19.14 HU), (b) MBIR with q-GGMRF with reduced regularization (noise: 14.43 HU), and (c) MBIR with adjusted GM-MRF with $p = 0.5$, $\alpha = 33$ HU (noise: 14.02 HU). Top row shows the full FOV of the reconstructed images, while the bottom row shows a zoomed-in FOV. Noise standard deviation is measured within an ROI in aorta, as illustrated in the FBP image, and is reported for each method. Display window is $[-110\ 190]$ HU. Note the reduced jagged appearance in the GM-MRF reconstruction.

Fig. 12 presents the GEPP reconstructions for 290 mA and 40 mA data, produced by using GM-MRF with three different sizes of patch models. As shown in the figure, larger patch is more robust to individual noise spots and can generate less grainy texture. This is more significant in 40 mA images due to the low SNR condition. It is also observed in the 290 mA images that increased patch size introduces some blurriness to the cyclic bars. We believe this is because larger patches belong to a higher-dimensional feature space and therefore contains more potential variations. While the GM-MRF with larger patch is able to provide more expressive patch models than that with smaller patch, it also requires more components to capture all possible variations. With limited number of components,

GM-MRFs with larger patches may compromise the modeling of fine structures, and therefore perform inferiorly for fine structures as compared to GM-MRFs with sufficient structure modeling, i.e., smaller patch.

The supplementary material includes 3-D GEPP reconstruction results produced by using GM-MRF with various number of components. The results show that the GM-MRF model tends to produce more enhanced fine structures with increasing number of components. This implies that additional GM components will capture the behavior of edges and structures.

In addition, the supplementary material presents the study on the impact of parameters p and α of GM covariance scaling on the reconstructed images. With matched noise
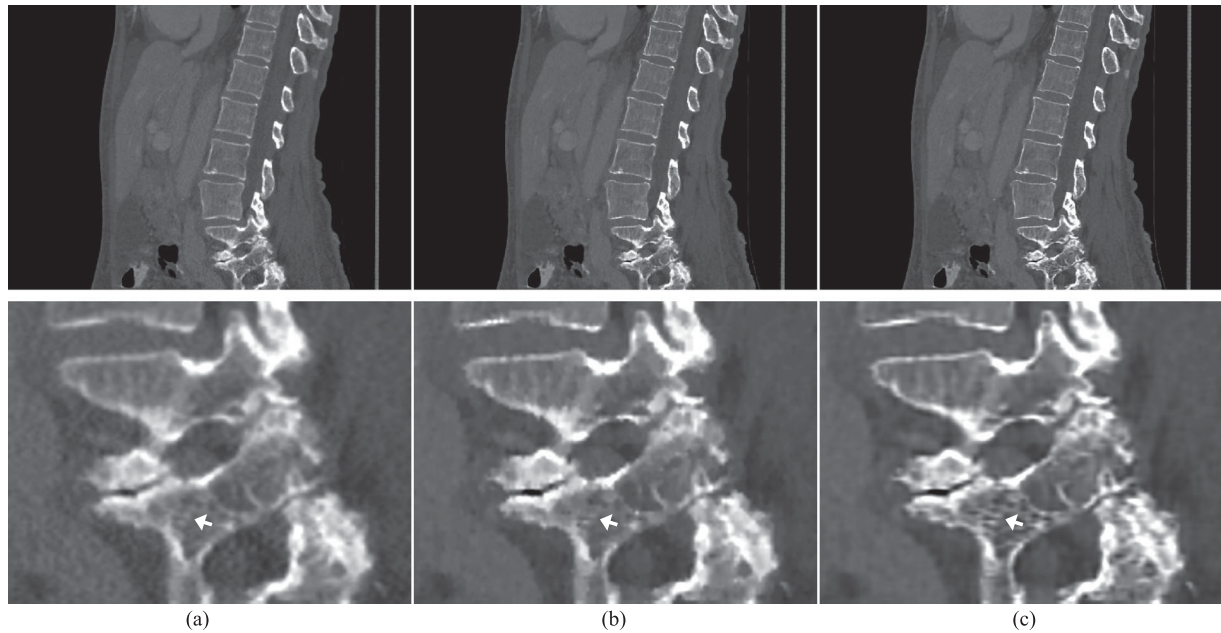
(a) (b) (c)

Fig. 14.   A sagittal view of the normal-dose clinical reconstruction in bone window. From left to right, the columns represent (a) FBP, (b) MBIR with q-GGMRF with reduced regularization, and (c) MBIR with adjusted GM-MRF with $p = 0.5, \alpha = 33$ HU. Top row shows the full FOV of the reconstructed images, while the bottom row shows a zoomed-in FOV. Display window is $[-300\ 900]$ HU. Note the honeycomb structure of the trabecular bone reconstructed by the GM-MRF prior, which is missing in other methods.
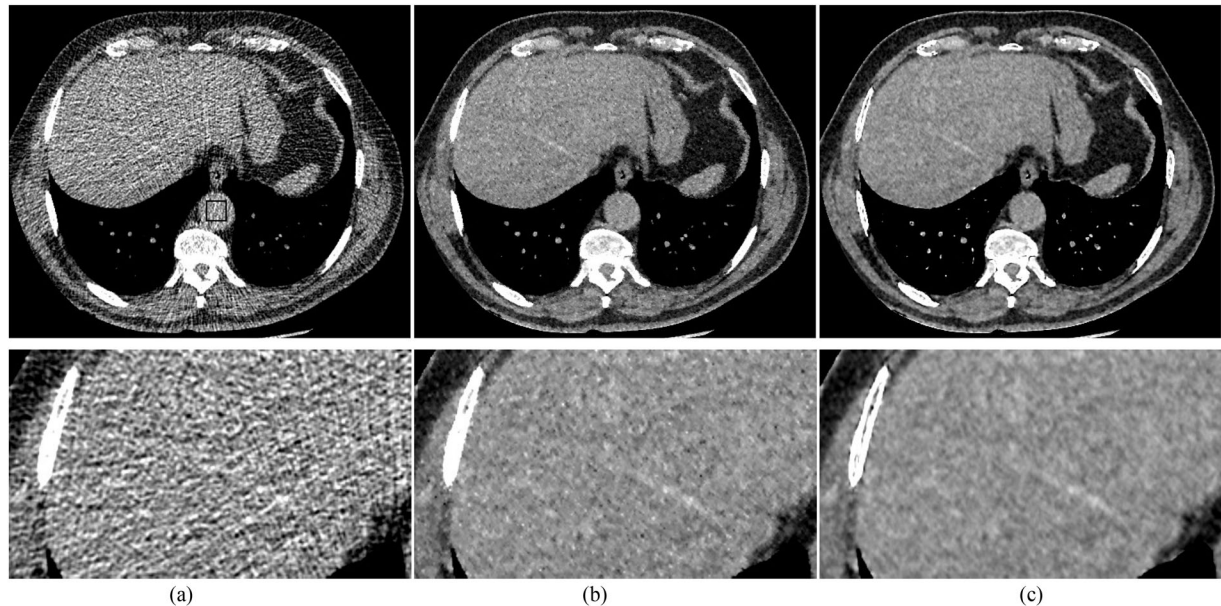

(a) (b) (c)

Fig. 15.   An abdominal axial slice of the low-dose clinical reconstruction. From left to right, the columns represent (a) FBP (noise: 62.15 HU), (b) MBIR with q-GGMRF with reduced regularization (noise: 25.46 HU), and (c) MBIR with adjusted GM-MRF with $p = 0.5, \alpha = 33$ HU (noise: 18.41 HU). Top row shows the full FOV of the reconstructed images, while the bottom row shows a zoomed-in FOV. Noise standard deviation is measured within an ROI in aorta, as illustrated in the FBP image, and is reported for each method. Display window is $[-160\ 240]$ HU. Note the suppression of speckle noise in soft tissue and improvement of sharpness in bone provided by MBIR with the GM-MRF prior.

level in homogeneous regions, increasing the value of p with fixed α leads to reduction of the covariances of high-contrast edges (group 5) and bones (group 6), and therefore results in decreasing high-contrast resolution. On the other hand, increasing the value of α with fixed p increases the covariances for all GM components, and therefore encourages distribution overlapping and may consequently reduce the specification of structures and edges, which leads to blurriness in cyclic bars.

C. 3-D Reconstruction for Clinical Datasets

For clinical datasets, we used the $5 \times 5 \times 3$ GM-MRF model with 66 components with covariances adjusted with parameters $p = 0.5$ and $\alpha = 33$ HU, to achieve balanced visual quality between low- and high-contrast regions.

Figs. 13–14 present the reconstructions with the normal-dose scan of the new patient whose data was not used for training.
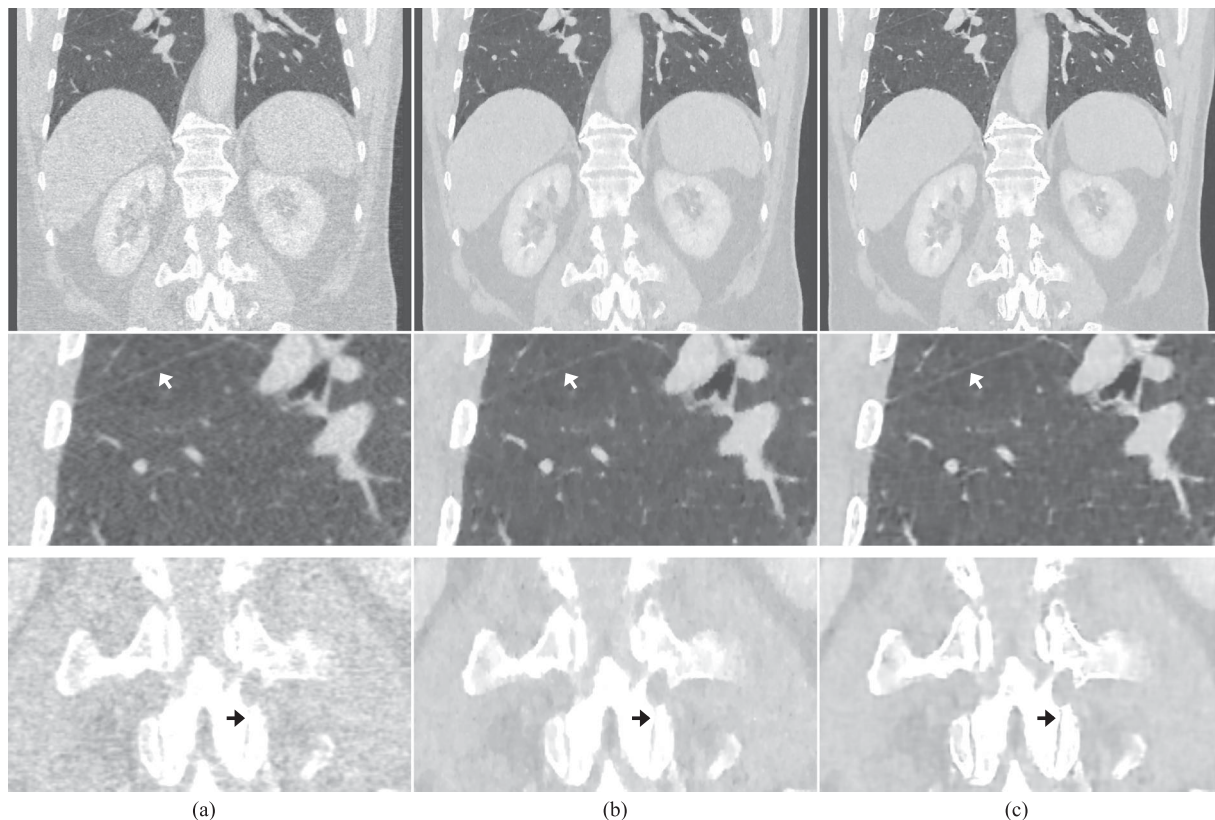
Fig. 16.   A coronal view of the normal-dose clinical reconstruction in lung window. From left to right, the columns represent (a) FBP, (b) MBIR with q-GGMRF with reduced regularization, and (c) MBIR with adjusted GM-MRF with $p = 0.5, \alpha = 33$ HU. Top row shows the full FOV of the reconstructed images, while the bottom row shows a zoomed-in FOV. Display window is $[-1400\ 400]$ HU. Note the lung fissure and bone structure are reconstructed much more clearly by using the GM-MRF prior as compared to the other methods.

As compared to FBP, MBIR with GM-MRF priors produce images with sharper bones, more lung details, as well as less noise in soft tissues. When compared to MBIR with traditional q-GGMRF prior with similar noise level, MBIR with GM-MRF priors reduce the jagged appearance in edges in Fig. 13. These improvements are due to better edge definition in the patch-based model over traditional pair-wise models. Moreover, MBIR with GM-MRF prior reveals more fine structures in bone, such as the honeycomb structure of trabecular bones in Fig. 14. This indicates that the GM-MRF model is also a flexible prior and inherently allows different regularization strategies for different tissues in CT images. This flexibility allows CT reconstructions with great soft-tissue quality while simultaneously preserving the resolution in regions with larger variation, such as bone and lung.

Figs. 15 and 16 present the reconstructions with the low-dose of the same patient whose normal-dose scan was used for training. It is shown that all the improvements revealed by experiment with normal-dose data can be observed more clearly in the low-dose situation, where the better image prior model is perhaps more valuable. Fig. 15 shows that the GM-MRF prior improves the texture in soft tissue without compromising the fine structures and details, as compared to the other methods. Particularly, when compared to MBIR with traditional q-GGMRF prior, MBIR with GM-MRF prior reduces the speckle noise in liver while still maintaining the normal texture and edge

definition. Fig. 16 shows the improved resolution in lung and bone as produced by the GM-MRF prior. More specifically, the zoomed-in images show that the lung fissure reconstructed by MBIR with GM-MRF have comparable resolution as that produced by FBP, which is blurred by MBIR with q-GGMRF. The GM-MRF prior also leads to much clearer bone structure as compared to other methods. These improvements demonstrate the material-specific regularization capability of the GM-MRF prior.

In addition, the supplementary material includes reconstruction results of a low-dose scan of another new patient. It is worth emphasizing that we only trained the GM-MRF model with data from a single patient scan, and then applied it to different scans on the same and different patients. However, we did not notice any systematic difference between the two cases. We believe that ideally this GM-MRF model can be trained off-line on a pool of normal-dose images from various patients and the resulting prior model can then be fixed and applied to future patients.

VII. CONCLUSION

In this paper, we introduced a novel Gaussian-mixture Markov random field (GM-MRF) image model along with the tools to use it as a prior for MBIR. The proposed method constructs an image model by seaming together GM patch models. In addition, we presented an analytical framework for

computing the MAP estimate with the GM-MRF prior using an exact surrogate function. We also proposed a systematic approach to adjust the covariances of the GM components of the GM-MRF model, in order to control the sharpness in low- and high-contrast regions of the reconstruction separately. The results in image denoising and multi-slice CT reconstruction experiments demonstrate improved image quality and material-specific regularization by the GM-MRF prior.

APPENDIX
PROOF OF LEMMA: SURROGATE FUNCTIONS FOR LOGS OF EXPONENTIAL MIXTURES

Lemma: Let $f : \Re^N \to \Re$ be a function of the form

$$f(x) = \sum_k w_k \exp\{-v_k(x)\} \, ,$$

where $w_k \in \Re^+$, $\sum_k w_k > 0$, and $v_k : \Re^N \to \Re$. Furthermore $\forall (x, x') \in \Re^N \times \Re^N$ define the function

$$q(x; x') \triangleq -\log f(x') + \sum_k \tilde{\pi}_k (v_k(x) - v_k(x')) \, ,$$

where $\tilde{\pi}_k = \frac{w_k \exp\{-v_k(x')\}}{\sum_l w_l \exp\{-v_l(x')\}}$. Then $q(x; x')$ is a surrogate function for $-\log f(x)$, and $\forall (x, x') \in \Re^N \times \Re^N$,

$$q(x'; x') = -\log f(x')$$
$$q(x; x') \geq -\log f(x)$$

Proof:

$$\log f(x) = \log f(x') + \log \left\{ \frac{f(x)}{f(x')} \right\}$$

$$= \log f(x') + \log \left\{ \sum_k \left(\frac{w_k}{f(x')} \right) \exp\{-v_k(x)\} \right\}$$

$$= \log f(x') + \log \left\{ \sum_k \left(\frac{w_k \exp\{-v_k(x')\}}{\sum_l w_l \exp\{-v_l(x')\}} \right) \right.$$
$$\left. \times \exp\{-v_k(x) + v_k(x')\} \right\}$$

$$= \log f(x') + \log \left\{ \sum_k \tilde{\pi}_k \exp\{-v_k(x) + v_k(x')\} \right\}$$

$$\geq \log f(x') + \sum_k \tilde{\pi}_k \{-v_k(x) + v_k(x')\}$$

where

$$\tilde{\pi}_k \triangleq \frac{w_k \exp\{-v_k(x')\}}{\sum_l w_l \exp\{-v_l(x')\}} \, .$$

The last inequality results from Jensen's inequality. Taking the negative of the final expression results in

$$-\log f(x) \leq -\log f(x') + \sum_k \tilde{\pi}_k \{v_k(x) - v_k(x')\} \triangleq q(x; x') \, ,$$

and evaluating this result at $x = x'$ results in

$$-\log f(x') = q(x'; x') \, .$$

REFERENCES

[1] D. Boas, D. Brooks, E. Miller, C. DiMarzio, M. Kilmer, R. Gaudette, and Q. Zhang, "Imaging the body with diffuse optical tomography," *IEEE Signal Process. Mag.*, vol. 18, no. 6, pp. 57–75, Nov. 2001.

[2] J. Qi and R. M. Leahy, "Iterative reconstruction techniques in emission computed tomography," *Phys. Med. Biol.*, vol. 51, no. 15, p. R541, 2006.

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

[4] J. A. Fessler, "Model-based image reconstruction for MRI," *IEEE Signal Process. Mag.*, vol. 27, no. 4, pp. 81–89, Jul. 2010.

[5] R. Zhang, J.-B. Thibault, C. A. Bouman, K. D. Sauer, and J. Hsieh, "Model-based iterative reconstruction for dual-energy X-ray CT using a joint quadratic likelihood model," *IEEE Trans. Med. Imag.*, vol. 33, no. 1, pp. 117–134, Jan. 2014.

[6] S. Venkatakrishnan, L. Drummy, M. Jackson, M. De Graef, J. Simmons, and C. A. Bouman, "Model-based iterative reconstruction for bright-field electron tomography," *IEEE Trans. Comput. Imag.*, vol. 1, no. 1, pp. 1–15, Jun. 2015.

[7] P. Jin, C. A. Bouman, and K. D. Sauer, "A model-based image reconstruction algorithm with simultaneous beam hardening correction for X-ray CT," *IEEE Trans. Comput. Imag.*, vol. 1, no. 3, pp. 200–216, Oct. 2015.

[8] A. K. Mohan, S. Venkatakrishnan, J. Gibbs, E. Gulsoy, X. Xiao, M. De Graef, P. Voorhees, and C. A. Bouman, "TIMBIR: A method for time-space reconstruction from interlaced views," *IEEE Trans. Comput. Imag.*, vol. 1, no. 2, pp. 96–111, Aug. 2015.

[9] R. C. Nelson, S. Feuerlein, and D. T. Boll, "New iterative reconstruction techniques for cardiovascular computed tomography: How do they work, and what are the advantages and disadvantages?" *J. Cardiovascular Comput. Tomography*, vol. 5, no. 5, pp. 286–292, 2011.

[10] Y. Yamada, M. Jinzaki, Y. Tanami, E. Shiomi, H. Sugiura, T. Abe, and S. Kuribayashi, "Model-based iterative reconstruction technique for ultralow-dose computed tomography of the lung: A pilot study," *Investigative Radiology*, vol. 47, no. 8, pp. 482–489, 2012.

[11] J. Besag, "Spatial interaction and the statistical analysis of lattice systems," *J. Roy. Statist. Soc. Series B (Methodological)*, vol. 36, pp. 192–236, 1974.

[12] K. D. Sauer and C. A. Bouman, "A local update strategy for iterative reconstruction from projections," *IEEE Trans. Signal Process.*, vol. 41, no. 2, pp. 534–548, Feb. 1993.

[13] B. De Man, J. Nuyts, P. Dupont, G. Marchal, and P. Suetens, "Reduction of metal streak artifacts in x-ray computed tomography using a transmission maximum a posteriori algorithm," *IEEE Trans. Nucl. Sci.*, vol. 47, no. 3, pp. 977–981, Jun. 2000.

[14] I. A. Elbakri and J. A. Fessler, "Statistical image reconstruction for polyenergetic X-ray computed tomography," *IEEE Trans. Med. Imag.*, vol. 21, no. 2, pp. 89–99, Feb. 2002.

[15] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1, pp. 259–268, 1992.

[16] E. Y. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," *Phys. Med. Biol.*, vol. 53, no. 17, pp. 4777–4807, 2008.

[17] J. Tang, B. E. Nett, and G.-H. Chen, "Performance comparison between total variation (TV)-based compressed sensing and statistical iterative reconstruction algorithms," *Phys. Med. Biol.*, vol. 54, no. 19, p. 5781, 2009.

[18] L. Ritschl, F. Bergner, C. Fleischmann, and M. Kachelrieß, "Improved total variation-based CT image reconstruction applied to clinical data," *Phys. Med. Biol.*, vol. 56, no. 6, p. 1545, Feb. 2011.

[19] Y. Liu, J. Ma, Y. Fan, and Z. Liang, "Adaptive-weighted total variation minimization for sparse data toward low-dose x-ray computed tomography image reconstruction," *Phys. Med. Biol.*, vol. 57, no. 23, p. 7923, 2012.

[20] T. Li, X. Li, J. Wang, J. Wen, H. Lu, J. Hsieh, and Z. Liang, "Nonlinear sinogram smoothing for low-dose x-ray CT," *IEEE Trans. Nucl. Sci.*, vol. 51, no. 5, pp. 2505–2513, Oct. 2004.

[21] P. J. La Riviere, "Penalized-likelihood sinogram smoothing for low-dose CT," *Med. Phys.*, vol. 32, no. 6, pp. 1676–1683, 2005.

[22] J. Wang, T. Li, H. Lu, and Z. Liang, "Penalized weighted least-squares approach to sinogram noise reduction and image reconstruction for low-dose X-ray computed tomography," *IEEE Trans. Med. Imag.*, vol. 25, no. 10, pp. 1272–1283, Oct. 2006.

[23] I.-T. Hsiao, A. Rangarajan, and G. Gindi, "Joint-MAP Bayesian tomographic reconstruction with a gamma-mixture prior," *IEEE Trans. Image Process.*, vol. 11, no. 12, pp. 1466–1477, Dec. 2002.

[24] I.-T. Hsiao, A. Rangarajan, and G. Gindi, "Bayesian image reconstruction for transmission tomography using deterministic annealing," *J. Electron. Imag.*, vol. 12, no. 1, pp. 7–16, 2003.

[25] G. Wang, L. Schultz, and J. Qi, "Statistical image reconstruction for muon tomography using a Gaussian scale mixture model," *IEEE Trans. Nucl. Sci.*, vol. 56, no. 4, pp. 2480–2486, Aug. 2009.

[26] A. Mehranian and H. Zaidi, "Joint estimation of activity and attenuation in whole-body TOF PET/MRI using constrained Gaussian mixture models," *IEEE Trans. Med. Imag.*, vol. 34, no. 9, pp. 1808–1821, Aug. 2015.

[27] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.

[28] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, vol. 2, 2005, pp. 60–65.

[29] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.

[30] S. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, "Plug-and-play priors for model based reconstruction," in *Proc. IEEE Global Conf. Signal Inform. Process.*, Dec. 2013, pp. 945–948.

[31] S. H. Chan, "Algorithm-induced prior for image restoration," (Feb. 2016). [Online]. Available: https://arxiv.org/abs/1602.00715

[32] A. M. Teodoro, J. M. Bioucas-Dias, and M. A. Figueiredo, "Image restoration and reconstruction using variable splitting and class-adapted image priors," (May 2016). [Online]. Available: http://arxiv.org/abs/1602.04052

[33] S. Ravishankar and Y. Bresler, "MR image reconstruction from highly undersampled k-space data by dictionary learning," *IEEE Trans. Med. Imag.*, vol. 30, no. 5, pp. 1028–1041, May 2011.

[34] Q. Xu, H. Yu, X. Mou, L. Zhang, J. Hsieh, and G. Wang, "Low-dose X-ray CT reconstruction via dictionary learning," *IEEE Trans. Med. Imag.*, vol. 31, no. 9, pp. 1682–1697, Aug. 2012.

[35] J. Huang, J. Ma, N. Liu, H. Zhang, Z. Bian, Y. Feng, Q. Feng, and W. Chen, "Sparse angular CT reconstruction using non-local means based iterative-correction POCS," *Comput. Biol. Med.*, vol. 41, no. 4, pp. 195–205, 2011.

[36] S. Sreehari, S. Venkatakrishnan, B. Wohlberg, L. F. Drummy, J. P. Simmons, and C. A. Bouman, "Plug-and-Play priors for bright field electron tomography and sparse interpolation," (Dec. 2015). [Online]. Available: http://arxiv.org/abs/1512.07331

[37] Y. Chen, J. Ma, Q. Feng, L. Luo, P. Shi, and W. Chen, "Nonlocal prior Bayesian tomographic reconstruction," *J. Math. Imag. Vis.*, vol. 30, no. 2, pp. 133–146, 2008.

[38] G. Wang and J. Qi, "Penalized likelihood PET image reconstruction using patch-based edge-preserving regularization," *IEEE Trans. Med. Imag.*, vol. 31, no. 12, pp. 2194–2204, Dec. 2012.

[39] J. Ma, H. Zhang, Y. Gao, J. Huang, Z. Liang, Q. Feng, and W. Chen, "Iterative image reconstruction for cerebral perfusion CT using a pre-contrast scan induced edge-preserving prior," *Phys. Med. Biol.*, vol. 57, no. 22, pp. 7519–7542, Nov. 2012.

[40] H. Zhang, J. Ma, J. Wang, Y. Liu, H. Lu, and Z. Liang, "Statistical image reconstruction for low-dose CT using nonlocal means-based regularization," *Comput. Med. Imag. Graph.*, vol. 38, no. 6, pp. 423–435, 2014.

[41] H. Y. Liao and G. Sapiro, "Sparse representations for limited data tomography," in *Proc. IEEE Int. Symp. Biomed. Imag. (ISBI): From Nano to Macro*, 2008, pp. 1375–1378.

[42] Y. Lu, J. Zhao, and G. Wang, "Few-view image reconstruction with dual dictionaries," *Phys. Med. Biol.*, vol. 57, no. 1, p. 173, 2012.

[43] L. Pfister and Y. Bresler, "Tomographic reconstruction with adaptive sparsifying transforms," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2014, pp. 6914–6918.

[44] D. Zoran and Y. Weiss, "From learning models of natural image patches to whole image restoration," in *Proc. Int. Conf. Comput. Vision*, 2011, pp. 479–486.

[45] G. Yu, G. Sapiro, and S. Mallat, "Solving inverse problems with piecewise linear estimators: From Gaussian mixture models to structured sparsity," *IEEE Trans. Image Process.*, vol. 21, no. 5, pp. 2481–2499, Apr. 2012.

[46] T. M. Nguyen and Q. Wu, "Fast and robust spatially constrained gaussian mixture model for image segmentation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 4, pp. 621–635, Apr. 2013.

[47] Y.-Q. Wang and J.-M. Morel, "SURE guided Gaussian mixture image denoising," *SIAM J. Imag. Sci.*, vol. 6, no. 2, pp. 999–1034, 2013.

[48] R. Zhang, J.-B. Thibault, C. A. Bouman, and K. D. Sauer, "Soft classification with Gaussian mixture model for clinical dual-energy CT reconstructions," in *Proc. 12th Int. Mtg. Fully Three-Dimensional Image Reconstruction Radiol. Nuclear Med.*, Jun. 2013, pp. 408–411.

[49] J. Yang, X. Yuan, X. Liao, P. Llull, D. J. Brady, G. Sapiro, and L. Carin, "Video compressive sensing using Gaussian mixture models," *IEEE Trans. Image Process.*, vol. 23, no. 11, pp. 4863–4878, Oct. 2014.

[50] Z. Nadir, M. S. Brown, M. L. Comer, and C. A. Bouman, "Gaussian mixture prior models for imaging of flow cross sections from sparse hyperspectral measurements," in *Proc. IEEE, Global Conf. Signal Inform. Process.*, Dec. 2015, pp. 527–531.

[51] R. Zhang, C. Bouman, J.-B. Thibault, and K. Sauer, "Gaussian mixture Markov random field for image denoising and reconstruction," in *Proc. IEEE Global Conf. Signal Inform. Process.*, Dec. 2013, pp. 1089–1092.

[52] R. Zhang, D. Pal, J.-B. Thibault, K. D. Sauer, and C. A. Bouman, "Model-based iterative reconstruction with a Gaussian mixture MRF prior for X-ray CT," in *Proc. 13th Int. Mtg. Fully Three-Dimensional Image Reconstruction Radiol. Nuclear Med.*, Jun. 2015, pp. 407–410.

[53] C. A. Bouman, "Cluster: An unsupervised algorithm for modeling Gaussian mixtures," Apr. 1997. [Online]. Available: http://engineering.purdue.edu/bouman.

[54] R. Salakhutdinov and G. Hinton, "An efficient learning procedure for deep Boltzmann machines," *Neural Comput.*, vol. 24, no. 8, pp. 1967–2006, 2012.

[55] D. R. Hunter and K. Lange, "A tutorial on MM algorithms," *Amer. Statistician*, vol. 58, no. 1, pp. 30–37, 2004.

[56] J. A. Fessler, "Mean and variance of implicitly defined biased estimators (such as penalized maximum likelihood): Applications to tomography," *IEEE Trans. Image Process.*, vol. 5, no. 3, pp. 493–506, Mar. 1996.

[57] J. D. Evans, D. G. Politte, B. R. Whiting, J. A. OSullivan, and J. F. Williamson, "Noise-resolution tradeoffs in x-ray CT imaging: A comparison of penalized alternating minimization and filtered backprojection algorithms," *Med. Phys.*, vol. 38, no. 3, pp. 1444–1458, 2011.

[58] K. Li, J. Garrett, Y. Ge, and G.-H. Chen, "Statistical model based iterative reconstruction (MBIR) in clinical CT systems. Part II. Experimental assessment of spatial resolution performance," *Med. Phys.*, vol. 41, no. 7, pp. 071 911-1–071 911-12, 2014.

[59] R. Rubinstein, M. Zibulevsky, and M. Elad, "Efficient implementation of the K-SVD algorithm using batch orthogonal matching pursuit," Feb. 2013. [Online]. Available: www.cs.technion.ac.il/ronrubin/software.

[60] K. Dabov, A. Danieyan, and A. Foi, "BM3D demo software for image/video restoration and enhancement," Jan. 2014. [Online]. Available: www.cs.tut.fi/foi/GCF-BM3D/index.html#ref_software.

[61] D.-J. Kroon, "Fast non-local means 1D, 2D color and 3D," Sep. 2010. [Online]. Available: http://www.mathworks.com/matlabcentral/fileexchange/27395-fast-non-local-mea ns-1d 2d-color-and-3d.

Authors' biographies and photographs not available at the time of publication.