Published in final edited form as:

Phys Med Biol. ; 63(15): 155021. doi:10.1088/1361-6560/aad356.

Iterative Spectral CT Reconstruction Based on Low Rank and Average-Image-Incorporated BM3D

Morteza Salehjahromi¹, Yanbo Zhang¹, and Hengyong Yu^{1,*}

¹Department of Electrical and Computer Engineering, University of Massachusetts Lowell, Lowell, MA, USA 01854

Abstract

In a photon counting detector integrated spectral CT scanner, the received photons are counted in several energy channels to generate the corresponding projections. Since the projection in each energy channel is generated using part of the received photons, the reconstructed channel image suffers from severe noises. Therefore, image reconstruction in spectral CT is considered as a big challenge. Because the inter-channel images are all from the same object but in different energy bins, there exists a strong correlation among these images. Moreover, it is suggested that there are similarities among various patches of CT images in spatial domain. In this work, we propose an average-image-incorporated block-matching and 3D (aiiBM3D) filtering along with a low rank regularization for iterative spectral CT reconstruction. The aiiBM3D method is based on filtered 3D data arrays formed by similar 2D blocks using the mapped version of the average image obtained from linear regression. The reconstruction procedure consists of two main steps. First, the alternating direction method of multipliers (ADMM) is employed to solve the problem with low rank regularization where the goal is to exploit the correlation in inter-channel images. Second, our proposed BM3D-based algorithm is applied to all the channel images to make use of the redundant information in spatial domain and inter-channel. The two steps repeat until the stopping criteria are satisfied. The proposed method is validated on numerically simulated and preclinical datasets. Our results confirm its high performance in terms of signal to noise ratio and structural preservation.

Keywords

Spectral CT (SCT); iterative reconstruction; BM3D; low rank

1. INTRODUCTION

There are several techniques to collect spectral computed tomography (SCT) projection datasets, such as fast kVp switching (FKS) [1], dual source CT (DSCT) [2] and using energy-resolving detectors [3, 4]. The energy-resolving photon-counting detectors are capable of counting the received photons and assigning them into several different energy bins/channels simultaneously. In this way, the multichannel projection datasets are generated where the number of received photon in each channel is quite low. In other words, the higher

*Corresponding author: Hengyong-yu@ieee.org.

the channel number, the lower the received photons in each channel. Because the obtained projection in each channel suffers from severe Poisson noise, it results in very noisy reconstructed image [5]. Therefore, developing a decent algorithm is a challenging task and one should make full use of the spatial similarity in channel images and also the inter-channel correlation.

Various methods have been developed in the field of spectral CT in recent years. In 2012, Xu *et al.* [6] developed an iterative reconstruction using the total variation (TV) regularization on each channel image independently. Later, the inter-channel image correlation was exploited using low rank (LR) regularization. Chu *et al.* [7] combined the TV and LR regularizations to further improve the quality of the channel images. The multichannel image in two dimensional (2D) spectral CT is a three dimensional (3D) data array where the first two dimensions represent the spatial domain and the third one is the number of energy channels which is also called the spectral domain. The concept of low rank can be generalized to higher order data arrays, such as the patches in the multichannel image in spectral CT. It is suggested that the materials in small patches are limited and the patches can be represented in a lower rank [7–9]. In 2015, Kim *et al.* [8] applied low rank to all the 3D patches obtained from the multichannel image for sparse-view reconstruction in spectral CT. In 2016, Zhang *et al.* [10] suggested a tensor-based dictionary learning method assuming existence of low rank along the spectral dimension. In 2016, Zeng *et al.* [11] applied a nonlocal means filter where the average image was used to compute the filtering weights of the pixels in each channel image. The average image is calculated by summing up all the channel images. This image is equivalent to the conventional CT where the detectors receive photons from the entire spectrum. The average image has a higher signal to noise ratio (SNR) and has been widely used to extract the redundant information in spectral CT. In 2017, Wang *et al.* [12] assumed a local linear relationship between the intended channel image and applied the average image as a regularization for material decomposition in dual spectral CT, and they came up with promising results.

Similar to natural images, it is widely accepted that there are similarities among various small patches in medical images [10, 11, 13]. Therefore, one should take advantage of these similarities of medical images in spatial domain. Some approaches were developed to denoise a group of similar blocks, such as Wiener filtering or hard thresholding on the coefficients of the corresponding group in a transform domain. The more similar patches exist and the more correlated they are in an image, the better the performance of the aforementioned group denoising is. In spectral CT, there exist highly correlated patches at the same position of different channel images because the inter-channel images are all from the same object. As a result, by incorporating similar patches from both the intended channel image and inter-channel images, it is more encouraged to use patched-based denoising in spectral CT compared with other denoising approaches such as TV, anisotropic diffusion [14], Huber [15], *etc.*

As one of the most powerful image denoising methods, the block-matching and 3D (BM3D) filtering was proposed by Dabov *et al.* [16, 17] in 2007. Very recently, Harrisson *et al.* [18] presented a block matching denoising algorithm referred as BM3D_PCCT for spectral CT. They applied a cross-channel decorrelation by adding the average image as a further spectral

dimension to exploit the inter-channel correlation in the BM3D algorithm. The original BM3D algorithm consists of two main steps [19]. The first step is the collaborative filtering where a hard thresholding filtration is performed on the 3D transformed blocks of similar patches. In the second step, the Wiener filtering is applied on the original image using the coefficients computed from the image obtained in the first step. Inspired by this denoising scheme, in this paper, an average-image-incorporated block-matching and 3D (aiiBM3D) filtering is proposed for spectral CT iterative reconstruction. In the proposed aiiBM3D filtering, the idea of average-image is not only used in the block matching (BM) part, it is also used in the Wiener filtering part with linear regression. Therefore, the similar blocks from other channels are also indirectly used in the filtering process.

The rest of this paper is organized as follows: the optimization problem along with the low rank regularization and the aiiBM3D filtering are described in section II. The experiment setup and results are presented in section III. Finally, the conclusion of the paper is drawn in section IV.

2. METHODOLOGY

In this section, the spectral CT reconstruction problem with low rank regularization and the aiiBM3D filtering process are presented. In each iteration, the channel images are first reconstructed by solving an optimization model with low rank regularization, and then the aiiBM3D filtering process is applied on each of the channel images.

2.1 Low rank regularization

The mode- k unfolding operation over the n dimensional data array $\mathfrak{B} \in \mathbb{R}^{L_1 \times L_2 \times \dots \times L_n}$ is defined as

$$\text{unfold}_k(\mathfrak{B}) = B_{(k)} \in \mathbb{R}^{L_k \times (L_1 \dots L_{k-1} L_{k+1} \dots L_n)}. \quad (1)$$

Suppose that $\mathfrak{X} \in \mathbb{R}^{I_1 \times I_2 \times S}$ is a 3D spectral image of 2D CT where I_1 and I_2 denote the image width and height, respectively, and S is the number of energy channels. We first formulate the low rank spectral CT minimization problem as

$$\underset{\mathfrak{X}}{\text{argmin}} \sum_{s=1}^S \|\mathbf{A}\mathbf{x}_s - \mathbf{y}_s\|_2^2 + \lambda_1 R_*(\mathfrak{X}), \quad (2)$$

where $\mathbf{x}_s \in \mathbb{R}^{(N \times 1)}$ ($N = I_1 I_2$) and $\mathbf{y}_s \in \mathbb{R}^{(M \times 1)}$ are the vectorized forms of image and projections in the s -th energy channel. Here, $\mathbf{A} \in \mathbb{R}^{(M \times N)}$ is the system matrix and $R_*(\mathfrak{X})$ represents the tensor nuclear norm of $\mathfrak{X}$ with the regularization parameter λ_1 . The nuclear norm serves as the low rank regularization[20], i.e., $R_*(\mathfrak{X}) = \sum_k \gamma_k \|\mathfrak{X}_{(k)}\|_*$, $\gamma_k = 1$, where $\|B\|_* = \sum_m \sigma_m(B)$ and σ_m is the m -th singular value of matrix B .

In general, the nuclear norm of a multi-dimensional data array is computed using all its mode- k unfolding matrices. In 2D spectral CT, however, since the reconstructed 3D images in various energy channels are highly correlated, only the mode-3 unfolding matrix is considered to be low rank (i.e., $\mathbf{T}_1 = \mathbf{T}_1 = 0$). Therefore (2) can be written as

$$\underset{\mathbf{x}}{\operatorname{argmin}} \sum_{s=1}^S \frac{1}{2} \|\mathbf{A}\mathbf{x}_s - \mathbf{y}_s\|_W^2 + \lambda_1 \|\mathbf{x}_{(3)}\|_* \quad (3)$$

In order to solve this optimization problem, it can be first reformulated as

$$\begin{aligned} &\underset{\mathbf{x}, \mathbf{D}}{\operatorname{argmin}} \sum_{s=1}^S \frac{1}{2} \|\mathbf{A}\mathbf{x}_s - \mathbf{y}_s\|_W^2 + \lambda_1 \|\mathbf{D}\|_*, \\ &\text{subject to } \mathbf{D} = \mathbf{x}_{(3)}. \end{aligned} \quad (4)$$

Therefore, the augmented Lagrangian can be formed as

$$L_\eta(\mathbf{x}, \mathbf{D}) = \sum_{s=1}^S \frac{1}{2} \|\mathbf{A}\mathbf{x}_s - \mathbf{y}_s\|_W^2 + \lambda_1 \|\mathbf{D}\|_* + \langle \mathbf{H}, \mathbf{x}_{(3)} - \mathbf{D} \rangle_F + \frac{\eta}{2} \|\mathbf{x}_{(3)} - \mathbf{D}\|_F^2, \quad (5)$$

where $\mathbf{H}$ is a dual variable, $\eta > 0$ and $\langle \cdot \rangle_F$ is the Frobenius inner product. Using the ADMM, $\mathbf{x}$ and $\mathbf{D}$ can be minimized alternatively and the dual variable $\mathbf{H}$ can be updated afterwards as

$$\begin{aligned} \mathbf{x}^{g+1/2} &= \underset{\mathbf{x}}{\operatorname{argmin}} \sum_{s=1}^S \frac{1}{2} \|\mathbf{A}\mathbf{x}_s - \mathbf{y}_s\|_W^2 + \frac{\eta}{2} \|\mathbf{x}_{(3)} - \mathbf{D}^g + \frac{\mathbf{H}}{\eta}\|_F^2, \\ \mathbf{D}^{g+1/2} &= \underset{\mathbf{D}}{\operatorname{argmin}} \lambda_1 \|\mathbf{D}\|_* + \frac{\eta}{2} \|\mathbf{x}_{(3)}^{g+1/2} - \mathbf{D} + \frac{\mathbf{H}}{\eta}\|_F^2 \\ \mathbf{H}^{g+1/2} &= \mathbf{H}^g + \eta \left(\mathbf{x}_{(3)}^{g+1/2} - \mathbf{D}^{g+1/2} \right), \end{aligned} \quad (6)$$

where $\mathbf{x}$ can be updated using various methods such as Gauss-Seidel[21] or Fourier transform, and $\mathbf{D}$ is updated using the proximity operator as

$$\mathbf{D}^{g+1/2} = \operatorname{prox}_{\lambda_1 \|\cdot\|_*} \left(\mathbf{x}_{(3)}^{g+1/2} + \frac{\mathbf{H}}{\eta} \right). \quad (7)$$

Eq.(7) has a closed-form solution using the singular value shrinkage operator,

$$\text{prox}_{\frac{\lambda_1}{\eta} \|\cdot\|_*}(\mathbf{Z}) = \mathbf{U} \Sigma_{\frac{\lambda_1}{\eta}} \mathbf{V}^T, \quad (8)$$

where $\mathbf{Z} = \mathbf{U} \mathbf{V}^T$ is the singular value decomposition of $\mathbf{Z}$, and

$$\Sigma_{\frac{\lambda_1}{\eta}} = \text{diag} \left(\max \left\{ \sigma_m - \frac{\lambda_1}{\eta}, 0 \right\} \right).$$

2.2 AiiBM3D filtering

Without loss of generality, we assume that $\mathbf{X}$ is one of the S channel images ($\mathbf{x}_s$) and $\mathbf{\Xi}$ is the obtained image from summing up all the channel images (denoted as the average image). Here, our goal is to employ the average image in the BM3D algorithm to improve its performance in the field of spectral CT reconstruction. The denoising process of BM3D algorithm is performed in the neighborhoods (called sliding blocks) where the similar blocks are chosen for further process within these neighborhoods. When the filtering process is performed in a specific sliding block in each channel image, its corresponding sliding block from the average image $\mathbf{\Xi}$ (which is obtained in each iteration) is also incorporated. Then, a linear regression approach is used to map the attenuation coefficients (pixel values) of each sliding block in the average image into the intended channel image using the least-square fit as:

$$\mathbf{X}^{sb} = a \mathbf{\Xi}^{sb} + b, \quad (9)$$

where $\mathbf{X}^{sb}$ and $\mathbf{\Xi}^{sb}$ are the vectorized forms of the sliding blocks in the channel and average images, respectively, the superscript “sb” refers to sliding block and the slope a and the intercept b are computed by linear regression. Therefore, a mapped version of sliding block from the average image can be computed (refers to as mapped sliding block in the rest of this paper) using a and b . Similar to the original BM3D algorithm, the proposed BM3D-based algorithm for spectral CT consists of two main steps as follows.

2.2.1 Collaborative filtering (Step 1)

Step 1a. Block-matching (BM): In this step, similar blocks are to be found within the sliding block. The proposed algorithm is to be utilized in an iterative image reconstruction. Because the reconstructed channel images have very low SNR, the obtained average-image cannot be used for BM in the first iterations. Here, we use an image (denoted as $\mathbf{\nabla}$) obtained from averaging all the channel images reconstructed by FBP to perform BM[22]. In this way, a reference $N_I \times N_I$ block (denoted as ∇_{2D}^{ref}) is selected from $\mathbf{\nabla}$ as a center of a larger sliding block ($N_e \times N_e$ neighborhood). Similar $N_I \times N_I$ blocks within this sliding block are stacked together if their distances to the reference block are lower than a specific threshold (denoted as ζ_{BM}). The block-distance is defined as

$$d(\nabla_{2D}^{ref}, \nabla_{2D}) = \frac{\|\zeta_{2D}(\nabla_{2D}^{ref}) - \zeta_{2D}(\nabla_{2D})\|_2^2}{N_1^2}, \quad (10)$$

where $\|\cdot\|_2$ denotes the L_2 norm, ∇_{2D} is a 2D $N_I \times N_I$ block in the sliding block, and ζ_{2D} is a normalized 2D linear transform such as 2D discrete cosine transform (DCT). Therefore, the coordinates of the similar blocks within the sliding block are obtained. Consequently, a block (denoted as $X_{3D}^{sb} \in \mathbb{R}^{N_1 \times N_1 \times N_2}$) is formed by stacking together the corresponding N_2 similar blocks in the intended channel image using the coordinates obtained from ∇ . The corresponding similar blocks in the mapped sliding block (denoted as $\tilde{X}_{3D}^{sb}$) is also used to form a 3D block $[X_{3D}^{sb} \tilde{X}_{3D}^{sb}] \in \mathbb{R}^{N_1 \times N_1 \times 2N_2}$ for the next sub step.

Step 1b. Collaborative filtering.: In this sub step, a 3D transform such as DCT is applied to the obtained block from previous sub section, and then a hard-thresholding of the coefficients is applied to suppress the noise. The 3D transform is obtained by performing a separate 2D and 1D transform along the first two spatial dimensions and the third spectral dimension of the obtained block, respectively. The hard thresholding filtration is performed on the 3D transformed block. The final estimated block is computed by inverting the 3D transform. The number of non-zero coefficients (denoted as $N_{x,ref}$) after hard-thresholding is also stored for the next sub-step. The collaborating filtering in a sliding block is computed as

$$[\hat{X}_{3D}^{1sb} \hat{\tilde{X}}_{3D}^{sb}] = \zeta_{3D}^{-1}(\mathcal{T}(\zeta_{3D}([X_{3D}^{sb} \tilde{X}_{3D}^{sb}]), \lambda_{3D} \sigma_X)), \quad (11)$$

where $\hat{X}_{3D}^{1sb}$ and $\hat{\tilde{X}}_{3D}^{sb}$ are respectively the filtered similar 3D blocks within the sliding block in the intended channel image and the mapped sliding block, the superscript “1” on $\hat{X}_{3D}^{1sb}$ refers to step 1, ζ_{3D} is a 3D transform and $\mathcal{T}$ is the hard-thresholding operator with a threshold set to $\lambda_{3D} \sigma_X$ where λ_{3D} is a constant coefficient in BM3D algorithm and σ_X is assumed as the standard deviation of the noisy channel image X and is defined as

$$\sigma_X = C \hat{\sigma}_X \quad (12)$$

where $\hat{\sigma}_X$ is the estimated standard deviation of noise by using a fast noise estimation method [23] and C is a constant. The noise estimation method works well for the images corrupted by additive Gaussian noise. However, the noise distribution of the reconstructed channel images is more complicated and not Gaussian. Therefore, the parameter C is used as a constant to compensate for the misestimated $\hat{\sigma}_X$. The parameter C will also be used in the second step serving for the same purpose. Finally, only the filtered block $\hat{X}_{3D}^{1sb}$ is separated for further process in the next sub step.

Step 1c. Aggregation.: This sub step is exactly similar to the one in the original BM3D algorithm. After computing the estimated similar blocks in all the sliding blocks, there could be more than one single value for a pixel. Therefore, a weighted averaging process is needed to compute the final value of those pixels with more than one value associated with. In each sliding block, the weight of the estimated blocks and the final estimated value of a pixel are calculated as

$$W_{X^{sb}}^1 = \frac{1}{\sigma_{X^{ref}}^2 N_{X^{ref}}},$$

$$X_{ij}^B = \frac{\sum_{sb} w_{X^{sb}}^1 \cdot \hat{X}_{ij}^{1sb}}{\sum_{sb} w_{X^{sb}}^1}. \quad (13)$$

where $\hat{X}_{ij}^{1sb}$ is the estimated value of a pixel at coordinate (i, j) obtained from $\hat{X}_{3D}^{1sb}$ and the sum is over all the sliding blocks which has an estimated value for this pixel. Therefore, by computing all the pixel values, the filtered channel image (refers to as basic image X^B) is obtained where the number of sliding blocks is determined by the sliding step N_{step} .

2.2.2 Wiener filtering using the average-image (Step 2)

Step 2a. BM.: Since the average-image is used for the BM, the computed coordinates of similar blocks for each sliding block are fixed, and we only need to compute them once and use them in not only the two main steps but also for all the channel images and iterations. In this way, we can significantly speed up our proposed aiiBM3D denoising because the BM is the most time consuming part in this algorithm. Therefore, unlike the original BM3D which performs BM on the basic image, similar $N_I \times N_I$ blocks within the sliding block have been already found from ∇ in the previous step.

Step 2b. Wiener filtering.: Unlike the original BM3D which performs the Wiener filtering on original image using the coefficients computed from the basic image, in the proposed method, the Wiener coefficients are computed from the mapped version of the average-image and then apply on the basic image. In each sliding block, the 3D block obtained from the mapped sliding block is used to compute Wiener coefficients $W_{3D}^{\mathbf{x}}$ as

$$W_{3D}^{\mathbf{x}} = \frac{\left| \zeta_{3D}(\tilde{\Xi}_{3D}^{sb}) \right|^2}{\left| \zeta_{3D}(\tilde{\Xi}_{3D}^{sb}) \right|^2 + \sigma_B^2}, \quad (14)$$

where ζ_{3D} is a normalized 3D linear transform. Similar to the first step, $\sigma_B = C\hat{\sigma}_B$ is assumed as the standard deviation of the basic image, $\hat{\sigma}_B$ is the estimated standard deviation of noise using the method mentioned in the first step and C is a constant to compensate for the misestimated noise level of the basic image (see Eq. 12). In the next sub-step, the norm

of Wiener coefficients calculated here is used for aggregation. The estimated similar blocks in the channel image are computed as

$$\hat{\mathbf{X}}_{3D}^{2sb} = \zeta_{3D}^{-1}(\mathbf{W}_{3D}^{\mathbf{x}} \cdot * \zeta_{3D}(\mathbf{X}_{3D}^B)), \quad (15)$$

where $\mathbf{X}_{3D}^B$ is a 3D block obtained by stacking together similar 2D blocks from the basic image, the superscript “2” on the filtered 3D sliding block $\hat{\mathbf{X}}_{3D}^{2sb}$ refers to step 2, and “.” represents the element-wise multiplication.

Step 2c. Aggregation.: This part is similar to the previous aggregation part in the first step. However, the weights of estimated blocks (denoted as $W_{\mathbf{X}^{sb}}^2$) and the final estimated value of a pixel at coordinate (i, j) denoted as $\hat{\mathbf{X}}_{ij}$ are computed as

$$W_{\mathbf{X}^{sb}}^2 = \frac{1}{\sigma_B^2 \|\mathbf{W}_{3D}^{\mathbf{x}}\|_2^2},$$

$$\hat{\mathbf{X}}_{ij} = \frac{\sum_{sb} w_{\mathbf{X}^{sb}}^2 \cdot \hat{\mathbf{X}}_{ij}^{2sb}}{\sum_{sb} w_{\mathbf{X}^{sb}}^2}. \quad (16)$$

where $\hat{\mathbf{X}}_{ij}^{2sb}$ is the estimated value of a pixel at coordinate (i, j) obtained from $\hat{\mathbf{X}}_{3D}^{2sb}$.

The whole procedure to reconstruct $\mathbf{x}$ with the low rank regularization and aiiBM3D filtering is summarized below as Algorithm 1.

Algorithm 1:

Spectral CT reconstruction using LR and aiiBM3D

-
- 1: Input: $\mathbf{y}_s, \mathbf{A}, \mathbf{D}, \mathbf{H}, \eta, \lambda, \mu$; parameters used in aiiBM3D;
 - 2: Output: $\mathbf{x}$;
 - 3: Obtaining $\mathbf{V}$ by summing up all the channel images reconstructed by FBP;
 - 4: For $g=1, \dots, G$;
 - 5: Updating $\mathbf{x}^{g+1/2}$ using equations in (6);
 - 6: For $s=1, \dots, S$;
 - 7: Performing the aiiBM3D on $\mathbf{x}_s$
 - 8: End
 - 9: Updating $\mathbf{x}^{g+1}$
 - 10: End
-

3. EXPERIMENTS AND RESULTS

A performance analysis of the proposed method is made by using numerical simulations and a realistic preclinical mouse dataset. The image quality of the proposed method is compared with channel images reconstructed by the conventional FBP, iterative reconstruction with TV regularization[6], iterative reconstruction with TV and LR regularization [7, 24] and BM3D_PCCT algorithm. Two image evaluation indices are utilized to conduct the comparison study. Because the attenuation coefficients decreases as the emitted energy increases, higher energy channels tend to give lower root mean square error (RMSE). Therefore, it is not a good index to evaluate the performance. To overcome this problem, the relative root mean square error (rRMSE) is used for comparison. The rRMSE for each channel image $\mathbf{X}$ is defined as

$$rRMSE(\mathbf{X}, \mathbf{X}^{NF}) = \frac{\sqrt{\frac{1}{N_p} \sum_{i,j} (X_{ij}^{NF} - X_{ij})^2}}{\max(\mathbf{X}^{NF}(:))}, \quad (17)$$

where $\mathbf{X}^{NF}$ is the corresponding channel image obtained from noise-free projections (denoted as noise-free image), X_{ij}^{NF} and X_{ij} are the value of $\mathbf{X}^{NF}$ and $\mathbf{X}$ at coordinate (i, j), respectively, and $\max(\mathbf{X}^{NF}(:))$ indicates the maximum value of $\mathbf{X}^{NF}$ over all its pixels. In addition, structural similarity (SSIM) index[25] is used to evaluate the similarity between the channel image and its corresponding noise-free image.

3.1 Numerical Simulations

Here, an equi-distant fan-beam geometry is assumed where its parameters along with the corresponding ones in the aiiBM3D are listed in table 1. Using the SpectrumGUI software[26], a 50 kVp x-ray spectrum is generated with energy step of 1 keV and divided into eight non-overlapping energy channels. The energy channels (see figure 1) are [16,22), [22,25), [25,28), [28,31), [31,34), [34,37), [37,41), and [41,50) keV. In this way, the channels can have close SNRs. A phantom of mouse thorax with iodine contrast agent (see figure 2) is used. The concentration of iodine is 1.2% in regions A, B and C and 0.6% in region E, and the Gaussian distribution varying from 0.32% to 1.2% is used in region D. Each detector bin receives a total number of P photons in a blank scan where the proportional Poisson noise is added to each channel projection.

The effects of varying the LR regularization parameter λ_1 and the coefficient C for correcting the standard deviation of noise in aiiBM3D are examined. In all the experiments, we set $\eta=0.4$. The default values are $\lambda_1 = 0.06$ and $C = 2.5$. In each experiment, only one of the variables varies and the other is fixed. In figure 3, the mean of rRMSE and SSIM are plotted with respect to different λ_1 and C . As the aforementioned, the optimal parameter is chosen when the smallest mean of rRMSE is achieved. From figure 3, it can be seen that the smallest mean of rRMSE is obtained at $\lambda_1=0.06$ where the SSIM index remains almost the same. Moreover, the smallest rRMSE is obtained at $C = 2.5$. Because the aiiBM3D filtering is applied separately on each channel image after solving the convex low rank problem, the

proposed method does not guarantee to converge to the global solution. To verify the practical convergence of our method, the convergence curve is plotted in figure 4. As we can see, the error decreases monotonically and it remains stable after around 30 iterations.

The reconstructed images of three representative channels are displayed in figure 5(a). The reconstructed images from noise-free projections are references and shown in the first column, and the other columns are the images reconstructed from the FBP, TV, TV-LR, BM3D_PCCT and our method. A region-of-interest (ROI) with fine structures is chosen within the highlighted square in the top left image. This magnified ROI represents the thoracic vertebra bones surrounded by low density tissues. Since the last channel receives the smallest number of photons, it has the lowest SNR and a lot of fine structures including the zoomed-in ROI are lost in the FBP and TV reconstructed images. The reconstructed images from the TV-LR are slightly better. The BM3D_PCCT gives relatively high quality images where the fine structures are well preserved. However, there exist some artifacts in some regions (see upper right of the iodine region with the Gaussian distribution). By looking at the reconstructed images, our method also delivers close images to the noise-free ones where no distinguished artifact can be observed. In figure 5(b), the absolute value of the difference images between the images shown in figure 5(a) and their corresponding reference images are shown. Since the absolute value is used, the display window starts from zero and darker regions indicate better reconstructions. By comparison, the bony structures are better reconstructed in our method while the number of bright structures is considerable in TV and TV-LR. Moreover, the difference images obtained from our method are darker as a whole compared to the TV and TV-LR. The BM3D_PCCT gives slightly brighter regions compared to our method.

The rRMSE and SSIM indices for the proposed method and other competing methods are evaluated. The indices for each channel and the mean value for each method are listed in tables 2 and 3 where the bold values indicate the best collected value among all the competing methods. It can be seen that the FBP has the worst image quality indices for each energy channel, followed by the TV and TV-LR. Our method provides the best image quality index in terms of rRMSE followed by the BM3D_PCCT. In terms of SSIM index, our method provides the highest value followed by the TV-LR and BM3D_PCCT. Moreover, the relative mean values of bone, soft tissue and iodine with respect to noise-free images are displayed in figure 6 to compare the competitive methods. The mean value for each material is computed by applying its corresponding 2D binary mask on the intended channel image and taking the mean value. The relative mean values are calculated by subtracting the computed mean values from the mean values of reference images. One should note that because the binary mask is applied on only the regions of each material, the computed relative mean values tend to be positive. The TV and TV-LR tend to smooth the bony structures and they have the greatest biases for bone. Our method and FBP provide the minimum biases followed by the BM3D_PCCT for bone regions. Regarding the soft tissue, the mean values are relatively closer to one another. For these regions, the TV and our method provide the closest mean values to the reference images. Finally, the BM3D_PCCT and FBP followed by our method give the lowest biases for iodine regions.

The capability of our proposed method is also validated in terms of material decomposition. Three basis materials (bone, soft tissue and iodine contrast agent) are obtained from the multichannel images using a post-reconstruction material decomposition approach [27]. The material decomposition results for various methods are shown in figure 7 where the first row images are the color representation of the decomposed materials with a highlighted ROI. From this row, we can see that many pixel values are misclassified by the TV method and some by the TV-LR (see the soft tissue regions surrounded by bone structures in the magnified ROI). Our method and the BM3D_PCCT provide the closest material decomposition images to the reference image.

To examine the decomposition, the root mean square error (RMSE) values of the three decomposed basis materials are displayed in figure 8(a). The RMSE for FBP is not provided in the figure because of relatively large values. The largest RMSE is observed in the FBP for all the three materials followed by the TV and TV-LR. Our method and the BM3D_PCCT provide the smallest RMSE for the basis materials where our method provides the minimum RMSE for the bone and soft tissue regions, whereas the BM3D_PCCT provides the minimum RMSE for iodine regions. One should note that the pixel values of the material decomposition results are no longer the attenuation coefficients of the corresponding materials. In fact, they can be viewed as the percentile of each material in the corresponding position. Therefore, the pixel value corresponding to each material is normalized between [0 1] and the relative mean value of each material (with respect to noise-free images) in figure 8(b) is computed using the same method for figure 6. The references were chosen as the mean values from the images obtained from noise-free projection data. By looking at figure 8(b), our method delivers the closest mean value to the reference image for bone and soft tissue indicating the minimum biases for those tissue types. In terms of iodine, our method gives closer mean than the TV-LR and FBP while the BM3D_PCCT and TV deliver the minimum bias. The TV also misclassifies iodine in some other regions (see figures 7 and 8(a) for the RMSE of iodine for TV regularization). Generally, the TV-LR does not provide accurate images in terms of biases due to the soft thresholding operation which changes the mean value of the reconstructed images, whereas our method and the BM3D_PCCT Provide the minimum biases.

The performance of the proposed method is also evaluated under different noise levels by varying the total number of emitted photons P . As a result, the corresponding channel projections are generated under different Poisson noise levels. In tables 4 and 5, the mean values of rRMSE and SSIM are listed for different P values. The mean values correspond to the last columns of tables 2 and 3. The range of P is varied from 3×10^3 to 9×10^3 with a step size of 2×10^3 . Our method provides the best image quality indices in terms of both rRMSE and SSIM followed by the BM3D_PCCT, TV-LR, TV and FBP. It should be noted that since the received number of photons in each energy channel is different and also the contrast of the channel images decreases for higher energy channels, the comparison is in fact performed for a wide range of projections corresponding to different SNRs.

3-2 Preclinical Mouse Study

In this sub-section, our proposed method is evaluated using a preclinical mouse study. In this study, a Medipix All Resolution System (MARS) micro-CT with Medipix MXR cadmium telluride detectors was used to scan a mouse [28]. Eight multichannel projections were collected with the source tube operated at 120kVp and 175 mA. The sonogram of each energy channel is generated based on the received photons whose their energies are higher than a threshold. In fact, the threshold level increases as the index of the energy channel increases. In other words, by increasing the index number of a channel, higher energy photons are collected. Therefore, the first channel receives photons from the widest energy range and subsequently has the highest SNR while the last channel has the lowest one. Theoretically speaking, these channels can be converted to ideal ones by subtracting every two successive channels. However, the received data are pre-processed so that subtraction option is not possible. The distance from source to the detector and the system origin to the source were 255 and 158 mm, respectively. The number of projections and detector bins were 371 and 1024, respectively. In each channel, every two neighboring bins were merged to give a sinogram of size 371×512 . Then, a Wavelet-Fourier filtering method [29] was adopted to alleviate ring artifacts caused by detector defects.

The reconstructed images in three representative channels 1, 5, and 8 using the previous methods are displayed in figure 9 where strong noise can be seen in images reconstructed by FBP. For example, many tissue structures are not distinguishable from noise in the last channel. The TV regularization could slightly reduce the noise. However, blocky structures are imposed in various parts of the channel images. The TV-LR provides clearer images than those reconstructed from TV. In the images reconstructed from BM3D_PCCT and our method, the soft tissue regions are less noisy than the ones in TV-LR. In addition, the bony and fine structures are better preserved by BM3D_PCCT and our method (see the gap between the bony structures indicated by blue arrows).

Figure 10 displays the magnified thoracic vertebra indicated in figure 9. The magnified ROI reconstructed from FBP suffers from strong noise. The blocky structures mentioned in TV can be seen clearly in its magnified ROI. The bone structure in thoracic vertebra can be distinguished from noise after using TV combined with the LR regularization. In both BM3D_PCCT and the proposed methods, some structural features such as the soft tissue gap between the bony structures are better preserved. Moreover, two basis materials as bone and soft tissue are obtained from the multichannel images using the same material decomposition approach in the previous subsection. The material decomposition results for various methods with two highlighted ROIs are shown in figure 11. As we can see in the ROIs, it is encouraging to say that the bony structures can be clearly observed in our method compared to other competing methods.

4. DISCUSSION AND CONCLUSION

For simplicity, we used non-overlapping spectrums for the numerical simulations. In practical experiments, due to degradation factors in photon counting detectors, such as pulse pileup, charge sharing, etc., the photon counting detectors do not have an ideal response, and the spectral broadening could be quite large for high energy pixel detectors made of CdTe

and CdZnTe [30]. A more realistic overlapped spectrum model can be considered in the future.

All experiments were implemented in MATLAB (version R2017a) and performed on an Intel(R) Core(TM) i5-4670 CPU, 3.40 GHz with 16 GB RAM. The forward and backward projections in each iteration of the OS-SART take 105 seconds for each channel image which is the most time consuming procedure compared with the LR regularization and the aiiBM3D filtering. The noise estimation method and the LR regularization take 0.05 and 0.08 second, respectively. The original BM3D algorithm takes less than 1.97 seconds where it uses internal C++ subroutines using the MEX files. The computational time of the proposed aiiBM3D filtering is comparable to the original BM3D with the additional cost of computing the slope a and the intercept b for each sliding block and its corresponding 3D transform where the number of sliding blocks is approximately N/N_e^2 . The overall additional computational cost is not high and it is less than the block matching procedure. The 3D transform of the 3D similar blocks in the mapped sliding blocks $\zeta_{3D}(\tilde{\Xi}_{3D}^{sb})$ can be efficiently found from the sliding blocks of the average image Ξ^{sb} . In this way, for all the sliding blocks of the average image, Similar $N_I \times N_I$ blocks are stacked together and their 3D transformed $\zeta_{3D}(\Xi_{3D}^{sb})$ would be saved in advance in each iteration. Then, using the corresponding slope a and the intercept b , it is easy to check that $\zeta_{3D}(\tilde{\Xi}_{3D}^{sb})$ can be efficiently obtained as [31] (see Eq. 9)

$$\begin{aligned}\zeta_{3D}(\tilde{\Xi}_{3D}^{sb})(:, :, :) &= a[\zeta_{3D}(\Xi_{3D}^{sb})(:, :, :)] \\ \zeta_{3D}(\tilde{\Xi}_{3D}^{sb})(1, 1, :) &= bN_1 + [\zeta_{3D}(\Xi_{3D}^{sb})(1, 1, :)]\end{aligned}\quad (18)$$

where ‘:’ is the MATLAB notation which indicates all the elements in the intended dimension. In the block matching sub-steps, the coordinates of the similar blocks are found using the image ∇ . These coordinates are fixed and they are used for all the channel images and for all the iterations. Therefore, the overall computational cost can be significantly reduced because the block matching is the most time consuming part in aiiBM3D algorithm.

From figure 3, we observe that the performance of the proposed filtering method without the LR regularization i.e., OS-SART with aiiBM3D outperforms the BM3D_PCCT and TV-LR as the mean rRMSE and SSIM are 1.0×10^{-2} and 0.965, respectively. They are better than those in the BM3D_PCCT where the mean of rRMSE is 1.04×10^{-2} and SSIM is 0.946 (see table 2 and 3).

In summary, in this paper, an average-image-incorporated block-matching and 3D (aiiBM3D) filtering along with low rank regularization was proposed for iterative spectral CT reconstruction. Numerical simulation results show that our method is superior in terms of rRMSE and SSIM. In addition, in terms of visual assessment, the proposed method and the BM3D_PCCT show better preservation of detailed structures than other methods such as the TV and TV-LR in both numerical simulations and preclinical mouse study. However, the

proposed method can better reduce artifacts in numerical simulations compared to the BM3D_PCCT. In the preclinical mouse study, the performance of the proposed method and BM3D_PCCT are comparable, and they outperform the FBP, TV and TV-LR.

ACKNOWLEDGMENTS

This work was supported in part by the NIH/NIBIB U01 grant EB017140.

REFERENCES

- [1]. Zou Y, and Silver MD, "Analysis of fast kV-switching in dual energy CT using a pre-reconstruction decomposition technique." 6913, 691313 (2008).
- [2]. Graser A, Johnson TR, Chandarana H et al., "Dual energy CT: preliminary observations and potential clinical applications in the abdomen," *European radiology*, 19(1), 13 (2009). [PubMed: 18677487]
- [3]. Heismann BJ, Schmidt BT, and Flohr T, "Spectral computed tomography." SPIE (2012)
- [4]. McCollough CH, Leng S, Yu L et al., "Dual-and multi-energy CT: principles, technical approaches, and clinical applications," *Radiology*, 276(3), 637–653 (2015). [PubMed: 26302388]
- [5]. Liu R, Fu L, De Man B et al., "GPU-based Branchless Distance-Driven Projection and Backprojection," *IEEE Transactions on Computational Imaging*, (2017).
- [6]. Xu Q, Yu H, Bennett J et al., "Image reconstruction for hybrid true-color micro-CT," *IEEE transactions on biomedical engineering*, 59(6), 1711–1719 (2012). [PubMed: 22481806]
- [7]. Chu J, Li L, Chen Z et al., "Multi-energy CT reconstruction based on Low Rank and Sparsity with the Split-Bregman Method (MLRSS)." 2411–2414.
- [8]. Kim K, Ye JC, Worstell W et al., "Sparse-view spectral CT reconstruction using spectral patch-based low-rank penalty," *IEEE transactions on medical imaging*, 34(3), 748–760 (2015). [PubMed: 25532170]
- [9]. Zhang Y, Yu H, Mou X et al., "Dictionary Learning and Low Rank based Multi-energy CT Reconstruction," (2014).
- [10]. Zhang Y, Mou X, Wang G et al., "Tensor-based dictionary learning for spectral ct reconstruction," *IEEE Transactions on Medical Imaging*, (2016).
- [11]. Zeng D, Huang J, Zhang H et al., "Spectral CT Image Restoration via an Average Image-Induced Nonlocal Means Filter," *IEEE Transactions on Biomedical Engineering*, 63(5), 1044–1057 (2016). [PubMed: 26353358]
- [12]. Wang Q, Zhu Y, and Yu H, "Locally linear constraint based optimization model for material decomposition," *Physics in Medicine & Biology*, 62(21), 8314 (2017). [PubMed: 28933350]
- [13]. Xu Q, Yu H, Mou X et al., "Low-dose X-ray CT reconstruction via dictionary learning," *IEEE Transactions on Medical Imaging*, 31(9), 1682–1697 (2012). [PubMed: 22542666]
- [14]. Yang Z, Silver MD, and Noshi Y, "Adaptive weighted anisotropic diffusion for computed tomography denoising," (2011).
- [15]. Elbakri IA, and Fessler JA, "Statistical image reconstruction for polyenergetic X-ray computed tomography," *IEEE transactions on medical imaging*, 21(2), 89–99 (2002). [PubMed: 11929108]
- [16]. Dabov K, Foi A, Katkovnik V et al., "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Transactions on image processing*, 16(8), 2080–2095 (2007). [PubMed: 17688213]
- [17]. Dabov K, Foi A, Katkovnik V et al., "Color image denoising via sparse 3D collaborative filtering with grouping constraint in luminance-chrominance space." 1, I-313–I-316.
- [18]. Harrison AP, Xu Z, Pourmorteza A et al., "A multichannel block-matching denoising algorithm for spectral photon-counting CT images," *Medical physics*, 44(6), 2447–2452 (2017). [PubMed: 28332211]
- [19]. Lebrun M, "An analysis and implementation of the BM3D image denoising method," *Image Processing On Line*, 2, 175–213 (2012).

- [20]. Semerci O, Hao N, Kilmer ME et al., "Tensor-based formulation and nuclear norm regularization for multienergy computed tomography," *IEEE Transactions on Image Processing*, 23(4), 1678–1693 (2014). [PubMed: 24808339]
- [21]. Goldstein T, and Osher S, "The split Bregman method for L1-regularized problems," *SIAM journal on imaging sciences*, 2(2), 323–343 (2009).
- [22]. Salehjahromi M, Zhang Y, and Yu H, "A spectral CT denoising algorithm based on weighted block matching 3D filtering." 10391, 103910G.
- [23]. Tai S-C, and Yang S-M, "A fast method for image noise estimation using laplacian operator and adaptive edge detection." 1077–1081.
- [24]. Chu J, Cong W, Li L et al., "Combination of current-integrating/photon-counting detector modules for spectral CT," *Physics in medicine and biology*, 58(19), 7009 (2013). [PubMed: 24043342]
- [25]. Wang Z, Bovik AC, Sheikh HR et al., "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, 13(4), 600–612 (2004). [PubMed: 15376593]
- [26]. [<http://spectrumgui.sourceforge.net/>].
- [27]. Granton P, Pollmann S, Ford N et al., "Implementation of dual-and triple-energy cone-beam micro-CT for postreconstruction material decomposition," *Medical physics*, 35(11), 5030–5042 (2008). [PubMed: 19070237]
- [28]. Butler A, Anderson N, Tipples R et al., "Bio-medical X-ray imaging with spectroscopic pixel detectors," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 591(1), 141–146 (2008).
- [29]. Münch B, Trtik P, Marone F et al., "Stripe and ring artifact removal with combined wavelet—Fourier filtering," *Optics express*, 17(10), 8567–8591 (2009). [PubMed: 19434191]
- [30]. Taguchi K, and Iwanczyk JS, "Vision 20/20: Single photon counting x-ray detectors in medical imaging," *Medical physics*, 40(10), (2013).
- [31]. Petrou M, and Petrou C, [Image processing: the fundamentals] John Wiley & Sons, (2010).

Figure 1.
Eight energy channels assumed in X-ray spectrum

Figure 2.
The mouse phantom (left) and the contrast agent (right)

Figure 3. The means of rRMSE and SSIM indices with respect to the parameters λ_1 and C ($P = 5 \times 10^3$).

Figure 4. A typical convergence curve of our reconstruction method ($P = 5 \times 10^3$).

Figure 5.

Reconstructed images from different methods in numerical simulations ($P = 5 \times 10^3$). (a) From top to bottom rows, images correspond to channels 1, 6 and 8 and the display windows are $[0 \ 2.5]$, $[0 \ 1.1]$ and $[0 \ 0.9] \text{ cm}^{-1}$, respectively. The first column images are reconstructed from noise-free projections, and from second to fifth columns, the images are reconstructed from noisy projections using FBP, TV, TV-LR, BM3D_PCCT and our method, respectively, (b) The absolute value of difference images shown in (a) with respect to their corresponding noise-free images. From top to bottom, the display windows are $[0 \ 0.5]$, $[0 \ 0.3]$ and $[0 \ 0.25] \text{ cm}^{-1}$, respectively.

Figure 6.
The relative mean values of the three types of tissues in each channel (cm^{-1}).

Figure 7.

Material decomposition results of numerical simulations. From left to right columns, images are the material decomposition results of noise-free images, FBP, TV, TV-LR, BM3D_PCCT and our method, respectively. The first row is the color representation of the decomposed materials where the color red, green and blue correspond to bone, soft tissue and iodine contrast agent, respectively. These three basis materials are also separately shown in the second to forth rows, respectively ($P = 5 \times 10^3$).

Figure 8.

The RMSE and mean values of different decomposed materials ($P = 5 \times 10^3$). (a) The RMSEs of three materials with respect to the noise-free images. (b) The relative mean values of three materials (cm^{-1}).

Figure 9.

Reconstructed images from different methods in preclinical mouse study. From top to bottom rows, images correspond to channels 1, 5 and 8. From first to fourth columns, the images are reconstructed by the TV, TV-LR, BM3D_PCCT and our method, respectively. The display windows is $[0 \ 1] \text{ cm}^{-1}$.

Figure 10.
Magnified bone ROI indicated in figure 9.

Figure 11.

Color representation of the decomposed images and two magnified bone ROIs for the mouse study.

Table 1.

Parameters in System Geometry for Numerical Simulations (left) and Parameters in the iiBM3D method (right).

PARAMETERS	VALUE	FIRST STEP		SECOND STEP	
Pixel size	$0.075 \times 0.075 \text{ mm}^2$	N_1	8	N_1	8
Source-object-distance	132 mm				
Source-detector-distance	180 mm	N_e	39	N_e	39
Detector bin spacing	0.10 mm	N_{step}	3	N_{step}	3
Number of detector bins	512	N_2	16	N_2	16
Number of views per rotation	640	ζ_{BM1}	5000	ζ_{BM2}	3500
Image size	512×512×8	λ_{3D}	2.8		

Table 2.

The rRMSE for Numerical Simulation Results ($\times 10^{-2}$), ($P = 5 \times 10^3$).

Method	Channel number								Mean
	# 1	# 2	# 3	# 4	# 5	# 6	# 7	# 8	
FBP	4.612	5.901	6.850	8.435	10.66	13.64	16.02	19.85	10.75
TV	1.126	1.235	1.359	1.500	1.697	2.075	2.274	2.511	1.722
TV-LR	1.072	0.984	0.908	0.888	0.985	1.403	1.460	1.755	1.182
BM3D_PCCT	0.789	0.771	0.807	0.873	0.975	1.166	1.357	1.628	1.046
Our method	0.705	0.685	0.681	0.733	0.853	1.109	1.181	1.382	0.916

Table 3.

The SSIM for Numerical Simulation Results ($P = 5 \times 10^3$).

Method	Channel number								Mean
	# 1	# 2	# 3	# 4	# 5	# 6	# 7	# 8	
FBP	0.285	0.210	0.169	0.126	0.098	0.072	0.068	0.047	0.134
TV	0.971	0.970	0.968	0.960	0.944	0.921	0.905	0.868	0.938
TV-LR	0.971	0.977	0.976	0.971	0.958	0.935	0.925	0.903	0.952
BM3D_PCCT	0.988	0.979	0.977	0.965	0.949	0.922	0.908	0.878	0.946
Our method	0.989	0.985	0.983	0.977	0.966	0.947	0.937	0.917	0.963

Table 4.

Mean of rRMSE for different noise levels.

	FBP	TV	TV-LR	BM3D_PCCT	Our method
$P = 3.0 \times 10^3$	12.65	1.86	1.38	1.13	1.00
$P = 5.0 \times 10^3$	10.75	1.72	1.18	1.05	0.92
$P = 7.0 \times 10^3$	9.10	1.63	1.07	0.94	0.86
$P = 9.0 \times 10^3$	7.68	1.46	1.01	0.87	0.76

Table 5.

Mean of SSIM for different noise levels.

	FBP	TV	TV-LR	BM3D_PCCT	Our method
$P = 3.0 \times 10^3$	0.109	0.925	0.940	0.936	0.953
$P = 5.0 \times 10^3$	0.134	0.938	0.952	0.946	0.963
$P = 7.0 \times 10^3$	0.162	0.951	0.964	0.961	0.972
$P = 9.0 \times 10^3$	0.194	0.960	0.968	0.967	0.978