Union of Learned Sparsifying Transforms Based Low-Dose 3D CT Image Reconstruction

Xuehang Zheng, Saiprasad Ravishankar, Yong Long*, and Jeffrey A. Fessler

***Abstract*—We propose a new penalized weighted-least squares (PWLS) reconstruction method that exploits regularization based on an efficient Union of Learned TRAnsforms (PWLS-ULTRA). The union of square transforms is pre-learned from numerous 3D patches extracted from a dataset of CT volumes. The proposed PWLS-based cost function is optimized by alternating between a CT image reconstruction step, and a sparse coding and clustering step. The CT image reconstruction step is accelerated by a relaxed linearized augmented Lagrangian method with ordered-subsets that reduces the number of forward and backward projections. Simulations with 3D axial CT scans of the XCAT phantom show that for low-dose levels, the proposed method significantly improves the quality of reconstructed images compared to PWLS reconstruction with a nonadaptive edge-preserving regularizer (PWLS-EP). PWLS with regularization based on a union of learned transforms leads to better image reconstructions than using a single learned square transform.**

I. Introduction

The development of computed tomography (CT) image reconstruction methods that significantly reduce patient radiation exposure while maintaining high image quality is an important area of research in low-dose CT (LDCT) imaging.

The analytical filtered back-projection (FBP) image reconstruction method typically produces unacceptable image quality when the radiation dose is reduced. Model-based image reconstruction (MBIR) methods, aka statistical image reconstruction methods, can provide high-quality reconstructions from low-dose scans. In MBIR methods, penalized weighted-least squares (PWLS) cost functions with a statistically weighted quadratic data-fidelity term and a penalty term (regularizer) modeling prior knowledge of the underlying unknown object are commonly used [1].

Extracting prior information from big datasets of regular dose CT images has great potential to enable MBIR methods to produce improved reconstructions from LDCT measurements. For example, dictionary learning has been recently applied to CT image reconstruction and was shown to have better performance over total variation-based PWLS method [2]. The sparse coding step in both synthesis dictionary learning and analysis [3] dictionary learning is NP-Hard and methods such as K-SVD [4], [5] involve expensive computations. Recently, Ravishankar et al. [6], [7] proposed a generalized analysis dictionary learning approach, called sparsifying transform learning, to more efficiently learn a square transform (ST). Pfister et al. [8] showed the promise of PWLS reconstruction with ST based regularization by combining the MBIR technique with an adaptive ST based regularizer, to jointly estimate the ST and the image. More recently, Wen et al. extended the single ST learning method to learning an overcomplete transform with block cosparsity (OCTOBOS) [9]. This approach jointly learns a union of square transforms and a clustering of image patches or textures.

Incorporating the OCTOBOS model, we propose a new PWLS reconstruction method with regularization based on a Union of Learned TRAnsforms (PWLS-ULTRA) for 3D (e.g., cone beam) LDCT reconstruction. The union of square transforms is pre-learned from numerous 3D patches extracted from a dataset of CT volumes. Experiments with 3D axial CT scans of the XCAT phantom show that for low-dose levels, the proposed PWLS-ULTRA method significantly improves the quality of reconstructed images compared to PWLS reconstruction with a nonadaptive edge-preserving regularizer (PWLS-EP). While we recently proposed a PWLS method based on a pre-learned ST (PWLS-ST) for 2D CT [10], here we study its extension to 3D CT reconstruction. PWLS-ULTRA that uses a union of learned transforms is shown to lead to better image reconstructions than PWLS-ST that uses a learned ST. The image reconstruction step of the proposed PWLS based methods is accelerated by a relaxed linearized augmented Lagrangian method with ordered-subsets (relaxed OS-LALM) [11].

II. Problem Formulation

We reconstruct an image volume $\mathbf{x} \in \mathbb{R}^{N_p}$ from noisy sinogram data $\mathbf{y} \in \mathbb{R}^{N_d}$ using a union of K pre-learned square transform matrices $\{\mathbf{\Omega}_k \in \mathbb{R}^{l \times l}, k = 1, \dots, K\}$, by solving the following optimization problem:

$$\min_{\mathbf{x} \in \mathcal{C}} \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_{\mathbf{W}}^2 + \beta \mathsf{R}(\mathbf{x}) \tag{P0}$$

where the regularizer $\mathsf{R}(\mathbf{x})$ is based on a union of sparsifying transforms, and is defined as

$$\mathsf{R}(\mathbf{x}) \triangleq \min_{\{\mathbf{z}_j, C_k\}} \sum_{k=1}^{K} \left\{ \sum_{j \in C_k} \left\{ \|\mathbf{\Omega}_k \mathbf{P}_j \mathbf{x} - \mathbf{z}_j\|_2^2 + \gamma^2 \|\mathbf{z}_j\|_0 \right\} \right\} \quad \text{s.t.} \quad \{C_k\} \in \mathcal{G}, \tag{1}$$

This work was supported in part by the SJTU-UM Collaborative Research Program, NSFC (61501292), Shanghai Pujiang Talent Program (15PJ1403900), NIH grant U01 EB018753, ONR grant N00014-15-1-2141, DARPA Young Faculty Award D14AP00086, and ARO MURI grants W911NF-11-1-0391 and 2015-05174-05.

X. Zheng and *Y. Long are with the University of Michigan - Shanghai Jiao Tong University Joint Institute, Shanghai Jiao Tong University, Shanghai 200240, China (email: yong.long@sjtu.edu.cn)

S. Ravishankar and J. A. Fessler are with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI, 48109 USA

where $\mathbf{W} = \mathsf{diag}\{w_i\} \in \mathbb{R}^{N_d \times N_d}$ is the diagonal weighting matrix with elements being the estimated inverse variance of y_i [12], $\mathbf{A} \in \mathbb{R}^{N_d \times N_p}$ is the system matrix of a CT scan, and $\mathcal{C} \triangleq \{\mathbf{x} | x_j \geq 0, \forall\, j\}$. The operator $\mathbf{P}_j \in \mathbb{R}^{l \times N_p}$ extracts the jth patch of l voxels of $\mathbf{x}$ as $\mathbf{P}_j\mathbf{x}$, vector $\mathbf{z}_j \in \mathbb{R}^l$ denotes the transform-sparse representation of $\mathbf{P}_j\mathbf{x}$, and $\|\mathbf{z}_j\|_0$ is the ℓ_0 "norm" that counts the number of nonzero elements in $\mathbf{z}_j$. (P0) assigns patches to K clusters, and the patches in the kth cluster are matched to the transform $\mathbf{\Omega}_k$. The set C_k contains the indices j corresponding to the patches $\mathbf{P}_j\mathbf{x}$ in the kth cluster, and $\mathcal{G}$ is the set of all possible partitions of the set of integers $[1:N] \triangleq \{1, 2, ..., N\}$ into K disjoint subsets. The parameter $\beta > 0$ controls the noise and resolution trade-off, and $\gamma > 0$ is a weight that controls the sparsity in the model.

In (P0), the patches of the underlying image are assumed to be approximately sparse in the learned union of sparsifying transforms domain. We estimate the image $\mathbf{x}$, the sparse coefficients $\{\mathbf{z}_j\}$, and the cluster assignments $\{C_k\}$ from LDCT data $\mathbf{y}$.

III. Algorithms

A. Learning a Union of Sparsifying Transforms

We pre-learn the union of sparsifying transforms from $\tilde{N}$ 3D patches extracted from a dataset of regular dose CT volumes by solving the following training optimization problem:

$$\min_{\{\mathbf{\Omega}_k, \mathbf{Z}_i, C_k\}} \sum_{k=1}^{K} \sum_{i \in C_k} \left\{ \|\mathbf{\Omega}_k \mathbf{Y}_i - \mathbf{Z}_i\|_2^2 + \eta^2 \|\mathbf{Z}_i\|_0 \right\} + \sum_{k=1}^{K} \lambda_k Q(\mathbf{\Omega}_k) \qquad \text{s.t.} \quad \{C_k\} \in \mathcal{G} \tag{P1}$$

where $\eta > 0$ is a scalar parameter, and $\{\mathbf{Z}_i\}_{i=1}^{\tilde{N}}$ denote the sparse codes of the training signals (vectorized patches) $\{\mathbf{Y}_i\}_{i=1}^{\tilde{N}}$. We set $\lambda_k = \lambda_0 \|\mathbf{Y}_{C_k}\|_F^2$ [9], where $\lambda_0 > 0$ is a constant and $\mathbf{Y}_{C_k}$ is a matrix whose columns are the training signals of $\{\mathbf{Y}_i\}_{i=1}^{\tilde{N}}$ in the kth cluster. Regularizer $Q(\mathbf{\Omega}_k) \triangleq \|\mathbf{\Omega}_k\|_F^2 - \log|\det \mathbf{\Omega}_k|$ prevents trivial solutions and fully controls the condition number of $\mathbf{\Omega}_k$ [6].

We use an alternating minimization algorithm for (P1) that alternates between a *transform update step* (solving for $\{\mathbf{\Omega}_k\}$) and a *sparse coding and clustering step* (solving for $\{\mathbf{Z}_i, C_k\}$). The transform update step is performed the same way (using singular value decompositions) as in [9]. The sparse coding and clustering step has a simple closed-form solution as follows. First, it is clear that if the cluster memberships of the patches are fixed, then the optimal sparse codes are $\mathbf{Z}_i = H_\eta(\mathbf{\Omega}_k \mathbf{Y}_i), \forall i \in C_k$, where the *hard-thresholding* operator $H_\eta(\cdot)$ zeros out vector entries with magnitude less than η. Using this result, it follows that the optimal cluster membership for each $\mathbf{Y}_i$ is $\hat{k}_i = \underset{1 \leq k \leq K}{\arg\min} \{ \|\mathbf{\Omega}_k \mathbf{Y}_i - H_\eta(\mathbf{\Omega}_k \mathbf{Y}_i)\|_2^2 + \eta^2 \|H_\eta(\mathbf{\Omega}_k \mathbf{Y}_i)\|_0 + \lambda_0 \|\mathbf{Y}_i\|_2^2 Q(\mathbf{\Omega}_k) \}$, and the optimal sparse code is $\hat{\mathbf{Z}}_i = H_\eta(\mathbf{\Omega}_{\hat{k}_i} \mathbf{Y}_i)$.

B. Optimization Algorithm for (P0)

We propose an alternating algorithm for (P0) that alternates between updating $\mathbf{x}$ (*image update step*), and $\{\mathbf{z}_j, C_k\}$ (*sparse coding and clustering step*).

1) Image Update Step: With $\{\mathbf{z}_j, C_k\}$ fixed, (P0) reduces to the following weighted least squares problem:

$$\min_{\mathbf{x} \in \mathcal{C}} \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_{\mathbf{W}}^2 + \beta \sum_{k=1}^{K} \sum_{j \in C_k} \|\mathbf{\Omega}_k \mathbf{P}_j \mathbf{x} - \mathbf{z}_j\|_2^2. \tag{2}$$

Here, we denote the second term as $\mathsf{R}_2(\mathbf{x})$. Using relaxed OS-LALM [11], we solve (2) by iterating over the following updates, where for each k, we further iterate over $1 \leq m \leq M$, i.e., the M ordered subsets:

$$\begin{cases} \mathbf{s}^{(k+1)} = \rho(\mathbf{D}_\mathbf{A}\mathbf{x}^{(k)} - \mathbf{h}^{(k)}) + (1-\rho)\mathbf{g}^{(k)} \\ \mathbf{x}^{(k+1)} = [\mathbf{x}^{(k)} - (\rho \mathbf{D}_\mathbf{A} + \mathbf{D}_\mathsf{R})^{-1}(\mathbf{s}^{(k+1)} + \nabla \mathsf{R}_2(\mathbf{x}^{(k)}))]_\mathcal{C} \\ \boldsymbol{\zeta}^{(k+1)} \triangleq M \mathbf{A}_m^T \mathbf{W}_m (\mathbf{A}_m \mathbf{x}^{(k+1)} - \mathbf{y}_m) \\ \mathbf{g}^{(k+1)} = \frac{\rho}{\rho+1}(\alpha \boldsymbol{\zeta}^{(k+1)} + (1-\alpha)\mathbf{g}^{(k)}) + \frac{1}{\rho+1}\mathbf{g}^{(k)} \\ \mathbf{h}^{(k+1)} = \alpha(\mathbf{D}_\mathbf{A}\mathbf{x}^{(k+1)} - \boldsymbol{\zeta}^{(k+1)}) + (1-\alpha)\mathbf{h}^{(k)} \end{cases} \tag{3}$$

where $\mathbf{D}_\mathbf{A} \succeq \mathbf{A}^T\mathbf{W}\mathbf{A}$ is a diagonal majorizing matrix of $\mathbf{A}^T\mathbf{W}\mathbf{A}$ (e.g., $\mathbf{D}_\mathbf{A} \triangleq \mathsf{diag}\{\mathbf{A}^T\mathbf{W}\mathbf{A}\mathbf{1}\} \succeq \mathbf{A}^T\mathbf{W}\mathbf{A}$ [13]), $\nabla \mathsf{R}_2(\mathbf{x}) = 2\beta \sum_{k=1}^{K} \sum_{j \in C_k} \mathbf{P}_j^T \mathbf{\Omega}_k^T (\mathbf{\Omega}_k \mathbf{P}_j \mathbf{x} - \mathbf{z}_j)$, the operator $[\cdot]_\mathcal{C}$ projects the input vector onto the convex set $\mathcal{C}$, $\mathbf{A}_m$ is the subset forward projection matrix, and $\mathbf{W}_m$ is a sub-matrix of $\mathbf{W}$. The initial $\mathbf{g}^{(0)} = \boldsymbol{\zeta}^{(0)} = M\mathbf{A}_M^T\mathbf{W}_M(\mathbf{A}_M\mathbf{x}^{(0)} - \mathbf{y}_M)$ and $\mathbf{h}^{(0)} = \mathbf{D}_\mathbf{A}\mathbf{x}^{(0)} - \boldsymbol{\zeta}^{(0)}$, with $\mathbf{x}^{(0)}$ being the most recent estimate of the image volume (warm start). The (over-)relaxation parameter $\alpha \in [1, 2)$, $\rho > 0$ is the AL penalty parameter decreasing gradually with iteration [10], and $\mathbf{D}_\mathsf{R} \succeq \nabla^2 \mathsf{R}_2(\mathbf{x}) = 2\beta \sum_{k=1}^{K} \sum_{j \in C_k} \mathbf{P}_j^T \mathbf{\Omega}_k^T \mathbf{\Omega}_k \mathbf{P}_j$ is a diagonal majorizing matrix of the Hessian of the regularizer $\mathsf{R}_2(\mathbf{x})$, specifically,

$$\mathbf{D}_\mathsf{R} \triangleq 2\beta \big\{ \max_k \lambda_{\max}(\mathbf{\Omega}_k^T \mathbf{\Omega}_k) \big\} \sum_{k=1}^{K} \sum_{j \in C_k} \mathbf{P}_j^T \mathbf{P}_j. \tag{4}$$

The term $\sum_{k=1}^{K} \sum_{j \in C_k} \mathbf{P}_j^T \mathbf{P}_j = \sum_{j=1}^{N_p} \mathbf{P}_j^T \mathbf{P}_j \in \mathbb{C}^{N_p \times N_p}$ above is a diagonal matrix with the diagonal entries corresponding to image pixel locations and their values being the number of contributing image patches [14]. If we assume periodically positioned overlapping image patches that wrap around at image boundaries, and a patch stride of 1 voxel along each (of x, y, z) direction, then $\sum_{j=1}^{N_p} \mathbf{P}_j^T \mathbf{P}_j = l\mathbf{I}$, where $\mathbf{I}$ is identity matrix. Since the $\mathbf{D}_\mathsf{R}$ in (4) is independent of $\mathbf{x}$, $\{\mathbf{z}_j\}$, and $\{C_k\}$, we precompute it prior to iterating.

2) Sparse Coding and Clustering Step: With $\mathbf{x}$ fixed, we solve the following problem to determine the optimal sparse codes and cluster assignments for each patch:

$$\min_{\{\mathbf{z}_j\}, \{C_k\} \in \mathcal{G}} \sum_{k=1}^{K} \left\{ \sum_{j \in C_k} \left\{ \|\mathbf{\Omega}_k \mathbf{P}_j \mathbf{x} - \mathbf{z}_j\|_2^2 + \gamma^2 \|\mathbf{z}_j\|_0 \right\} \right\}. \tag{5}$$

For each j, with $\mathbf{z}_j = H_\gamma(\mathbf{\Omega}_k \mathbf{P}_j \mathbf{x})$, the optimal cluster assignment is as follows:

$$\hat{k}_j = \underset{1 \leq k \leq K}{\arg\min} \|\mathbf{\Omega}_k \mathbf{P}_j \mathbf{x} - H_\gamma(\mathbf{\Omega}_k \mathbf{P}_j \mathbf{x})\|_2^2 + \gamma^2 \|H_\gamma(\mathbf{\Omega}_k \mathbf{P}_j \mathbf{x})\|_0.$$

Minimizing over k for each patch $\mathbf{P}_j\mathbf{x}$ above determines which cluster the patch belongs to. Then, the optimal sparse codes are $\hat{\mathbf{z}}_j = H_\gamma(\mathbf{\Omega}_{\hat{k}_j} \mathbf{P}_j \mathbf{x})$.

IV. Experimental Results

We evaluated the proposed PWLS-ULTRA method for axial cone-beam CT reconstruction, and compared the image reconstruction quality with those of the conventional Feldkamp-Davis-Kress (FDK) method with a Hanning window, and PWLS reconstruction with edge-preserving regularization (PWLS-EP). We also evaluated the proposed method with $K = 1$, i.e., the PWLS-ST method.

We pre-learned a union of square transforms from a $512 \times 512 \times 54$ XCAT phantom [15] using (P1). We extracted $8 \times 8 \times 8$ overlapping image patches with a patch stride $2 \times 2 \times 2$ ($\tilde{N} \approx 1.5 \times 10^6$ patches) for learning. To ensure convergence, we ran the alternating minimization algorithm for (P1) for 1000 iterations, and set $\lambda_0 = 31$ and $\eta = 50$.

We simulated an axial CT scan using a $840 \times 840 \times 96$ XCAT phantom that differs from the training phantom, with $\Delta_x = \Delta_y = 0.4883$ mm and $\Delta_z = 0.625$ mm. We used the "Poisson + Gaussian" model, i.e., $y_i \sim k\,\text{Poisson}\{I_0 \exp(-[\mathbf{Ax}]_i)\} + \text{Normal}\{0, \sigma^2\}$ to generate CT measurements, where the parameter $k = 1$ models the conversion gain from X-ray photons to electrons, and $\sigma^2 = 5^2$ is the variance of electronic noise [16]. We set $y_i = \max(y_i, 0.1)$ to clip negative values. We generated sinograms of size $888 \times 64 \times 984$ using GE LightSpeed cone-beam geometry corresponding to a monoenergetic source with $I_0 = 1 \times 10^4$ and 5×10^3 incident photons per ray and no scatter, respectively. We reconstructed a $420 \times 420 \times 96$ volume with a coarser grid, where $\Delta_x = \Delta_y = 0.9766$ mm and $\Delta_z = 0.625$ mm.

To compare the methods quantitatively, we calculated the Root Mean Square Error (RMSE) and Structural Similarity Index Measurement (SSIM) [17] of the reconstructions in a region of interest (ROI). In our experiments, the ROI consisted of the central 64 of 96 axial slices and a circular (around center) region in each slice (i.e., cylinder in 3D). RMSE in Hounsfield units (HU) is defined as RMSE $= \sqrt{\sum_{i=1}^{N_{p,\text{ROI}}} (\hat{x}_i - x_i^*)^2 / N_{p,\text{ROI}}}$, where $\mathbf{x}^*$ is the ground truth image and $N_{p,\text{ROI}}$ is the number of voxels in the ROI.

For the PWLS-EP method, we used the edge-preserving regularizer $\mathsf{R}(\mathbf{x}) = \sum_{j=1}^{N_p} \sum_{k \in N_j} \kappa_j \kappa_k \varphi(x_j - x_k)$, where $\varphi(t) \triangleq \delta^2(|t/\delta| - \log(1 + |t/\delta|))$ ($\delta = 10$ HU), N_j is the neighborhood, and κ_j and κ_k are the parameters encouraging noise uniformity [18]. Initialized with FDK reconstructions, we ran the PWLS-EP algorithm with regularization parameter 2^{14} for 50 iterations using relaxed OS-LALM with 24 subsets (for both photon intensities).

For the proposed PWLS-ST and PWLS-ULTRA methods, the patch size was set as $8 \times 8 \times 8$ and the overlapping stride was set as $2 \times 2 \times 2$ during reconstruction ($N \approx 2 \times 10^6$ patches). PWLS-EP reconstructions were used to initialize the image in the proposed methods. We empirically chose (β, γ) for $I_0 = 1 \times 10^4$ and 5×10^3 as follows: $\left(2.0 \times 10^5, 20\right)$ and $\left(1.5 \times 10^5, 20\right)$ for PWLS-ST; and $\left(2.5 \times 10^5, 20\right)$ and $\left(1.5 \times 10^5, 25\right)$ for PWLS-ULTRA. In each iteration of the proposed methods, we ran 2 iterations of the image update step with 4 subsets. We performed clustering once every 50 outer iterations, which worked well and saved computation.

TABLE I: RMSE (HU) and SSIM of reconstructions with FDK, PWLS-EP, PWLS-ST ($K = 1$) and PWLS-ULTRA ($K = 15$) for two incident photon intensities.

Intensity	FDK	PWLS-EP	PWLS-ST	PWLS-ULTRA
1×10^4	67.8	33.7	31.9	**31.5**
	0.536	0.917	0.976	**0.979**
5×10^3	89.0	39.9	37.3	**37.2**
	0.463	0.894	0.967	**0.969**

Figure 1(a) shows the reconstructions (shown for the central axial, sagittal, and coronal planes) by FDK, PWLS-EP and PWLS-ULTRA ($K = 15$). Compared to FDK and PWLS-EP, PWLS-ULTRA significantly improves image quality with respect to reducing noise and preserving structural details. Figure 1(b) shows the RMSE for each axial slice (in ROI) in the PWLS-EP and PWLS-ULTRA reconstructions. PWLS-ULTRA clearly provides large improvements in RMSE for many slices, with greater improvement near the central slice.

Figure 2 presents an example of the pixel-level clustering in the PWLS-ULTRA ($K = 5$) reconstruction. Since PWLS-ULTRA clusters image patches, we cluster individual pixels using a majority vote among the (3D) patches that overlap the pixel. Class 1 contains most of the soft tissues; Classes 2 and 3 have some high-contrast edges oriented along specific directions; Class 4 comprises most of the bones and blood vessels; and Class 5 mainly includes the low-contrast edges. Since the clustering step (both during learning and reconstruction) is unsupervised, i.e., different anatomical structures were not labeled manually, there are also a few edges with high pixel intensities included in Class 4.

Table I lists the RMSE and SSIM values in the ROI of the reconstructions with FDK, PWLS-EP, PWLS-ST ($K = 1$), and PWLS-ULTRA ($K = 15$). Both PWLS-ST and PWLS-ULTRA significantly improve the RMSE and SSIM achieved by the PWLS-EP method. PWLS-ULTRA ($K > 1$) using a union of learned transforms leads to better reconstructions than PWLS-ST with a single learned square transform.

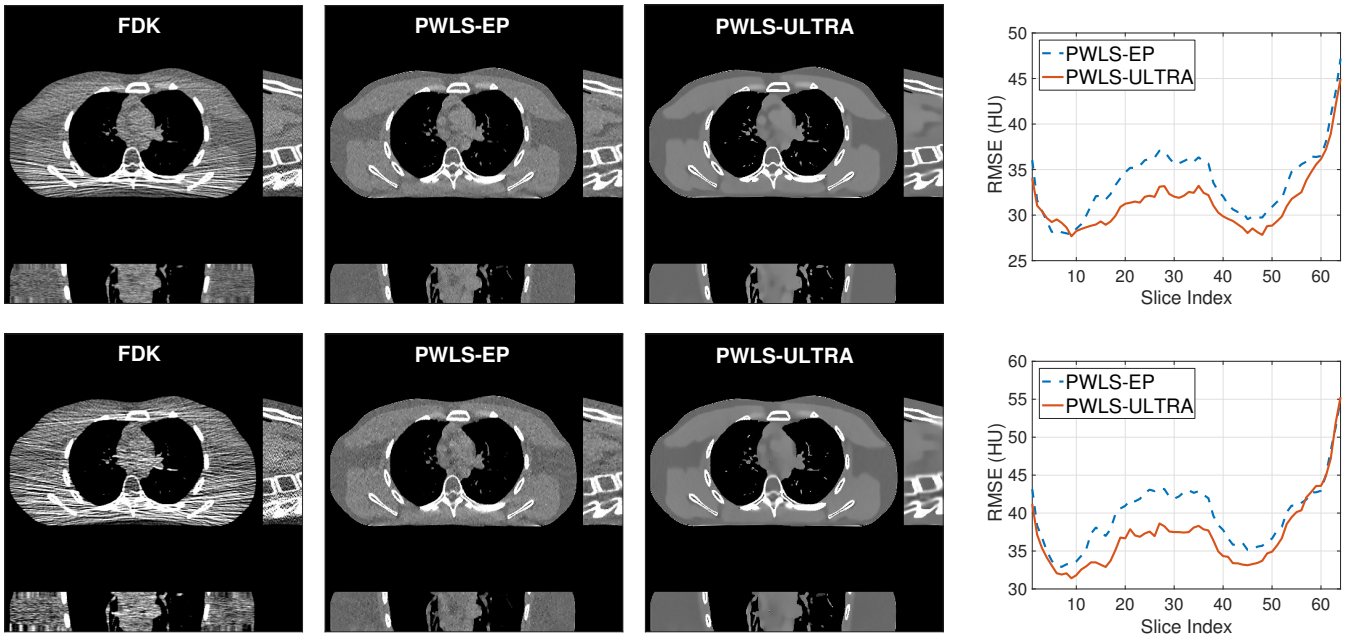
V. Conclusion

We presented the PWLS-ST and PWLS-ULTRA methods for low-dose 3D CT imaging, which combine conventional PWLS reconstruction with regularization based on learned sparsifying transforms. Simulations with 3D axial CT scans of the XCAT phantom show that the proposed methods help reduce X-ray dose to a low level while still providing high quality image reconstructions. For future work, we plan to compare the proposed methods to PWLS with regularization based on a learned synthesis dictionary [2]. We will also apply the proposed methods to clinical CT data.

References

[1] J.-B. Thibault, K. Sauer, C. Bouman, and J. Hsieh, "A three-dimensional statistical approach to improved image quality for multi-slice helical CT," *Med. Phys.*, vol. 34, pp. 4526–4544, Nov. 2007.

[2] 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, pp. 1682–1697, Sept. 2012.



(a) (b)

Fig. 1: Results (within ROI) for $I_0 = 1 \times 10^4$ (top row) and $I_0 = 5 \times 10^3$ (bottom row): (a) Comparison of the reconstructions (shown for the central axial, sagittal, and coronal planes) with the FDK, PWLS-EP, and PWLS-ULTRA ($K = 15$) methods (Display window: $[800, 1200]$ HU); (b) RMSE for each axial slice (in ROI) in the PWLS-EP and PWLS-ULTRA reconstructions.

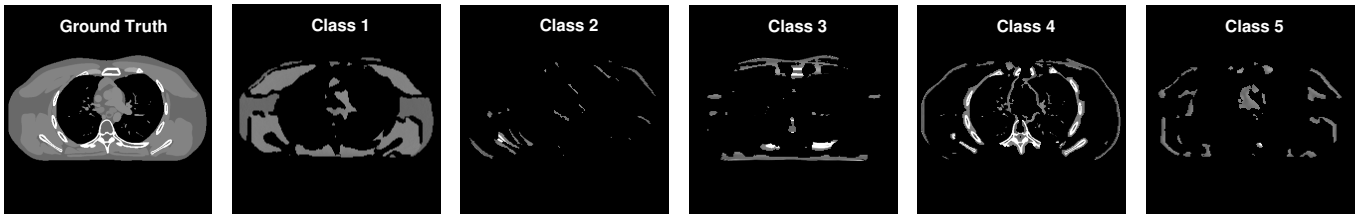


Fig. 2: Pixel-level clustering results for the central axial slice of the PWLS-ULTRA ($K = 5$) reconstruction at photon intensity 1×10^4. The pixels in each class are displayed using the intensities in the reconstruction. Display window: $[800, 1200]$ HU.

[3] R. Rubinstein, T. Peleg, and M. Elad, "Analysis K-SVD: A dictionary-learning algorithm for the analysis sparse model," *IEEE Trans. Sig. Proc.*, vol. 61, pp. 661–677, Feb. 2013.

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

[5] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Sig. Proc.*, vol. 54, pp. 4311–4322, Nov. 2006.

[6] S. Ravishankar and Y. Bresler, "Learning sparsifying transforms," *IEEE Trans. Sig. Proc.*, vol. 61, pp. 1072–1086, Mar. 2013.

[7] S. Ravishankar and Y. Bresler, "ℓ_0 sparsifying transform learning with efficient optimal updates and convergence guarantees," *IEEE Trans. Sig. Proc.*, vol. 63, pp. 2389–2404, May 2015.

[8] L. Pfister and Y. Bresler, "Adaptive sparsifying transforms for iterative tomographic reconstruction," in *Proc. 3rd Intl. Mtg. on image formation in X-ray CT*, pp. 107–110, 2014.

[9] B. Wen, S. Ravishankar, and Y. Bresler, "Structured overcomplete sparsifying transform learning with convergence guarantees and applications," *Intl. J. Comp. Vision*, vol. 114, no. 2-3, pp. 137–167, 2015.

[10] X. Zheng, Z. Lu, S. Ravishankar, Y. Long, and J. A. Fessier, "Low dose CT image reconstruction with learned sparsifying transform," in *Proc. IEEE Wkshp. on Image, Video, Multidim. Signal Proc.*, pp. 1–5, July 2016.

[11] H. Nien and J. A. Fessler, "Relaxed linearized algorithms for faster X-ray CT image reconstruction," *IEEE Trans. Med. Imag.*, vol. 35, pp. 1090–1098, Apr. 2016.

[12] J.-B. Thibault, C. A. Bouman, K. D. Sauer, and J. Hsieh, "A recursive filter for noise reduction in statistical iterative tomographic imaging," in *Proc. SPIE*, vol. 6065, pp. 60650X–1–60650X–10, 2006.

[13] H. Erdoğan and J. A. Fessler, "Ordered subsets algorithms for transmission tomography," *Phys. Med. Biol.*, vol. 44, pp. 2835–2851, Nov. 1999.

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

[15] W. P. Segars, M. Mahesh, T. J. Beck, E. C. Frey, and B. M. W. Tsui, "Realistic CT simulation using the 4D XCAT phantom," *Med. Phys.*, vol. 35, pp. 3800–3808, Aug. 2008.

[16] L. Fu, T. C. Lee, S. M. Kim, A. M. Alessio, P. E. Kinahan, Z. Chang, K. Sauer, M. K. Kalra, and B. D. Man, "Comparison between pre-log and post-log statistical models in ultra-low-dose CT reconstruction," *IEEE Trans. Med. Imag.*, vol. 36, pp. 707–720, Mar. 2017.

[17] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Im. Proc.*, vol. 13, pp. 600–612, Apr. 2004.

[18] J. H. Cho and J. A. Fessler, "Regularization designs for uniform spatial resolution and noise properties in statistical image reconstruction for 3D X-ray CT," *IEEE Trans. Med. Imag.*, vol. 34, pp. 678–689, Feb. 2015.