Published in final edited form as:

Phys Med Biol. 2015 April 7; 60(7): 2803–2818. doi:10.1088/0031-9155/60/7/2803.

Tensor-based Dictionary Learning for Dynamic Tomographic Reconstruction

Shengqi Tan^{1,2}, Yanbo Zhang^{3,4}, Ge Wang⁵, Xuanqin Mou³, Guohua Cao⁶, Zhifang Wu^{1,2}, and Hengyong Yu^{4,*}

¹Beijing Key Laboratory of Nuclear Detection & Measurement Technology, Beijing 100084, China

²Institute of Nuclear and New Energy Technology, Tsinghua University, Beijing 100084, China

³Institute of Image Processing and Pattern Recognition, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China

⁴Department of Electrical and Computer Engineering, University of Massachusetts Lowell, Lowell, MA, 91854, USA

⁵Biomedical Imaging Center/Cluster, CBIS, Dept. of BME, Rensselaer Polytechnic Institute, Troy, NY 12180, USA

⁶VT-WFU School of Biomedical Engineering and Sciences, Virginia Tech, Blacksburg, VA 24060, USA

Abstract

In dynamic computed tomography (CT) reconstruction, the data acquisition speed limits the spatio-temporal resolution. Recently, compressed sensing theory has been instrumental in improving CT reconstruction from far few-view projections. In this paper, we present an adaptive method to train a tensor-based spatio-temporal dictionary for sparse representation of an image sequence during the reconstruction process. The correlations among atoms and across phases are considered to capture the characteristics of an object. The reconstruction problem is solved by the alternating direction method of multipliers. To recover fine or sharp structures such as edges, the nonlocal total variation is incorporated into the algorithmic framework. Preclinical examples including a sheep lung perfusion study and a dynamic mouse cardiac imaging demonstrate that the proposed approach outperforms the vectorized dictionary-based CT reconstruction in the case of few-view reconstruction.

I. Introduction

X-ray computed tomography (CT) has been developed and widely used in hospitals and clinics for diagnosis and intervention. The CT image quality can be significantly degraded by respiratory motion, and accurate estimation of physiological and pathological parameters become difficult and unreliable. To solve this problem, the dynamic CT imaging technique was developed as a powerful tool for noninvasive visualization and analysis of changes

*Corresponding author: Hengyong Yu, hengyong-yu@ieee.org.

inside a patient [1–4]. During a dynamic CT imaging procedure, a sequence of projections are collected at multiple time points or different phases. However, the data acquisition speed of the current CT systems usually cannot collect a sufficient number of projections for each phase to reach the Shannon/Nyquist sampling rate required by the classic reconstruction methods. Over recent years, compressive sensing (CS) theory has become popular for CT reconstruction from far-few projections [5][6]. In the CS-based reconstruction framework, the key is to explore the prior information which generally focuses on the sparsity of an object to be reconstructed. In the CT reconstruction field, the discrete gradient transform (DGT) is the most popular sparsifying transform implicitly assuming a piecewise constant image model, and the corresponding reconstruction technique is called total variation (TV) minimization [7–10]. However, most of the medical images are not piecewise constant. Hence, the DGT cannot distinguish true structures from image noise. Therefore, some fine structures may be lost or distorted, and blocky artifacts are generated in reconstructed images [11]. Herman and Davidi showed that tumor-like structures may be removed by the TV-minimization based method for few-view reconstruction [12]. The inherent defect of the TV-minimization approach motivated us to investigate other sparse representations, such as dictionary learning (DL) based options [11]. The sparse representation in terms of a redundant dictionary has attracted a great interest and achieved significant improvements such as for denoising [13], restoration [14], image analysis [15], and magnetic resonance imaging [16]. In particular, the DL method has demonstrated a promising performance for CT reconstruction in terms of preserving fine structures and suppressing image noise [11]. A redundant dictionary is an over-complete basis. The atoms in this basis are learned from application-specific training images. The representative methods to optimize a dictionary include the method of optimal directions (MOD) [17] and K-SVD [18] which treat signals as vectors.

Dynamic volumetric imaging involves a time dimension. The correlation across different phases must be taken into account [1–4][19]. For example, a spatio-temporal dictionary has been trained to encode the whole dataset which is superior to the spatial dictionary [3][4] because the temporal coherence across respiration phases is usually not guaranteed otherwise [20]. Generally speaking, the dimensionality of this dynamic dictionary is 3-D or higher dimensions, and the tensor notion has been used to deal with multidimensional data as a powerful tool [21]–[24]. It has been shown that tensor-based DL methods effectively capture structures and sparsely represent signals [21]–[24]. Also, tensor tools were employed to deal with high-dimensional data including multi-energy or spectral CT reconstruction [25]–[27] and 4D CT reconstruction [28], [29]. Encouraging results were reported based on the prior knowledge of rank, sparsity, and structure learning.

There are two widely used tensor methods to decompose datasets: one is the Tucker decomposition, and the other is Canonical Decomposition (CANDECOMP, abbreviated CPD) also known as Parallel Factor Analysis (PARAFAC) [21]. In the Tucker decomposition, multidimensional signals are represented by sparse core tensors with block structures. In this paper, via the Tucker decomposition we propose a tensor-based DL method to train a 3-D spatio-temporal dictionary for sparse representation during the CT reconstruction [30][31]. The alternating direction method of multipliers (ADMM) scheme [32] is introduced to accelerate the iteration. To improve finer structures with sharp edges,

the nonlocal mean (NLM) method was well known as well. Buades *et al.* proposed a NLM approach to achieve better performance for image denoising [33]. Combined with the TV and NLM methods, the nonlocal total variation (NLTV) regularization was applied and proven helpful in enhancing the local strong edge information in reconstructed images [34]–[36]. The key idea behind these methods is to exploit the similarity between image patches.

The rest of this paper is organized as follows. In Section II, we introduce the tensor based DL and the spatio-temporal dictionary training. In Section III, the proposed tensor based DL reconstruction framework is presented, including the ADMM optimization scheme. The performances of two DL methods (tensor and vector based DL) are compared. In Section IV, the improved tensor DL method with NLTV is described. Representative experimental results are reported which are reconstructed from preclinical and clinical projections respectively. Finally, related issues are discussed in Section V.

II. Tensor Based Dictionary Learning

Sparse coding has attracted much attention in many fields. A representative method is the sparse representation by a well-trained over-complete dictionary. In 2006, Elad *et al.* proposed an K-SVD algorithm to train a dictionary for image denoising [18], which was extended by Mairal *et al.* for color image restoration [14]. In 2012, Xu *et al.* developed a low-dose x-ray CT reconstruction method using a dictionary learning technique [11]. However, the input signals/patches in the aforementioned methods were rearranged as vectors. This strategy can lose or distort the inherent spatial constraints of the original structures, especially for high dimensional images. The tensor factorization and decomposition have been popular in the signal processing community to deal with high dimensional signals/images, and the tensor based methods were investigated for DL to achieve better performance in the cases of multi-dimensional signals/images [21][22][23].

Briefly speaking, a vector is considered as a one-dimensional (1D) tensor and a matrix as a two-dimensional (2D) tensor. Thus, a three-dimensional (3D) signal e.g. $\underline{U} \in R^{I_1 \times I_2 \times I_3}$ is a 3D tensor, where R represents the real space and I_n ($n = 1, 2, 3$) are the dimensions of each mode. A 3D tensor $\underline{U}$ can be unfolded to a matrix $U_{(n)}$ according to a specific order, i.e. $U_{(1)} \in R^{I_1 \times I_2 I_3}$, $U_{(2)} \in R^{I_2 \times I_1 I_3}$ and $U_{(3)} \in R^{I_3 \times I_1 I_2}$. The Tucker tensor decomposition is defined as [37][38][39]:

$$\underline{U} = \underline{G} \times_1 D_1 \times_2 D_2 \times_3 D_3, \quad (1)$$

where $D_n \in R^{I_n \times R_n}$ are factor matrices, $\underline{G} \in R^{R_1 \times R_2 \times R_3}$ is the coefficient tensor (also known as core tensor) reflecting the affiliated interaction between factor matrices, R_n ($n = 1, 2, 3$) are the dimensions of each mode of $\underline{G}$, and $\underline{G} \times_n D_n$ expresses a product of tensor $\underline{G}$ with a matrix D_n in mode n ($n = 1, 2, 3$). The CPD can be seen as a special case of the Tucker decomposition where the coefficient tensor is diagonal. Given a group of 3-D training tensors $\underline{U}^{(t)}$, $t = 1, 2, \dots, T$, the dictionary learning problem can be formulated as the following minimization problem:

$$\operatorname{argmin}_{\underline{G}^{(t)}, D_n} (\Sigma_{t=1}^T \left\| \underline{U}^{(t)} - \underline{G}^{(t)} \times_1 D_1 \times_2 D_2 \times_3 D_3 \right\|_F^2) \text{ s.t. } \left\| \underline{G}^{(t)} \right\|_0 \leq K, \quad (2)$$

where K represent the sparsity level and it is the number of nonzero entries in the coefficient tensor $\underline{G}^{(t)}$. An alternating least squares method can be used to update D_n at a time while other variables are fixed. For example, to update D_1 , the function can be rewritten in mode-1 as follows [21]:

$$\Sigma_{t=1}^T \left\| \underline{U}^{(t)} - \underline{G}^{(t)} \times_1 D_1 \times_2 D_2 \times_3 D_3 \right\|_F^2 = \left\| \left[U_{(1)}^{(1)} \cdots U_{(1)}^{(T)} \right] - D_1 \left[G_{(1)}^{(1)} C_{(1)} \cdots G_{(1)}^{(T)} C_{(1)} \right] \right\|_F^2, \quad (3)$$

where $C_{(1)} = (D_3 \otimes D_2)^T$ and $\otimes$ represents the Kronecker product [37]. Denoting

$Us = \left[U_{(1)}^{(1)} \cdots U_{(1)}^{(T)} \right]$ and $W = \left[G_{(1)}^{(1)} C_{(1)} \cdots G_{(1)}^{(T)} C_{(1)} \right]$, the least square solution for D_1 can be simplified as

$$D_1 = (Us \cdot W^T)(WW^T)^\dagger, \quad (4)$$

where $\dagger$ represents the Moore-Penrose pseudo-inverse of the matrix. D_2 and D_3 can be computed following the same fashion. Once D_n ($n = 1, 2, 3$) are learned from the training dataset, we can apply the Kronecker-OMP [40] to update $\underline{G}^{(t)}$. The pseudo-codes for the Tensor-MOD DL approach can be summarized as the following Algorithm I. This process can be stopped when the predefined criteria are reached (e.g. the number of iterations, representation error).

Algorithm1:
Tensor?MOD
Dictionary
Learning

Input:	Training set of tensors $\underline{U}^{(1)}, \underline{U}^{(2)}, \dots, \underline{U}^{(T)}, \in R^{I_1 \times I_2 \times I_3}$; sparsity level K and tolerance ε .
Output:	matrices $D_n \in R^{I_n \times R_n}$ ($n = 1, 2, 3$).
1:	Initializing matrices D_n ($n = 1, 2, 3$) randomly;
2:	Computing global error $e = \Sigma_{t=1}^T \left\ \underline{U}^{(t)} - \underline{G}^{(t)} \times_1 D_1 \times_2 D_2 \times_3 D_3 \right\ _F^2$
3:	While the stopping criteria are not satisfied (e.g. $e > \varepsilon$) do
4:	for $n = 1$ to 3 do
5:	updating D_n using Eq.(4);
6:	end for
7:	Normalizing columns of matrices D_n ($n = 1, 2, 3$);
8:	Updating sparse core tensors $\underline{G}^{(t)} \in R^{R_1 \times R_2 \times R_3}$ ($t = 1, 2, \dots, T$) using the Kronecker-OMP with the sparsity level K ;
9:	Updating global error e ;
10:	end while
11:	Return D_n ($n = 1, 2, 3$).

III. Tensor Based DL for CT Reconstruction

3.1. Reconstruction Method

Let E_s be an operator to extract a spatio-temporal patch (a 3D tensor) from a dynamic image sequence u with the size of (s_x, s_y, s_t) , where x and y are two spatio-dimensions and t is temporal dimension. Based on the aforementioned DL model, the dynamic CT reconstruction problem can be formulated as follow:

$$\min_{u, D, \{\underline{G}^{(s)}\}} \Sigma_s \left(\frac{\lambda_1}{2} \left\| \underline{G}^{(s)} \times_1 D_1 \times_2 D_2 \times_3 D_3 - E_s u \right\|_2^2 + \left\| \underline{G}^{(s)} \right\|_0 \right) + \frac{\lambda_2}{2} \|Au - f\|_2^2 \quad (5)$$

where λ_1 and λ_2 are tuning parameters. While the last term enforces the data fidelity in projection domain with a system matrix A and projection measurements f , the first and second correspond to the tensor based DL model. For convenience, we use Dg_s to represent $\underline{G}^{(s)} \times_1 D_1 \times_2 D_2 \times_3 D_3$ and (5) can be simplified as

$$\min_{u, D, \{g_s\}} \Sigma_s \left(\frac{\lambda_1}{2} \|Dg_s - E_s u\|_2^2 + \|g_s\|_0 \right) + \frac{\lambda_2}{2} \|Au - f\|_2^2. \quad (6)$$

To solve the minimization problem Eq.(6), an alternating direction method of multipliers (ADMM) based on the augmented Lagrangian can be applied with two auxiliary variables $b = Eu - Dg$ and h . The ADMM method and its variants are widely applied in the field of image processing to solve convex minimization problems [41]. The detail iterative procedure in our framework is as follow [32]:

$$\{D^{k+1}, g_s^{k+1}, u^{k+1}, b_s^{k+1}\} = \operatorname{argmin}_{D, \{g_s\}, u, b} \Sigma_s \left(\|g_s\|_0 + \frac{\lambda_1}{2} \|b_s\|_2^2 + \frac{\beta}{2} \|Dg_s + b_s - E_s u - h_s/\beta\|_2^2 \right) + \frac{\lambda_2}{2} \|Au - f\|_2^2, \quad (7)$$

$$h_s^{k+1} = h_s^k + \beta \left(-D^{k+1} g_s^{k+1} - b_s^{k+1} + E_s u^{k+1} \right). \quad (8)$$

Problem (7) can be broken down into the following three subproblems.

1. Dictionary Learning and Sparse Coding
2. Fixing other variables, we update the dictionary D and sparse coding g_s to represent the spatio-temporal image u sparsely and accurately:

$$\{D^{k+1}, g_s^{k+1}\} = \operatorname{argmin}_{D, \{g_s\}} \Sigma_s \|Dg_s - E_s u\|_2^2, t. \|g_s\|_0 \leq K \forall s. \quad (9)$$

As mentioned in section 2, the tensor-MOD dictionary learning method can be applied to update D and the Kronecker-OMP can be used to compute g_s for all patches.

3. b -sub-problem
4. This sub-problem views b as the only free variable. As a result, Problem (6) becomes:

$$b_s^{k+1} = \operatorname{argmin}_{b_s} \Sigma_s \left(\frac{\lambda_1}{2} \|b_s\|_2^2 + \frac{\beta}{2} \|D^k g_s^k + b_s - E_s u^k - h_s / \beta\|_2^2 \right). \quad (10)$$

This least square problem (10) is solved by

$$b_s^{k+1} = \frac{\beta}{\lambda_1 + \beta} (E_s u^k - D^{k+1} g_s^{k+1} + h_s^k / \beta). \quad (11)$$

5. u -sub-problem

6. The optimization problem for variable u can be expressed as:

$$u^{k+1} = \operatorname{argmin}_u \Sigma_s \left(\frac{\beta}{2} \|D^k g_s^k + b_s^k - E_s u - h_s / \beta\|_2^2 \right) + \frac{\lambda_2}{2} \|Au - f\|_2^2. \quad (12)$$

The problem (12) can be solved by the separable paraboloid surrogate method [42]:

$$u^{k+1} = u^k - \frac{\lambda_2 A^T (Au^k - f) + \beta \Sigma_s E_s^T (E_s u^k - D^{k+1} g_s^{k+1} - b_s^{k+1} + h_s^k / \beta)}{\lambda_2 A A^T + \beta \Sigma_s E_s E_s^T}. \quad (13)$$

The pseudo-codes of the tensor based adaptive DL for dynamic CT reconstruction are summarized as the following Algorithm 2.

Algorithm2: Tensor Based Adaptive DL for Dynamic CT Reconstruction

Input:	projection data f , regularization parameters $\lambda_1, \lambda_2, \beta$; patch size $I_x \times I_y \times I_t$ and core tensor size $R_1 \times R_2 \times R_3$ and other parameters.
Output:	reconstructed image sequences u
1:	Initializing matrices D_n ($n = 1, 2, 3$), u and $b = 0$, $h = 0$;
2:	While the stopping criteria are not satisfied do
3:	Constructing a dictionary D_n ($n = 1, 2, 3$) and coefficients g using the Algorithm 1;
4:	Updating b using Eq. (11);
5:	Representing u using Eq.(13);
6:	Updating h using Eq. (8);
7:	end while
8:	Return the final reconstruction u .

The tensor dictionary is acquired adaptively in our scheme and updated during each iteration step. As discussed by Xu *et al.* [11], when the training images do not match a specific application closely, the trained global dictionary cannot reveal some details in the reconstruction. To reconstruct images stably, the tensor dictionary is trained adaptively in our following experiments, while the computational cost should be taken into account. When an excellent training data set is available, a global tensor dictionary can be prepared to boost the iteration. Besides, how to choose the tuning parameters is a common and interesting problem for almost all the regularized reconstruction algorithms. In our work,

these parameters were selected empirically to ensure the reconstructed image quality and reasonable computational cost.

3.2. Experimental Results

Projections acquired in a low-dose (80kV/17mAs) sheep lung perfusion study at 20 time points (20 phases) [43] were used to evaluate and validate the proposed algorithm. 1160 projections were collected uniformly over a 360° range for all phases and for each projection, 672 detector elements were equal-angularly distributed. First, 20 ground truth images of 512×512 were reconstructed using the fully sampled data for each phase by the conventional filtered backprojection (FBP), and a global dictionary was trained by the well-known K-SVD algorithm for the vector based DL. Readers can refer to [43] for details of data acquisition.

Throughout our experiments, the reconstructed images were matrixes of 512×512 pixels covering a 29.09×29.09cm² region and the sparsity constrain was forced on the entire lung region of 500×370 pixels. The regularization parameters were set to $\lambda_1 = 10^{-4}$, $\lambda_2 = 1$ and $\beta = 0.2$. Regarding the vector based DL, we empirically set the number of atoms as 512 and the sparsity level K as 6. The patch size was 8×8 and the image sequence was reconstructed frame-by-frame. For the adaptive tensor based DL, the default patch size was 4×4×4 covering the same amount of pixels in vector based DL and the sparsity level K was also 6. The core tensor size was 8×8×8. The image sequence was treated as a volume data. Because this method was developed for dynamic (sparse projections) and low dose CT (high level noise), the projections here for each phase were down-sampled from 1160 to 232 and 116 views respectively to test the method. The convergence of DL based method for CT reconstruction has been discussed in [11] and we did not repeat it in this paper. In our experiments, the iteration number was 70 and the reconstructed images changed little with further iterations for both of the methods. Two indexes were introduced to quantitatively evaluate the reconstructed image quality. The first one is the root mean square error

(RMSE): $RMSE = \sqrt{\sum_{j=1}^J (\hat{u}_j - u_j^r)^2 / J}$, where u_j^r is the r -th reconstructed value, $\hat{u}_j$ is the ground truth value of the j th pixel and J is the number of the pixels in an image. The second one is the widely used image quality assessment (IQA) index for structural similarity (SSIM) [44]. The closer to 1 the value of SSIM/FSIM is, the higher the structural similarity is achieved.

Fig. 1. shows some representative results of the sheep lung perfusion study for phase 10. In this case, the results reconstructed from the down-sampled projections were very similar and their differences were not obvious. Alternatively, the difference images with respect to the reference were plotted for better visualization and comparison. The quantitative evaluation results in terms of RMSE and SSIM are in Fig. 2.

Evidently, the proposed tensor based DL method achieves lower RMSEs and higher SSIMs/FSIMs than the vector based DL for dynamic CT reconstruction at all phases. Since both of the methods follow the same reconstruction framework, it is not easy to distinguish them from the reconstructed images visually although the images reconstructed by the tensor based DL are closer to the ground truth than the vector based DL. The white regions in Figs.

2(b) and (c) indicated the images reconstructed by the tensor based DL method are more accurate. Therefore, we conclude that tensor based DL outperforms the vector based DL method.

IV. Tensor-based DL Method with Nonlocal Total Variation

4.1. Reconstruction Approach

The DL method is not always effective to reconstruct some finer structures with strong edges, especially when the chosen image patch size is not suitable. The TV constraint processes the image globally which may generate blocky artifacts. In order to enhance the strong edge information and local finer structures, we incorporate the nonlocal total variation (NLTV) regularization [46] into the tensor based DL reconstruction framework. Nonlocal means schemes exploit the similarity between patches in the image which takes a mean of all pixels in the image, weighted by the similarities between these pixels and the target pixel [47].

Let $\Omega \in \mathbb{R}^2$, $x \in \Omega$, and $u(x)$ be a real function $\Omega \rightarrow \mathbb{R}$. The nonlocal gradient $\nabla_{NL}u(x)$ is defined as a vector of all partial differences $\nabla_{NL}u(x, \cdot)$ at x as

$$\nabla_{NL}u(x, y) = (u(y) - u(x)) \sqrt{w(x, y)}, \forall y \in \Omega, \quad (14)$$

where w is a nonnegative symmetric weight factor. The graph divergence of a vector $p: \Omega \times \Omega \rightarrow \mathbb{R}$ is defined as the standard adjoint relation with the NL gradient operator as

$$\langle \nabla_{NL}u, p \rangle = - \langle u, \text{div}_{NL}p \rangle, \forall p: \Omega \times \Omega \rightarrow \mathbb{R}, \quad (15)$$

where the graph divergence div_{NL} of $p: \Omega \times \Omega \rightarrow \mathbb{R}$ is obtained by

$$\text{div}_{NL}p(x) = \int_{\Omega} (p(x, y) - p(y, x)) \sqrt{w(x, y)} dy. \quad (16)$$

Then the L^1 norm of the weighted graph gradient $\nabla_{NL}u(x)$, that is NLTV, is defined:

$$\|\nabla_{NL}u\|_1 = \int_{\Omega} |\nabla_{NL}u(x)| dx. \quad (17)$$

We incorporate the NLTV into our tensor based DL method by solving the following problem:

$$\min_{u, D, \{g_s\}, d} \sum_s \left(\frac{\lambda_1}{2} \|Dg_s - E_s u\|_2^2 + \|g_s\|_0 \right) + \frac{\lambda_2}{2} \|Au - f\|_2^2 + \lambda_3 \|\nabla_{NL}d\|_1, s.t. d = u. \quad (18)$$

The scaled augmented Lagrangian function of Eq. (18) can be expressed as [4], [32]:

$$L_{\beta}(D, g_s, u, d, y) = \sum_s \left(\|g_s\|_0 + \frac{\lambda_1}{2} \|Dg_s - E_s u\|_2^2 \right) + \frac{\lambda_2}{2} \|Au - f\|_2^2 + \lambda_3 \|\nabla_{NL}d\|_1 + \frac{\beta}{2} \|d - u - y\|_2^2, \quad (19)$$

where y is an auxiliary variable. Similar to the iteration procedure in section III, the iteration solution for problem Eq.(19) can be summarized as follow:

$$\{D^{k+1}, g_s^{k+1}\} = \operatorname{argmin}_{D, \{g_s\}} \sum_s \|Dg_s - E_s u^k\|_2^2 \text{ s.t. } \|g_s\|_0 \leq K \forall s, \quad (20)$$

$$d^{k+1} = \operatorname{argmin}_d \lambda_3 \|\nabla_{NL} d\|_1 + \frac{\beta}{2} \|d - u^k - y^k\|_2^2, \quad (21)$$

$$u^{k+1} = \operatorname{argmin}_u \sum_s \left(\frac{\lambda_1}{2} \|D^{k+1} g_s^{k+1} - E_s u\|_2^2 \right) + \frac{\lambda_2}{2} \|Au - f\|_2^2 + \frac{\beta}{2} \|d^{k+1} - u - y^k\|_2^2, \quad (22)$$

$$y^{k+1} = y^k + u^{k+1} - d^{k+1}. \quad (23)$$

The solution for subproblem (20) has been discussed in section III, and u can be updated by :

$$u^{k+1} = u^k - \frac{\lambda_1 \sum_s E_s^T (E_s u^k - D^{k+1} g_s^{k+1}) + \lambda_2 A^T (Au^k - f) + \beta (u^k + y^k - d^{k+1})}{\lambda_1 \sum_s E_s E_s^T + \lambda_2 A A^T + \beta I}. \quad (24)$$

Therefore, here we focus on the sub-problem (21). Because there is no coupling between elements of d , we can apply the Split-Bregman method proposed by Zhang *et al.*[34] to deal with this nonlocal optimization problem, and the variable d was updated phase by phase during the iteration. The readers can refer to [34][46] for more details of the NLTV optimization problem. Because the nonlocal dictionary learning has achieved good results for image denoising [48], combination of nonlocal TV and nonlocal sub-dictionary learning will be an interesting follow-up topic.

4.2. Sheep-lung Perfusion Study

In order to evaluate the improved tensor based DL method with NLTV, we compared the reconstructed results by the tensor based DL with and without NLTV. The regularization parameters used in the tensor-DL NLTV were: $\lambda_1 = 0.2$, $\lambda_2 = 1$ and $\beta = 0.1$ and other parameters remain the same which were listed in section 3.2. The improvement in terms of difference image and quantitative evaluation are in Figures 3 and 4, respectively.

Because the datasets used for the above evaluations were acquired from a real CT scanner, the measurement noise has already been included. In order to evaluate the robustness of the proposed method at different noise levels, additive white Gaussian noise was manually added into the raw projections to simulate different signal-noise-ratios (SNR): 25dB (high noise) and 40dB (low noise). While the Poisson noise model was widely accepted to simulate the interactions between the imaging object and x-ray photons, the Gaussian noise model is approximately equivalent to the Poisson noise model with an idea bowtie filter to reach the same expected photon number at all the detector elements. The same parameters in

section 3 were used for the vector based and tensor based DL methods. As for the tensor based DL NLTVM method, the values of the regularization parameters were: $\lambda_1 = 0.2$ and $\lambda_2 = 1$. We choose different values of β for different SNRs: $\beta = 0.3$ (25dB) and $\beta = 0.1$ (40dB).

The results of phase 10 reconstructed from 116 down-sampled projections with SNR=40dB were shown in Fig. 5 and the corresponding magnified local regions were shown in Fig. 6. The quantitative evaluation results in terms of RMSE, SSIM and FSIM were shown in Fig. 7. In Fig. 7, the left column were for data with SNR=25dB and the right column were for SNR=40dB. Although the performance of all the methods degrades as the SNR and projection data quality decrease, the tensor based DL NLTVM method outperforms the other two methods for almost all the conditions with finer structures and less noise. It can be observed that the quality degradation in tensor based DL NLTVM is not as apparent as the other two methods. Therefore, we can conclude that the tensor based DL NLTVM method is robust. From figures 5–7, it can be concluded that the tensor based DL method outperforms the vector based DL method and the improved tensor based DL method with NLTVM reconstructs the best image quality with lowest RMSEs and highest SSIMs/FSIMs, especially for the situations of few-view and noisy projections. In Fig. 6, it can be seen that tensor based DL NLTVM method can recover local regions with stronger edges (see the regions indicated by arrow “A” and “C”) and finer details (see the region indicated by arrow “B”).

4.3. Mouse Cardiac Imaging Study

To demonstrate the potential application of the proposed algorithms for dynamic cardiac imaging, we collected a group of existing raw projections in a mouse cardiac CT studies (2 phases). 400 projections were equiangularly ranged from 0° to 199.5° . For each projection, 1200 detector elements were equidistantly distributed, and the detector size was 60mm. The radius of the scanning trajectory was 141.53mm. In this study, images were reconstructed in a matrix of 512×512 over a $39.168 \times 39.168 \text{ mm}^2$ region. Throughout this experiments, the parameters were chosen as $\lambda_1 = 0.2$, $\lambda_2 = 1$ and $\beta = 0.3$. For the vector based DL method, the patch size was 5×5 , and the number of atoms was set to 101 and the sparsity level K was set to 5. Both of the phases were reconstructed successively. For the tensor based DL with and without NLTVM methods, the core tensor size was $7 \times 7 \times 2$, the patch size was $5 \times 5 \times 2$ and the sparsity level K was 5. The global dictionary for vector based DL was trained from images reconstructed by the FBP using the fully sampled data. The reconstructed images were shown in Fig. 8. Because the raw projections were collected from limited angles, it can be seen that the FBP results had more streak artifacts. However, the images reconstructed from the DL based methods were much better. Meanwhile, the NLTVM regularization could enhance the edges of the reconstructed images. The results for phase 2 of the mouse cardiac CT from 200 views were shown in Fig. 9.

Clearly, the conventional FBP algorithm is the worst choice in this situation with few-view projections. Images reconstructed from the tensor based DL NLTVM method were much cleaner. By the limitation of the number of phases in this preclinical study, information lies in the temporal dimensional was not emphasized so much. Because the image quality reconstructed by the tensor based DL NLTVM method was the tradeoff between tensor based

DL and NLTV, the shortcomings of NLTV will also be introduced. Both of the DL and NLTV regularizations will affect the reconstructed image quality. The tradeoff between definition and cleanness for the reconstructed images needs to be carefully evaluated. Besides, the sizes of core tensor and patch are quite important for the reconstruction and they need to be selected according to the scanned imaging objects and the required spatial resolution.

V. Discussions and Conclusion

Because a tensor dictionary reflects correlations between pixels and across phases, the intrinsic features are emphasized. As a result, more information can be applied for image reconstruction. When the few-view and/or low-dose problems are dealt with, better results can be reconstructed with more details. When the dictionary learning approach is applied to CT reconstruction, atoms should match structural features as much as possible. Therefore, tensor atoms are preferred over vector atoms in which the temporal variation is not easy to learn.

From the experimental results in Sections III and IV, it is seen that the tensor-based dictionary has shown superiority over the vector-based dictionary, as evidenced by the fact that the structural information in the reconstructed results is richer with the former. The redundancy in the data was learned, and the sparsity constraints imposed on the tensor dictionary adaptively to exploit both temporal correlation and spatial structures. The improvements are evident in our experiments in terms of quantitative evaluation and visual comparison.

The proposed algorithms were solved with ADMM which works at a fast convergence rate. Several popular methods were also introduced to deal with the sub-problems in the iteration. By incorporating the non-local TV regularization to the tensor-based DL method, the images reconstructed from few-view and noisy projections are refined to give finer structures. Further research is needed to optimize the controlling parameters, especially the sizes of patch and core tensor. The combination of nonlocal TV and nonlocal sub-dictionary learning will be an interesting topic because both of them may benefit from non-locality. In the Tucker decomposition, multidimensional signals can be efficiently represented by sparse core tensors with block structures. The sparsity of core tensor (especially in the time dimension) could be coupled with a low-rank requirement, and a data structure oriented tensor dictionary learning (such as Quadtree decomposition, et al.) is also worth serious investigation. In the near future, we will extend this method for cardiac CT reconstruction and systematically evaluate the imaging performance through extensive numerical simulation and preclinical/clinical experiments.

Acknowledgment

This work was partially supported by the NSF CAREER Award CBET-1149679, the NSF collaborative grant DMS-1210967, and NIH/NIBIB U01 grant (EB011740).

References

1. Gao H, Cai J-F, Shen Z, Zhao H. Robust principal component analysis-based four-dimensional computed tomography. *Phys. Med. Biol.* 2011 Jun.56(11):3181–3198. [PubMed: 21540490]
2. Cai J-F, Jia X, Gao H, Jiang SB, Shen Z, Zhao H. Cine cone beam CT reconstruction using low-rank matrix factorization: algorithm and a proof-of-principle study. *ArXiv12043595 Phys.* 2012 Apr.
3. Caballero J, Price AN, Rueckert D, Hajnal JV. Dictionary Learning and Time Sparsity for Dynamic MR Data Reconstruction. *IEEE Trans. Med. Imaging.* 2014 Apr.33(4):979–994. [PubMed: 24710166]
4. Wang Y, Ying L. Compressed Sensing Dynamic Cardiac Cine MRI Using Learned Spatiotemporal Dictionary. *IEEE Trans. Biomed. Eng.* 2014 Apr.61(4):1109–1120. [PubMed: 24658236]
5. Donoho DL. Compressed sensing. *IEEE Trans. Inf. Theory.* 2006 Apr.52(4):1289–1306.
6. Candes EJ, Romberg J, Tao T. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inf. Theory.* 2006 Feb.52(2):489–509.
7. Yu H, Wang G. Compressed sensing based interior tomography. *Phys. Med. Biol.* 2009 May.54(9):2791. [PubMed: 19369711]
8. Yu H, Yang J, Jiang M, Wang G. Supplemental Analysis on Compressed Sensing Based Interior Tomography. *Phys. Med. Biol.* 2009 Sep.54(18):N425–N432. [PubMed: 19717891]
9. Xu Q, Mou X, Wang G, Sieren J, Hoffman EA, Yu H. Statistical Interior Tomography. *IEEE Trans. Med. Imaging.* 2011 May; 30(5):1116–1128. [PubMed: 21233044]
10. Ritschl L, Bergner F, Fleischmann C, Kachelrieß M. Improved total variation-based CT image reconstruction applied to clinical data. *Phys. Med. Biol.* 2011 Mar.56(6):1545. [PubMed: 21325707]
11. Xu Q, Yu H, Mou X, Zhang L, Hsieh J, Wang G. Low-Dose X-ray CT Reconstruction via Dictionary Learning. *IEEE Trans. Med. Imaging.* 2012 Sep.31(9):1682–1697. [PubMed: 22542666]
12. Herman GT, Davidi R. Image reconstruction from a small number of projections. *Inverse Probl.* 2008 Aug.24(4):045011.
13. Elad M, Aharon M. Image Denoising Via Sparse and Redundant Representations Over Learned Dictionaries. *IEEE Trans. Image Process.* 2006 Dec.15(12):3736–3745. [PubMed: 17153947]
14. Mairal J, Elad M, Sapiro G. Sparse Representation for Color Image Restoration. *IEEE Trans. Image Process.* 2008 Jan.17(1):53–69. [PubMed: 18229804]
15. Wright J, Yang AY, Ganesh A, Sastry SS, Ma Y. Robust face recognition via sparse representation. *IEEE Trans. Pattern Anal. Mach. Intell.* 2009 Feb.31(2):210–227. [PubMed: 19110489]
16. Ravishanker S, Bresler Y. MR image reconstruction from highly undersampled k-space data by dictionary learning. *IEEE Trans. Med. Imaging.* 2011 May; 30(5):1028–1041. [PubMed: 21047708]
17. Engan K, Aase SO, Hakon Husoy J. Method of optimal directions for frame design. 1999 IEEE International Conference on Acoustics, Speech, and Signal Processing, 1999. Proceedings. 1999; 5:2443–2446. vol.5.
18. Aharon M, Elad M, Bruckstein A. K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation. *IEEE Trans. Signal Process.* 2006 Nov.54(11):4311–4322.
19. Jia X, Tian Z, Lou Y, Sonke J-J, Jiang SB. Four-dimensional cone beam CT reconstruction and enhancement using a temporal nonlocal means method. *Med. Phys.* 2012 Sep.39(9):5592–5602. [PubMed: 22957625]
20. Wu G, Wang Q, Lian J, Shen D. Estimating the 4D respiratory lung motion by spatiotemporal registration and super-resolution image reconstruction. *Med. Phys.* 2013 Mar.40(3):031710. [PubMed: 23464305]
21. Caiafa CF, Cichocki A. Multidimensional compressed sensing and their applications. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* 2013 Nov.3(6):355–380.

22. Duan, G.; Wang, H.; Liu, Z.; Deng, J.; Chen, Y-W. K-CPD: Learning of overcomplete dictionaries for tensor sparse coding; 2012 21st International Conference on Pattern Recognition (ICPR); 2012. p. 493-496.
23. Zubair, S.; Wang, W. Tensor dictionary learning with sparse TUCKER decomposition; 2013 18th International Conference on Digital Signal Processing (DSP); 2013. p. 1-6.
24. Peng Y, Meng D, Xu Z, Gao C, Yang Y, Zhang B. Decomposable Nonlocal Tensor Dictionary Learning for Multispectral Image Denoising. CVPR. 2014
25. Li L, Chen Z, Wang G, Chu J, Gao H. A tensor PRISM algorithm for multi-energy CT reconstruction and comparative studies. J. X-Ray Sci. Technol. 2014 Jan.22(2):147-163.
26. Chen G-H, Tang J, Leng S. Prior image constrained compressed sensing (PICCS): a method to accurately reconstruct dynamic CT images from highly undersampled projection data sets. Med. Phys. 2008 Feb.35(2):660-663. [PubMed: 18383687]
27. Gao H, Yu H, Osher S, Wang G. Multi-energy CT based on a prior rank, intensity and sparsity model (PRISM). Inverse Probl. 2011 Nov.27(11):115012. [PubMed: 22223929]
28. Gao H, Li R, Lin Y, Xing L. 4D cone beam CT via spatiotemporal tensor framelet. Med. Phys. 2012 Nov.39(11):6943-6946. [PubMed: 23127087]
29. Zhou W, Cai J-F, Gao H. Adaptive tight frame based medical image reconstruction: a proof-of-concept study for computed tomography. Inverse Probl. 2013 Dec.29(12):125006.
30. Römer F, Galdo GD. Tensor-Based Dictionary Learning for Multidimensional Sparse Recovery. [Online]. Available: http://workshops.fhr.fraunhofer.de/cosera/pdf/A1_3t.pdf.
31. Goldstein T, Osher S. The Split Bregman Method for L1-Regularized Problems. SIAM J. Imaging Sci. 2009 Jan.2(2):323-343.
32. Liu Q, Liang D, Song Y, Luo J, Zhu Y, Li W. Augmented Lagrangian-Based Sparse Representation Method with Dictionary Updating for Image Deblurring. SIAM J. Imaging Sci. 2013 Jan.6(3):1689-1718.
33. Buades A, Coll B, Morel J. A Review of Image Denoising Algorithms, with a New One. Multiscale Model. Simul. 2005 Jan.4(2):490-530.
34. Zhang X, Burger M, Bresson X, Osher S. Bregmanized Nonlocal Regularization for Deconvolution and Sparse Reconstruction. SIAM J. Imaging Sci. 2010 Jan.3(3):253-276.
35. Jung C, Ju J, Jiao L, Yang Y. Enhancing dictionary-based super-resolution using nonlocal total variation regularization. Opt. Eng. 2013; 52(1):017005-017005.
36. Yang Z, Jacob M. Nonlocal Regularization of Inverse Problems: A Unified Variational Framework. IEEE Trans. Image Process. 2013 Aug.22(8):3192-3203. [PubMed: 23014745]
37. Kolda T, Bader B. Tensor Decompositions and Applications. SIAM Rev. 2009 Aug.51(3):455-500.
38. Bader, BW.; Kolda, TG. MATLAB Tensor Toolbox Version 2.5. 2012. [Online]. Available: <http://www.sandia.gov/~tgkolda/TensorToolbox/>.
39. Bader BW, Kolda TG. Efficient MATLAB Computations with Sparse and Factored Tensors. SIAM J. Sci. Comput. 2008 Jan.30(1):205-231.
40. Caiafa CF, Cichocki A. Computing Sparse Representations of Multidimensional Signals Using Kronecker Bases. Neural Comput. 2012 Sep.25(1):186-220. [PubMed: 23020110]
41. Boyd S, Parikh N, Chu E, Peleato B, Eckstein J. Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. Found Trends Mach Learn. 2011 Jan.3(1):1-122.
42. Elbakri IA, Fessler JA. Statistical image reconstruction for polyenergetic X-ray computed tomography. IEEE Trans. Med. Imaging. 2002 Feb.21(2):89-99. [PubMed: 11929108]
43. Yu H, Zhao S, Hoffman EA, Wang G. Ultra-low Dose Lung CT Perfusion Regularized by a Previous Scan. Acad. Radiol. 2009 Mar.16(3):363-373. [PubMed: 19201366]
44. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. 2004 Apr.13(4):600-612. [PubMed: 15376593]
45. Zhang L, Zhang D, Mou X, Zhang D. FSIM: A Feature Similarity Index for Image Quality Assessment. IEEE Trans. Image Process. 2011 Aug.20(8):2378-2386. [PubMed: 21292594]

46. Gilboa G, Osher S. Nonlocal Operators with Applications to Image Processing. *Multiscale Model. Simul.* 2008 Nov.7(3):1005–1028.
47. Buades A, Coll B, Morel J-M. A non-local algorithm for image denoising. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005. CVPR 2005.* 2005; 2:60–65. vol. 2.
48. Yan R, Shao L, Liu Y. Nonlocal Hierarchical Dictionary Learning Using Wavelets for Image Denoising. *IEEE Trans. Image Process.* 2013 Dec.22(12):4689–4698. [PubMed: 23955752]

Fig. 1.

Tensor dictionary learning based reconstruction results in the sheep lung perfusion study. (a) The reference image reconstructed from a full dataset (1160 views) in a display window $[-700 \text{ HU}, 800 \text{ HU}]$; (b) with respect to that reference the difference image between the absolute errors of the vector and tensor dictionary learning methods from 116 down-sampled projections in a display window $[-5 \text{ HU}, 5 \text{ HU}]$, where the white regions show greater reconstruction errors for the vector dictionary learning method; and (c) the counterpart of (b) for 232 projections.

Fig. 2. Quantitative evaluation in the sheep lung perfusion study in terms of (a) RMSE and (b) SSIM, where the horizontal axis indicates the phase of the dynamic reconstruction.

(a)

(b)

Fig. 3.

Representative results in the sheep lung perfusion study for the phase 10. (a) The difference image between the absolute errors (relative to the reconstruction from the full dataset) of the tensor DL and tensor DL NLTV methods from 116 down-sampled projections in a display window $[-5\text{HU}, 5\text{HU}]$, where the tensor DL NLTV method produced more details in the white regions; and (b) the counterpart of (a) for 232 projections.

Fig. 4. Quantitative evaluation for the tensor method with/without NLTV in terms of (a) RMSE, and (b) SSIM in the sheep lung perfusion study, where the horizontal axis indicates the phase of the dynamic reconstruction.

Fig. 5. The advantages of tensor based DL and NLTV are naturally combined into a new framework to improve the reconstructed images eventually
Reconstructed images at the phase 10 with 40dB white Gaussian noise. (a) The reference image reconstructed from the full dataset (1160 views), and (b)–(d) the reconstructed images using the vector DL, tensor DL and tensor DL NLTV methods from 116 down-sampled projections, respectively. The display window is $[-700 \text{ HU}, 800 \text{ HU}]$.

Fig. 6.
Magnified portions in the corresponding rectangular regions in Figure 5 in a display window of $[-650\text{HU}, 600\text{HU}]$.

Fig. 7. Quantitative evaluation results in the sheep lung perfusion study with additive noise in terms of RMSE and SSIM. The projection SNRs for the left and right columns are 25dB and 40dB, respectively.

Fig. 8. Reconstructed images of the mouse cardiac imaging study (phase 1). From left to right, the images were reconstructed by the FBP, vector based DL, tensor based DL and tensor based DL NLTV, respectively. From top to bottom, the imagers were reconstructed from projection data down-sampled to 100, 200 and 400 views. The display window is $[-600 \text{ HU}, 1500 \text{ HU}]$.

Fig. 9.

Reconstructed images of the mouse cardiac CT (phase 2). From left to right, the images were reconstructed by the FBP, vector based DL, tensor based DL and tensor based DL NLTV, respectively. The original projections were down-sampled to 200 views and the display window is $[-600 \text{ HU}, 1500 \text{ HU}]$.