Volumetric Computed Tomography Reconstruction with Dictionary Learning

Ti Bai, Hao Yan, Xun Jia, Steve Jiang, Ge Wang, Xi Chen and Xuanqin Mou

***Abstract*—Despite the rapid developments of x-ray cone-beam CT (CBCT), image noise still remains a major issue for the low-dose CBCT. Iterative reconstruction algorithms with 2D dictionary learning (DL) were validated for fine structures and suppressed noise in the case of low-dose CT reconstruction. However, an enhanced version for volumetric CBCT is absent. Besides, it is recognized that representation efficiencies of the sparsity-promotion regularizers are of primary importance for the success of the image processing tasks. In this work, a sparse constraint based on the 3D dictionary is incorporated into a statistical iterative reconstruction, defining the 3D-DL reconstruction framework. From a statistical viewpoint, the distributions of the representation coefficients associated with the 2D/3D dictionaries are analyzed to compare their efficiencies in representing volumetric images. The whole program is implemented on graphic processor units (GPU) to boost the computation efficiency. Experiments demonstrated that the 3D dictionary allows a much higher representation efficiency and a better image quality compared to the 2D dictionary case. Regarding the tested radiation therapy datasets, with a volume of $512 \times 512 \times 512$ and a projection dataset of $512 \times 384 \times 363$, the whole reconstruction process can be finished within 5 minutes.**

***Index Terms*—Dictionary learning, sparse representation, low-dose CT, cone-beam CT, noise suppression, GPU**

I. Introduction

As a powerful tool to visualize internal structures of an object in a non-invasive fashion, x-ray computed tomography (CT) has experienced rapid developments over the past decades. Due to volumetric coverage during fast gantry rotation, isotropic spatial resolution and high tube efficacy, cone-beam CT (CBCT)[1] plays an important role in many scenarios, such as patient setup in radiation therapy[2], intra-operative imaging[3], [4], and maxillofacial visualization[5]. In preclinical research, micro-CBCT is often used to map organs of small animals[6].

Despite the aforementioned applications, there are increased demands for the radiation dose reduction as low as reasonably achievable (ALARA)[7], [8]. Basically, low-dose CBCT can be achieved by either collecting fewer projections (few-view protocol) or reducing the exposure level (low-exposure protocol). In this work, we will focus on the low-exposure protocol, as it can be simply implemented by reducing the tube current and is advantageous sampling-wise. Low-exposure protocol, however, would inevitably result in noisy projection data, and data noise would be propagated into reconstructed images, possibly rendering the images less useful or useless.

A great effort has been devoted to image noise reduction. Specifically, by accommodating the measurement statistics, modeling the data acquisition geometry, and enforcing physical constraints, the regularized iterative reconstruction algorithms often produce superior image quality with low SNR measurements, and hence having become increasingly popular. In the context of iterative reconstruction, an appropriate physical constraint about the underlying image, i.e., the regularizer, is regarded as being of primary importance (e.g., references[9], [10], [11], [12], [13], [14]). Thanks to the rapid development of compressive sensing theory[15], the sparsity-promotion regularizers have been successful, most of which could be applied to both the few-view and low-exposure protocols. For example, Yu *et al.* and Sidky *et al.* proposed an iterative reconstruction algorithm by minimizing the total variation (TV) of the image[16], [17]. Provided a high-quality image which resembles the image under reconstruction, Chen *et al.* developed a method referred to as prior image constrained compressive sensing (PICCS) for accurate reconstruction of dynamic CT images[11]. In the low-dose CBCT domain, many iterative reconstruction algorithms were published, with an emphasis on the design of the regularizer. Sidky *et al.* developed a 3D-TV minimization method for volumetric image reconstruction from a circular CBCT scan[10]. Jia *et al.* constructed an iterative CBCT reconstruction framework regularized by the tight frame (TF) based sparse representation[18], attaining competitive performance, if not better, compared to the TV minimization method.

Recently, learning based image processing techniques gained significant interest, with a primary example known as dictionary learning. The basic idea is a well-accepted assumption that in the natural scene images, there exist abundant structures which could be sparsely represented with a redundant dictionary. The dictionary can be learned from images sharing similar spatial structures. This property was helpful in the image restoration tasks, such as denoising, super-resolution, and deblurring[19]. Inspired by the successes, the dictionary learning based image restoration techniques were introduced for low-dose CT imaging. For example, Xu *et al.* incorporated a dictionary learning based sparse constraint into the statistical x-ray CT iterative reconstruction framework. It enhances the image quality such that image noise is effectively reduced while subtle structures are well retained[12]. Li *et al.* combined the dictionary learning based sparse constraint and the TV minimization based constraint together to facilitate dual-energy CT reconstruction[20].

Currently, most of the dictionary learning based sparse representation techniques are for 2D cases. Intuitively, 3D structures in volumetric images should be directly targeted by training a 3D dictionary, which consists of 3D atoms. In this paper, we report a 3D dictionary learning (3D-DL) based reconstruction framework for low-dose volumetric CT

This work is partly supported by the National Natural Science Foundation of China (NSFC) (No. 61571359), the National Key Research and Development Program of China (No. 2016YFA0202003) and the National Natural Science Foundation of China (NSFC) (No. 61401349).

Ti Bai, Xi Chen and Xuanqin Mou are with the Institute of Image processing and Pattern recognition, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China (e-mail: xi_chen@mail.xjtu.edu.cn, xqmou@mail.xjtu.edu.cn)

Hao Yan is with Cyber Medical Corporation, Shannxi 710049, China (e-mail: hao.yan@cybermedical.com)

Xun Jia and Steve Jiang are with Department of Radiation Oncology, UT Southwestern Medical Center, Dallas, TX 75390, USA (e-mail: Xun.Jia@utsouthwestern.edu, Steve.Jiang@utsouthwestern.edu)

Ge Wang is with Biomedical Imaging Center/Cluster, CBIS, Dept. of BME, Rensselaer Polytechnic Institute, Troy, NY, 12180 (email: wangg6@rpi.edu))

DOI: 10.12059/Fully3D.2017-11-3201024

and perform a systematic investigation on this new technique. First, based on two realistic experiments, we qualitatively and quantitatively compare the performance of the proposed 3D-DL method to other competing algorithms, such as the previous 2D dictionary learning method and the TF method. In the field of natural scene statistics, it is recognized that the sparser (more efficient) an image could be represented with a group of basis (e.g., the dictionary in our context), the more efficient the later processing stages could be (e.g., deliver higher-quality reconstructions in our context)[21], [22], [23]. Motivated by this knowledge, we then evaluate the representation efficiencies of the 2D/3D dictionaries for structures in volumetric images. Finally, to make the proposed method clinically practical, we parallelize the whole program on the GPU using several algorithmic tricks.

II. METHODS AND MATERIALS

A. Formulation

For completeness, we will first review the iterative x-ray CT reconstruction and introduce related notations. For more details, please refer to the references[12], [24]. Basically, the objective of an image reconstruction is to find the unknown true image $\hat{\mathbf{x}} \in \mathbb{R}^{N\times 1}$ from observed measurements $\mathbf{y} \in \mathbb{R}^{M\times 1}$ (the transmission data through log transform) defined by $\mathbf{y} = \mathbf{A}\hat{\mathbf{x}} + \epsilon$, where $\mathbf{A} \in \mathbb{R}^{M\times N}$ is the system matrix, $\epsilon \in \mathbb{R}^{M\times 1}$ denotes the measured noise which can be modeled as a zero-mean Gaussian distribution with respect to ray-dependent variances[25], and M and N are integers. By incorporating some physical constraints, the regularized statistical iterative reconstruction is formulated as

$$\min_{\mathbf{x}} ||\mathbf{A}\mathbf{x} - \mathbf{y}||_{\mathbf{W}}^2 + \beta \mathbf{R}(\mathbf{x}), \quad (1)$$

where $||\mathbf{u}||_{\mathbf{W}}^2 = \mathbf{u}^{\mathrm{T}}\mathbf{W}\mathbf{u}$ and $\mathbf{W} = \mathrm{diag}\{w_{ii}\} \in \mathbb{R}^{M\times M}$ is a diagonal matrix consisting of the statistical weights that are inversely proportional to the measurements variances[25], β denotes the regularization parameter controlling the relative weight between the fidelity term $||\mathbf{A}\mathbf{x} - \mathbf{y}||_{\mathbf{W}}^2$ and the regularization term $\mathbf{R}(\mathbf{x})$.

Specifically, the dictionary learning based statistical iterative reconstruction can be written as

$$\min_{\mathbf{x}} ||\mathbf{A}\mathbf{x} - \mathbf{y}||_{\mathbf{W}}^2 + \beta \sum_s (||\mathbf{E}_s\mathbf{x} - \mathbf{D}\alpha_s||_2^2 + \gamma ||\alpha_s||_p), \quad (2)$$

where $\mathbf{E}_s$ denotes an extraction operator for the s^{th} data block which can be sparsely represented by a learned dictionary $\mathbf{D}$, and the associated coefficients are α_s, γ is the Lagrangian multiplier, and $||\alpha_s||_p$ denotes the ℓ_p norm of α_s with $p \geq 0$.

In Eq. (2), the sparse constraint $||\alpha_s||_p$ can be enforced in several ways. A direct way is to minimize the quasi-ℓ_0-norm of the sparse coefficients with $p = 0$, that is, the number of nonzero elements. An alternative way is to choose $p = 1$, so as to minimize the ℓ_1-norm which is a best convex approximation of the quasi-ℓ_0-norm. Indeed, both the quasi-ℓ_0-norm and the ℓ_1-norm have been widely employed, and they result in comparable image qualities[12], [26], [27]. Particularly, by setting $p = 1$, Eq. (2) can be readily understood in a Bayesian viewpoint, where the sparse coefficients are supposed to be modeled with a zero-mean Laplacian distribution, which is highly peaked around zero with heavy tails[28].

B. 3D dictionary

Equation (2) is the general framework for 2D/3D dictionary learning based statistical iterative reconstruction. When $\mathbf{y}$ is the sinogram of a 2D CT slice image, $\mathbf{D} \in \mathbb{R}^{K\times L}$ is a 2D dictionary consisting of L 2D atoms of $P\times Q$ with $K = P\times Q$. Similarly, when $\mathbf{y}$ is composing of various 2D projections with respect to a volumetric image, $\mathbf{D} \in \mathbb{R}^{K\times L}$ is a 3D dictionary comprising of L 3D atoms of $P \times Q \times R$ with $K = P \times Q \times R$, which represent the local structures of the volumetric image. Note that each column of the dictionary $\mathbf{D}$ is a vector rearranging from the corresponding 2D/3D atom. A simple illustration about the 2D/3D dictionary learning based sparse representation is given in Fig. (1).

C. Optimization algorithm

An alternating minimization scheme is employed to solve Eq. (2), where $p = 0$ is adopted in this work. It allows us to obtain the solution by alternatingly solving the following two sub-problems:

$$\min_{\mathbf{x}} ||\mathbf{A}\mathbf{x} - \mathbf{y}||_{\mathbf{W}}^2 + \beta \sum_s (||\mathbf{E}_s\mathbf{x} - \mathbf{D}\alpha_s||_2^2), \quad (3)$$

$$\min_{\alpha_s} ||\mathbf{x}_s - \mathbf{D}\alpha_s||_2^2 + \xi ||\alpha_s||_0. \quad (4)$$

The sub-problem (3) is of simple quadratic form, which can be optimized by the OS based separable quadratic surrogate (OS-SQS) method[24]:

$$\mathbf{x}^{j+1} = \mathbf{x}^j - \frac{V\mathbf{A}_C^{\mathrm{T}}\mathbf{W}_C(\mathbf{A}_C\mathbf{x}^j - \mathbf{y}_C) + \beta\sum_s \mathbf{E}_s^{\mathrm{T}}(\mathbf{E}_s\mathbf{x}^j - \mathbf{D}\alpha_s)}{\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A}\mathbf{I} + \beta\sum_s \mathbf{E}_s^{\mathrm{T}}\mathbf{E}_s\mathbf{I}}, \quad (5)$$

where V is the number of the subsets, subscript C denotes one subset of the projections, superscript j denotes the j^{th} iteration, $\mathbf{I}$ is the unity vector, T denotes the transpose operator. Physically speaking, the second term in the numerator of Eq. (5) means that given any extracted data block $\mathbf{E}_s\mathbf{x}^j$, find its sparse representation and then put the denoised version back to serve as the prior information. To accelerate convergence of the above algorithm, two additional algorithmic tricks are employed, i.e., Nesterov's weighting strategy[29] and double surrogates strategy[30]. The Nesterov's weighting strategy could achieve a nearly optimal convergence rate for the first-order methods by utilizing the previous iterations to adjust the current update direction. The double surrogates strategy allows us to update the gradient of the regularization term less frequently compared to the number of subsets, and hence reduce the computational cost per iteration.

To optimize the sub-problem (4), we rewrite it as the following equivalent problem:

$$\min_{\alpha_s} ||\alpha_s||_0, \mathrm{s.t.}, ||\mathbf{E}_s\mathbf{x} - \mathbf{D}\alpha_s||_2^2 \leq \epsilon. \quad (6)$$

Problem (6) is a typical sparse coding task, and a lot of solvers have been proposed[31], [32], [33]. Here, a Cholesky decomposition based orthogonal matching pursuit (OMP) algorithm is employed[34].

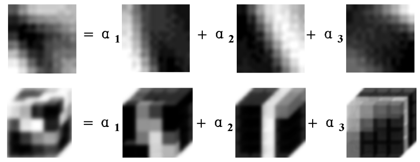

Fig. 1. Illustration of 2D (first row) and 3D (second row) dictionary learning based sparse representations.

D. Experiments

To evaluate the 3D-DL method, clinical applications in radiation therapy are carried out. We first perform comparison studies among different regularizers. Then, a statistical analysis is conducted for the representation efficiencies of the 2D/3D dictionaries on the structures in the CBCT images. Finally, we conduct the computational cost and convergence analysis for the whole program. All the algorithms are implemented in the CUDA 7.0 programming environment on a NVIDIA GeForce GTX 980 video card which is installed on a personal computer (Intel i5-4460 CPU and 8GB RAM).

1) Experimental data: Two realistic datasets were collected including one head-neck (HN) full-fan scan and one prostate half-fan scan in the image guided radiation therapy (IGRT). The HN and prostate patient datasets were collected from an on board imager integrated in a TrueBeam medical accelerator (Varian Medical System, Palo Alto, CA), where the source-to-axis and source-to-detector distances are $1000\mathrm{mm}$ and $1500\mathrm{mm}$, respectively. The collected projection data were deidentified in the Radon space. They were rebinned using a 2×2 mode, resulting an imager of 512×384 with a detector size of $0.776 \times 0.776\mathrm{mm}^2$. To be specific, the HN patient was scanned in a full-fan mode to acquire 363 projections in a 200 degrees arc, the exposure level was 0.4 mAs per projection. The prostate patient was scanned in a half-fan mode with a $160\mathrm{mm}$ lateral shift, acquiring 656 projections in 360 degrees with an exposure level of 1.25 mAs per projection. To evaluate the potential for dose reduction, Poisson noise with levels of 1×10^4 and 3×10^4 photon incidents per ray were superimposed into the above datasets. This generated low-dose HN and prostate patient cases, respectively. The reconstructed images were of $512 \times 512 \times 512$ and $512 \times 512 \times 256$ with voxel sizes of $0.6 \times 0.6 \times 0.6\mathrm{mm}^3$ and $1 \times 1 \times 1\mathrm{mm}^3$, respectively.

2) Comparison studies among different regularizers: In this study, we qualitatively and quantitatively compare the proposed 3D-DL method with two existing methods, namely, 2D dictionary learning based method [12] and TF method[18]. Specifically, for the 2D dictionary learning based method, three 2D dictionary learning based sparse constraints are consecutively applied for the transversal, coronal and sagittal views for a fair comparison, though this strategy suffers from heavy computation burden. Furthermore, to illustrate the inherent 2D property of the 2D dictionary learning based method for volumetric CT reconstruction, three additional strategies are employed to utilize the 2D dictionary learning based regularizer in a slicewise fashion for each of the three different views. The employed 2D dictionary contains 256 atoms of 8×8 trained from the planning volumetric CT image of a third prostate patient. For a fair comparison, in the 3D-DL method, the employed 3D dictionary also contains 256 3D atoms of $4 \times 4 \times 4$ trained from the same sample source. For simplicity, we will use the following abbreviations to represent the different methods:

3D-DL: the proposed 3D-DL method
2D-DL: three 2D dictionary learning based sparse constraints are consecutively enforced on all the three views
TF: the TF method
881-DL: the 2D dictionary learning based sparse constraint is only enforced on the transversal view in a slicewise method
818-DL: the 2D dictionary learning based sparse constraint is only enforced on the coronal view in a slicewise method
188-DL: the 2D dictionary learning based sparse constraint is only enforced on the sagittal view in a slicewise method

The OS acceleration technique is applied on all the cases by evenly distributed the datasets into several subsets. To be specific, 11 subsets with 33 projections per subset, denoted as $11(\times 33)$ OS protocol hereafter, is used for the HN patient case. $8(\times 82)$ OS protocols is used for the prostate patient case. The whole optimization program is terminated after 10 iterations for both cases. Both datasets are also reconstructed by the FDK algorithm to benchmark the regularized iterative reconstructions.

For both cases, because the regular dose reconstructions are available, the results are quantified with the root mean squared errors (RMSE) and the structure similarity (SSIM) index (the closer to 1, the better the image is)[35]. The RMSE is calculated based on the whole 3D volume. Because the SSIM index is devised for the 2D images, in this work, the SSIM index is calculated based on a single presented transversal slice.

3) Analysis for the representation efficiencies of the 2D/3D dictionaries: In the field of neuroscience, it has been found that the receptive fields of the simple cells in human primary visual cortex can be characterized as being spatially localized, oriented and bandpass[36]. These are consistent with the statistical structures of the natural scene images if they are efficiently coded[37]. Olshausen *et al.* suggested that a learning algorithm attempting to maximize the sparseness (and hence maximize the representation efficiency) of the representation coefficients can produce a complete set for these localized, oriented and bandpass structures[21]. The philosophy behind this conjecture is that all the trained patterns should be able to represent the structures of the natural scene images as efficiently as possible by reducing the statistical dependences. That is, the sparser the representation coefficients are, the more the trained patterns like the characteristics of the receptive fields[22], [23]. Moreover, the resultant sparser coefficients of the image can facilitate the later processing stages more efficiently.

Regarding these facts, it is important to explore the representation efficiencies of the 2D/3D dictionaries for the structures of the volumetric images. The result can be used to justify the 3D-DL method in a natural scene statistical viewpoint. Specifically, it is well-accepted that the representation coefficients of the natural scene images shall follow a zero-mean Laplacian distribution that is highly peaked around zero with heavy tails[28]. Therefore, we conduct the representation efficiency analysis by investigating the resultant Laplacian distributions. It is expected that the narrower the Laplacian distribution is, the more efficient the trained dictionary can represent the structures, and hence facilitate the later processing stages better, such as the reconstruction task in our context.

In this study, the regular dose FDK reconstructions of the HN patient and prostate patient are employed to explore the representation efficiencies. In details, for the HN patient case, a group of 1×10^5 data samples of $4 \times 4 \times 4$ are randomly extracted from the regular dose FDK reconstruction. Then, three groups of 1×10^5 2D data samples of $8 \times 8 \times 1$, $8 \times 1 \times 8$ and $1 \times 8 \times 8$ are randomly extracted from the transversal, coronal and sagittal views, respectively. The same rules are also applied on the prostate patient case. The above data samples are fed into the following equation to obtain the corresponding representation coefficients

$$\min_{\alpha} ||\mathbf{X} - \mathbf{D}\alpha||_2^2 + \gamma||\alpha||_1, \tag{7}$$

where each column of $\mathbf{X}$ is a vector stacking representation of the 2D/3D data sample, $\mathbf{D}$ is the 2D/3D dictionary, α denotes the resultant representation coefficients, and γ is a Lagrangian multiplier. Equation (7) is solved with the SPAMS software package in this work. We will analyze the distributions of the resultant coefficients and calculate the associated variance of each distribution to quantify the representation efficiencies.

4) Computational cost and convergence analysis: One of the main drawbacks of an iterative reconstruction algorithm is the heavy computational cost. The computationally intensive sparse coding stage in the dictionary learning based methods aggravates this situation further. Without loss of generality, we will take the HN and prostate patient cases as examples to conduct the computational cost and convergence analysis. It is believed that similar observations can be also achieved from the other cases. The time overhead is calculated as the averaged time consumption of the regularization term update among all the 10 iterations, i.e., the averaged time consumption for all the extracted data blocks to be sparsely coded. In addition, to investigate the computational cost of the 3D-DL method, we will also conduct computational cost analysis in terms of the data fidelity term update and the regularization term update, and compare with the other involved methods in this work. The computational costs for these two parts are calculated as the averaged time consumptions among all the 10 iterations. Lastly, to experimentally study the convergence property of the 3D-DL method, we will plot the curves of the change amount of the images between consecutive iterations vs the iteration number, i.e., $\Delta\mathbf{x} = ||\mathbf{x}^{j+1} - \mathbf{x}^j||_2^2$ vs j, where j is the iteration number. Generally speaking, if $\Delta\mathbf{x}$ is smaller than a pre-specified tiny tolerance, then the algorithm could be regarded as converged.

III. RESULTS

Figure (2) demonstrates the transversal view of the reconstructed images for the HN patient case. It is observed that the low-dose FDK reconstruction is overwhelmed by the noise, while the noise is substantially suppressed by the regularized iterative reconstruction algorithms. As indicated by the zoomed-in ROIs in Fig. (2), if the 2D dictionary learning based sparse constraint is enforced on only one view, the unprocessed views exhibit directional streak artifacts, such as the horizontal and vertical streak artifacts corresponding to the 818-DL and 188-DL sub-figures, respectively. Regarding the processed views, the structures are distorted. For example, as indicated by the arrow in the 881-DL zoomed-in ROI of Fig. (2), part of the soft bone is missing compared to the regular dose FDK reconstruction. The directional streak artifacts can be alleviated if the 2D dictionary learning based sparse constraint is enforced on all the three views consecutively, i.e., the 2D-DL method. However, compared to the 3D-DL method, the 2D-DL method exhibits lower spatial resolution and higher image noise, as indicated by the zoomed-in ROIs in Fig. (2). Another disadvantage of the 2D-DL method is the inherent high computational cost, considering the fact that three individual sparse coding stages are required for regularization. Regarding the TF method, one cannot well distinguish the subtle structures which are blurred due to the reduced resolution and the remained noise in the high contrast region, as indicated by the TF sub-figure in Fig. (2). On the other hand, it can be seen that the 3D-DL method achieves promising results in enhancing the anatomical structures and in removing the noise effectively, and hence validates its efficacy. Quantitatively, with the regular dose FDK reconstruction as the reference, the RMSE and SSIM values are listed in Table I. The lowest RMSE and the highest SSIM further verify that the 3D-DL method outperforms other algorithms.

The transversal views of the reconstructed images corresponding to the prostate patient case are in Fig. (3). Figure (4) demonstrates the zoomed-in ROIs for the reconstructions in Fig. (3) with respect to different methods. It is shown that the 3D-DL method efficiently suppresses noise and well retains anatomical structures both for the low contrast and high contrast regions. Regarding the TF method, the structures are contaminated by the remained pepper-like noise, as shown by the TF sub-figure in Figs. (3) and (4). From Fig. (3), it can be seen that the 2D-DL method exhibits stronger noise with comparable resolution, if not inferior, compared to the 3D-DL method, this phenomenon could be more clearly observed from the zoomed-in ROIs in Fig. (4), as indicated by the arrow in the 2D-DL sub-figure. On the other hand, directional streak artifacts are observed from the reconstructions corresponding to the 818-DL/188-DL methods (such as the 818-DL/188-DL zoomed-in ROIs in Fig. (4)), while the structures are distorted for the processed view (as indicated by the arrow in the 881-DL zoomed-in ROI of Fig. (4)). One also could

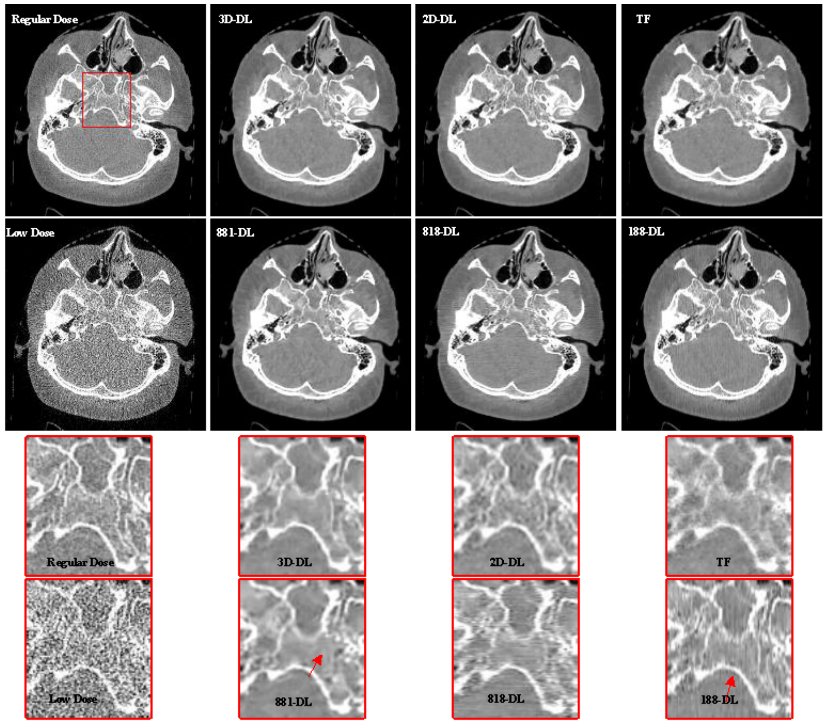

Fig. 2. Transversal views of the HN patient images reconstructed by different methods. From left to right in the first row, the images are regular dose FDK reconstruction, reconstructions from the 3D-DL, 2D-DL and TF methods, respectively. From left to right in the second row, the images are reconstructed by the FDK, 881-DL, 818-DL and 188-DL methods, respectively. The last two rows show the corresponding zoomed-in ROIs of the red box in the first two rows. The display window is [-750 750] HU.

TABLE I
THE RMSE (UNIT: HU) AND SSIM VALUES OF DIFFERENT METHODS FOR THE HN PATIENT CASE.

Methods	3D-DL	2D-DL	TF	881-DL	818-DL	188-DL	Low Dose	Regular Dose
RMSE	**71.97**	74.78	82.68	79.29	78.99	76.61	193.61	0
SSIM	**0.8756**	0.8709	0.8587	0.8628	0.8621	0.8656	0.6308	1

TABLE II
THE RMSE (UNIT: HU) AND SSIM VALUES OF DIFFERENT METHODS FOR THE PROSTATE PATIENT CASE.

Methods	3D-DL	2D-DL	TF	881-DL	818-DL	188-DL	Low Dose	Regular Dose
RMSE	**22.57**	25.54	23.46	27.35	25.68	25.09	68.78	0
SSIM	**0.9840**	0.9814	0.9808	0.9798	0.9790	0.9795	0.8378	1

TABLE III
VARIANCES OF THE DISTRIBUTIONS AMONG DIFFERENT DICTIONARIES AND DIFFERENT CASES.

	3D-444	2D-881	2D-818	2D-188
HN	0.0005	0.1643	0.1674	0.1611
Prostate	0.0002	0.1541	0.1524	0.1469

find that the 881-DL result is less sharp compared to the 3D-DL result. The calculated RMSE/SSIM values are in Table II. As expected, the 3D-DL method quantitatively outperforms the other competitors in terms of the lowest RMSE and the highest SSIM measures, which are consistent with the visual observations that the 3D-DL method leads to more naturally and visually pleasant denoising results by better preserving the image texture areas.

Figure (5) plots the distributions of the representation coefficients among different dictionaries. Note that the y-axis is set to be the logarithmic probabilities to show the Laplacian nature of these probability distributions more clearly[28]. It is obvious that for both cases, the 3D dictionary based sparse coefficients consistently have much narrower Laplacian distributions compared to the 2D dictionary based sparse coefficients, suggesting the higher representation efficiencies of the 3D dictionary. The variances of these distributions are summarized in Table III. It is shown that the variances associated with the 3D dictionary are much smaller compared to those with respect to the 2D dictionaries.

Table IV lists the time overheads of the fidelity term update and the regularization term update for all the methods considered in this work. The time overhead of the fidelity term update is the total time required by all the subsets. It is observed that the time consumption for the fidelity term update is quite stable among different methods, however, for the regularization term update, the time consumption is highly correlated with the choice of the regularizer. It is shown that similar computational cost is required for the 3D-DL method and the 881-DL/818-DL/188-DL methods, while the 2D-DL method suffers from significantly higher computational cost. It also can be seen

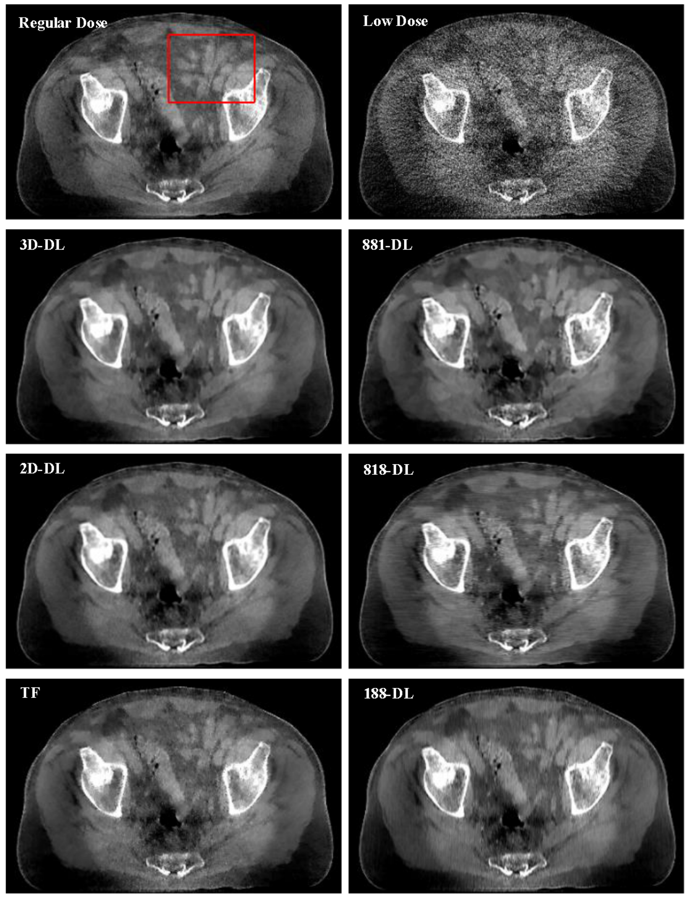

Fig. 3. Transversal views of the prostate patient images reconstructed by different methods. From top to down in the first column, the images are regular dose FDK reconstruction, reconstructions from the 3D-DL, 2D-DL and TF methods, respectively. From top to down in the second column, the images are reconstructed by the FDK, 881-DL, 818-DL and 188-DL methods, respectively. The display window is [-300 150] HU.

TABLE IV
TIME OVERHEADS OF DIFFERENT METHODS FOR THE HN AND PROSTATE PATIENT CASES (UNIT: SECONDS)

cases		3D-DL	2D-DL	TF	881-DL	818-DL	188-DL
HN	fidelity term	13.04	13.99	13.02	13.91	13.02	13.17
HN	regularization term	20.49	61.62	5.46	21.29	20.86	20.81
Prostate	fidelity term	9.33	9.76	9.28	9.53	9.52	9.9
Prostate	regularization term	14.93	50.16	2.25	20.6	16.84	17.94

from Table IV that one of the biggest advantages of the TF method is its low computation complexity.

Figure 6 plots the convergence curves for the HN and prostate patient cases. It can be seen that 10 iterations are sufficient for the whole program to converge in both cases.

IV. Discussions and Conclusions

In this study, a 3D dictionary learning based CBCT reconstruction algorithm has been proposed for low-dose CBCT, being validated in multiple realistic data experiments. The comparison has indicated that the 3D-DL method can deliver superior image quality in terms of well-preserved structures and effectively-suppressed noise. If the 2D dictionary learning based sparse constraint was enforced on all the three views consecutively, i.e., the 2D-DL method, the reconstructed images appeared blurry while there was still remaining noise. If the 2D dictionary learning based sparse constraint was enforced on a single view, such as with the 881-DL/818-DL/188-DL methods, directional streak artifacts were induced into the unprocessed views, while the structures on the processed view could be distorted. The TF method would result in noisy high contrast regions. To understand and optimize the performance of the 3D-DL method, a statistical analysis on the representation efficiencies of the 2D/3D dictionary were carried out, suggesting the higher representation efficiency of the 3D dictionary. Moreover, the whole program was well parallelized by employing several algorithmic tricks, attaining a high computational efficiency.

The cause of the directional streak artifacts is that the noise in the processed view is smoothened out when the 2D dictionary based processing is applied on only one view. As a result, the intersections of the unprocessed views with the processed view would exhibit directional streak artifacts, such as the horizontal/vertical streak artifacts in the transversal

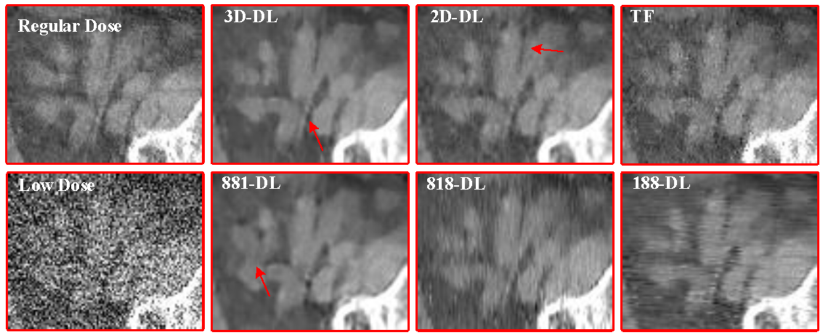

Fig. 4. The zoomed-in ROIs of the red box with respect to different reconstructions in Fig. (3). From left to right in the first row, the images are regular dose FDK reconstruction, reconstructions from the 3D-DL, 2D-DL and TF methods, respectively. From left to right in the second row, the images are reconstructed by the FDK, 881-DL, 818-DL and 188-DL methods, respectively. The display window is [-300 150] HU.

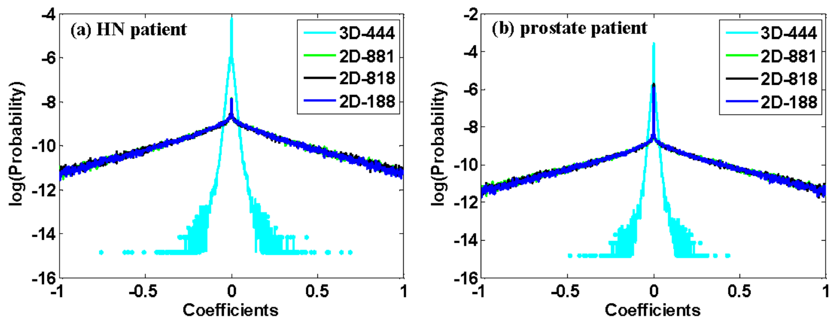

Fig. 5. Distributions of the representation coefficients among different dictionaries for the HN patient (a) and prostate patient (b). The x-axis is the values of the coefficients, the y-axis is the logarithmic probabilities. 3D-444 denotes the distributions of the coefficients for the 3D data samples represented by the 3D dictionary of dimension $4 \times 4 \times 4$. 2D-881/2D-818/2D-188 denote the distributions of the coefficients with the 2D dictionary for the 2D data samples extracted from the transversal/coronal/sagittal views, respectively.

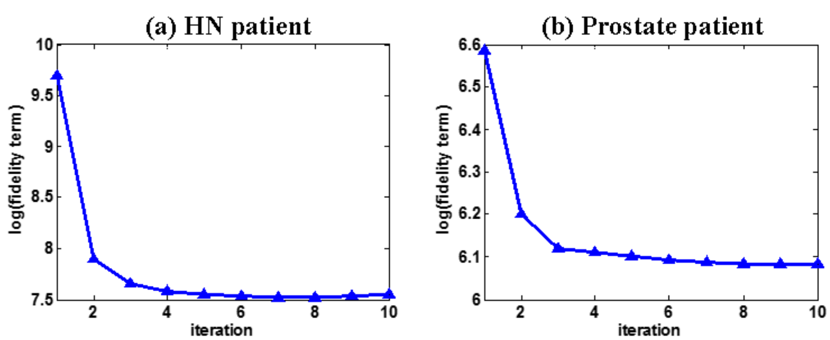

Fig. 6. The convergence curves of the 3D-DL method corresponding to the HN (a) and prostate (b) patient cases, respectively. The x-axis is the iteration number, the y-axis is the change amount of the images between consecutive iterations.

views if the coronal/ sagittal views are processed, as illustrated by the 818-DL/188-DL sub-figures in Fig. (2). The reason of the distorted structures in the processed view may be explained by the fact that the directional streak artifacts are spread out through the cone beam forward projection. If the sparse constraint are applied on all the three views successively, i.e., the 2D-DL method, the directional streak artifacts could be alleviated. However, the 2D-DL method may incorrectly interpret the directional streak artifacts from the previous steps as the potential structures. To avoid this side effect, a large tolerance is required, however, it would result in blurred structures as demonstrated in Sec. (III). Moreover, one may need to carefully select the suitable tolerance for each of the three steps in the 2D-DL method; in this work, the tolerances were set to be the same for all the three steps. Another disadvantage of the 2D-DL method is the high computational cost, as indicated by Table IV. The TF method employs a group of piecewise linear TF basis functions consisting of low pass filters for low frequency components, as well band pass and high pass filters for edges. As a consequence, in processing high contrast regions, the high pass filters are required to represent the structures, and may result in the noise-like artifacts.

The higher representation efficiency of the 3D dictionary over the 2D dictionary may be explained by the fact that the 3D dictionary could sufficiently capture spatial correlations in all the three dimensions simultaneously, while the 2D dictionary could only make use of the planar spatial correlations. As mentioned in Sec. (II-D3), a more efficient representation could facilitate the later processing stages. Indeed, this has been experimentally validated in our comparison studies as described in Sec. (III).

In this work, the volume dimensions of the HN and prostate patient cases are $512 \times 512 \times 512$ and $512 \times 512 \times 256$, the corresponding projection datasets are $512 \times 384 \times 363$ and $512 \times 384 \times 656$, respectively. As shown in Table IV, with the 3D-DL method, the time consumption per iteration in both cases is 30s, as illustrated in Fig. (6), the whole program would converge in 10 loops. Therefore, without loss of generality, a high-quality CBCT image is expected within 5 minutes from comparable amounts of data with similar computing devices (e.g., NVIDIA GeForce GTX 980 video card) relative to those used in this work.

In summary, a 3D-DL based sparse constraint has been incorporated into the iterative reconstruction framework to facilitate the low-dose CBCT. The 3D dictionary has exhibited a higher representation efficiency over the 2D dictionary, demonstrating a potential of enhancing the image quality. To offer the clinical utility, our whole program has been implemented on GPU with several algorithmic tricks.

References

[1] D. A. Jaffray and J. H. Siewerdsen, "Cone-beam computed tomography with a flat-panel imager: Initial performance characterization," *Medical Physics*, vol. 27, no. 6, pp. 1311–1323, 2000.

[2] D. A. Jaffray, J. H. Siewerdsen, J. W. Wong, and A. A. Martinez, "Flat-panel cone-beam computed tomography for image-guided rt," *International Journal of Radiation Oncologybiologyphysics*, vol. 48, no. 3, pp. 194–194, 2000.

[3] S. Lee, G. L. Gallia, D. D. Reh, S. Schafer, A. Uneri, D. J. Mirota, S. Nithiananthan, Y. Otake, J. W. Stayman, and W. Zbijewski, "Intraoperative c-arm cone-beam ct: Quantitative analysis of surgical performance in skull base surgery," *Laryngoscope*, vol. 122, no. 9, p. 19251932, 2012.

[4] A. S. Wang, J. W. Stayman, Y. Otake, S. Vogt, G. Kleinszig, and J. H. Siewerdsen, "Accelerated statistical reconstruction for c-arm cone-beam ct using nesterov's method," *Medical Physics*, vol. 42, no. 5, pp. 2699–2708, 2015.

[5] T. Kiljunen, T. Kaasalainen, A. Suomalainen, and M. Kortesniemi, "Dental cone beam ct: A review." *Physica Medica*, vol. 31, no. 8, pp. 844–860, 2015.

[6] C. L. Sang, H. Kyungkim, I. Konchun, M. Hyecho, S. Y. Lee, and H. C. Min, "A flat-panel detector based micro-ct system: performance evaluation for small-animal imaging," *Physics in Medicine and Biology*, vol. 48, no. 24, pp. 4173–4185, 2004.

[7] A. S. Brody, D. P. Frush, W. Huda, and R. L. Brent, "Radiation risk to children from computed tomography." *Pediatrics*, vol. 120, no. 3, pp. 677–682, 2007.

[8] B. Sinnott, E. Ron, and A. B. Schneider, "Exposing the thyroid to radiation: a review of its current extent, risks, and implications." *Endocrine Reviews*, vol. 31, no. 5, pp. 756–773, 2010.

[9] J. Song, Q. H. Liu, G. A. Johnson, and C. T. Badea, "Sparseness prior based iterative image reconstruction for retrospectively gated cardiac micro-ct," *Medical Physics*, vol. 34, no. 11, pp. 4476–4483, 2007.

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

[11] G.-H. Chen, J. Tang, and S. Leng, "Prior image constrained compressed sensing (piccs): a method to accurately reconstruct dynamic ct images from highly undersampled projection data sets," *Medical Physics*, vol. 35, no. 2, pp. 660–663, 2008.

[12] Q. Xu, H. Yu, X. Mou, L. Zhang, J. Hsieh, and G. Wang, "Low-dose x-ray ct reconstruction via dictionary learning," *IEEE Transactions on Medical Imaging*, vol. 31, no. 9, pp. 1682–1697, 2012.

[13] J. Wang, T. Li, Z. Liang, and L. Xing, "Dose reduction for kilovotage cone-beam computed tomography in radiation therapy," *Physics in Medicine and Biology*, vol. 53, no. 11, pp. 2897–2909, 2008.

[14] X. Jia, Y. Lou, R. Li, W. Y. Song, and S. B. Jiang, "Gpu-based fast cone beam ct reconstruction from undersampled and noisy projection data via total variation," *Medical Physics*, vol. 37, no. 4, pp. 1757–1760, 2010.

[15] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[16] G. Yu, L. Li, J. Gu, and L. Zhang, "Total variation based iterative image reconstruction," *Proc. Computer Vision for Biomedical Image Applications: First International Workshop*, vol. 3765, pp. 526–534, 2005.

[17] E. Sidky, C. Kao, and X. Pan, "Accurate image reconstruction from few-views and limited-angle data in divergent-beam ct," *Journal of X-ray Science and Technology*, vol. 14, no. 2, pp. 119–139, 2006.

[18] X. Jia, B. Dong, Y. Lou, and S. B. Jiang, "Gpu-based iterative cone-beam ct reconstruction using tight frame regularization." *Physics in Medicine and Biology*, vol. 56, no. 13, pp. 3787–3807, 2011.

[19] M. Aharon, M. Elad, and A. Bruckstein, "K-svd : An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, 2006.

[20] J. P. Li L, Chen Z, "Dual-energy ct reconstruction based on dictionary learning and total variation constraint," in *Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, 2012 IEEE, pp. 2358–2361.

[21] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images." *Nature*, vol. 381, no. 6583, pp. 607–609, 1996.

[22] M. S. Lewicki and T. J. Sejnowski, "Learning overcomplete representations." *Neural Computation*, vol. 12, no. 2, pp. 337–365, 2000.

[23] Q. Liu, S. Wang, and J. Luo, "A novel predual dictionary learning algorithm," *Journal of Visual Communication and Image Representation*, vol. 23, no. 1, pp. 182–193, 2012.

[24] I. A. Elbakri and J. A. Fessler, "Statistical image reconstruction for polyenergetic x-ray computed tomography," *IEEE Transactions on Medical Imaging*, vol. 21, no. 2, pp. 89 – 99, 2002.

[25] J. Wang, H. Lu, Z. Liang, D. Eremina, G. Zhang, S. Wang, J. Chen, and J. Manzione, "An experimental study on the noise properties of x-ray ct sinogram data in radon space," *Physics in Medicine and Biology*, vol. 53, no. 12, pp. 3327–3341, 2008.

[26] H. Y. J. Wu, X. Mou and G. Wang, "Statistical interior tomography via dictionary learning without assuming an object support," *Proc. Fully3D*, pp. 440–443, 2013.

[27] X. Mou, J. Wu, Q. Xu, H. Yu, and G. Wang, "Dictionary learning based low-dose x-ray ct reconstruction using a balancing principle," *Proc. SPIE*, vol. 9212, pp. 921 207–921 207–15, 2014.

[28] J. Huang and D. Mumford, "Statistics of natural images and models," *cvpr*, p. 1541, 1999.

[29] N. Y., "A method for unconstrained convex minimization problem with the rate of convergence o (1/k2)," *Doklady an SSSR*, vol. 269, no. 3, pp. 543–547, 1983.

[30] J. H. Cho and J. A. Fessler, "Accelerating ordered-subsets image reconstruction for x-ray ct using double surrogates," *Proc. SPIE*, vol. 8313, no. 2, pp. 145–150, 2012.

[31] D. Needell and J. A. Tropp, "Cosamp: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, 2009.

[32] D. L. Donoho, Y. Tsaig, I. Drori, and J. L. Starck, "Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 58, no. 2, pp. 1094 – 1121, 2006.

[33] T. Goldstein and S. Osher, "The split bregman method for l1-regularized problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 2, pp. 323–343, 2009.

[34] B. L. Sturm and M. G. Christensen, "Comparison of orthogonal matching pursuit implementations," *Proc. the 20th European Signal Processing Conference*.

[35] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[36] J. P. Jones and L. A. Palmer, "The two-dimensional spectral structure of simple receptive fields in cat striate cortex." *Journal of Neurophysiology*, vol. 58, no. 58, pp. 1212–32, 1987.

[37] D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells." *Journal of the Optical Society of America A Optics and Image Science*, vol. 4, no. 4, pp. 2379–2394, 1987.