Low-Dose X-ray CT Reconstruction via Dictionary Learning

Qiong Xu,
Institute of Image Processing and Pattern Recognition, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China, and also with the Biomedical Imaging Division, VT-WFU School of Biomedical Engineering and Sciences, Wake Forest University Health Sciences, Winston-Salem, NC 27157 USA

Hengyong Yu [Senior Member, IEEE],
Biomedical Imaging Division, VT-WFU School of Biomedical Engineering and Sciences, and the Department of Radiology, Division of Radiologic Sciences, Wake Forest University Health Sciences, Winston-Salem, NC 27157 USA

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

Lei Zhang [Member, IEEE],
Department of Computing, Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, China

Jiang Hsieh [Senior Member, IEEE], and
GE Healthcare Technology, Waukesha, WI 53188 USA

Ge Wang [Fellow, IEEE]
Biomedical Imaging Division, VT-WFU School of Biomedical Engineering and Sciences, Virginia Tech, Blacksburg, VA 24061 USA, and also with the Wake Forest University Health Sciences, Winston-Salem, NC 27157 USA

Qiong Xu: xjtuxqiong@gmail.com; Hengyong Yu: hengyong-yu@ieee.org; Xuanqin Mou: xqmou@mail.xjtu.edu.cn; Lei Zhang: cslzhang@comp. polyu.edu.hk; Jiang Hsieh: jiang.hsieh@med.ge.com; Ge Wang: ge-wang@ieee.org

Abstract

Although diagnostic medical imaging provides enormous benefits in the early detection and accuracy diagnosis of various diseases, there are growing concerns on the potential side effect of radiation induced genetic, cancerous and other diseases. How to reduce radiation dose while maintaining the diagnostic performance is a major challenge in the computed tomography (CT) field. Inspired by the compressive sensing theory, the sparse constraint in terms of total variation (TV) minimization has already led to promising results for low-dose CT reconstruction. Compared to the discrete gradient transform used in the TV method, dictionary learning is proven to be an effective way for sparse representation. On the other hand, it is important to consider the statistical property of projection data in the low-dose CT case. Recently, we have developed a dictionary learning based approach for low-dose X-ray CT. In this paper, we present this method in detail and evaluate it in experiments. In our method, the sparse constraint in terms of a redundant dictionary is incorporated into an objective function in a statistical iterative reconstruction framework. The dictionary can be either predetermined before an image reconstruction task or adaptively defined during the reconstruction process. An alternating minimization scheme is developed to minimize the objective function. Our approach is evaluated with low-dose X-ray

projections collected in animal and human CT studies, and the improvement associated with dictionary learning is quantified relative to filtered backprojection and TV-based reconstructions. The results show that the proposed approach might produce better images with lower noise and more detailed structural features in our selected cases. However, there is no proof that this is true for all kinds of structures.

Index Terms

Compressive sensing (CS); computed tomography (CT); dictionary learning; low-dose CT; sparse representation; statistical iterative reconstruction

I. Introduction

Nowadays, X-ray computed tomography (CT) is widely used in hospitals and clinics for diagnosis and intervention. It is well known that X-ray radiation can be harmful which may induce genetic, cancerous, and other diseases [1]–[3]. Therefore, the radiation risk issue is receiving more and more attention. As a result, the well-known ALARA (as low as reasonably achievable) principle is applied to avoid excessive radiation dose in the medical community. Since X-ray imaging is a quantum accumulation process, the signal-to-noise ratio (SNR) depends on the X-ray dose quadratically. Given other conditions being identical, reducing the X-ray dose will degrade image quality. Consequently, how to reconstruct adequate CT images at a minimum radiation dose level is a hot topic in the CT field.

There are two strategies for radiation dose reduction: the first one is to reduce the X-ray flux towards each detector element, and the second one is to decrease the number of X-ray attenuation measurements across a whole object to be reconstructed. The former is usually implemented by adjusting the operating current, the operating potential and exposure time of an X-ray tube, leading to noisy projections. The latter necessarily produces insufficient projection data, suffering from few-view, limited-angle, interior scan, or other problems. These problems can co-exist in one dataset, representing a major opportunity for algorithmic research.

To reconstruct images from noisy projections, various reconstruction algorithms were proposed. These algorithms can be categorized into two classes in terms of the variables used in the optimization process.

The first class focuses on preprocessing projection data before image reconstruction. The optimization variables are projection data. Hsieh proposed an adaptive filtering approach where the filter parameterization was adjusted according to the noise property [4]. Kachelriess *et al.* developed a multi-dimensional adaptive filtering approach to refine projection data along the detector row, view, and longitudinal directions [5]. La Riviere developed a penalized likelihood technique to smooth a sinogram [6]. Wang *et al.* presented a penalized weighted least-squares approach to reduce sinogram noise [7].

Different from these preprocessing methods, the second class is referred to as statistical iterative reconstruction (SIR). It optimizes the maximum-likelihood or penalized-likelihood function formulated according to the statistical characteristics of projection data. Its optimization variables are image pixels or voxels. Kamphues and Beekman proposed an ordered-subsets convex algorithm [8]. Nuyts *etal.* studied a gradient ascent algorithm [9]. Erdogan and Fessler developed a monotonic algorithm based on separable paraboloidal surrogate [10]. Recently, Thibault *et al.* applied the SIR schemeto real multi-slice helical data [11]. These results show that the statistical framework promises an optimal reconstruction quality from noisy projection data.

Compressive sensing (CS) theory has now become popular, and has been instrumental for image reconstruction from incomplete and noisy datasets. CS theory allows a sparse signal to be accurately reconstructed from samples far less than what is required by the Shannon/ Nyquist sampling theorem [12], [13]. The key for the success of CS is the sparsity of a signal under study. Although an object is not sparse in general, often times a sparsifying transform can be used to convert it into a domain in which the signal has a sparse representation. One common sparsifying transform is the discrete gradient transform (DGT) whose coefficients can be summed up to form the so-called total variation (TV). Inspired by CS theory, various TV minimization algorithms were suggested to solve the few-view, limited-angle, and interior problems. For example, Chen *et al.* proposed a prior image constrained compressed sensing (PICCS) algorithm for dynamic CT application [14]. Yu and Wang proved that a piecewise constant interior region of interest (ROI) can be uniquely reconstructed by a TV minimizing technique [15], [16]. Xu *et al.* extended this CS based interior tomography formulation into a SIR framework [17]. Ritschl *et al.* proposed an improved TV method within the ASD-POCS framework for clinical applications [18]. Although these TV-based algorithms are successful in a number of cases, the power of the TV minimization constraint is still limited. First, the TV constraint is a global requirement, which cannot directly reflect structures of an object. Second, the DGT operation cannot distinguish true structures and image noise. Consequently, images reconstructed with the TV constraint may lose some fine features and generate a blocky appearance in incomplete and noisy cases. Hence, it is necessary to investigate superior sparsifying methods for CS-inspired image reconstruction.

Recently, the sparse representation in terms of a redundant dictionary [19] has attracted an increasing interest in the image processing, imaging analysis and magnetic resonance imaging (MRI) fields. Such a dictionary is an over-complete basis. The elements in this basis are called atoms, which are learned from application-specific training images. Then, an object image can be sparsely represented as a linear combination of these atoms. Usually, an object image is decomposed into small overlapped patches. The dictionary learning approach acts on these patches, and an average of the corresponding values in the overlapped patches is computed at a given location. Since the dictionary is learned from training images, it is expected to have a better sparsifying capability than any generic sparse transform. Also, the redundancy of the atoms facilitates a sparser representation. More importantly, the dictionary tends to capture local image features effectively because of the patch-based analysis and most importantly the structural self-similarity in many cases.

Given the aforementioned advantages, dictionary learning based techniques have led to excellent results in image/video denoising, inpainting, restoration, face recognition, texture classification, MRI, and other areas. In 2006, Elad and Aharon addressed the image denoising problem using a dictionary learned using the K-SVD algorithm [20]. Mairal *et al.* extended this method for color image restoration [21]. Then, Mairal *et al.* generalized it into a multi-scale framework for image and video restoration [22]. Wright *et al.* used the dictionary learning based sparse representation in human face recognition [23]. As far as medical imaging is concerned, the dictionary learning based work just began to emerge, initially for MRI. In 2010, Chen *et al.* combined a dictionary learning method and a TV-based MRI scheme to improve image quality [24]. In 2011, Ravishankar and Bresler proposed an adaptive dictionary learning based method to reconstruct MRI images from highly under-sampled *k*-space data [25].

It is our hypothesis that the dictionary learning technique can be used for low-dose CT reconstruction. Recently, we proposed a dictionary learning based reconstruction method to address the low-dose CT reconstruction problem [26]. Our method consists of two components. The first component is the SIR process that enforces the statistical knowledge.

The second component is the dictionary-based sparsification whose role is similar to the TV minimization [17]. In our work, two types of dictionaries are used. The first type is a global dictionary learned from a pre-specified set of training images, and it will not change during the reconstruction process. The second type is an adaptive dictionary learned from an intermediate image and updated iteratively in the reconstruction process.

The rest of this paper is organized as follows. In Section II, we will introduce the background on dictionary learning and SIR. In Section III, we will describe our reconstruction approach and its implementation. In Section IV, we will report representative results from preclinical and clinical low-dose projections, and quantify the performance of our proposed methodology. Finally, in Section V we will discuss relevant problems and conclude the paper.

II. Background

A. Dictionary Learning and Sparse Representation

Let N and K be integers, and $\mathbb{R}$ be the real space. A dictionary is a matrix $\mathbf{D} \in \mathbb{R}^{N \times K}$ whose column $\mathbf{d}_k \in \mathbb{R}^{N \times 1}$ ($k = 1, \ldots, K$) is a N dimensional vector, which is called an atom. Usually, the dictionary is redundant or over-complete; that is, $N \ll K$. An image patch of $\sqrt{N} \times \sqrt{N}$ pixels can be expressed as a N dimensional vector $\mathbf{x} \in \mathbb{R}^{N \times 1}$. Suppose that a patch $\mathbf{x}$ can be exactly or approximately represented as a sparse linear combination of the atoms in the dictionary $\mathbf{D}$; that is

$$\left\| \mathbf{x} - \mathbf{D}\alpha \right\|_2^2 \leq \varepsilon \quad (1)$$

where $\varepsilon \geq 0$ is a small error bound, and the representation vector $\alpha \in \mathbb{R}^{K \times 1}$ has few nonzero entries, $\| \alpha \|_0 \ll N \ll K$ with $\|\cdot\|_0$ being the l_0-norm. The dictionary redundancy implies that the number of atoms is larger than the length of an atom.

Finding a sparse representation $\alpha \in \mathbb{R}^{K \times 1}$ of an image patch $\mathbf{x} \in \mathbb{R}^{N \times 1}$ with respect to a given dictionary $\mathbf{D} \in \mathbb{R}^{N \times K}$ is equivalent to solve the following optimization problem:

$$\min_{\alpha} \left\| \alpha \right\|_0 \quad s.t. \quad \left\| \mathbf{x} - \mathbf{D}\alpha \right\|_2^2 \leq \varepsilon. \quad (2)$$

By the Lagrange method, (2) can be rewritten in an unconstraint form

$$\min_{\alpha} \left\| \mathbf{x} - \mathbf{D}\alpha \right\|_2^2 + \nu \left\| \alpha \right\|_0 \quad (3)$$

where ν is the Lagrange multiplier. It is pointed out that the above two problems are equivalent when a suitable ν is chosen. Since solving (2) or (3) directly is NP-hard, an approximate alternative strategy is desirable. In this regard, there are several greedy algorithms available, such as matching pursuit (MP) [27] and orthogonal matching pursuit (OMP) [28], [29] algorithms. Also, the l_0-norm can be replaced by the l_1-norm to make the problem convex and manageable using a basis pursuit (BP) [30] algorithm.

Given a training set of S patches, the so-called dictionary learning is to seek a dictionary that makes each patch in the training set be sparsely represented by the atoms in this dictionary. Denote the given patch set as a matrix $\mathbf{X} \in \mathbb{R}^{N \times S}$ with a patch $\mathbf{x}_s \in \mathbb{R}^{N \times 1}$ ($s = 1, \ldots, S$) being a column vector of $\mathbf{X}$, and the corresponding sparse representation vector as a matrix $\alpha \in$

$\mathbb{R}^{K \times S}$ with the representation $\alpha_s \in \mathbb{R}^{K \times 1}$ of a patch being a column vector of $\alpha \in \mathbb{R}^{K \times S}$. Then, the dictionary learning is to solve

$$\min_{\mathbf{D},\alpha} \sum_{s=1}^{S} \left(\left\| \mathbf{x}_s - \mathbf{D}\alpha_s \right\|_2^2 + \nu_s \left\| \alpha_s \right\|_0 \right). \quad (4)$$

Equation (4) is basically equivalent to either of the following problems:

$$\min_{\mathbf{D},\alpha} \left\| \mathbf{X} - \mathbf{D}\alpha \right\|_2^2 \quad s.t. \quad \forall s, \quad \left\| \alpha_s \right\|_0 \leq L_0 \quad (5)$$

$$\min_{\mathbf{D},\alpha} \sum_{s=1}^{S} \left\| \alpha_s \right\|_0 \quad s.t. \quad \left\| \mathbf{X} - \mathbf{D}\alpha \right\|_2^2 \leq \varepsilon \quad (6)$$

where L_0 and ε are the sparsity and precision of the sparse representation, respectively. The l_0-norm in (4)–(6) can be replaced by the l_1-norm to make them easier to solve. There are many algorithms available for dictionary learning, including the classical K-SVD method [31] and the fast online learning technique [32].

A successful application of the dictionary learning technique is image denoising [20]. Let a vector $\mathbf{z} \in \mathbb{R}^{M \times 1}$ represent a noise image of $H \times W$ pixels and a vector $\mathbf{x} \in \mathbb{R}^{M \times 1}$ denote its corresponding filtered version, $M = H \times W$. A set of small overlapping patches can be extracted from the image. With a sliding distance of one pixel, we will have $(H - \sqrt{N} + 1) \times (W - \sqrt{N} + 1)$ patches. It is assumed that the patches extracted from the filtered image can be sparely represented in terms of a dictionary, and the filtered image should be close to the original noisy image. Hence, the denoising procedure is to minimize the following objective function:

$$\min_{\mathbf{x},\alpha,(\mathbf{D})} \left\| \mathbf{x} - \mathbf{z} \right\|_2^2 + \lambda \sum_{s=1}^{S} \left(\left\| \mathbf{E}_s \mathbf{x} - \mathbf{D}\alpha_s \right\|_2^2 + \nu_s \left\| \alpha_s \right\|_0 \right) \quad (7)$$

where $\mathbf{E}_s \in \mathbb{R}^{N \times M}$ is the matrix to extract a patch from the image $\mathbf{x}$, $S = \left(H - \sqrt{N} + 1 \right) \times \left(W - \sqrt{N} + 1 \right)$, and λ is a regularization parameter related to the noise level of $\mathbf{z}$. The dictionary $\mathbf{D}$ in (7) can be determined in two ways. One way is to predetermine it from a training set, which should contain representative structures in the image to be filtered. The other way is to construct the dictionary during the denoising procedure.

B. Statistical Reconstruction and Objective Function

Without loss of generality, we only assume a monochromatic source. Approximately, measured data follow the Poisson distribution:

$$y_i \sim \text{Poisson} \{ b_i e^{-l_i} + r_i \}, i = 1, \ldots, I \quad (8)$$

where y_i is the measurement along the ith X-ray path, and b_i is the blank scan factor, $l_i = \sum_{j=1}^{J} a_{ij} \mu_j = [\mathbf{A}\mu]_i$ is the integral of the X-ray linear attenuation coefficients, $\mathbf{A} = \{ a_{ij} \}$ is the system matrix, $\mu = (\mu_1, \ldots \mu_J)^T$ is a linear attenuation coefficient distribution, r_i accounts for read-out noise, I and J are the number of projections and pixels, respectively.

Assuming that the noise distributions along different paths are statistically independent, the Poisson log-likelihood function of the joint probability distribution can be written as [33]

$$L(\mathbf{y}|\mu) = \ln P(\mathbf{y}|\mu) = \ln\left(\prod_{i=1}^{I}\left(\frac{e^{-\bar{y}_i}\bar{y}_i^{y_i}}{y_i!}\right)\right) \quad (9)$$

where $P(\cdot)$ is the probability function, and $\bar{y}_i = b_i e^{-l_i} + r_i$ is the expected value of y_i. Ignoring the constant terms, we obtain $L(\mathbf{y}|\mu) = -\sum_{i=1}^{I}(\bar{y}_i - y_i \ln \bar{y}_i)$.

From the statistical perspective, an image can be reconstructed by maximizing *a posteriori* (MAP) of the function $P(\mu|y)$. According to the Bayesian rule $P(\mu|y) = P(y|\mu) P(\mu)/P(\mathbf{y})$ and the monotonic increment property of the natural logarithm, the reconstruction is equivalent to maximize the following objective function:

$$\tilde{\Phi}(\mu) = L(\mathbf{y}|\mu) + \ln P(\mu) \quad (10)$$

where $\ln P(\mu)$ is a regularization term based on the prior knowledge. Let $R(\mu) = -\ln P(\mu)$, the task can be converted into minimizing the following objective function:

$$\Phi(\mu) = \sum_{i=1}^{I}\left(\left(b_i e^{-l_i} + r_i\right) - y_i \ln\left(b_i e^{-l_i} + r_i\right)\right) + R(\mu). \quad (11)$$

Performing a second-order Taylor expansion of $g_i(l) = (b_i e^{-l} + r_i) - y_i \ln(b_i e^{-l} + r_i)$ with respect to an estimated line integral $l_i \triangleq \ln(b_i/(y_i - r_i))$ [34], (11) becomes

$$\Phi(\mu) = \sum_{i=1}^{I}\frac{w_i}{2}\left([\mathbf{A}\mu]_i - \widehat{l_i}\right)^2 + R(\mu) \quad (12)$$

where $w_i = (y_i - r_i)^2/y_i$ is the statistical weight for each X-ray path.

While the simultaneous algebraic iterative technique (SART) minimizes a least square function, SIR deals with a statistically weighted least square function defined by (12). The statistical weight represents the confidence of the projection measurement along each path. The projection data through denser paths would have lower signal-to-noise-ratios (SNR). While methods were developed based on physical experiments to calculate statistical weights [7], [35], [36], in this feasibility study we use the Poisson monochromatic model for SIR. For practice applications, the measured realistic data may not fully satisfy the Poisson model. The compound Poisson model may be more accurate [37]. However, the Poisson model is well accepted in the CT field [38].

III. Methodology

A. GDSIR and ADSIR

In Section II-B the regularization term $R(\mu)$ in (12) represents prior information on reconstructed images. Various assumptions about the prior information lead to different reconstruction algorithms. For example, the assumption of smooth variation over adjacent pixels suggests a regularization in terms of quadratic differences between adjacent pixels. The piecewise constant assumption supports a TV regularization. A more general form of regularization is the *q-generalized Gaussian Markov field (q-GGMRF) prior* [39], which has

two adjustable parameters. The quadratic and TV regularization functions are special cases of *q-GGMRF*. Because all these regularization means are based on the relationship of adjacent pixels, it is difficult for them to distinguish weak structures and strong noise.

On the other hand, the dictionary learning and sparse representation techniques perform well in sensing structures and suppressing noise. Here, we propose to use the sparsity constraint in terms of a redundant dictionary as the regularization term of (12). In reference to (7), we have the following minimization problem:

$$\min_{\mu,\alpha,(\mathbf{D})} \sum_{i=1}^{I} \frac{w_i}{2} \left([\mathbf{A}\mu]_i - \widehat{l_i} \right)^2 + \lambda \left(\sum_s \left\| \mathbf{E}_s\mu - \mathbf{D}\alpha_s \right\|_2^2 + \sum_s \nu_s \left\| \alpha_s \right\|_0 \right) \quad (13)$$

where $\mathbf{E}_s = \{e_{nj}^s\} \in \mathbb{R}^{N \times J}$ is an operator to extract patches from an image.

Similar to the comments we made on (7), the dictionary $\mathbf{D}$ in (13) can be either predetermined or dynamically defined. According to these two options, we can perform either global dictionary based statistical iterative reconstruction (GDSIR) or adaptive dictionary based statistical iterative reconstruction (ADSIR). Because structures in various subjects are quite similar in a given application, a training set of patches for construction of a global dictionary can be extracted from images of similar objects. On the other hand, the training set of patches for construction of an adaptive dictionary can be extracted from an intermediate image and dynamically updated during the reconstruction process.

Algorithm Description—For GDSIR, the image reconstruction process is equivalent to solve the following optimization problem:

$$\min_{\mu,\alpha} \sum_{i=1}^{I} \frac{w_i}{2} \left([\mathbf{A}\mu]_i - \widehat{l_i} \right)^2 + \lambda \left(\sum_s \left\| \mathbf{E}_s\mu - \mathbf{D}\alpha_s \right\|_2^2 + \sum_s \nu_s \left\| \alpha_s \right\|_0 \right) \quad (14)$$

where there are two variables μ and α. An alternating minimization scheme can be used to optimize the two variables. First, an intermediate reconstructed image μ with a fixed sparse representation $\tilde{\alpha}_s$ is updated to reduce the data discrepancy. Thus, the objective function (14) becomes

$$\min_{\mu} \sum_{i=1}^{I} \frac{w_i}{2} \left([\mathbf{A}\mu]_i - \widehat{l_i} \right)^2 + \lambda \sum_s \left\| \mathbf{E}_s\mu - \mathbf{D}\tilde{\alpha}_s \right\|_2^2. \quad (15)$$

By the separable paraboloid surrogate method [34], (15) can be iteratively solved as

$$\mu_j^t = \mu_j^{t-1} - \frac{\sum_{i=1}^{I} (a_{ij} w_i ([\mathbf{A}\mu^{t-1}]_i - \widehat{l_i})) + 2\lambda \sum_s \sum_{n=1}^{N} e_{nj}^s ([\mathbf{E}_s\mu^{t-1}]_n - [\mathbf{D}\tilde{\alpha}_s]_n)}{\sum_{i=1}^{I} \left(a_{ij} w_i \sum_{k=1}^{J} a_{ik} \right) + 2\lambda \sum_s \sum_{n=1}^{N} e_{nj}^s \sum_{k=1}^{J} e_{nk}^s} \quad j=1,\ldots,J \quad (16)$$

where the superscript $t = 1, 2, \ldots, T$ is the iteration index. Second, the intermediate image μ is re-expressed in terms of the dictionary for a better sparse representation. Thus, the objective function (14) becomes

$$\min_{\alpha} \sum_s \left\| \mathbf{E}_s \mu^t - \mathbf{D}\alpha_s \right\|_2^2 + \sum_s \nu_s \left\| \alpha_s \right\|_0 \quad (17)$$

which is a sparse representation problem. As an example, the OMP algorithm can be used to find the sparse representation of each patch. The above two steps are alternately performed until a stopping criterion is satisfied.

For ADSIR, the image reconstruction process is equivalent to solve the following optimization problem:

$$\min_{\mu, \mathbf{D}, \alpha} \sum_{i=1}^I \frac{w_i}{2} \left([\mathbf{A}\mu]_i - \widehat{l_i} \right)^2 + \lambda \left(\sum_s \left\| \mathbf{E}_s \mu - \mathbf{D}\alpha_s \right\|_2^2 + \sum_s \nu_s \left\| \alpha_s \right\|_0 \right) \quad (18)$$

where there are three variables μ, $\mathbf{D}$, and α. Similar to GDSIR, an alternating minimization scheme is used to optimize the three variables. First, an intermediate reconstructed image μ is updated to reduce the data discrepancy with a fixed sparse representation α and a current dictionary $\mathbf{D}$. This step is exactly the same as that for GDSIR expressed by (16). Second, keeping the intermediate image μ unchanged, $\mathbf{D}$ and α are estimated by

$$\min_{\mathbf{D}, \alpha} \sum_s \left\| \mathbf{E}_s \mu^t - \mathbf{D}\alpha_s \right\|_2^2 + \sum_s \nu_s \left\| \alpha_s \right\|_0. \quad (19)$$

Equation (19) is the generic dictionary learning and sparse representation problem, which can be solved with respect to α and $\mathbf{D}$ alternatingly using the classic K-SVD algorithm. For fast convergence, we first use the fast online algorithm [32], [40] to train a dictionary from the patches extracted from an intermediate image μ. Then, we fix the dictionary to update the sparse representation as in (17) using the OMP algorithm. The above procedures are alternatingly performed until a stopping criterion is satisfied.

B. Parameter Selection

Similar to other regularized iterative reconstruction algorithms, the regularization parameter λ is to balance the data fidelity and prior information terms. The final reconstruction depends on the parameter λ. It is an interesting problem how to choose an optimal parameter. Usually, it is empirically selected in practice. For GDSIR and ADSIR, we can also empirically select λ. Because the data fidelity term is proportional to the noise standard deviation in the projection domain, λ should be increased with the noise increment. Specifically, our parameter selection problem can be reformulated as follows. After some variable exchanges and constant additions, we can rewrite (15) as

$$\min_{\mu} \left\| \mathbf{B}\mu - \mathbf{L} \right\|_2^2 + \lambda N \left\| \mu - \tilde{\mu} \right\|_2^2 \quad (20)$$

where $\mathbf{B} = \mathbf{wA}$, $\mathbf{L} = \mathbf{wl}$, $\mathbf{w} = \mathrm{diag}(\sqrt{w_i/2})$, $\mathbf{l} = (l_1, \ldots, l_I)^T$, $\tilde{\mu} = \{(1/N) \sum_{s \in \Omega_j} \mathbf{F}_{sj} \mathbf{D}\alpha_s\}$, N is the number of pixels in a patch, Ω_j denotes the set of patches covering a pixel j, and $\mathbf{F}_{sj} \in \mathbb{R}^{1 \times N}$ extracts a pixel j from a patch s. Note that in (19) we can discard the pixels in margins and only consider the pixels in the intersection between the rows from $\sqrt{N}$ to $\left(H - \sqrt{N} + 1\right)$ and the columns from $\sqrt{N}$ to $\left(W - \sqrt{N} + 1\right)$ for an $H \times W$ image. Let $\tilde{\mu} = \tilde{\mu} - \mu$ and $\tilde{\mathbf{L}} = \mathbf{L} - \mathbf{B}\mu$. Then, (20) becomes

$$\min_{\mu}\|\mathbf{B}\boldsymbol{\xi}-\tilde{\mathbf{L}}\|_2^2+\lambda N\|\boldsymbol{\xi}\|_2^2. \quad (21)$$

Equation (21) is a classical regularization problem, which can be solved using existing methods [41], such as the Miller method [42] and the generalized cross-validation method [43]. Nevertheless, λ is empirically selected in this feasibility study.

The dictionary redundancy improves the sparsity of representation. To ensure the redundancy, the number of atoms in a dictionary should be much greater than that of pixels in a patch, that means $K \gg N$. In the image processing field, $K = 4N$ is a conventional choice. On the one hand, a larger patch size corresponds to a larger number of atoms in a dictionary, which would increase the computational cost. On the other hand, if the patch size is too small, it could not effectively catch features in an image. In [44], it was pointed out that there was no significant difference between the results with $N = 8\times8$, $K = 256$ and $N = 16\times16$, $K = 1024$, and a larger patch size may lead to an over-smoothed image. Hence, $N = 8\times8$, $K = 256$, are used in this paper. In the dictionary learning process, we solve the optimization problem (5) by minimizing the representation error with a fixed sparsity level L_0^D. The sparsity level is selected as 5~10 atoms. The aforementioned parameter selection for N, K and L_0^D has been applied in many applications [20], [22], and also worked well in our experiments to be described below.

Besides, there are two parameters in the sparse representation step of image reconstruction, which are the sparsity level L_0^S and the precision level ε. The OMP process will stop when either $\|\alpha_s\|_0 \geq L_0^S$ or $\|\mathbf{E}_s\mu-\mathbf{D}\alpha_s\|_2^2 \leq \varepsilon$. The sparsity level L_0^S is the number of atoms involved in representing a patch, which is empirically determined according to the complexity of an image to be reconstructed and the property of the dictionary. Usually, L_0^S is greater than or equal to the L_0^D in dictionary learning, and smaller than $N/2$ for sparsity. The precision level ε represents the tolerance of the difference between the reconstructed and original images, which is determined by the image noise level and the property of the dictionary. The stronger the noise is, the greater ε is. We can estimate the image noise level over a flat region in a reconstructed image.

Based on the above discussions, the parameter selection guidelines are summarized in Table I.

C. Monotonic Decrement of the Objective Function

Since the objective function is monotonically minimized in each part of the alternating minimization process, the value of the objective function is also monotonically reduced using the proposed algorithms.

In the GDSIR case, let μ^t denote an intermediate image and α^t the sparse representation after t ($t = 1,2, \ldots, T$) iterations. Since the updating formula (16) monotonically minimizes the objective function (15) [34], we have

$$\sum_{i=1}^{I}\frac{w_i}{2}\left([\mathbf{A}\mu^t]_i-\widehat{l_i}\right)^2+\lambda\sum_s\left\|\mathbf{E}_s\mu^t-\mathbf{D}\alpha_s^{t-1}\right\|_2^2$$
$$\leq \sum_{i=1}^{I}\frac{w_i}{2}\left([\mathbf{A}\mu^{t-1}]_i-\widehat{l_i}\right)^2+\lambda\sum_s\left\|\mathbf{E}_s\mu^{t-1}-\mathbf{D}\alpha_s^{t-1}\right\|_2^2 \quad (22)$$
$$\forall t=1, 2, \ldots, T.$$

Then, in the sparse representation step using the OMP method, the following problem will be optimized:

$$\min_{\alpha_s} \left\| \mathbf{E}_s\mu - \mathbf{D}\alpha_s \right\|_2^2 \quad s.t. \quad \overset{\text{when } \left\| \mathbf{E}_s\mu - \mathbf{D}\alpha_s \right\|_2^2 > \varepsilon}{\left\| \alpha_s \right\|_0 \le L_0}. \quad (23)$$

Clearly, no matter $\left\| \mathbf{E}_s\mu^{t-1} - \mathbf{D}\alpha_s^{t-1} \right\|_2^2 \le \varepsilon$ or not, the solution of (23) satisfies

$$\left\| \mathbf{E}_s\mu^t - \mathbf{D}\alpha_s^t \right\|_2^2 \le \left\| \mathbf{E}_s\mu^t - \mathbf{D}\alpha_s^{t-1} \right\|_2^2, \forall s = 1, 2, \ldots, S. \quad (24)$$

Therefore, we have

$$\sum_{i=1}^{I} \frac{w_i}{2} \left([\mathbf{A}\mu^t]_i - \widehat{l_i} \right)^2 + \lambda \sum_s \left\| \mathbf{E}_s\mu^t - \mathbf{D}\alpha_s^t \right\|_2^2$$
$$\le \sum_{i=1}^{I} \frac{w_i}{2} \left([\mathbf{A}\mu^{t-1}]_i - \widehat{l_i} \right)^2 + \lambda \sum_s \left\| \mathbf{E}_s\mu^{t-1} - \mathbf{D}\alpha_s^{t-1} \right\|_2^2 \quad (25)$$
$$s.t. \quad \forall s, \; \left\| \alpha_s \right\|_0 \le L_0, \; \forall t = 1, 2, \ldots, T$$

which shows that GDSIR monotonically decreases its objective function. In the ADSIR case, we can follow similar steps to reach the same conclusion.

The above-mentioned monotonic decrement in the objective function value does not mean the convergence of the reconstruction process for neither GDSIR nor ADSIR. The convergence of GDSIR or ADSIR is much more difficult to prove, and is considered beyond the scope of this paper. However, our numerical and experimental studies to be reported below seem suggesting the convergence of our GDSIR and ADSIR algorithms.

D. Stopping Criterion and Computational Cost

In practice, we can simply stop the reconstruction process after a number of iterations when the change of the objective function or the reconstructed image becomes very small. This stopping rule has been widely used in the iterative reconstruction community. As far as the computational cost is concerned, after image updating GDSIR needs one OMP step for sparse representation, and ADSIR not only needs one OMP step but also one dictionary learning step. The computational cost of OMP and dictionary learning depends on the sparse representation level, precision level, patch size, dictionary size, etc. The computational cost of the OMP step is usually less than that for image updating which involves forward and backward projections. Therefore, the computational cost of GDSIR is comparable to other iterative methods. In this work, the online dictionary learning method was used [32], [40], which is significantly faster than other dictionary learning methods. As a result, the computational time of ADSIR is comparable to that of GDSIR.

E. Overall Work flow

The specific flowcharts for GDSIR and ADSIR are presented in Tables II and III, respectively.

IV. Experiments

Low-dose raw projections from sheep lung perfusion and patient cardiac angiography studies were used to evaluate and validate our proposed algorithms. Two numerical

simulation studies were designed to evaluate the performance of the proposed algorithms. Also, the popular TV regularization method and a non-statistical version of GDSIR were compared with GDSIR and ADSIR. Furthermore, low-count and few-view datasets were used to demonstrate the potential of the proposed techniques.

A. Sheep Lung Study

1) Data Acquisition—In a sheep lung perfusion study, an anesthetized sheep was scanned at normal and low doses respectively on a SIEMENS Somatom Sensation 64-slice CT scanner in a circular cone-beam scanning mode. A scan protocol was developed for low-dose studies with ECG gating: time point 1 for a normal X-ray dose scan (100kV/150 mAs) before a contrast agent injection, and time points 2–21 for low-dose scans (80 kV/17 mAs) after the contrast agent injection. All the sinograms of the central slice were extracted, which were in a fan-beam geometry. The radius of the trajectory was 57 cm. Over a 360° range, 1160 projections were uniformly collected. For each projection, 672 detector elements were equi-angularly distributed defining a field-of-view (FOV) of 25.05 cm in radius. In this experiment, the reconstructed images were matrixes of 768×768 pixels covering a 43.63×43.63 cm^2 region. The sparsity constraint was enforced on the entire lung region of 500×370 pixels.

2) Global Dictionary Learning—In this study, a baseline image was reconstructed from the normal-dose sinogram using the filtered backprojection (FBP) algorithm to construct a global dictionary, and the performance of the proposed techniques was evaluated with other low-dose sinograms. Because of the discrepancy in the normal and low-dose X-ray kVp settings and injection of the contrast agent, the attenuation maps of the low-dose images were quite different from that of the baseline image. Also, the physiological motion of the sheep most likely introduced structural differences. As such, this group of sinograms actually offers a challenging opportunity to evaluate the robustness of GDSIR.

First, a set of overlapping patches were extracted from the lung region in the baseline image. The patch size was of 8×8 pixels. The patches with very small variance were removed from the extracted patch set. The direct current (dc) component was removed from each patch. Then, a global dictionary of 256 atoms was constructed using the online dictionary learning method with a fixed sparsity level $L_0^D = 5$ [40]. The lung region and the final dictionary are shown in Fig. 1. Finally, a dc atom was added to the dictionary.

3) Low-Dose Results—For comparison, five reconstruction techniques were applied to the aforementioned low-dose sinograms. As the benchmark, low dose images were reconstructed using the FBP method. The corresponding reconstructions using the other four reconstruction techniques were described as follows.

First, the GDSIR algorithm with a prelearned global dictionary (Section IV-A2) was employed to reconstruct low-dose images with the following empirical parameters: $\nu = 0.04$, $\lambda = 2.5 \times 10^{-5}$, and $L_0^S = 25$. The initial image was from the FBP method. An ordered-subset strategy was used [8]. The number of subsets was 40. The iterative process was stopped after 50 iterations.

Then, the ADSIR algorithm was tested with the same low-dose sinograms. In each iteration, the dictionary was learned in real-time from the set of patches extracted from an intermediate image. The parameters for dictionary learning were the same as those in Section IV-A2. Because there was strong noise in an intermediate image, the atoms in this dictionary are noisy. Therefore, the error control item for ADSIR was made smaller than that for GDSIR in which the dictionary was learned from the normal-dose image. On the

other hand, since the dictionary was learned from the reconstructed image itself, there was no need to use many atoms to capture the structures. The sparsity level parameter L_0^S was made much smaller than that for GDSIR.

Taking all these factors into account, the parameter were empirically chosen as $\lambda = 0.04$, $\nu = 1 \times 10^{-5}$, and $L_0^S = 5$. The number of subsets was again 40. The iterative process was also stopped after 50 iterations.

Third, the popular TV regularization algorithm was included to demonstrate the merits of the proposed methods. For that purpose, the TV minimization constraint wasused as the regularization term in (12), and enforced using the soft-threshold filtering based alternating minimization algorithm [17], [45]. We denote this method as TVSIR.

Fourth, to evaluate the effect of the statistical reconstruction technique in this dictionary learning based reconstruction framework, we replaced the log-likelihood term $\sum_{i=1}^{I} (w_i/2)\left([\mathbf{A}\mu]_i - \widehat{l_i}\right)^2$ in (13) with an unweighted l_2-norm data fidelity term $(1/2)\|\mathbf{A}\mu - \widehat{l}\|_2^2$. The global dictionary based algorithm was modified with this constant weighting scheme, which we denote as GDNSIR. The adaptive dictionary based algorithm can be modified in a similar way. The regularization parameter was empirically set to $\lambda = 0.35$ given the weight change of the data fidelity term in the objective function.

The results from a representative low-dose sinogram are in Fig. 2. It can be seen that there is strong noise in the FBP reconstruction, and streak artifacts along high attenuation structures, such as around bones. This kind of streak artifacts can be easily identified from the difference between the FBP image and the results with the SIR methods. The dictionary learning based algorithms generally performed well with low-dose data. While GDSIR did better in preserving structures and suppressing noise, ADSIR kepts lightly more structures than GDSIR (see the region indicated by the arrow "A"). ADSIR generated a little less uniformity than GDSIR in the whole image (see the region indicated by the arrow "B"), and some edges were obscurer than those with GDSIR. The performance of GDNSIR was not much different from that of GDSIR. However, there were some streak artifacts with GDNSIR as in the FBP reconstruction (see the difference from the FBP image), especially around the bone (see the region indicated by the arrow "C"). The image reconstructed by TVSIR had much less noise than the FBP result, but it was a little blocky and had an inferior visibility compared to the dictionary learning based methods (see the regions indicated by the arrows "D" and "E"). Some bony structures in the TVSIR result were obscure or invisible (see the region indicated by the arrow "F").

4) Few-View Test—Reducing the number of projection views is an important strategy to reduce image time and radiation dose, giving the few-view problem. To evaluate the proposed dictionary learning based algorithms for few-view tomography, the number of low-dose views was down-sampled from 1160 to 580, 290 and 116, respectively. The GDSIR and ADSIR methods were then applied. Also, the FBP and TVSIR methods were performed for comparison. The results are in Fig. 3.

It is seen that the FBP reconstruction results became worse and worse when the number of views was gradually decreased from 1160 to 116. The GDSIR, ADSIR and TVSIR results were much better than the FBP reconstruction. In the case of 580 views, the GDSIR and ADSIR results were almost as good as that reconstructed from 1160 views in Fig. 2. However, in the cases of 290 and 116 views, some details were lost. The TVSIR results always had more noise and fewer structures than the dictionary learning based reconstructions (see the regions indicated by the arrows).

5) Plots of the Terms in the Objective Function—To monitor the convergence of the proposed dictionary learning based algorithms, we took GDSIR as an example. Fig. 4 plots the convergence curves of the log-likelihood term $\sum_{i=1}^{I} w_i([\mathbf{A}\mu]_i - \widehat{l_i})^2$ and the sparse representation error term $\sum_s \|\mathbf{E}_s\mu - \mathbf{D}\alpha_s\|_2^2$ with respect to the iteration number. These curves show that both the log-likelihood term and the sparse representation error term would decrease monotonically. After ~40 iterations, both the two terms changed little with further iterations. Practically, the reconstruction process can be stopped after a fixed number of iterations when the log-likelihood term and sparse representation error term do not decrease insignificantly.

B. Numerical Simulations

To evaluate the performance of the proposed algorithms, we conducted two comprehensive numerical studies. One is the sheep lung simulation to evaluate the performance of the proposed methods and demonstrate the parameter selection. The other is the human thorax simulation focused on the robustness of GDSIR.

1) Sheep Lung Simulation

Data Acquisition: In this study, the FBP reconstruction from the normal-dose sinogram in Section IV-A2 [see Fig. 1(a)] was chosen as a numerical phantom. A fan-beam geometry was defined with the same parameters as presented in Section IV-A1. A monochromatic source was used. The Poisson noise was superimposed onto the raw projection data to synthesize low-dose projections assuming 1.0×10^4 photons from the X-ray source towards each detector element. Furthermore, this set of low-dose data was down-sampled from 1160 views to 580, 290, and 116 views, respectively.

Image Reconstruction: The FBP, GDSIR, ADSIR, and TVSIR algorithms were all performed on each of the four datasets. For GDSIR, the global dictionary was learned from the patches extracted from the TVSIR result in Fig. 2. The key details for dictionary learning were the same as that in Section IV-A2. The numerical phantom, the image for dictionary learning and the learned dictionary are shown in Fig. 5.

All the reconstructed images and associated differences are shown in Fig. 6. The performances of these algorithms can be seen from not only the reconstructed images (the regions indicated by the arrows) but also the difference images. Clearly, the FBP algorithm had the worst results, and TVSIR had more noise and was more blocky than the dictionary learning based methods. When the number of views was decreased, heavy noise and artifacts compromised adaptive dictionary learning, which caused the loss of some fine structures in the ADSIR results. As a result, it seems that ADSIR works better for low-dose datasets while GDSIR is preferred in the cases of ultra-low-dose datasets.

Quantitative Indexes Evaluation: The aforementioned algorithms were quantitatively evaluated using two indexes. One is the root mean square error (RMSE)

$$\text{RMSE} = \sqrt{\frac{\sum_{j=1}^{J}(\mu_j^r - \mu_j^*)^2}{J}} \quad (26)$$

where μ_j^r is a reconstructed value, μ_j^* is the phantom value of the jth pixel, and J is the number of pixels in an image. The other is the image quality assessment (IQA) index for

structural similarity (SSIM) (the closer to 1, the higher structural similarity) [46], which is proved to be consistent with visual perception. With the image phantom as the reference, we evaluated all the reconstructed images in terms of the RMSE and SSIM indexes.

The RMSE and SSIM indexes of these results are shown in Tables IV and V, respectively. Quantitatively, GDSIR and ADSIR outperformed TVSIR, all of which outperformed FBP. GDSIR and ADSIR had basically the same performance. In the cases of 1160 and 580 views, ADSIR kept slightly less noise and more features than GDSIR, but in the cases of 290 and 116 views GDSIR did better than ADSIR. In brief, GDSIR and ADSIR would produce less noise and more structures than the TVSIR and FBP.

Parameter Comparison: To evaluate how the parameter selection affects the performance of GDSIR, here we describe our analysis on the reconstructed results from 290 views as an example. The regularization parameter λ, the sparsity level L_0^S and the precision level were first empirically selected as Case 1 in Table VI. Then, these three parameters were individually perturbed around that in Case 1 to form Cases 2–7, respectively. Finally, we kept the parameters as Case 1, but changed the global dictionary to the dictionary learned from the phantom itself (that is, the dictionary in Section IV-A2) to form Case 8.

The reconstructed results and difference images are in Fig. 7. The RMSE and SSIM indexes are in Table VI. It can be seen that there are not much differences among Case 1–5 and 7. Case 6 is the worst one, which was caused by the smaller precision level ε. A too small precision level would introduce fake structures, and keep more noise. Case 8 is the best one. The reason is that the global dictionary in this case was learned from the phantom image itself, and most useful structures were kept. From these results, we see that decent results can be obtained as long as the parameters do not deviate from the proper ranges as summarized in Table I.

Computational Cost Comparison: Our codes ran on a PC with one i7–2600 CPU and a 4 GB RAM. In Table VII, we recorded the average computational times for the image updating (including forward and backward projections), OMP, dictionary learning and TV minimization steps in one iteration with 1160 views. It can be seen that an OMP step cost about half of the time for image updating, while the TV minimization only took one percent of the time for OMP. For TVSIR, most of the time was consumed in image updating. GDSIR took almost 1.5 times of the overhead of TVSIR. For ADSIR, the dictionary learning step needed almost twice of the OMP overhead.

2) Human Thorax Simulation

Data Acquisition: In this study, a human thorax perfusion image was downloaded from [47] to generate a numerical phantom. The thorax region in the original image was kept, including 351×512 pixels. The pixel values were mapped onto [0 2], and the phantom covered a region of 20.57×30 cm^2 (Fig. 8). A fan-beam geometry with equi-angularly distributed detectors was assumed. The distance from the X-ray source to the system origin was 55cm. Over a 360° full scan, 1100 projections were uniformly collected. For each projection, 600 detector elements spanned an FOV of 19.71 cm in radius. The Poisson noise was superimposed onto raw projections to synthesize low-dose data. Furthermore, this low-dose dataset was down-sampled from 1100 views to 550, 275, and 110 views, respectively.

The reconstructed images were in the same matrix as the original phantom. The sparsity constraint was enforced. In this simulation, we did not use any patient specific information, and the dictionary in the sheep lung study [Fig. 1(b)] was taken as the global dictionary for GDSIR reconstruction. Because the sheep lung image is significantly different from the

human thorax image, the robustness of the GDSIR algorithm was actually over-tested in this arrangement.

Reconstructed Results: The results reconstructed by the FBP and the proposed dictionary learning based algorithms as well as the associated differences images are shown in Fig. 9. It can be seen that the dictionary learning based reconstructions suppressed noise (such as the region indicated by the arrow "a" in Fig. 8) and the streak artifacts (the region indicated by the arrow "b" in Fig. 8), and preserved structures effectively (such as the region indicated by the arrow "c" in Fig. 8, which is a superior vena cava highlighted by the contrast agent, and can be clearly reconstructed using the GDSIR and ADSIR algorithms in the case of 110 views). Surprisingly, the global dictionary learned from the sheep lung image performed well in this human thorax study. Also, GDSIR had similar results as those with ADSIR. In the case of 110 views, due to the degraded dictionary quality, ADSIR gave worse results than GDSIR.

C. Clinical Study

1) Data Acquisition: Under the approval of the Institutional Review Board (IRB), Wake Forest University Health Science, the proposed algorithms were evaluated with a group of existing raw projections collected in cardiac perfusion CT studies for other purposes. The patient was scanned by a state-of-the-art GE discovery CT750 HD scanner in a helical mode to examine the coronary artery. After appropriate pre-processing, we obtained 64-slice fan-beam sinograms. These slices had substantial structural differences. The radius of the scanning trajectory was 53.852 cm. Over a 360° range, 2200 projections were uniformly acquired. For each projection, 888 detector elements were equi-angularly distributed over an FOV of 24.92 cm in radius. In this study, images were reconstructed in a matrix of 600×800 pixels over a 45×60 cm^2 region.

2) Global Dictionary Learning: To evaluate the robustness of the GDSIR, we randomly selected slice #20 for global dictionary learning and use the resultant dictionary to reconstruct other slices without loss of generality. To provide an excellent training set for construction of a global dictionary, we first reconstructed slice #20 using the TVSIR algorithm from all the 2200 views, and removed speckle noise by a 3×3 median filter. Then, the patches were extracted from the refined image to construct a global dictionary, as described in Section IV-A2. The image used to extract patches and construct the dictionary is shown in Fig. 10.

3) Few-View Test: All the 64 sinograms were down-sampled from 2200 views to 440 and 220 views, respectively. Images were reconstructed from the down-sampled datasets using the FBP, TVSIR, GDSIR, and ADSIR algorithms, respectively. As an example, here we present the results from slice #32.

In the case of 440 views, the parameters for GDSIR were chosen as $\lambda = 0.3$, $\nu = 5 \times 10^{-8}$, $L_0^S = 10$, and the parameters for ADSIR were as $\lambda = 0.3$, $\nu = 3 \times 10^{-8}$, $L_0^S = 8$. Fig. 11 shows the reconstructed results for slice #32 after 100 iterations with 10 subsets. It is seen that while the FBP result had more noise and streak artifacts, the TVSIR result was substantially better. However, the TVSIR result was slightly blocky and had more noise than the GDSIR and ADSIR counterparts.

In the case of 220 views, the parameters for GDSIR were chosen as $\lambda = 0.4$, $\nu = 5 \times 10^{-8}$, $L_0^S = 10$, and the parameters for ADSIR were as $\lambda = 0.4$, $\nu = 5 \times 10^{-8}$, $L_0^S = 8$. The reconstructed images for slice #32 after 200 iterations with five subsets are in Fig. 12. Our study seems indicating that useful results can be obtained from 1/10 of the original views

using the proposed dictionary learning based reconstruction algorithms. The magnified cardiac regions are in Fig. 13. Clearly, GDSIR and ADSIR outperformed FBP and TVSIR, more effectively suppressing image noise while keeping subtle structures.

V. Discussions and Conclusion

Based on the CS theory, the TV regularization method was widely used for CT reconstruction, and produced good results from incomplete and noisy data. Our results in Section IV have also verified that the TV regularization method can generate better results than the conventional FBP algorithm from low-dose and/or few-view datasets. However, because the TV regularization method is based on a piecewise constant image model, it may produce blocky results in practical applications when there is too much noise. Moreover, the TV constraint uniformly penalizes the image gradient, and is not capable of distinguishing structural details from noise and artifacts. These problems dampen the enthusiasm for the clinical application of the TV regularization method. While many efforts have been devoted to improving the TV based CT reconstruction [14], [17], [18], [45], [48], the soft-threshold method [17], [45] was selected as an example to produce the TVSIR results in our experiments for comparison with our proposed dictionary learning based algorithms.

Different from the TV regularization method, the dictionary learning approach aims at capturing localized structural information and suppressing image noise. The sparse representation in terms of a redundant dictionary is able to keep the atoms reflecting structural features and avoid the other atoms. Use of the dictionary-based sparse representation as the regularization term for SIR is a new mechanism to improve image quality. Moreover, any missed structural information due to the enforcement of the sparsity constraint will be compensated for in the subsequent updating steps. SIR is very effective to eliminate streak artifacts, and works well with a dictionary. In principle, the dictionary learning process should lead to a sparser representation of an underlying image in a specific application. Our simulation results for monoenergetic imaging and for one imaging region have shown that both the GDSIR and ADSIR algorithms outperformed TVSIR from low-dose and/or few-view data. However, it should be noted that there remain differences between the true image and the GDSIR and ADSIR results, as seen in Section IV-B. The basic idea of the dictionary learning based approach is to find a best match to a true image from the dictionary-spanned image space. When the true image is outside the dictionary image space, the reconstructed image can be viewed as its projection on the dictionary image space with an unavoidable error. Therefore, some structures may be lost while artifacts may be introduced although the reconstructed results often have less noise and more structural information. A proper dictionary should represent the structural information of an object as much as possible. In this way, the reconstruction with a sparse representation in terms of the dictionary can perform well. With a global dictionary, the structural differences between its training images and a true image would affect the final reconstruction quality. Practically, both the aforementioned real data and simulation studies have demonstrated that GDSIR perform robustly well with a dictionary learned from a quite different image. Usually, it is not difficult to prepare an excellent training set with sufficiently many structural features and less noise for construction of a global dictionary. Since GDSIR does not need to update the dictionary in each iteration step, it is much faster than ADSIR. On the other hand, it is necessary to use an adaptive dictionary when a global dictionary does not match a specific application closely. As indicated in Section IV, ADSIR can reveal details which are invisible in the GDSIR reconstruction. However, the image noise from dictionary learning may affect the final reconstruction. When the sinogram is rather noisy, the dictionary learning process will no longer be able to extract high-quality structural information, and will degrade the image quality.

In our proposed dictionary learning based algorithms, there are some parameters to be determined. How to optimize parameters is an interesting problem for almost all the regularized reconstruction algorithms. Similar to other regularized methods, the regularization parameter λ is very important. A simple and effective way is to select λ empirically based on heuristic guidelines. For a fixed scanning geometry, λ should be increased with the decrement of the projection SNR. Besides, the proposed objective function can be transformed into a classical regularization problem as shown in (21), which makes it possible to use established general methods in our context. In this paper, the OMP method was used for sparse representation with two parameters: sparsity L_0^S and precision ε. These two parameters are related to not only the dictionary but also the image features and noise level. A too small L_0^S or too large ε could not represent the structures sufficiently. On the other hand, a too large L_0^S or too small ε may introduce fake structures. Usually, L_0^S is related to the sparsity level of the dictionary learning L_0^D, and ε depends on the image noise level. As far as the dictionary is concerned, the dictionary size K and patch size N should be determined. The empirical values $N = 8 \times 8$, $K = 256$ have been well applied in many applications and also validated in our experiments. Additionally, it appears possible to analyze the relationship between noise and structural information in each iteration and develop an adaptive parameter selection method.

Some image preprocessing techniques can be involved in the proposed algorithms. Because a predetermined dictionary is not required before the ADSIR reconstruction, and the dictionary is learned from intermediate images during the iterations, there is little image preprocessing issue. For the GDSIR reconstruction, we need to predetermine a global dictionary from a training set and keep the dictionary unchanged during the iterations. To achieve a best performance, the dictionary is generally learned from well-reconstructed images with structures similar to objects of interest. In practice, the images for dictionary training may contain non-structured noise or structure-like artifacts such as speckle noise. For the non-structured noise, we do not need any image preprocessing because the dictionary learning process is inherently good at suppressing noise (e.g., the sheep lung study in Section IV-A). For artifacts, an image preprocessing step is helpful to avoid unexpected atoms. In the clinical study for this paper, there are speckle noises in the TVSIR results. This kind of noise would lead to speckle-like atoms in the global dictionary, which may match speckle noise well in the sparse representation and lead to speckle noise in the GDSIR results. Since median filtering is a conventional way to remove speckle noise without losing important structures, a 3×3 median filter was used to improve the image quality for global dictionary learning.

The computational cost is a common problem for all the iterative reconstruction methods. The total cost of any iterative algorithm can be expressed as the product between the computation time for each iteration and the total iteration number. In our implementation the image updating step adopted a finite-detector-based high-accuracy area model for the forward and backward projections [49]. If there is no requirement for high resolution, the well-known distance-driven method can be used to replace the time-consuming area model [50], which can speed up the image updating step by an order of magnitude. The OMP and dictionary learning operations take comparable time as the image updating step. Rapidly, GPU and other hardware based acceleration methods have been developed, which can be applied to our algorithms.

The results from low-dose and/or few-view datasets in this paper are only exemplary applications of the proposed dictionary learning based reconstruction methods. We believe that they will perform well for other CT applications, such as limited-angle, interior tomography, and so on. Besides, more clinical applications are possible, such as ultra-low-

dose lung cancer screening. While our work has been presented in the context of X-ray CT, our results could be extended to other modalities such as phase-contrast CT, PET, and SPECT.

In conclusion, we have proposed two novel dictionary learning based SIR algorithms for low-dose CT and other applications. The sparsity constraint has been introduced into the SIR framework in terms of either a global or adaptive dictionary. Our approach has produced promising results in terms of preserving structural details and suppressing image noise. Especially, it has been verified that the proposed approach has outperformed the TV minimization method for low-dose CT. However, there is no proof that this is true for all kinds of structures. Further improvement and extension are underway.

Acknowledgments

References

1. Brenner DJ, Elliston CD, Hall EJ, Berdon WE. Estimated risks of radiation-induced fatal cancer from pediatric CT. Am J Roentgenol. 2001; 176:289–296. [PubMed: 11159059]

2. de Gonzalez AB, Darby S. Risk of cancer from diagnostic X-rays: Estimates for the UK and 14 other countries. Lancet. 2004; 363:345–351. [PubMed: 15070562]

3. Brenner DJ, Hall EJ. Computed tomography–An increasing source of radiation exposure. New Eng J Med. 2007; 357:2277–2284. [PubMed: 18046031]

4. Hsieh J. Adaptive streak artifact reduction in computed tomography resulting from excessive X-ray photon noise. Med Phys. 1998; 25:2139–2147. [PubMed: 9829238]

5. Kachelriess M, Watzke O, Kalender WA. Generalized multidimensional adaptive filtering for conventional and spiral single-slice, multi-slice, and cone-beam CT. Med Phys. 2001; 28:475–490. [PubMed: 11339744]

6. La Rivière P. Penalized-likelihood sinogram smoothing for low-dose CT. Med Phys. 2005; 32:1676. [PubMed: 16013726]

7. Wang J, Li T, Lu H, Liang Z. Penalized weighted least-squares approach to sinogram noise reduction and image reconstruction for low-dose X-ray computed tomography. IEEE Trans Med Imag. Oct; 2006 25(10):1272–1283.

8. Kamphuis C, Beekman F. Accelerated iterative transmission CT reconstruction using an ordered subsets convex algorithm. IEEE Trans Med Imag. Dec; 1998 17(6):1101–1105.

9. Nuyts J, De Man B, Dupont P, Defrise M, Suetens P, Mortel-mans L. Iterative reconstruction for helical CT: A simulation study. Phys Med Biol. 1998; 43:729–737. [PubMed: 9572499]

10. Erdogan H, Fessler J. Monotonic algorithms for transmission tomography. IEEE Trans Med Imag. Sep; 1999 18(9):801–814.

11. Thibault J, Sauer K, Bouman C, Hsieh J. A three-dimensional statistical approach to improved image quality for multislice helical CT. Med Phys. 2007; 34:4526. [PubMed: 18072519]

12. Donoho DL. Compressed sensing. IEEE Trans Inf Theory. 2006; 52:1289–1306.

13. Candes EJ, Romberg J, Tao T. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. IEEE Trans Inf Theory. Feb; 2006 52(2):489–509.

14. Chen GH, 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; 35:660. [PubMed: 18383687]

15. Yu HY, Wang G. Compressed sensing based interior tomography. Phys Med Biol. 2009; 54:2791–2805. [PubMed: 19369711]

16. Yu HY, Yang JS, Jiang M, Wang G. Supplemental analysis on compressed sensing based interior tomography. Phys Med Biol. 2009; 54:N425–N432. [PubMed: 19717891]

17. Xu Q, Mou X, Wang G, Sieren J, Hoffman E, Yu H. Statistical interior tomography. IEEE Trans Med Imag. May; 2011 30(5):1116–1128.

18. Ritschl L, Bergner F, Fleischmann C, Kachelrieß M. Improved total variation-based CT image reconstruction applied to clinical data. Phys Med Biol. 2011; 56:1545. [PubMed: 21325707]

19. Donoho DL, Elad M. Maximal sparsity representation via l1 minimization. Proc Nat Aca Sci. 2003; 100:2197–2202.

20. Elad M, Aharon M. Image denoising via sparse and redundant representations over learned dictionaries. IEEE Trans Image Process. Dec; 2006 15(12):3736–3745. [PubMed: 17153947]

21. Mairal J, Elad M, Sapiro G. Sparse representation for color image restoration. IEEE Trans Image Process. Jan; 2007 17(1):53–69. [PubMed: 18229804]

22. Mairal J, Sapiro G, Elad M. Learning multiscale sparse representations for image and video restoration. SIAM Multiscale Model Simulat. 2008; 7:214–241.

23. Wright J, Yang AY, Ganesh A, Sastry SS, Ma Y. Robust face recognition via sparse representation. IEEE Trans Pattern Anal Mach Intell. Feb; 2008 31(2):210–227. [PubMed: 19110489]

24. Chen Y, Ye X, Huang F. A novel method and fast algorithm for MR image reconstruction with significantly under-sampled data. Inverse Problems Imag. 2010; 4:223–240.

25. Ravishankar S, Bresler Y. MR image reconstruction from highly undersampled -space data by dictionary learning. IEEE Trans Med Imag. May; 2011 30(5):1028–1041.

26. Xu, Q.; Yu, H.; Mou, X.; Wang, G. Dictionary learning based low-dose X-ray CT reconstruction. Proc. 11th Int. Meeting Fully Three-Dimensional Image Reconstruct. Radiol. Nucl. Med.; Potsdam, Germany. 2011. p. 258-261.

27. Mallat SG, Zhang Z. Matching pursuits with time-frequency dictionaries. IEEE Trans Signal Process. 1993; 41:3397–3415.

28. Chen S, Billings S, Luo W. Orthogonal least squares methods and their application to non-linear system identification. Int J Control. 1989; 50:1873–1896.

29. Pati YC, Rezaiifar R, Krishnaprasad P. Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. Proc 27th Asilomar Conf Signals Syst Comput. 1993; 1:40–44.

30. Chen SS, Donoho DL, Saunders MA. Atomic decomposition by basis pursuit. SIAM J Sci Comput. 1999; 20:33–61.

31. Aharon M, Elad M, Bruckstein A. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. IEEE Trans Signal Process. Nov; 2006 54(11):4311–4322.

32. Mairal J, Bach F, Ponce J, Sapiro G. Online learning for matrix factorization and sparse coding. J Mach Learn Res. 2010; 11:19–60.

33. Sauer K, Bouman C. A local update strategy for iterative reconstruction from projections. IEEE Trans Signal Process. Feb; 1993 41(2):534–548.

34. Elbakri IA, Fessler JA. Statistical image reconstruction for polyenergetic X-ray computed tomography. IEEE Trans Med Imag. Feb; 2002 21(2):89–99.

35. Li T, Li X, Wang J, Wen J, Lu H, Hsieh J, Liang Z. Nonlinear sinogram smoothing for low-dose X-ray CT. IEEE Trans Nucl Sci. Oct; 2004 51(5):2505–2513.

36. Little K, Vargas P, Riviere PL. Weight choice in PWLS algorithms for emission and transmission tomography. IEEE Nucl Sci Symp Conf Rec. 2010:M17–3.

37. Whiting BR, Massoumzadeh P, Earl OA, O'Sullivan JA, Snyder DL, Williamson JF. Properties of preprocessed sinogram data in X-ray computed tomography. Med Phys. 2006; 33:3290. [PubMed: 17022224]

38. Thibault JB, Sauer KD, Bouman CA, Hsieh J. A three-dimensional statistical approach to improved image quality for multislice helical CT. Med Phys. 2007; 34:4526. [PubMed: 18072519]

39. Bouman C, Sauer K. A generalized Gaussian image model for edge-preserving MAP estimation. IEEE Trans Image Process. Jul; 1993 2(3):296–310. [PubMed: 18296219]

40. [Online]. Available: http://www.di.ens.fr/willow/SPAMS/

41. Bertero, M.; Boccacci, P. Introduction to Inverse Problems in Imaging. Vol. 5.6. Oxford, U.K: Taylor Francis; 1998.

42. Miller K. Least squares methods for ill-posed problems with a prescribed bound. SIAM J Math Anal. 1970; 1:52.

43. Golub GH, Heath M, Wahba G. Generalized cross-validation as a method for choosing a good ridge parameter. Technometrics. 1979; 21:215–223.

44. Mairal J, Sapiro G, Elad M. Learning multiscale sparse representations for image and video restoration. SIAM Multiscale Model Simulat. 2008; 7:214–241.

45. Yu HY, Wang G. A soft-threshold filtering approach for reconstruction from a limited number of projections. Phys Med Biol. 2010; 55:3905–3916. [PubMed: 20571212]

46. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: From error visibility to structural similarity. IEEE Trans Image Process. Apr; 2004 13(4):600–612. [PubMed: 15376593]

47. [Online]. Available: http://www.radiolog.at/start.php?bereich=ct_mr&nav=untersuchungen&lw=computertomographie

48. Song J, Liu QH, Johnson GA, Badea CT. Sparseness prior based iterative image reconstruction for retrospectively gated cardiac micro-CT. Med Phys. 2007; 34:4476–4483. [PubMed: 18072512]

49. Yu HY, Wang G. Finite detector based projection model for high spatial resolution. J X-Ray Sci Technol. 2012; 20(2):229–238.

50. De Man B, Basu S. Distance-driven projection and backprojection in three dimensions. Phys Med Biol. 2004; 49:2463–2475. [PubMed: 15248590]
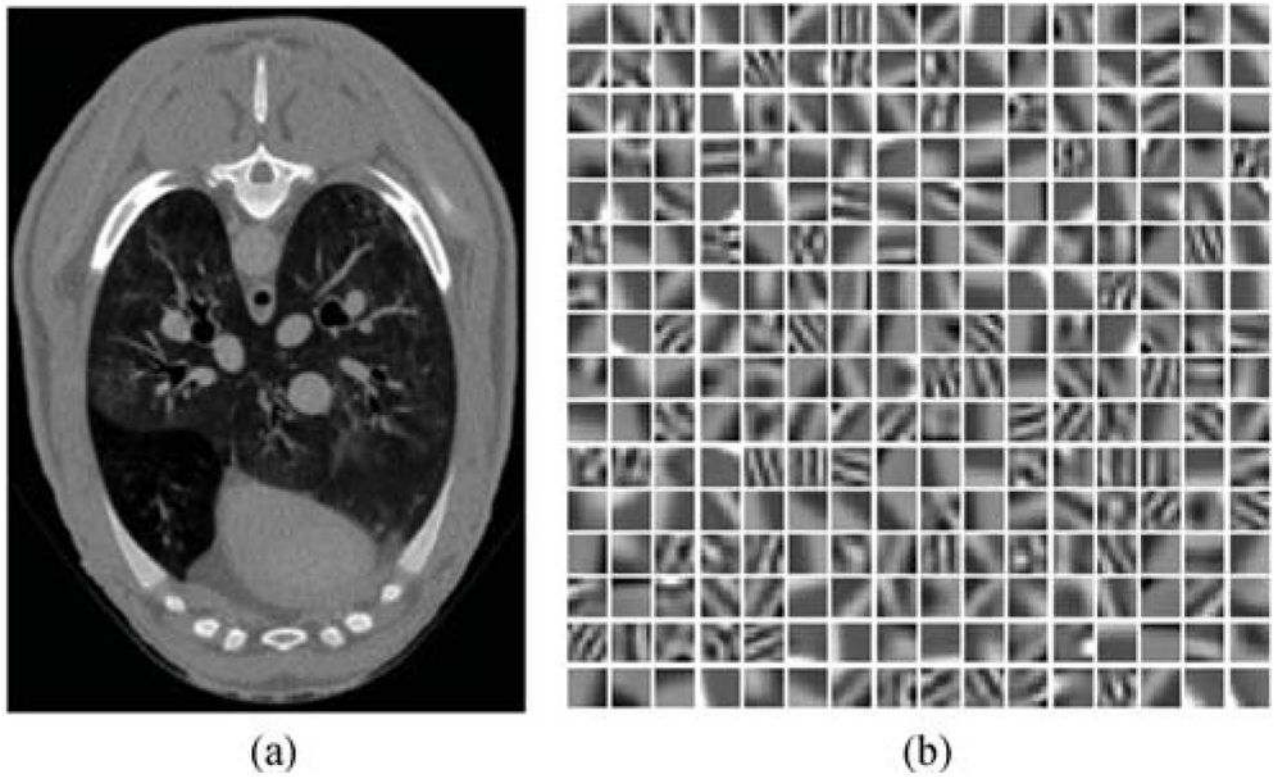
Fig. 1.
Construction of a global dictionary. (a) Sheep lung image with a display window [−700, 800] HU reconstructed from the normal-dose sinogram by the FBP method, which is used to extract the training patches. (b) Learned dictionary consisting of 256 atoms. The attenuation coefficient of water was assumed as 0.018 to convert the reconstructed image (a) into HU and other images throughout this paper.
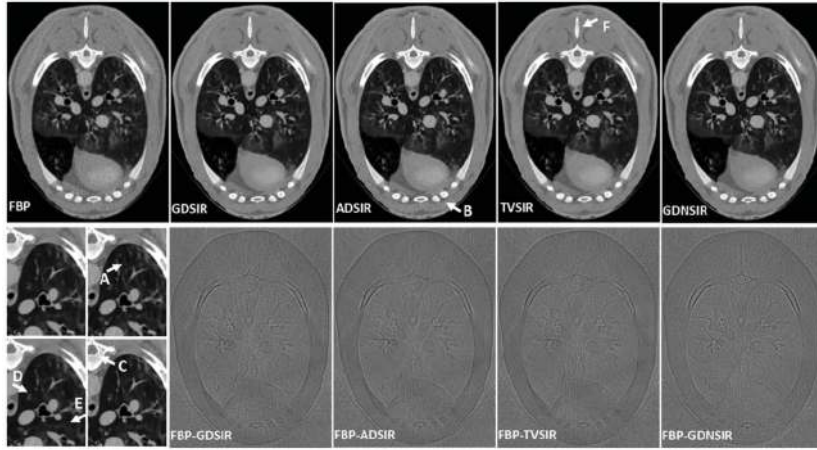
Fig. 2.
Reconstructed images from a representative low-dose sinogram collected in the sheep lung CT perfusion study. The upper row are the images reconstructed using the FBP, GDSIR, ADSIR, TVSIR, and GDNSIR methods (from left to right), respectively. The magnified local regions of them are shown below the FBP results (upper left, upper right, lower left, and lower right correspond to GDSIR, ADSIR, TVSIR, and GDNSIR, respectively). The display window is [−700, 800] HU. The second to fifth images of the bottom column are the difference images between the FBP image and the results by the GDSIR, ADSIR, TVSIR, and GDNSIR with a display window [−556, 556] HU.
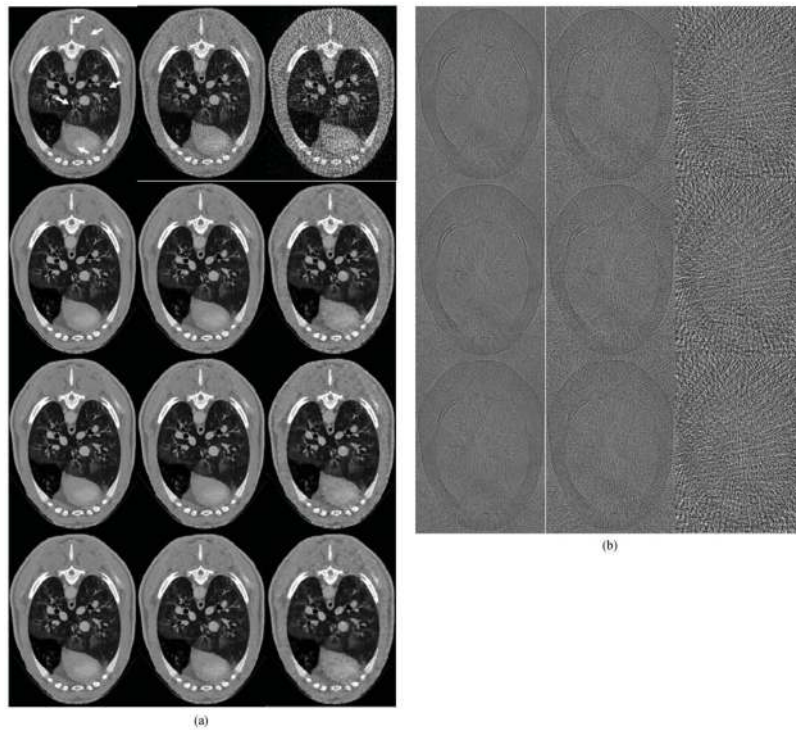
Fig. 3.
Reconstructed images from the few-view low-dose sinograms. From top to bottom of (a), the images are reconstructed by the FBP, GDSIR, ADSIR, and TVSIR methods, respectively. The display window is [−700, 800] HU. From top to bottom of (b), the images are the differences between the FBP results and the corresponding results by the GDSIR, ADSIR, and TVSIR methods, respectively. The display window is [−556, 556] HU. From left to right, the images are reconstructed from 580, 290, and 116 views, respectively.
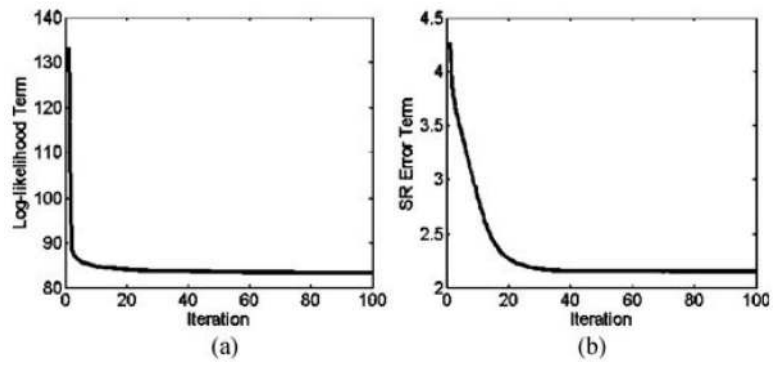
Fig. 4.
Plots of objective functional terms associated with the GDSIR method applied to a representative low-dose sinogram. (a) and (b) are the curves for the log-likelihood and sparse representation errors, respectively.
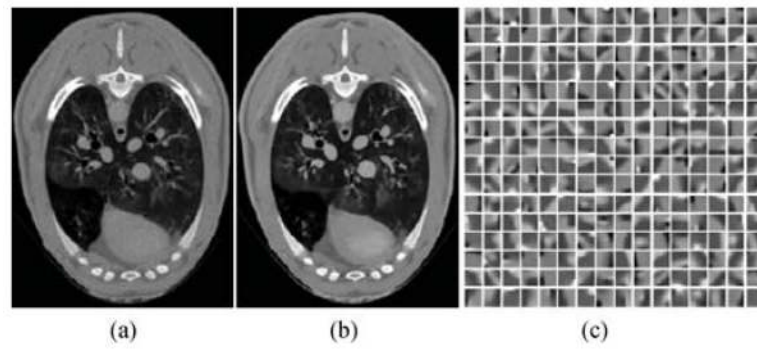
Fig. 5.
Sheep lung simulation study. (a) FBP result shown in Fig. 1(a), which is used as the numerical phantom. (b) TVSIR result shown in Fig. 2, which is used to extract the training patches for global dictionary. The display window is [−700, 800] HU. (c) Global dictionary learned from the extracted patches.
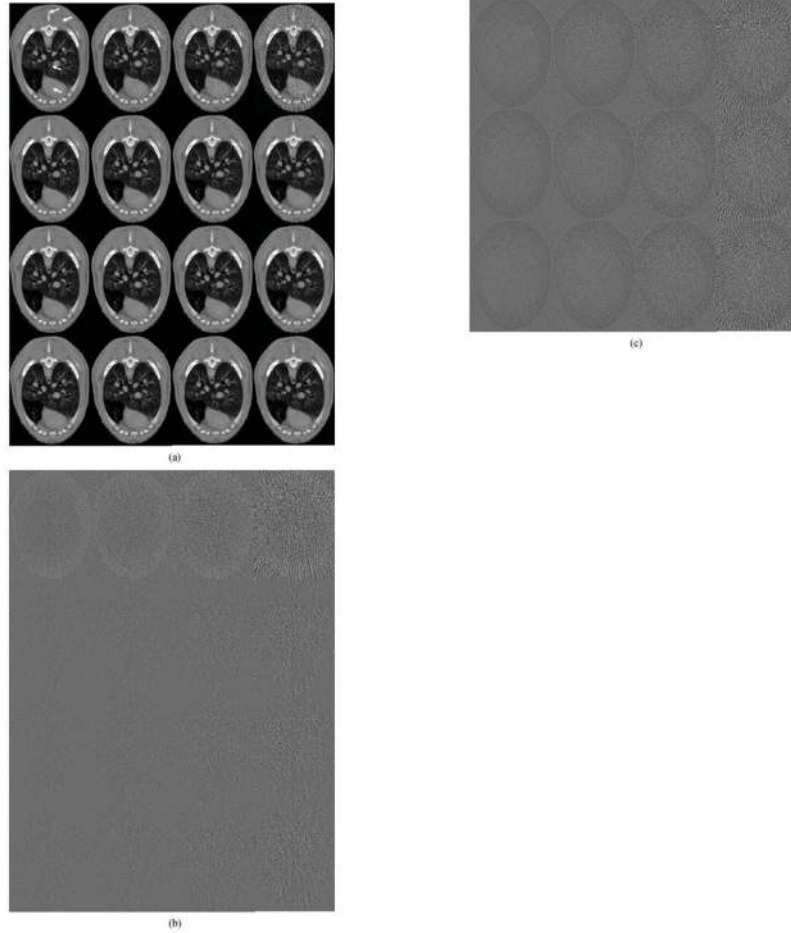
Fig. 6.
Reconstructed images of the sheep lung simulation study. From top to bottom of (a), the results are reconstructed by the FBP, GDSIR, ADSIR, and TVSIR algorithms, respectively. The display window is [−700, 800] HU. From top to bottom of (b), the images are the phantom minus the results by the FBP, GDSIR, ADSIR, and TVSIR algorithms, respectively.
Reconstructed images of the sheep lung simulation study. From top to bottom of (c), the images are the FBP results minus the corresponding results by the GDSIR, ADSIR, and TVSIR algorithms, respectively. The display window is [−556, 556] HU. From left to right, the images are reconstructed from 1160, 580, 290, and 116 views, respectively.
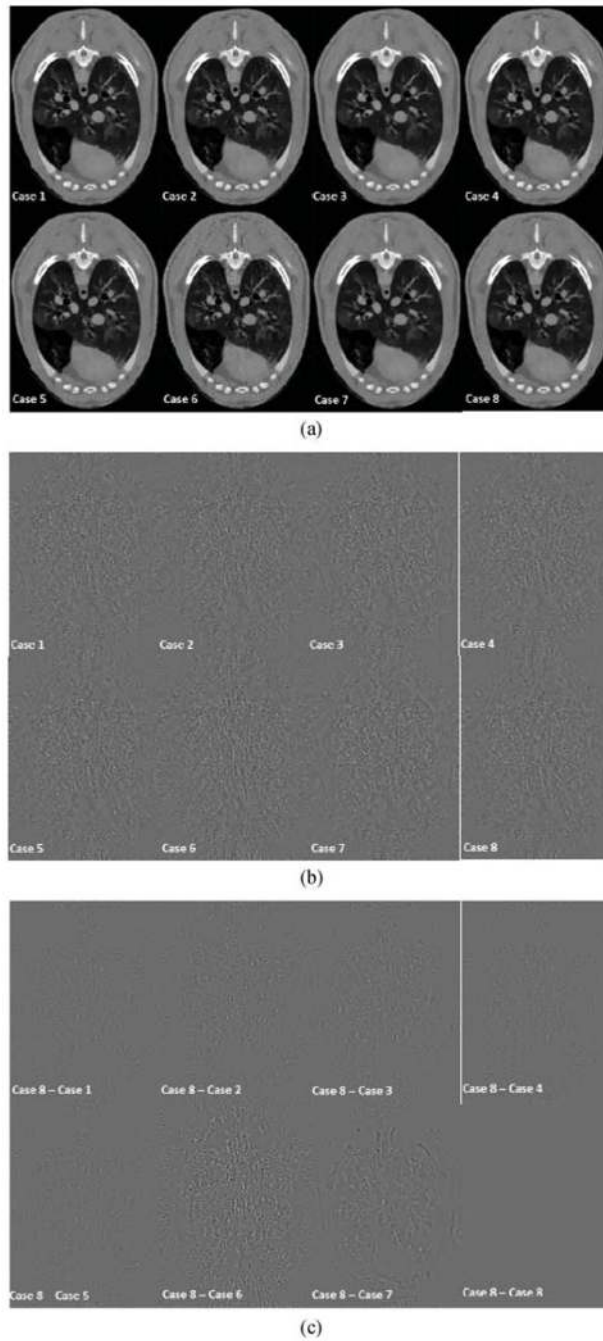
Fig. 7.
The results of numerical simulation from 290 views by the GDSIR algorithm under different parameter selections listed in Table V. (a) Reconstructed images with a display window [−700, 800] HU. (b) Differences images related to the original numerical phantom with a display window [−556, 556] HU. (c) Differences images related to the case 8 with a display window [−228, 228] HU.
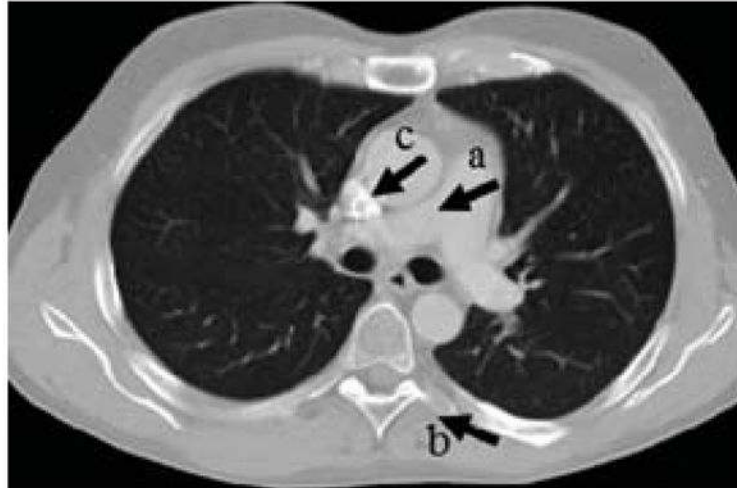
Fig. 8.
Numerical phantom for human thorax simulation study with a display window [0 2].
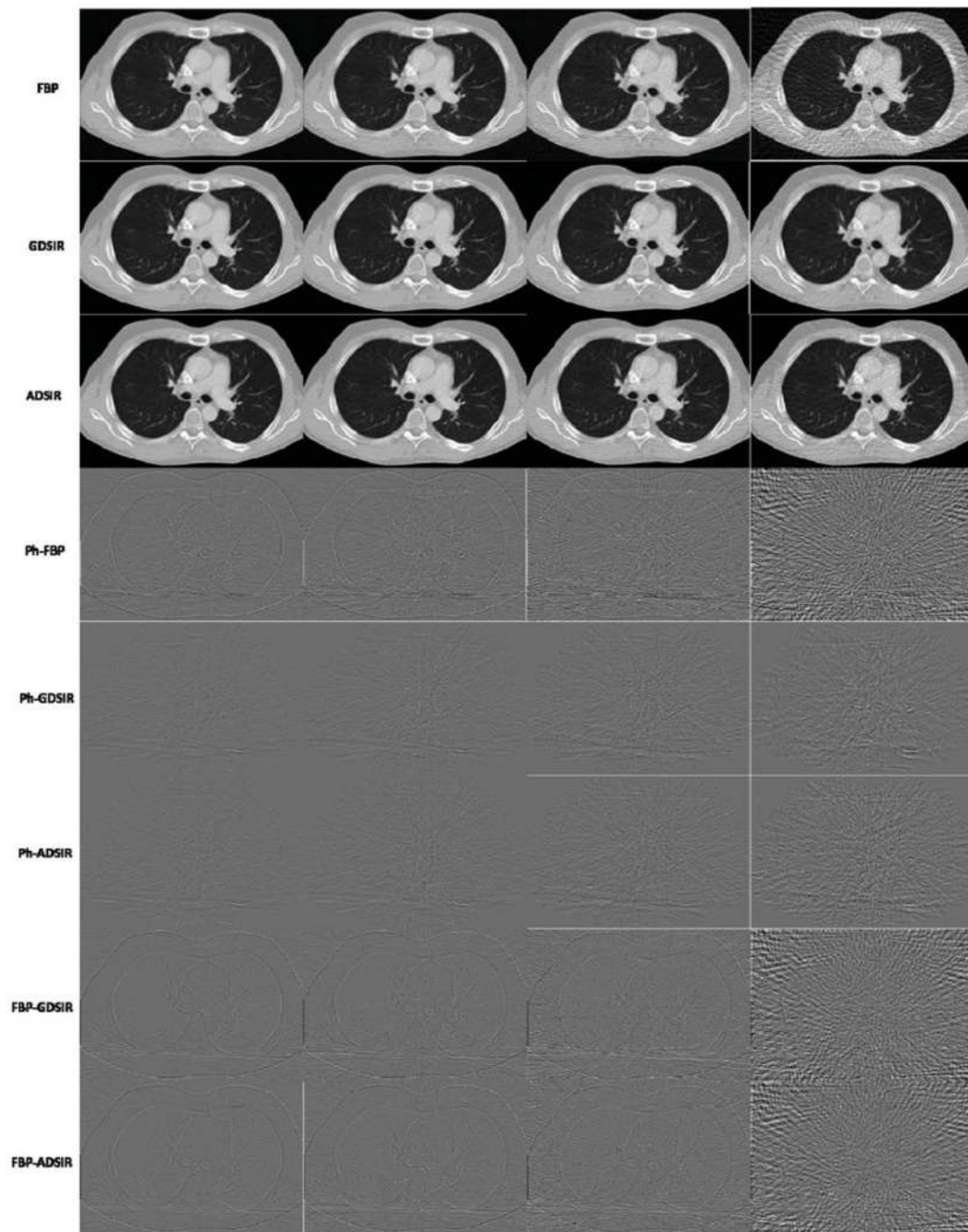
Fig. 9.
The results of human thorax simulation study. The first to third are the reconstructed images with a display window [0 2], the fourth to eighth are the differences images related to the numerical phantom and FBP results with a display window [−0.5 0.5]. From left to right, the images are reconstructed from 1100, 550, 275, and 110 views, respectively.
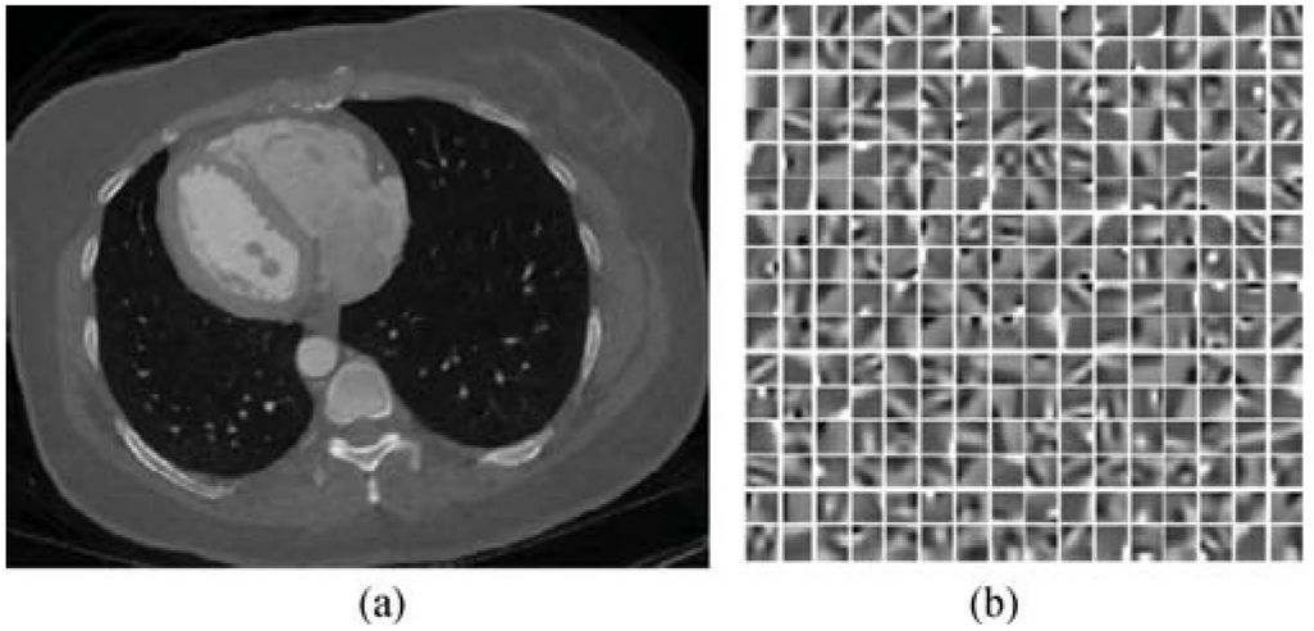
(a)

(b)

Fig. 10.
Construction of a global dictionary based on the slice #20 from a clinical CT scan. (a) is the refined image for the slice #20 for patch extraction with a display window [−1000 1000] HU, and (b) is the learned dictionary consisting of 256 atoms.
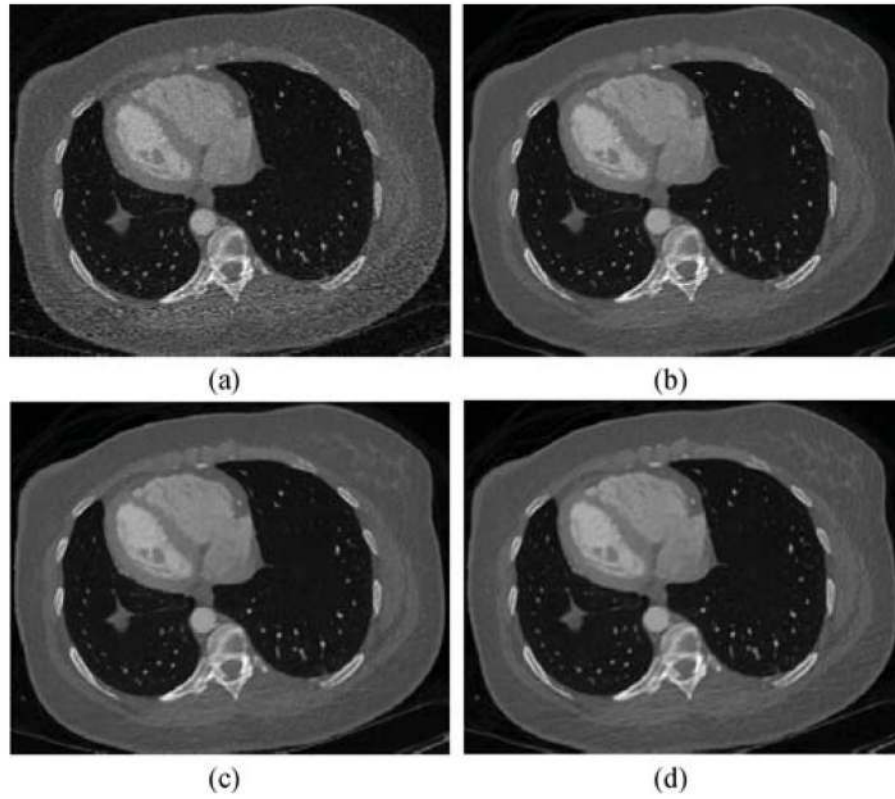
Fig. 11.
Reconstructed images for the Slice #32 from 440 views in the clinical CT scan. (a)–(d) Results reconstructed using FBP, TVSIR, GDSIR, and ADSIR methods, respectively. The display window is [−1000 1000] HU.
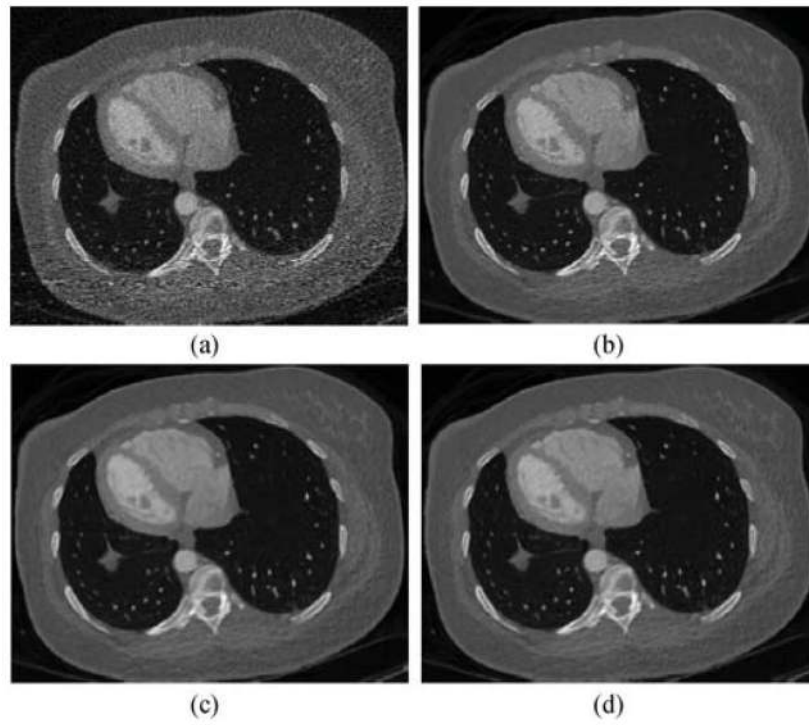
Fig. 12.
Reconstructed images corresponding to those in Fig. 11 but only from 220 views.
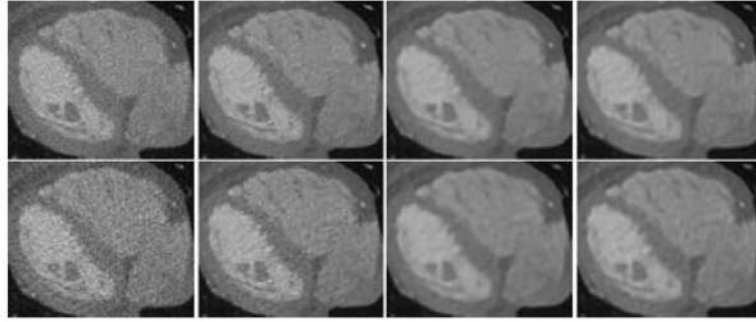
Fig. 13.
Zoomed-in views. The first and second rows are the magnified cardiac region in Figs. 11 and 12, respectively. The first to fourth columns are the results reconstructed by the FBP, TVSIR, GDSIR, and ADSIR methods, respectively.

TABLE I

Summary of the Parameter Selection Guidelines

No.	Variable	Meaning	Criterion/Range
1	λ	Parameter to balance the data fidelity and prior information terms	It should be increased with the noise level. It is usually empirically selected in practice
2	K, N	Number of atoms and number of pixels in a patch	They are selected as N=64 and K=256 in the image processing field, and work well in our study.
3	L_0^D	Sparsity level L_0 for dictionary learning	It is usually set to 5~10 atoms.
4	L_0^S	Sparsity level L_0 for image reconstruction	It is determined by the complexity of an image and the power of a dictionary; $N/2 > L_0^s \geq L_0^D$.
5	ε	Tolerance of the difference between the reconstructed and original images	It is determined by the image noise and the dictionary capability, and normally comparable to the image noise level.

TABLE II

Workflow for the GDSIR Algorithm

Global Dictionary Learning

   Choose parameters for dictionary learning:

   Extract patches to form a training set;

   Construct a global dictionary [40].

Image Reconstruction

   Initialize $\Box^0$, $\Box^0$ and $t = 0$.

   Set parameters $\Box$ $\Box$ and L_0^S;

   While a stopping criterion is not satisfied

     1. Update $\Box^{-1}$ to $\Box$ using (16);

     2. Represent $\Box$ with a sparse $\Box$ using OMP;

Output the final reconstruction.

TABLE III

Workflow for the ADSIR Algorithm

Choose $\square$ $\square$ L_0^S and other parameters;

Initialize $\square^0$, $\mathbf{D}^0$, $\square^0$ and $t = 0$;

While a stopping criterion is not satisfied

   1. Update $\square^{-1}$ to $\square$ using (16);

   2. Extract patches from $\square$ to form a training set;

   3. Construct a dictionary $\mathbf{D}^t$ from the training set [40];

   4. Represent $\square$ with a sparse $\square$ in terms of the dictionary $\mathbf{D}^t$ using OMP;

Output the final reconstruction.

TABLE IV

RMSE Values of the Results Reconstructed by the FBP, TVSIR, GDSIR, and ADSIR Methods, Respectively, in the Sheep Lung Simulation Study (Unit: HU)

Views	FBP	GDSIR	ADSIR	TVSIR
1160	87.71	43.44	**43.01**	46.97
580	117.61	47.27	**46.84**	51.04
290	162.72	**52.93**	53.41	57.48
116	259.86	**67.70**	69.11	70.39

TABLE V

SSIM Values of the Results Reconstructed by the FBP, TVSIR, GDSIR, and ADSIR Methods, Respectively, in the Sheep Lung Simulation Study

Views	FBP	GDSIR	ADSIR	TVSIR
1160	0.5907	0.8468	**0.8485**	0.8457
580	0.4557	0.8303	**0.8321**	0.8292
290	0.3220	**0.8100**	0.8093	0.8047
116	0.1765	0.7708	0.7666	0.7599

TABLE VI

RMSE and SSIM Values of the Results Reconstructed Using GDSIR From 290 Views With Different Parameter Settings in the Sheep Lung Simulation Study

Case No.	σ	L_0^S	σ	Global Dictionary	RMSE (Unit: HU)	SSIM
1	0.17	10	2×10^{-5}	Fig. 5 (c)	53.48	0.8094
2	0.14	10	2×10^{-5}	Fig. 5 (e)	55.07	0.8042
3	0.20	10	2×10^{-5}	Fig. 5 (c)	53.00	0.8107
4	0.17	7	2×10^{-5}	Fig. 5 (c)	52.93	0.8100
5	0.17	13	2×10^{-5}	Fig. 5 (c)	53.83	0.8090
6	0.17	10	1×10^{-5}	Fig. 5 (c)	**60.90**	**0.7856**
7	0.17	10	3×10^{-5}	Fig. 5 (c)	53.27	0.8072
8	0.17	10	2×10^{-5}	Fig. 1 (b)	**51.81**	**0.8127**

TABLE VII

Comparison of the Computational Costs in the Sheep Lung Simulation Study (Unit: s)

Methods	Image updating	OMP	Dictionary learning	TV minimization
GDSIR	6.2	2.9	/	/
ADSIR	6.2	2.6	5.2	/
TVSIR	6.2	/	/	0.032