Abstract

The concerns of the effects of ionizing radiation on patients in medical imaging environments has motivated new low-dose image reconstruction techniques for computed tomography (CT). Since the radiation detection process is a quantum process, lowering the dose to patients inherently reduces the quality of the reconstructed image. A dictionary learning based image reconstruction method is proposed to reduce the noise of low-dose CT images. The proposed method is a statistical image reconstruction process that reconstructs the image from low-dose projection data using a sparsity constraint in terms of a redundant dictionary. A study using a numerical phantom is performed and images generated with the proposed method are compared to traditional filtered backprojection (FBP) and a total-variation (TV) based reconstruction method.

1 Introduction

The paper that this group has chosen to review is titled *Low-Dose X-ray CT Reconstruction via Dictionary Learning* by Xu, et al. [1]. This paper focuses on a novel dictionary learning approach for reconstruction of low-dose CT data. The dictionary learning procedure is implemented along with statistical iterative reconstruction (SIR) to reduce noise and enhance detailed structural features in low-dose x-ray CT data.

X-ray CT is a commonly used medical imaging modality used for diagnosis and intervention. Since ionizing radiation can be harmful to patients [2], one would like to minimize the dose a patient receives during a CT scan. The radiation detection process is a quantum process and the signal-to-noise ratio (SNR) of CT data quadratically depends on the x-ray dose given to the patient. Therefore, reducing the patient dose will inherently reduce the quality of the image. Two strategies are commonly employed to reduce radiation dose. The first strategy is to reduce the incident x-ray flux a patient receives during a CT scan. This is done by reducing the operating current and potential of the x-ray tube, leading to noisy projection data. The second strategy is to reduce the number of projections across the imaging domain. This strategy could lead to insufficient projection data which have limited-angle, few-view or other problems. Both strategies are tested by the authors in [1].

The authors of the paper aim to reduce noise in low-dose CT images and enhance structural details by implementing an SIR process that incorporates dictionary-based learning. A sparsity constraint is introduced in terms of a redundant dictionary. The proposed method is inspired by compressive sensing theory and TV based minimization. Several TV minimization algorithms have been proposed to solve the de-noising problem associated with low dose CT [3, 4, 5, 6, 7]. The success of such TV algorithms is limited for low-dose scenarios because the TV constraint is based on the relationship of adjacent pixels which makes it difficult for TV minimization to distinguish between true structures of an object and noise in low-dose scenarios. This results in a blocky appearance in the reconstructed images.

Instead, the authors propose enforcing a sparsity constraint via a redundant dictionary. The dictionary is an over-complete basis in which an object image can be sparsely represented as a linear combination of the basis elements (also known as atoms) of the dictionary [8]. The atoms are learned from application specific training images. Here, an object image is partitioned into overlapping patches and the dictionary learning decomposes each patch into a linear combination of sparse atoms. The dictionary is designed to capture local features effectively even in the presence of noise from low-dose projections because of this patch-based analysis and sparsity constraint.

The proposed method consists of two components. The first is an SIR procedure that optimizes a maximum-likelihood function according to the statistical properties of the CT data. The second component is the dictionary-based sparsification, somewhat similar to TV minimization [7]. Two types of dictionaries are considered: a global or pre-determined dictionary learned from a pre-defined training image which does not change
with the image reconstruction process and an adaptive dictionary learned from an intermediate image and updated with each iteration in the reconstruction process.

2 Quantitative Performance Prediction

In this section, we first describe the proposed method in detail. Then we will present a quantitative assessment of the results achieved by the proposed method compared with other image reconstruction methods.

2.1 Dictionary Learning and Sparse Representation

Let us denote a dictionary \( D \in \mathbb{R}^{N \times K} \) with columns \( d_k \in \mathbb{R}^{N \times 1} \) for \( k = 1, ..., N \), where \( N \) is the number of elements in a patch and \( K \) is the number of atoms in the dictionary. Then we can describe the dictionary learning process as solving

\[
\min_{D, \alpha} \sum_{s=1}^{S} (\|x_s - D\alpha_s\|_2^2 + \nu_s\|\alpha_s\|_0),
\]

where we are given a training set of \( S \) patches and Lagrange multiplier \( \nu_s \). Denote the given patch set as a matrix \( X \in \mathbb{R}^{N \times S} \) with a patch \( x_s \in \mathbb{R}^{N \times 1} \) being a column vector of \( X \) 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 \( \alpha \). The authors of [1] mistakenly equate (1) with the following equations:

\[
\min_{D, \alpha} \|X - D\alpha\|_2^2 \quad \text{s.t.} \quad \|\alpha_s\|_0 \leq L_0
\]

\[
\min_{D, \alpha} \sum_{s=1}^{S} \|\alpha_s\|_0 \quad \text{s.t.} \quad \|X - D\alpha\|_2^2 \leq \epsilon
\]

where \( L_0 \) and \( \epsilon \) are the sparsity and precision of sparse representation respectively. It is pointed out however by Nikolova that (1), (2), and (3) are not equivalent in general because all these problems are non-convex [9]. For the purposes of this study, we choose to emulate the methodology described in [1] and perform the minimization described by (2) for the dictionary learning process.

An application of dictionary learning is image de-noising. Let \( z \in \mathbb{R}^{M \times 1} \) be a noisy image of \( M = H \times W \) pixels and a vector \( x \in \mathbb{R}^{M \times 1} \) be its corresponding filtered version. A set of small overlapping patches can be extracted from the image. Using a sliding distance of one pixel and patch size of \( \sqrt{N} \times \sqrt{N} \) pixels, we have \( S = (H - \sqrt{N} + 1) \times (W - \sqrt{N} + 1) \) patches. Assuming that the patches extracted from the filtered image can be sparsely represented in terms of a dictionary, and the filtered image is close to the original noisy image, the de-noising procedure can be written as:

\[
\arg \min_x \min_{\alpha(D)} \|x - z\|_2^2 + \lambda \sum_{s=1}^{S} (\|E_s x - D\alpha_s\|_2^2 + \nu_s\|\alpha_s\|_0),
\]

where \( E_s \in \mathbb{R}^{N \times M} \) is the matrix that extracts the \( s^{th} \) patch from the image \( x \) and \( \lambda \) is the regularization parameter. Dictionary \( D \) can either be predetermined from a training image or learned adaptively during the de-noising procedure.

2.2 Statistical Iterative Reconstruction

Following x-ray attenuation and assuming a quadratic approximation of the Poisson statistics describing the logarithm of the detector data, we can derive the following objective function that we want to minimize:

\[
\Phi(\mu) = \sum_{i=1}^{l} \frac{w_i}{2} (|A_i \mu|_1 - \hat{l}_i)^2 + R(\mu)
\]

where \( \mu = (\mu_1, ..., \mu_J)^T \) is the estimate image of linear attenuation coefficients, \( |A_i \mu|_1 = \sum_{j=1}^{J} a_{ij} \mu_j \) is the line integral of the mono-energetic x-ray linear attenuation coefficients, \( A = \{a_{ij}\} \) is the system matrix, \( I \) and \( J \) are the number of projections and pixels respectively, and \( R(\mu) \) is a regularization term. The term \( w_i = \frac{(y_i - r_i)^2}{y_i} \) is a statistical weight for each x-ray path, where \( r_i \) is readout noise. \( \hat{l}_i = \ln(\frac{b_i}{y_i - r_i}) \) is an estimated integral, where \( b_i \) is the blank scan factor.

The objective function (5) is a statistically weighted least squares function with statistical weights representing the quality of the projection measurement along each path. Although detector statistics for energy integrating detectors are not accurately described by Poisson statistics, the Poisson model is commonly used in the CT field [10]. The authors of [1] used a monochromatic Poisson model for their feasibility study.
2.3 Proposed Algorithm

Using a sparsity constraint in terms of redundant dictionary as the regularization term, we have the following minimization problem:

$$\arg \min_{\mu} \min_{\alpha(D)} \sum_{i=1}^{I} \frac{w_i}{2} (|A\mu|_i - \hat{l}_i)^2 + \lambda \sum_{s=1}^{S} (\|E_s\mu - D\alpha_s\|_2^2 + \nu_s \|\alpha_s\|_0).$$  \hfill (6)

The first approach for image reconstruction is to predetermine a global dictionary using a training image. This global dictionary is then used for sparse reconstruction. This method is called the global dictionary based statistical iterative reconstruction (GDSIR). For GDSIR, an alternating minimization scheme is used to optimize two variables, $\mu$ and $\alpha$, while keeping global dictionary $D$ fixed. After fixing $\alpha$ and $D$, the objection function becomes

$$\arg \min_{\mu} \sum_{i=1}^{I} \frac{w_i}{2} (|A\mu|_i - \hat{l}_i)^2 + \lambda \sum_{s=1}^{S} (\|E_s\mu - D\alpha_s\|_2^2).$$  \hfill (7)

The authors use the separable paraboloid surrogate (SPS) method [11] to minimize $\mu$ for a fixed $\alpha$. After $\mu$ is minimized, we get an intermediate image, $\mu^t$. Then the orthogonal matching pursuit (OMP) algorithm [12, 13] is used to minimize $\alpha$ in the objective function shown below to find the sparse representation $\alpha$:

$$\min_{\alpha} \sum_{s=1}^{S} (\|E_s\mu^t - D\alpha_s\|_2^2) \quad \text{s.t.} \quad \forall s, \quad \|\alpha_s\|_0 \leq L_0$$  \hfill (8)

Note that this minimization is actually performed instead of that described by (1). These two steps are performed alternately until the stopping condition is met.

The second approach for image reconstruction is to adaptively learn a new dictionary from an intermediate image with each iteration of the SIR process. This method is called the adaptive dictionary based statistical iterative reconstruction (ADSIR). For ADSIR, there are three variables, $\mu$, $\alpha$ and $D$. First, minimize $\mu$ for a fixed $\alpha$ and $D$ to generate an intermediate image, $\mu^t$, like the first step in GDSIR. The minimization problem now becomes

$$\min_{D, \alpha} \sum_{s=1}^{S} (\|E_s\mu^t - D\alpha_s\|_2^2) \quad \text{s.t.} \quad \forall s, \quad \|\alpha_s\|_0 \leq L_0.$$  \hfill (9)

For fast convergence, we first train a new dictionary using patches from the intermediate image using the fast online dictionary learning method [13, 14]. Then, we fix the dictionary $D$ and intermediate image $\mu^t$, and minimize $\alpha$ using the OMP algorithm. These steps are repeated until the stopping condition is met.

The table below shows the workflow for both the GDSIR and ADSIR algorithms respectively.

<table>
<thead>
<tr>
<th>TABLE II</th>
<th>WORKFLOW FOR THE GDSIR ALGORITHM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global Dictionary Learning</td>
<td>Choose parameters for dictionary learning; Extract patches to form a training set; Construct a global dictionary [40].</td>
</tr>
<tr>
<td>Image Reconstruction</td>
<td>Initialize $\mu'$, $a'$, and $t = 0$; Set parameters $\lambda$, $\epsilon$, and $L_0^a$; While a stopping criterion is not satisfied 1. Update $\mu^{-t}$ to $\mu'$ using (16); 2. Represent $\mu'$ with a sparse $a'$ using OMP; Output the final reconstruction.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE III</th>
<th>WORKFLOW FOR THE ADSIR ALGORITHM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Choose $\lambda$, $\epsilon$, $L_0^a$ and other parameters;</td>
<td>Initialize $\mu'$, $D'$, $a'$ and $t = 0$; While a stopping criterion is not satisfied 1. Update $\mu^{-t}$ to $\mu'$ using (16); 2. Extract patches from $\mu'$ to form a training set; 3. Construct a dictionary $D'$ from the training set [40]; 4. Represent $\mu'$ with a sparse $a'$ in terms of the dictionary $D'$ using OMP; Output the final reconstruction.</td>
</tr>
</tbody>
</table>

Figure 1: Workflow for GDSIR and ADISR algorithms [1].
2.4 Quantitative Assessment of Proposed Method

The proposed GDSIR and ADSIR methods are compared by Xu, et al. to the performances of FBP and TV minimization based SIR (TVSIR) [1]. To evaluate whether or not GDSIR or ADSIR outperform TVSIR and FBP, two figures of merit (FOM) are calculated. The root mean square error (RMSE) and the image quality assessment (IQA) index for structural similarity (SSIM) [15] of the reconstructed images are compared for each method, where RMSE is defined as

\[
\text{RMSE} = \sqrt{\sum_{j=1}^{J} (\mu^r_j - \mu^*_j)^2}.
\]

(10)

where \(\mu^r_j\) is the reconstructed value, \(\mu^*_j\) is the true value and \(J\) is the number of pixels in an image.

The authors of [1] performed a numerical simulation study by using a FBP image of a sheep lung as a numerical phantom. A fan-beam geometry was defined and a monochromatic source was used. Details of the specific parameters used are found in Section IV-A2 of [1]. A description of each parameter is provided in the table below.

<table>
<thead>
<tr>
<th>No.</th>
<th>Variable</th>
<th>Meaning</th>
<th>Criterion/Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(\lambda)</td>
<td>Parameter to balance the data fidelity and prior information terms</td>
<td>It should be increased with the noise level. It is usually empirically selected in practice.</td>
</tr>
<tr>
<td>2</td>
<td>(K,N)</td>
<td>Number of atoms and number of pixels in a patch</td>
<td>They are selected as (N=64) and (K=256) in the image processing field, and work well in our study.</td>
</tr>
<tr>
<td>3</td>
<td>(l_0^i)</td>
<td>Sparsity level (l_0) for dictionary learning</td>
<td>It is usually set to 5-10 atoms.</td>
</tr>
<tr>
<td>4</td>
<td>(l_0^s)</td>
<td>Sparsity level (l_0) for image reconstruction</td>
<td>It is determined by the complexity of an image and the power of a dictionary: (N/2 &gt; l_0^s \geq l_0^i).</td>
</tr>
<tr>
<td>5</td>
<td>(\epsilon)</td>
<td>Tolerance of the difference between the reconstructed and original images</td>
<td>It is determined by the image noise and the dictionary capability, and normally comparable to the image noise level.</td>
</tr>
</tbody>
</table>

Low dose projection data was synthesized by superimposing Poisson noise onto the raw sinogram data assuming \(1.0 \times 10^4\) photons from the source towards each detector element. Three sets of data were generated with the number of projection views set as 580, 290 and 116 respectively.

Fig. 3 shows the performance of the GDSIR and ADSIR methods compared with FBP and TVSIR in the numerical simulation study performed by the authors. Fig. 4 shows the low-dose reconstructed images of the numerical phantom and the difference images between the FBP image and other algorithms. We plan to replicate the authors' work in 2D with the specific goal of assessing whether or not the relative improvement in RMSE and SSIM are as substantive as shown, particularly in the low view cases. To do this, we will be conducting a simulation study using the XCAT phantom [16, 17] as a numerical phantom and generating projection data. FBP, TVSIR, GDSIR, and ADSIR methods will be used to perform image reconstruction. The MIRT [18] will help us generate the projection data and implement the SPS step, TVSIR method, and FBP method. Additionally, the SPAMS toolbox [19] will help use perform the fast online dictionary learning and the OMP step.
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)

<table>
<thead>
<tr>
<th>Views</th>
<th>FBP</th>
<th>GDSIR</th>
<th>ADSIR</th>
<th>TVSIR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1160</td>
<td>87.71</td>
<td>43.44</td>
<td><strong>43.01</strong></td>
<td>46.97</td>
</tr>
<tr>
<td>580</td>
<td>117.61</td>
<td>47.27</td>
<td><strong>46.84</strong></td>
<td>51.04</td>
</tr>
<tr>
<td>290</td>
<td>162.72</td>
<td><strong>52.93</strong></td>
<td>53.41</td>
<td>57.48</td>
</tr>
<tr>
<td>116</td>
<td>259.86</td>
<td><strong>67.70</strong></td>
<td>69.11</td>
<td>70.39</td>
</tr>
</tbody>
</table>

TABLE V
SSIM VALUES OF THE RESULTS RECONSTRUCTED BY THE FBP, TVSIR, GDSIR, AND ADSIR METHODS, RESPECTIVELY, IN THE SHEEP LUNG SIMULATION STUDY

<table>
<thead>
<tr>
<th>Views</th>
<th>FBP</th>
<th>GDSIR</th>
<th>ADSIR</th>
<th>TVSIR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1160</td>
<td>0.5907</td>
<td>0.8468</td>
<td><strong>0.8485</strong></td>
<td>0.8457</td>
</tr>
<tr>
<td>580</td>
<td>0.4557</td>
<td>0.8303</td>
<td><strong>0.8321</strong></td>
<td>0.8292</td>
</tr>
<tr>
<td>290</td>
<td>0.3220</td>
<td><strong>0.8100</strong></td>
<td>0.8093</td>
<td>0.8047</td>
</tr>
<tr>
<td>116</td>
<td>0.1765</td>
<td><strong>0.7708</strong></td>
<td>0.7666</td>
<td>0.7599</td>
</tr>
</tbody>
</table>

Figure 3: Results of Low-Dose CT Dictionary Learning [1]

Figure 4: Reconstructed images from low-dose projection data. (a) images reconstructed with FBP, GDSIR, ADSIR, and TVSIR from top to bottom. (b) difference images between FBP result and GDSIR, ADSIR and TVSIR methods respectively. From left to right, images contain 580, 290, and 116 views respectively [1].
3 Methodology

This section describes the methodology implemented by this group in our attempt to replicate the results of the original paper [1]. We describe the dictionary generation method as well as our implementation of the statistical image reconstruction. A description of the a TV minimization method is implemented for comparison.

3.1 XCAT Phantom and Projection Data

The XCAT phantom is a 4D extended cardio-torso phantom that was developed to provide an accurate representation of the human anatomy [16]. For our study, slice 215 of the XCAT phantom, which is a cross sectional image of the human thorax as shown in Fig. 5, is used to train a global dictionary. We will attempt to reconstruct both slice 200 and slice 275 of the XCAT phantom. These two slices are chosen to test the robustness of the dictionary learning method.

![Figure 5: From left to right: slice 215, 200 and 275 of XCAT phantom respectively.](image)

Projection data is generated using the above numerical phantoms assuming an incident photon flux of \(1.0 \times 10^5\) per detector element for the normal-dose case and \(1.0 \times 10^4\) for the low-dose case. The MIRT [18] is the main tool used generate all our projection data. A fan-beam geometry with GE Lightspeed system [20] having 888 detector elements, field-of-view (FOV) of 500 mm, and source-to-detector distance of 94.9 cm is modeled. Over a 360° range, 984 projections are measured using uniform angular spacing in our full-view case. Poisson noise is then added to the projection data. Unlike [1], no electronic readout noise, \(r_i\), is added. Reconstructed images are matrices of size \(512 \times 512\) pixels. During reconstruction, only the pixels inside the central circle with radius 256 pixels are considered. In addition to this data set, we also generate a few-view data set using 246 projections and low-dose incident photon flux. This data set is prepared to demonstrate the potential of the algorithms.

3.2 Filtered Backprojection

In our study, the conventional filtered backprojection algorithm is used to provide images for dictionary learning, initialization of statistical image reconstruction process and to provide a traditional reconstruction method to be compared against. To perform FBP reconstruction, the \texttt{fbp2} function with an ideal ramp filter in MIRT is used.

3.3 Online Dictionary Learning

The SPAMS toolbox is used to train a global dictionary using the fast online dictionary learning algorithm [19]. The \texttt{mexTrainDL} function is used to perform the minimization procedure described by (2). To train the dictionary, we set the patch size to \(8 \times 8\) pixels with unit sliding distance and set the sparsity constraint to \(L_D^0 = 5\). The dictionary generated from this process is composed of 256 atoms. Additionally, we add an additional DC atom to the dictionary.

Patch extraction is performed by using the \texttt{im2col} MATLAB function. For completeness and debugging purposes, we wrote our own function to perform the inverse operation of extracting an image from a set of patches since the MATLAB function \texttt{col2im} does not perform this exact operation. We also wrote a function that performs the transpose or backward operation of patch extraction operation, which is equivalent to put all patches back and summing them up for each image pixel.

The global dictionary is trained from a normal-dose/full-view FBP reconstructed image of slice 215 of the XCAT phantom. Similar to the authors of [1], we de-mean each patch and remove patches of the FBP image with low variance before learning the global dictionary. This process is discussed later in 4.1.
3.4 Penalized Weighted Least Square Solver and SPS implementation

The first step of the GDSIR and ADSIR algorithm is to minimize $\mu$ for a fixed $\alpha$. To perform this the authors of [1] used the SPS algorithm. For simplicity, we rewrite (7) into a weighted least square formulation. We do this by stacking the $A$ and $E_s$ matrices into a single matrix $A_2$, and similarly stacking the $l \in \mathbb{R}^{I \times 1}$ and $D_{\alpha} \in \mathbb{R}^{N \times 1}$ terms into a single matrix $b$, where $l = (l_1, l_2, ..., l_I)^T$. The new weight matrix, $W_2$, is a block diagonal matrix with statistical weight matrix, $W \in \mathbb{R}^{I \times 1}$, and the regularization parameter $\lambda$. The new objective function can be written as

$$\arg\min_{\mu} \frac{1}{2} \| A_2 \mu - b \|_W^2.$$ 

(11)

where $I$ is the identity matrix of size $\mathbb{R}^{N \times N}$, and $E = (E_1, E_2, ..., E_s)^T \in \mathbb{R}^{(N \times S) \times J}$.

Using this formulation, our problem becomes a weighted least square problem without a penalty term. We use the `pwls_sps_os` function in MIRT to emulate the SPS methodology performed by the authors of [1] to solve this problem. We use an FBP image to initialize $\mu$. To accelerate the SPS algorithm, we use 40 ordered-subsets. We chose to store $E$, as a sparse matrix because we do not have a proper method to extract patches for ordered-subsets. The $A$ and $E$ matrices first converted to $G_{\text{block}}$ objects with 40 blocks, and then stacked to form the large matrix $A_2$. For each iteration of GDSIR, we set the number of SPS iterations to be 10. This value was chosen after examining the value of the cost function (11) over many SPS iterations.

3.5 OMP Implementation

The OMP algorithm is implemented using the `mexOMP` function in the SPAMS toolbox [19]. As indicated by (8), two parameters are required by the OMP algorithm: the error tolerance of the sparse representation $\epsilon$, and the sparsity constraint of the dictionary representation $L_0^S$. Note that the sparsity constraint $L_0^S$ here is a different parameter from the sparsity control $L_0^S$ when we do dictionary training.

In our initial attempt, we implemented the OMP step after the initial SPS step as described in [1]. We used a random $\alpha$ for the first SPS step but found that this resulted in poor reconstructed images. To achieve a better initial guess for $\alpha$, we switched the order of the SPS and OMP steps. Performing the OMP step first to get an initial guess of $\alpha$ and then minimizing $\mu$ was found to produce images with less noise in fewer iterations.

3.6 Parameter Selection

Parameter selection plays an important role in the quality of the reconstructed image. We first set the number of GDSIR iterations to be 50 after examining the plots shown Fig. 6 using some arbitrary empirical parameters. Then, the three parameters, $\lambda, \epsilon$, and $L_0^S$, are perturbed one at a time while keeping the others fixed and the resulting images are compared to determine the optimal values for each parameter. The RMSE and SSIM are calculated for each parameter combination as well as a visual inspection of each image.

Note that we only tuned parameters for full-view/low-dose and few-view/low-dose GDSIR images of slice 200. We chose the best parameter combinations for these two cases and used them for the corresponding reconstructions of slice 275.

For the choice of $L_0^S$, the authors of [1] suggests $L_0^S < L_0^S \leq \frac{N}{2}$. We tried the values $L_0^S = 5, 10, 15$. For the choice of $\epsilon$, we tried $\epsilon = \sigma^2$ and $\frac{1}{2} \sigma^2$ where $\sigma^2$ is the variance of the flat region of the heart in the FBP reconstructed image. We set $\epsilon$ according to $\sigma^2$ because $\epsilon$ should be comparable to the image noise level in a flat region, and we want the OMP step to stop when the sparse representation error is less than the variance of a flat region of the image. There is no baseline for $\lambda$ and the tuning range of $\lambda$ can be very large. We tried $\lambda = 16, 32, 64, 128$ but note that the upper bound can be larger. The parameter combination $L_0^S = 5, \epsilon = \sigma^2, \lambda = 128$ was found to be the best for both cases. This parameter combination is also shown as Case 1 in Table 3.
Figure 6: Left: sparse representation error vs. number of iterations calculated at the end of each OMP iteration, where the sparse representation error is defined as $\sum_{s=1}^{S} \| E_s \mu^t - D \alpha_s \|^2_2$. Right: Cost function in (11) vs. iteration calculated at the end of each SPS iteration.

3.7 ADSIR Implementation

Sections 3.1 - 3.5 describe the steps in the GDSIR algorithm. The ADSIR algorithm is essentially the same with the additional step of adaptively updating the dictionary from the intermediate image in each iteration. To accelerate the ADSIR procedure, 10 iterations of GDSIR are performed initially to generate a good initial $\mu$ and reduce some noise. In the adaptive dictionary learning step, instead of removing patches with low variance, we use the MATLAB `edge` function and remove patches that do not contain edges. We use the Canny edge detector [21] option and set a low threshold of 0 and a high threshold of 0.2. These values were arbitrary chosen via a non-rigorous guess and check method and visualization. We found these parameters to work well enough to learn a clean dictionary. No parameter tuning is performed for ADSIR due to time constraints. The parameters that gave the best GDSIR results are used for ADSIR and we assume that these are good enough to generate preliminary results.

3.8 TV Regularization Method

A statistical image reconstruction with TV based regularizer is implemented for comparison. To perform the TVSIR reconstruction, we use a splitting-based ordered-subset (OS) algorithm, OS-LALM, described in [22]. Code was provided by Naveen Murthy, who also helped us understand and use the code. In our implementation, all parameters are set to values as described in [7]. We used the $l_1$ norm and finite differences in the horizontal, vertical and diagonal directions to create the `Reg1` object, as specified in MIRT. The $\kappa$ value is set as equal to the mask and the FBP image is used as the initial guess. The number of ordered-subsets is set to 10. To find the best value of the regularization parameter $\beta$, 25 iterations of OS-LALM are performed for $\beta$ values ranging from $2^{-4}$ to $2^4$. The $\beta$ value for the image with lowest RMSE value with respect to the ground truth is chosen as the optimal value.

4 Results

This section describes the results of our implementation of the global dictionary generation, GDSIR and ADSIR algorithms. Numerical simulation studies using the XCAT phantom are designed to evaluate the performance of the algorithms. We compare the GDSIR and ADSIR images to FBP, TVSIR and ground truth images. All images are shown with a display window [0, 0.045] $mm^{-1}$, and all difference images are shown with a display window [-0.015, 0.015] $mm^{-1}$.
4.1 Global Dictionary Generation

To learn the global dictionary, we first generate a full-dose (1 × 10^5 photons incident per detector element) and full-view (984 views) FBP image of slice 215 of the XCAT phantom. Using this FBP image, we remove all patches with low variance using a threshold based on a quick numerical study. Initially, we did not remove low variance patches and this led to a noisy dictionary and poorly reconstructed images. Fig. 7 and 8 show the dictionaries generated without and with low variance patch removal respectively. For visualization purposes, the DC atom is not shown.

Figure 7: Left: FBP image of slice 215 of XCAT Phantom. Right: Dictionary learned from full FBP image without removing low-variance patches.

Figure 8: Left: Remaining patches of FBP image of slice 215 of XCAT phantom after removal of low variance patches. Right: Dictionary learned from FBP image with low-variance patches removed.

As seen in Fig. 7, using the entire FBP image to train the dictionary results in a very noisy dictionary. This leads to very poor reconstructed images using GDSIR. Removing low-variance patches removed much of the noise and preserved most of the patches containing structural features of the thorax as seen in Fig. 8, resulting in a much less noisy dictionary. This dictionary performed much better in GDSIR.
4.2 GDSIR and ADSIR Performance for Slice 200

The GDSIR and ADSIR reconstructed images of slice 200 of the XCAT phantom are compared to the ground truth, FBP and TVSIR images. Two sets of cases are compared: a full-view/low-dose case and a few-view/low-dose case, where low dose means $1 \times 10^4$ incident photons per detector element, full views means 984 views and few views means 246 views.

The full-view/low-dose case models a practical low-dose situation in real world. The results are shown in Fig. 9. The TVSIR method performs better than GDSIR or ADSIR in terms of noise in uniform regions of the image. However, TVSIR does have blocky artifacts, while GDSIR and ADSIR preserve the natural look of the images. It is not clear if TVSIR or GDSIR/ADSIR represent structures and edges better which would be valuable in a diagnostic setting.
Figure 9: Low-dose/full-view image reconstruction comparisons for slice 200 of XCAT phantom. The differences are taken with respect to the ground truth.
The few-view/low-dose case is meant to be a challenge to test the robustness of the reconstruction methods. The resulting images are all worse than the full-view case. All images show noisy reconstruction with blurred edges. Here, GDSIR and ADSIR perform better than TVSIR in terms of image denoising. We attribute this gain to the small value of the sparsity constraint, $L_0^S$. Because the dictionary is trained with a full-dose/full-view reconstruction, denoising is possible as long as $L_0^S$ is not too large. Additionally, the denoising capability of GDSIR/ADSIR is also dependent on the value of $\epsilon$ which demonstrates the importance of parameter selection on a case-by-case basis.

Comparing the full-view and few-view GDSIR reconstructions, the few-view case show less noise but more blur. We believe this observation can be attributed to the fact that $\lambda$ is kept the same for both full-view and few-view cases. As the number of projections decreases, the relative weight of the data fit term is smaller. Thus with the same $\lambda$, the relative regularizer penalty is greater. This leads to more image denoising at the expense of image sharpness. Perhaps if we increase the sparsity constraint, $L_0^S$, the reconstructed images will be sharper while keeping the noise level down. Due to time constraints, we have not found a best case parameter combination.
Figure 10: Low-dose/few-view image reconstruction comparisons for slice 200 of XCAT phantom. The differences are taken relative to the ground truth.
Table 1 shows the FOM for the reconstructions. GDSIR and ADSIR performed better than the FBP method for the full-view case and slightly worse than the TVSIR method in terms of both RMSE and SSIM. However, for the few-view case GDSIR and ADSIR outperformed both FBP and TVSIR in terms of both RMSE and SSIM. In both of these cases GDSIR outperformed ADSIR slightly. However, it should be noted that we did not have time to find the best case parameter combination for the ADSIR method and simply used the best case parameter combination taken from the GDSIR method.

Our results differ from those obtained by the authors of [1]. Their GDSIR and ADSIR methods outperformed FBP and TVSIR in terms of both RMSE and SSIM. We suspect that they carefully tuned their parameters to achieve this result.

4.3 GDSIR and ADSIR Performance for Slice 275

To test the robustness of the dictionary learning method, particularly the adaptive dictionary method, in this section we reconstruct slice 275 which contains some different structures than slice 215. The reconstructed images for full-view and few-view and the corresponding difference images are shown in Fig. 11. TV reconstruction was not implemented for slice 275 due to time constraints.
As expected, GDSIR and ADSIR perform much better than FBP, in terms of both RMSE and SSIM shown in Table 2, with GDSIR slightly outperforming ADSIR. GDSIR and ADSIR produced similar images. This may be due to insufficient parameter tuning. We expected ADSIR to perform better when reconstructing images containing different features than the training image, although that result was not observed here. In the authors original work [1], when their training data and testing data were significantly different, both GDSIR and ADSIR performed better than TVSIR.
Table 2: RMSE and SSIM values for image reconstructions of slice 275 of XCAT phantom.

<table>
<thead>
<tr>
<th>Views</th>
<th>FOM</th>
<th>FBP</th>
<th>GDSIR</th>
<th>ADSIR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMSE (mm$^{-1}$)</td>
<td>0.004108</td>
<td>0.001865</td>
<td>0.002151</td>
</tr>
<tr>
<td></td>
<td>SSIM</td>
<td>0.9945</td>
<td>0.9988</td>
<td>0.9984</td>
</tr>
<tr>
<td>Few</td>
<td>RMSE (mm$^{-1}$)</td>
<td>0.008192</td>
<td>0.001630</td>
<td>0.002163</td>
</tr>
<tr>
<td></td>
<td>SSIM</td>
<td>0.9795</td>
<td>0.9990</td>
<td>0.9984</td>
</tr>
</tbody>
</table>

4.4 Parameter Comparison

To evaluate how the parameters affect the performance of GDSIR, we describe our analysis of the reconstructed results of slice 200 for full-view/low-dose case as an example. Table 3 shows our best parameter combination (Case 1) and the perturbations (Case 2-4). The reconstructed images and difference images relative to ground truth are shown in Fig. 12.

A comparison of the RMSE values of Cases 1 and 2 shows that a smaller $\epsilon$ results in a noisier image since the sparse representation must fit to noise to reduce the sparse representation error below $\epsilon$. A comparison between Cases 1 and 3 shows that a larger $\lambda$ leads to smoother but more blurred image. This is because a larger $\lambda$ increases the penalty term of the cost function and normally smoothness comes from the penalty. A comparison of Cases 1 and 4 shows that increasing the sparsity constraint leads to noisier but sharper images. This occurs because using more atoms allows for reconstruction of more complicated structures but can also lead to fitting noisy data. Some of these results may not be statistically significant; a more rigorous analysis with confidence intervals would be required.

Table 3: Parameter perturbation and FOM of the result images for full-view/low-dose GDSIR of slice 200.

<table>
<thead>
<tr>
<th>Case No.</th>
<th>$\lambda$</th>
<th>$L^0_S$</th>
<th>$\epsilon$</th>
<th>RMSE (mm$^{-1}$)</th>
<th>SSIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>128</td>
<td>5</td>
<td>$1 \sigma^2$</td>
<td>0.001327</td>
<td>0.9994</td>
</tr>
<tr>
<td>2</td>
<td>128</td>
<td>5</td>
<td>$\frac{1}{2} \sigma^2$</td>
<td>0.001345</td>
<td>0.9994</td>
</tr>
<tr>
<td>3</td>
<td>16</td>
<td>5</td>
<td>$1 \sigma^2$</td>
<td>0.002543</td>
<td>0.9979</td>
</tr>
<tr>
<td>4</td>
<td>128</td>
<td>15</td>
<td>$1 \sigma^2$</td>
<td>0.001889</td>
<td>0.9988</td>
</tr>
</tbody>
</table>

5 Conclusion and Future Work

The goal of this project was to replicate the results *Low-dose x-ray CT reconstruction via dictionary learning* [1]. We used a statistical image reconstruction framework and implemented a sparsity constraint based on redundant dictionary. The dictionary was learned from a global training image in the GDSIR method and adaptively from an intermediate image in the ADSIR method. These methods are compared to a total variation based minimization method and traditional filtered backprojection. The authors of [1] showed that ADSIR performs slightly better than GDSIR for full-view cases while GDSIR performs better in few-view cases. Our results show that for low-dose/full-view reconstructions, TVSIR performs slightly better than GDSIR and ADSIR in terms of both RMSE and SSIM. However, for the low-dose/few-view cases, GDSIR and ADSIR reconstruct images with less noise than TVSIR and thus a better RMSE and SSIM. GDSIR performance was better than ADSIR overall, although we believe this may be because of insufficient parameter tuning for the ADSIR cases. It is not clear if GDSIR/ADSIR can outperform TVSIR in terms of representing structures and edges, which is important for diagnosis.

We have shown that this dictionary learning method can be effectively used in image de-noising applications. However, successfully application of GDSIR and ADSIR depends on carefully selecting parameters on a case-by-case basis. Future work should focus on the robustness of this method and on developing a framework to select appropriate parameters that generate de-noised images in real-world CT scenarios. Additionally, streamlining the code to reduce computation time is necessary for these methods to become of practical use in the real world.
Figure 12: The reconstructed results and difference images of slice 200 for full-view/low-dose case with parameters shown in Table 3. Difference is taken relative to ground truth.

References


