LOW-DOSE CT RECONSTRUCTION WITH MULTICLASS ORTHOGONAL DICTIONARIES

Hiryu Kamoshita† Taïsuke Shibata† Daichi Kitahara† Ken‘ichi Fujimoto‡ Akira Hirabayashi†

†Graduate School of Information Science and Engineering, Ritsumeikan University, Shiga, Japan
‡Faculty of Engineering and Design, Kagawa University, Kagawa, Japan

ABSTRACT
We propose a high-accuracy CT image reconstruction from low-dose X-ray projection data. A state-of-the-art method exploits dictionary learning for image patches. This method generates an overcomplete dictionary from patches of standard-dose CT images and reconstructs low-dose CT images by minimizing the sum of date fidelity and regularization terms based on sparse representations with the dictionary. However, this method does not take characteristics of each patch into account, such as texture and edges. In this paper, we propose to divide all patches into several classes, and use an individual dictionary with an individual regularization parameter for each class. Moreover, for fast computation, we introduce the orthogonality for each dictionary. Since clustering collects similar patches, accuracy degradation by the orthogonality hardly occurs. Simulation shows the proposed method outperforms the state-of-the-art one in terms of accuracy and speed.

Index Terms— Low-dose CT, image reconstruction, sparse representation, fast dictionary learning, clustering.

1. INTRODUCTION
X-ray computed tomography (CT) is widely used for diagnosis and detection of various diseases because it scans the inside of the human body noninvasively in a few seconds. However the induction of cancerous and genetic diseases by the X-ray radiation is concerned [1]. Hence, it is desired to suppress the X-ray dose to as low as possible. If we apply the standard reconstruction method, filtered back projection (FBP), to low-dose projection data, then large noise and unnatural artifacts appear in the reconstructed image. As a result, we would fail to detect diseased tissues from the reconstructed image [2].

On the basis of the compressed sensing theory [3], we exploit the sparsity to reconstruct a high-quality image from the low-dose projection data. In the image processing field, the patch-based dictionary learning is used to acquire sparse representations of target signals [1], [2], [4]–[8]. A dictionary is generated as a matrix from training images, and a target signal is supposed to be expressed by a linear combination of a few column vectors of the dictionary. In [1], Xu et al. combined the patch-based dictionary learning with statistical iterative reconstruction (SIR) [9] from the low-dose projection data. Although this method can remove streak artifacts well, detailed structures appear in the reconstructed image. As a result, we would fail to detect diseased tissues from the reconstructed image [2].

To obtain more sparse representations, in this paper, we propose to extend the reconstruction method of Xu et al. [1] to a multiclass version. In the proposed method, all patches of training images are divided into some classes, and a dictionary is created for each class. Then, patches of the target image are also classified and expressed sparsely by using the dictionary for each class. In this case, it is assumed that the approximation errors, for the patches, of the sparse representations differ depending on the class. Therefore, we propose to use a different regularization parameter for each class. Moreover, for fast reconstruction, we replace the overcomplete dictionaries with orthogonal matrices. In general, if the number of column vectors of a dictionary is reduced, then the accuracy of each sparse representation degrades, which leads to low-quality reconstructed images. On the other hand, in the proposed method, since a dictionary has only a few columns, then the accuracy of each sparse representation hardly degrades. Simulation using real CT images shows that the proposed method achieves high accuracy and fast computation.

2. SIGNAL MODEL AND DICTIONARY-BASED SIR
In CT, nonnegative attenuation coefficients $\mu \in \mathbb{R}_+^d$ are observed by $l = R\mu + \xi$, (1)

where $R \in \mathbb{R}^{I \times J}$ is a Radon transform matrix based on X-ray paths, $l \in \mathbb{R}^I$ is the projection data, and $\xi \in \mathbb{R}^I$ is the observation error. There are two approaches for reducing the X-ray dose: (i) lowering the X-ray intensity and (ii) reducing the number of projections. In the first approach, however, the noise level $\|\xi\|_2$ becomes large. In the second approach, the number $I$ of row vectors of $R$ decreases, which results in rank($R$) $< J$. As a result, in either case, the standard reconstruction method, FBP based on the Fourier slice theorem, generates large noise and unnatural artifacts in the reconstructed images.

Sauer et al. proposed a method that reconstructs $\mu$ by maximum a posteriori (MAP) estimation [9]. This method is called SIR since it reconstructs $\mu$ by iteratively minimizing a statistical cost function

$$\|R\mu - l\|^2_{2,w} + \lambda_P(\mu) := \sum_{i=1}^{I} w_i (r_i^T \mu - l_i)^2 + \lambda_P(\mu).$$

(2)

The weight for the $i$th projection $r_i^T \mu$ can be statistically defined as $w_i := b_i \exp(l_i)$ from the X-ray intensity $b_i > 0$ and the observed data $l_i$. $\Psi : \mathbb{R}^J \rightarrow \mathbb{R}$ is a regularization term based on a prior information on $\mu$, and $\lambda > 0$ is a regularization parameter which controls the balance between the data fidelity and the regularization.

Xu et al. combined (2) and the patch-based dictionary learning [1]. In this method, a CT image $\mu$ is decomposed into small images $H_s \mu \in \mathbb{R}^P (s = 1, 2, \ldots, S)$, called patches, of size $\sqrt{P} \times \sqrt{P}$, where $H_s \in \{0, 1\}^{P \times J}$ is the $s$th patch extraction matrix and $S$ is the number of the patches. Then, $\mu$ is reconstructed by using an assumption that each patch can be expressed by a linear combination of a few column vectors of an appropriate dictionary $D \in \mathbb{R}^{P \times K}$, where $K$ is the number of column vectors $d_k \in \mathbb{R}^P$ s.t. $\|d_k\|_2 = 1 (k = 1, 2, \ldots, K)$. By defining $c_s \in \mathbb{R}^K$ as sparse representations (coefficients) for $H_s \mu$, this method solves the following problem

minimize $\|R\mu - l\|^2_{2,w} + \lambda \sum_{s=1}^{S} (\|H_s \mu - Ds\|^2_2 + \nu_s \|c_s\|_0).$

(3)
where \( C := (c_1, c_2, \ldots, c_S) \in \mathbb{R}^{K \times S} \) and \( \nu_q > 0 \) is a Lagrangian multiplier for each patch. On (3), there are two cases where (i) \( D \) is fixed to a matrix pre-learned from training images, and (ii) \( D \) is adaptively updated by using the estimated target image \( \mu \). The former and the latter are called global dictionary-based SIR (GDSIR) and adaptive dictionary-based SIR (ADSIR), respectively.

Since it is difficult to directly solve the optimization problem in (3), the sparse coefficient \( C \) (and the dictionary \( D \)) and the CT image \( \mu \) are alternately updated from the initial image \( \mu^0 \) obtained by the FBP. The matrix \( C \) (and \( D \)) is updated by solving the problem

\[
\min_{(D),C} \sum_{s=1}^{S} \left( \| H_s \mu^{m-1} - D c_s \|_2^2 + \nu_s \| c_s \|_0 \right),
\]

where \( \mu \) is fixed to the estimated value \( \mu^{m-1} \). In GDSIR, only \( C \) is updated by orthogonal matching pursuit (OMP) [10]. In ADSIR, after updating \( D \) by an online learning algorithm in [11], \( C \) is updated by OMP. Hence ADSIR requires a lot of computational time compared to GDSIR. The image \( \mu \) is updated by solving the problem

\[
\min_{\mu \in \mathbb{R}^{m+1}} \| R \mu - l \|_2^2 + \lambda \sum_{s=1}^{S} \| H_s \mu - D c_s \|_2^2
\]

with fixed \( C \) (and \( D \)). For the problem in (5), GDSIR and ADSIR use the majorization-minimization (MM) [12] to avoid the computation of a large inverse matrix. Specifically, \( C \) is updated from the previous one \( \mu^{m-1} \) as in (6), where \( W := \text{diag}(w), 0 \& 1 \) are respectively vectors whose all components are 0 and \( \in \) is the component-wise division, and \( max \) returns a larger value in each component [13].

### 3. SIR WITH MULTICLASS DICTIONARY LEARNING

Recently, in the high-speed MRI and low-dose CT fields, image reconstructions with multiclass dictionary learning have been proposed for higher compression ratio, noise reduction, and artifact suppression [2], [7], [8]. In these methods, image patches are classified into multiple classes by using geometric directions or pixel values. Then, since dictionaries are generated for each class, the accuracy of sparse representations for the patches is improved, which leads to good reconstruction results. The methods in [7], [8] are proposed for high-speed MRI and hence cannot be directly applied to low-dose CT. The method in [2] uses the \( \| \mu - \mu^i \|_2^2 \) as the data fidelity term, and hence the reconstruction results depend largely on \( \mu^0 \) obtained by the FBP.

On the other hand, since GDSIR uses \( \| R \mu - l \|_2^2 \) as the data fidelity term, the reconstruction results are robust against the initial image \( \mu^0 \). Therefore, in this paper, we propose to extend GDSIR to a multiclass version in order to reconstruct high-quality CT images.

#### 3.1. Multiclass GDSIR

Let \( Q \) be the number of classes and \( q = 1, 2, \ldots, Q \) be the index of the class. A dictionary \( D_q \in \mathbb{R}^{P \times K_q} \) for the \( q \)th class is generated in advance from standard-dose images, where \( K_q \) is the number of column vectors of \( D_q \). Patches \( H_s \mu \) of a target image are classified, and we define \( S_q \) s.t. \( \bigcup_{q=1}^{Q} S_q = \{ 1, 2, \ldots, S \} \) and \( \bigcap_{q=1}^{Q} S_q = \emptyset \) (\( q \neq q' \)) as a patch index set for the \( q \)th class. We reconstruct \( \mu \) by solving

\[
\min_{\mu \in \mathbb{R}^{m+1}} \| R \mu - l \|_2^2 + \sum_{q=1}^{Q} \lambda_q \sum_{s \in S_q} \left( \| H_s \mu - D_q c_s \|_2^2 + \nu_s \| c_s \|_0 \right)
\]

which is a multiclass version of (3). Note that here we use a different regularization parameter \( \lambda_q > 0 \) for each class. By fixing the sets \( S_q \), the problem in (7) can be solved by a similar update to that of GDSIR.

Moreover, to reduce the computational time, we propose to introduce orthogonal dictionaries \( \hat{D}_q \in \mathbb{R}^{P \times P} \) s.t. \( \hat{D}_q^T \hat{D}_q = E \) to (7), where \( E \) is the identity matrix of order \( P \). In general, if an overcomplete dictionary is replaced with an orthogonal matrix, then calculation speed increases but the accuracy of each sparse representation degrades. On the other hand, in the proposed method, since similar patches are collected by clustering, the degradation of the representation accuracy hardly occurs even if we use orthogonal dictionaries.

#### 3.2. The Proposed Reconstruction Algorithm

The proposed algorithm reconstructs \( \mu \) from the initial estimate \( \mu^0 \) by alternately executing the sparse coefficient update step and the image update step \( M \) times after executing the preparation one time.

**Preparation:** First, patches are extracted from training images and classified into \( Q \) classes by K-means algorithm in the same manner as [8]. To also classify patches of the target images later, we store the clustering centers. Then, an orthogonal dictionary \( \hat{D}_q \) is created for each class as an approximate solution to the optimization problem

\[
\min_{\hat{D}_q, C_q^{tr}} \sum_{s \in S_q} \left( \| H_s \mu^{tr} - \hat{D}_q c_s^{tr} \|_2^2 + \nu_s \| c_s^{tr} \|_0 \right) \quad \text{s.t.} \quad \hat{D}_q^T \hat{D}_q = E, \quad \mu^{tr} = \text{standard-dose training image}, \quad c_s^{tr} = \text{spare coefficient vector for the \( s \)th training patch \( H_s \mu^{tr} \)}, \quad S_q \text{ is a patch index set for the training patches of class } q.
\]

For solving the problem in (8), we alternately update \( \hat{D}_q \) and \( C_q^{tr} \) according to [5]. By fixing \( C_q^{tr} \) in (8), \( \hat{D}_q \) is updated as

\[
\hat{D}_q^* = \text{argmin}_{C_q^{tr}} \| \Phi_q^{tr} - \hat{D}_q C_q^{tr} \|_F^2 \quad \text{s.t.} \quad \hat{D}_q^T \hat{D}_q = E, \quad \Phi_q^{tr} = \text{training matrix whose column vectors are the training patches of class } q, \quad C_q^{tr} \text{ is the }\nu_q \text{-sparse coefficient vector for the } q \text{th training patch } H_q \mu^{tr} \text{, } S_q \text{ is a patch index set for the training patches of class } q.
\]

Then, by fixing \( \hat{D}_q \) in (8), \( C_q^{tr} \) is updated as

\[
C_q^{tr*} = \text{argmin}_{C_q^{tr}} \sum_{s \in S_q} \left( \| H_s \mu^{tr} - \hat{D}_q c_s^{tr} \|_2^2 + \nu_s \| c_s^{tr} \|_0 \right). \quad (11)
\]

From the property of \( \hat{D}_q^T \hat{D}_q = E \), the solution to (11) is quickly computed for each column vector \( c_s^{tr*} \) by

\[
c_s^{tr*} = \text{Hard}_{\nu_q}(\nabla_{\nu_q}(\hat{D}_q^T H_s \mu^{tr})) \quad (12)
\]

with the use of the hard thresholding operator

\[
\text{Hard}_{\nu_q}(c)[i] = \begin{cases} c[i] & \text{if } |c[i]| \geq \sqrt{\nu}, \\ 0 & \text{if } |c[i]| < \sqrt{\nu}, \end{cases}
\]

where \( [i] \) denotes the \( i \)th component of a vector. By repeating (10)
and (12) until a convergence condition is satisfied, $\tilde{D}_q$ is generated.

To solve (7) with the multiclass orthogonal dictionaries $\tilde{D}_q$, first, the initial estimate $\mu^0$ is obtained by applying the FBP to the low-dose projection data $I$. Then, by using the stored clustering centers in the preparation, the initial patches $H_s\mu^0$ are classified into $Q$ classes by the nearest neighbor method, and hereafter the patch index sets $S_q$ are fixed. We reconstruct a target CT image $\mu$ by repeating the sparse coefficient update and image update steps for $m = 1, 2, \ldots, M$.

### Sparse Coefficient Update Step

At the $m$th update of $C$, in the same manner as (12), each column vector $c_s$ is updated by

$$c_s = \text{Hard}_{\alpha}(\tilde{D}_q^TH_s\mu^{m-1}),$$

where $\mu$ is fixed to the estimated value $\mu^{m-1}$ and $q(s)$ denotes the class of the $s$th patch. Differently from the problem in (4), since we use the orthogonal dictionaries, we can quickly compute the exact closed form solution as shown in (13) without using OMP.

### Image Update Step

At the $m$th update of $\mu$, we solve the problem

$$\min_{\mu \in \mathbb{R}^m} \| R\mu - I \|^2_F + \sum_{q=1}^{Q} \lambda_q \sum_{s \in S_q} \| H_s\mu - \tilde{D}_q c_s \|^2_F$$

with fixed $C$. For the problem in (14), by using the same MM technique as [13], $\mu$ is updated from the previous one $\mu^{m-1}$ as in (15).

### 3.3. Importance of the Regularization Parameters $\lambda_q$

In multiclass dictionary learning, the number of patches is different for each class. In addition, there is a possibility that the ratio of the number of patches in each class is quite different between the training image and the target image. As a result, the approximation errors of the sparse representations by dictionaries differ depending on the class. In order to obtain better reconstruction results, it is desirable to set Lagrangian multipliers $\lambda_q$ (and the numbers of column vectors $K_q$ if overcomplete dictionaries are used) to different values for each class. However, it becomes difficult to achieve uniform representation accuracy among all classes as the number of classes increases.

For avoiding the above problem, in the proposed method, we also extend the regularization parameter $\lambda$ in (3) to the multiclass ones $\lambda_q$ in (7). This extension allows us to reconstruct high-quality images by proper tuning of each $\lambda_q$ even if we set $\nu_q$ to the same value for all classes. Specifically, we only have to set $\lambda_q$ to a large value for a class where the representation accuracy is high, and set $\lambda_q$ to a small value for a class where the representation accuracy is low.

## 4. NUMERICAL SIMULATION

In this section, we show the effectiveness of the proposed method in (7) with overcomplete (Proposed 1) and orthogonal (Proposed 2) dictionaries by numerical simulation based on actual CT images in [14]. We used a female head image f$_{1601}$ as a training image and used female head images f$_{1610}$ & f$_{1615}$ and male head images m$_{1114}$ & m$_{1132}$ as target images. All image sizes were $256 \times 256$ ($J = 65,536$). We created attenuation coefficients $\mu$ from CT values with the attenuation coefficient of water $\mu_{\text{water}} = 0.2059$ [cm$^{-1}$]. In (1), a Radon transform matrix was created on the basis of parallel pencil beams to 367 X-ray detectors in 60 views $\theta_g = 3(g - 1)$ [deg] ($g = 1, 2, \ldots, 60$ and $I = 22,020$), where we assumed that the width and height of one pixel were 0.09 [cm] and X-ray intensity was $b = 10^6$.

In all methods (GDSIR, ADSIR, and Proposed 1 & 2), we reconstructed images by $M = 1,000$ iterations and fixed parameters $K = 256$ (for overcomplete dictionaries) and $\nu_q = 0.2$. For overcomplete dictionaries, we used the program in [15] for fast implementation of dictionary learning and OMP. We extracted $S = 62,001$ patches of size $8 \times 8$ ($P = 64$) by shifting the areas pixel by pixel. Dictionaries of GDSIR and Proposed 1 & 2 were learned from the training image in Fig. 1(a). In Proposed 1 & 2, we divided the training patches into $Q = 7$ classes, which could achieve the highest PSNR among $Q \in \{2, 3, \ldots, 16\}$, and each clustering center was given as in Fig. 1(b). We set the regularization parameters to $\lambda_1 = 60$, $\lambda_2 = \lambda_3 = 60$, $\lambda_4 = \lambda_5 = \lambda_6 = \lambda_7 = 0.6$ in Proposed 1 & 2. The simulation environment was MacBook Pro, OS Mojave ver. 10.14, CPU 3.1 GHz Intel Core i5, and memory 8 GB 2133 MHz LPDDR3, where platform was MATLAB R2018a 64-bit.

Table 1 summarizes PSNR of the reconstructed images and the calculation time per one iteration for each method. In Figs. 2 and 3, (a), (b), (c), (d), (e), and (f) show the original image, the initial estimate obtained by the FBP, the reconstructed images by GDSIR, the reconstructed images by ADSIR, and the reconstructed images by Proposed 1 & 2, respectively. From Table 1, we found that Proposed 2 achieved both of high reconstruction accuracy and fast computation. From Figs. 2 and 3, we found that GDSIR & ADSIR lost detailed structures by over-smoothing while Proposed 1 & 2 could reconstruct the detailed structures.

## 5. CONCLUSION

In this paper, we proposed a low-dose CT reconstruction using multiclass dictionary learning. Since the propose method performs clustering for patches of a target only once, the computational cost hardly increases from GDSIR. Moreover, by introducing multiclass regularization parameters and orthogonal dictionaries, the proposed method can achieve high accuracy and fast computation. Simulation showed that the proposed method reconstructs detailed structures of targets.
Fig. 2. Reconstruction results of f_1615 from its low-dose projection data by each method [PSNR / SSIM].

Fig. 3. Reconstruction results of m_1132 from its low-dose projection data by each method [PSNR / SSIM].
6. REFERENCES


