

## PAPER

# Multiclass Dictionary-Based Statistical Iterative Reconstruction for Low-Dose CT\*

Hiryu KAMOSHITA<sup>†</sup>, Nonmember, Daichi KITAHARA<sup>†a)</sup>, Ken'ichi FUJIMOTO<sup>††b)</sup>, Members,  
Laurent CONDAT<sup>†††c)</sup>, Nonmember, and Akira HIRABAYASHI<sup>††d)</sup>, Member

**SUMMARY** This paper proposes a high-quality computed tomography (CT) image reconstruction method from low-dose X-ray projection data. A state-of-the-art method, proposed by Xu et al., 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 a data fidelity and a regularization term based on sparse representations with the dictionary. However, this method does not take characteristics of each patch, such as textures or edges, into account. In this paper, we propose to classify all patches into several classes and utilize an individual dictionary with an individual regularization parameter for each class. Furthermore, for fast computation, we introduce the orthogonality to column vectors of each dictionary. Since similar patches are collected in the same cluster, accuracy degradation by the orthogonality hardly occurs. Our simulations show that the proposed method outperforms the state-of-the-art in terms of both accuracy and speed.

**key words:** low-dose CT, image reconstruction, sparse representation, fast dictionary learning, clustering

## 1. Introduction

X-ray computed tomography (CT) [2], [3] 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 X-ray exposure is a concern [4]. Therefore, it is desirable to decrease the X-ray dose as much as possible. If we apply the standard reconstruction method, *filtered back projection (FBP)*, to low-dose projection data, then unnatural artifacts and large noise appear in reconstructed images, which leads to overlooking diseased tissues in the diagnosis [5].

To reconstruct a high-quality CT image from the low-dose projection data, we exploit the sparsity on the basis of

the compressed sensing theory [6]. In the image processing field, patch-based dictionary learning is often used to acquire sparse representations of target images [4]–[9]. A dictionary is generated as a matrix from training images, and the target image patches are supposed to be expressed sparsely as linear combinations of a few column vectors of the dictionary. In [4], Xu et al. utilized the patch-based dictionary learning with *statistical iterative reconstruction (SIR)* [10]–[12] from the low-dose projection data. Although this method can well remove *streak artifacts*, detailed structures of the target images are lost by over-smoothing. In order to reconstruct the detailed structures while suppressing the artifacts and noise, it is necessary to express each image patch more sparsely.

To this end, this paper proposes to extend the reconstruction method of Xu et al. [4] to a multiclass version. In the proposed method, all training image patches are divided into several classes, and a dictionary is created for each class. Then, target image patches are also classified and supposed to be expressed sparsely by the dictionary for each class. In this case, approximation errors of the sparse representations for the patches would differ depending on the class. Therefore, we also change the regularization parameter value for each class. Moreover, for fast reconstruction, we introduce the orthogonality to each dictionary, i.e., we replace overcomplete dictionaries with orthogonal matrices. In general, reduction of column vectors of the dictionary causes accuracy degradation of the sparse representations, and thus low image quality. On the other hand, in the proposed method, since the dictionary is utilized only for similar patches because of clustering, the accuracy hardly deteriorates. Our simulations based on actual CT images show that the proposed method can obtain reconstructed results better than those by the state-of-the-art method [4] in a shorter calculation time.

## 2. Basic Principles of X-ray CT

We formulate the X-ray CT image reconstruction in the  $xy$ -coordinate system as shown in Fig. 1. Let  $\mu(x, y) \geq 0$  be the X-ray attenuation coefficient distribution in a field of view. X-ray projection is described with the  $uv$ -coordinate system:

$$\begin{cases} u = x \cos \theta + y \sin \theta, \\ v = -x \sin \theta + y \cos \theta, \end{cases}$$

which is given by rotating the  $xy$ -coordinate system through an angle  $\theta \in [0, \pi)$ . The ideal projection data  $I_{\theta, n}^*$  are defined

Manuscript received February 13, 2020.

Manuscript revised August 3, 2020.

Manuscript publicized October 6, 2020.

<sup>†</sup>The authors are with the Dept. of Information Science and Engineering, Ritsumeikan University, Kusatsu-shi, 525-8577 Japan.

<sup>††</sup>The author is with the Dept. of Engineering and Design, Kagawa University, Takamatsu-shi, 761-0396 Japan.

<sup>†††</sup>The author is with the Visual Computing Center, King Abdulah University of Science and Technology, Thuwal, Saudi Arabia.

\*This paper is an extended version of our previous paper presented at the IEEE ICIP, Taipei, Taiwan, September 2019 [1]. The current paper describes the proposed method more precisely and numerical simulations are conducted in a more realistic setting.

a) E-mail: d-kita@media.ritsumei.ac.jp (Corresponding author)

b) E-mail: fujimoto.kenichi@kagawa-u.ac.jp

c) E-mail: laurent.condat@kaust.edu.sa

d) E-mail: akirahr@media.ritsumei.ac.jp

DOI: 10.1587/transfun.2020EAP1020

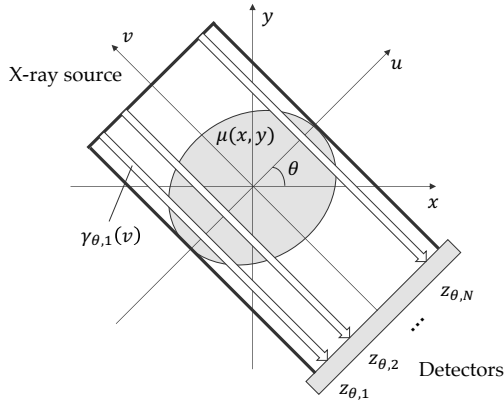


Fig. 1 X-ray projection in a parallel-beam CT system.

by line integrals, that are parallel to the  $v$ -axis, of the distribution  $\mu(x, y)$  as  $l_{\theta,n}^* := \int \mu(\gamma_{\theta,n}(v)) dv$  ( $n = 1, 2, \dots, N$ ), where  $\gamma_{\theta,n}(v)$  is a straight path from X-ray source to the  $n$ th detector. These line integrals are called the Radon transform. The Radon transform is discretized by numerical integration, as

$$l_{\theta,n}^* = \mathbf{r}_{\theta,n}^T \boldsymbol{\mu} + \xi_{\theta,n}^*, \quad (1)$$

where  $\boldsymbol{\mu} \in \mathbb{R}_+^J$  is a nonnegative vector corresponding to a CT image of size  $\sqrt{J} \times \sqrt{J}$ ,  $\mathbf{r}_{\theta,n} \in \mathbb{R}_+^J$  is a nonnegative vector corresponding to weights in the numerical integration,  $(\cdot)^T$  is the transpose operator, and  $\xi_{\theta,n}^* \in \mathbb{R}$  is the discretization error.

The observed data  $z_{\theta,n} \in \mathbb{Z}_+$  at the  $n$ th X-ray detector is a nonnegative integer following a Poisson distribution of expected value  $E[z_{\theta,n}] = b \exp(-l_{\theta,n}^*)$ , where  $b > 0$  is the X-ray radiation intensity of the X-ray source [4]. From this relation, the ideal projection data is expressed as  $l_{\theta,n}^* = \ln(b/E[z_{\theta,n}])$ . Actually, since  $E[z_{\theta,n}]$  is unknown,  $l_{\theta,n} := \ln(b/z_{\theta,n})$  is utilized, instead of  $l_{\theta,n}^*$ , as the projection data computed from the observed data  $z_{\theta,n}$ . If we acquire the projection data with  $N$  detectors for  $G$  angles  $\theta_g \in [0, \pi)$  ( $\theta_1 < \theta_2 < \dots < \theta_G$ ), the observation models in (1) are combined to a matrix form

$$\mathbf{l} = \mathbf{R}\boldsymbol{\mu} + \boldsymbol{\xi}, \quad (2)$$

where  $\mathbf{l} \in \mathbb{R}^I$  is a vector concatenating all  $l_{\theta_g,n}$ ,  $\mathbf{R} \in \mathbb{R}_+^{I \times J}$  is a matrix whose row vectors are  $\mathbf{r}_{\theta_g,n}^T$ ,  $\boldsymbol{\xi} \in \mathbb{R}^I$  is an error vector, and the number of the observations is  $I := GN$ . In standard-dose CT, since  $I \gg J$  and the noise level  $\|\boldsymbol{\xi}\|_2$  is small,  $\boldsymbol{\mu}$  is reconstructed with high accuracy by the FBP method [2], [3].

### 3. Dictionary-Based SIR from Low-Dose Projection

There exist two approaches for low-dose CT: (i) reducing the number  $G$  of projection views and (ii) lowering the X-ray radiation intensity  $b$ . In the first approach, the number  $I$  of row vectors of  $\mathbf{R}$  decreases, implying  $\text{rank}(\mathbf{R}) < J$ . In the second approach, the noise level  $\|\boldsymbol{\xi}\|_2$  in (2) becomes larger. As a result, in either case, the standard FBP method generates streak artifacts and large noise in the reconstructed images [5].

Sauer et al. proposed a method that reconstructs  $\boldsymbol{\mu}$  by a maximum a posteriori (MAP) estimation [10]. This method

reconstructs  $\boldsymbol{\mu}$  by iteratively minimizing a cost function

$$\|\mathbf{R}\boldsymbol{\mu} - \mathbf{l}\|_w^2 + \lambda \Psi(\boldsymbol{\mu}) := \sum_{i=1}^I w_i (\mathbf{r}_i^T \boldsymbol{\mu} - l_i)^2 + \lambda \Psi(\boldsymbol{\mu}), \quad (3)$$

where the *statistical weight*  $w_i := z_i = b \exp(-l_i)$  for the  $i$ th projection data  $l_i$  is derived by the second-order Taylor series approximation for the log-likelihood of the Poisson distribution, and  $\Psi: \mathbb{R}^J \rightarrow \mathbb{R}$  is some regularization term based on a prior distribution of  $\boldsymbol{\mu}$ . The regularization parameter  $\lambda > 0$  controls the balance between the data fidelity and regularization terms. The optimization algorithm of Sauer et al. and its accelerated version which uses a majorization-minimization (MM) technique [12] are called SIR methods [11].

Xu et al. combined (3) and patch-based dictionary learning [4]. In this method, a target image  $\boldsymbol{\mu}$  is decomposed into small images  $H_s \boldsymbol{\mu} \in \mathbb{R}_+^P$  ( $s = 1, 2, \dots, S$ ), called *patches*, of size  $\sqrt{P} \times \sqrt{P}$ , where  $H_s \in \{0, 1\}^{P \times J}$  denotes the  $s$ th patch extraction matrix and  $S$  is the number of the patches. Then,  $\boldsymbol{\mu}$  is reconstructed under 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  $\mathbf{d}_k \in \mathbb{R}^P$  s.t.  $\|\mathbf{d}_k\|_2 = 1$  ( $k = 1, 2, \dots, K$ ). Let  $\mathbf{c}_s \in \mathbb{R}^K$  be sparse representations (coefficients) for  $H_s \boldsymbol{\mu}$ , and this method reconstructs the CT image  $\boldsymbol{\mu}$  by solving

$$\underset{\boldsymbol{\mu}, (D), C}{\text{minimize}} \|\mathbf{R}\boldsymbol{\mu} - \mathbf{l}\|_w^2 + \lambda \sum_{s=1}^S (\|H_s \boldsymbol{\mu} - D\mathbf{c}_s\|_2^2 + \nu_s \|\mathbf{c}_s\|_0), \quad (4)$$

where  $C := (\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_S) \in \mathbb{R}^{K \times S}$  is a coefficient matrix and  $\nu_s > 0$  is a Lagrangian multiplier for each patch. There are two setups for the problem in (4): (i)  $D$  is fixed to a matrix pre-learned from training images, and (ii)  $D$  is updated adaptively from a current estimate of the target image  $\boldsymbol{\mu}$ . The former is called *global dictionary-based SIR (GDSIR)*, and the latter is called *adaptive dictionary-based SIR (ADSIR)*.

Since it is difficult to directly solve the problem in (4), the coefficient matrix  $C$  (and the dictionary  $D$  in ADSIR) and the CT image  $\boldsymbol{\mu}$  are alternately updated from the initial image  $\boldsymbol{\mu}^{(0)}$ . In this method, the initial image  $\boldsymbol{\mu}^{(0)}$  is set to the FBP result, in which negative pixel values are replaced with zeros. The matrix  $C$  (and  $D$  in ADSIR) is updated by solving

$$\underset{(D), C}{\text{minimize}} \sum_{s=1}^S (\|H_s \boldsymbol{\mu}^{(m-1)} - D\mathbf{c}_s\|_2^2 + \nu_s \|\mathbf{c}_s\|_0), \quad (5)$$

where  $\boldsymbol{\mu}$  is fixed to the current estimate  $\boldsymbol{\mu}^{(m-1)}$ . In GDSIR,  $C$  is updated by *orthogonal matching pursuit (OMP)* [13] (see Sect. 4.2 in detail). In ADSIR, after  $D$  is updated by an online dictionary learning method e.g. [14],  $C$  is updated by OMP. Thus ADSIR requires computational cost more than GDSIR.

The CT image  $\boldsymbol{\mu}$  is updated by solving

$$\underset{\boldsymbol{\mu} \in \mathbb{R}_+^J}{\text{minimize}} \|\mathbf{R}\boldsymbol{\mu} - \mathbf{l}\|_w^2 + \lambda \sum_{s=1}^S \|H_s \boldsymbol{\mu} - D\mathbf{c}_s\|_2^2 \quad (6)$$

with fixed  $C$  and  $D$ . The optimal solution  $\boldsymbol{\mu}^*$  to (6) is given by

$$\boldsymbol{\mu}^* = \left( R^T W R + \lambda \sum_{s=1}^S H_s^T H_s \right)^{-1} \left( R^T W \mathbf{I} + \lambda \sum_{s=1}^S H_s^T D \mathbf{c}_s \right)$$

with a diagonal matrix  $W = \text{diag}(\mathbf{w}) \in \mathbb{R}_+^{I \times I}$ . However, if the image size is large, this linear system is difficult to solve. To avoid this, Xu et al. used the same MM technique as in [12]. Specifically,  $\boldsymbol{\mu}$  is updated from  $\boldsymbol{\mu}^{(m-1)}$  by (7), where  $\mathbf{0}$  and  $\mathbf{1}$  denote vectors with all components 0 and 1, respectively,  $\oslash$  is the component-wise division, and  $\max$  returns a larger value for each pair of corresponding components of two vectors.

#### 4. SIR with Multiclass Dictionary Learning

##### 4.1 Multiclass GDSIR (MGDSIR)

Recently, in high-speed MRI and low-dose CT fields, image reconstruction methods using *multiclass* dictionary learning have been proposed for higher compression ratio, noise reduction, and artifact reduction [5], [15], [16]. In these methods, all image patches are divided into multiple classes using geometric directions or pixel values. Since an individual dictionary is generated for each class, sparse representations for the patches become more accurate, and better reconstruction results are obtained than when we utilize only one dictionary. The methods in [15], [16] are proposed for high-speed MRI and thus cannot be directly applied to low-dose CT images. The method in [5], which is proposed for low-dose CT, uses  $\|\boldsymbol{\mu} - \boldsymbol{\mu}^{(0)}\|_2^2$  as the data fidelity term, and hence the reconstruction results largely depend on  $\boldsymbol{\mu}^{(0)}$  obtained by the FBP method. On the other hand, since GDSIR uses  $\|R\boldsymbol{\mu} - \mathbf{I}\|_w^2$  in (4) as the data fidelity term, the reconstruction results are robust against the initial image  $\boldsymbol{\mu}^{(0)}$ . Thus, in this paper, we propose to extend GDSIR to a multiclass version in order to reconstruct higher-quality images from low-dose projection.

Let  $Q$  be the number of classes and  $q = 1, 2, \dots, Q$  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 CT images, where  $K_q$  is the number of column vectors of  $D_q$ . Each patch  $H_s \boldsymbol{\mu}$  is classified into one of  $Q$  classes, and a patch index set for the  $q$ th class is denoted by  $\mathcal{S}_q$  s.t.  $\bigcup_{q=1}^Q \mathcal{S}_q = \{1, 2, \dots, S\}$  and  $\mathcal{S}_q \cap \mathcal{S}_{q'} = \emptyset$  ( $q \neq q'$ ). We suppose  $\mathcal{S}_q$  is fixed hereafter. We reconstruct  $\boldsymbol{\mu}$  by solving a multiclass version of (4):

$$\underset{\boldsymbol{\mu}, C}{\text{minimize}} \quad \|R\boldsymbol{\mu} - \mathbf{I}\|_w^2 + \sum_{q=1}^Q \lambda_q \sum_{s \in \mathcal{S}_q} (\|H_s \boldsymbol{\mu} - D_q \mathbf{c}_s\|_2^2 + \nu_s \|\mathbf{c}_s\|_0), \quad (8)$$

where  $\lambda_q > 0$  is a regularization parameter for the  $q$ th class. The problem in (8) can be solved by a similar update to that of GDSIR. We call this method *multiclass GDSIR (MGDSIR)*.

##### 4.2 Significance of Multiple Regularization Parameters $\lambda_q$

In this part, we explain the merit of using multiple regularization parameters  $\lambda_q$ . When updating coefficient matrix  $C$  by OMP in a similar manner to (5) of GDSIR, there are mainly three types of update-rules [17]. Type 1 update-rule fixes the Lagrangian multipliers as  $\nu_s = \nu$  ( $s = 1, 2, \dots, S$ ) and solves

$$\underset{C}{\text{minimize}} \quad \sum_{s=1}^S (\|H_s \boldsymbol{\mu}^{(m-1)} - D_{q(s)} \mathbf{c}_s\|_2^2 + \nu \|\mathbf{c}_s\|_0),$$

where  $q(s)$  denotes the class of the  $s$ th patch. Type 2 update-rule restricts the sparse level of each  $\mathbf{c}_s$  under  $T$  and solves

$$\underset{C}{\text{minimize}} \quad \sum_{s=1}^S \|H_s \boldsymbol{\mu}^{(m-1)} - D_{q(s)} \mathbf{c}_s\|_2^2 \quad \text{s.t. } \forall s \|\mathbf{c}_s\|_0 \leq T.$$

Type 3 update-rule restricts each error under  $\varepsilon$  and solves

$$\underset{C}{\text{minimize}} \quad \sum_{s=1}^S \|\mathbf{c}_s\|_0 \quad \text{s.t. } \forall s \|H_s \boldsymbol{\mu}^{(m-1)} - D_{q(s)} \mathbf{c}_s\|_2^2 \leq \varepsilon.$$

Note that Type 2 and 3 problems can be expressed in Type 1 form as in (8) by using different Lagrangian multipliers  $\nu_s$ .

In multiclass dictionary learning, the number of patches is different for each class. Further, there is a possibility that the ratio of the number of patches in each class is quite different between training images and a target image. As a result, the approximation errors of the sparse representations, with the same sparse level, differ depending on the class. For high-quality image reconstruction,  $K_q$ ,  $\nu$ ,  $T$ , and  $\varepsilon$  should be set to different values for each class, and then we should choose the best update-rule. However, such a parameter tuning becomes more burdensome as the number  $Q$  of the classes increases.

To avoid the above burden in the multiclass dictionary learning, in the proposed method, we also extended the regularization parameter  $\lambda$  in (4) to the multiclass ones  $\lambda_q$  in (8). This extension allows us to reconstruct high-quality images by an appropriate tuning of each  $\lambda_q$  even if we set  $K_q$ ,  $\nu$ ,  $T$ , and  $\varepsilon$  to the same values for all classes. Specifically, we set  $\lambda_q$  to a large value for a class of high representation accuracy, and to a small value for a class of low accuracy. Indeed,  $\lambda_q$  removes the burdensome tuning in the dictionary learning.

#### 5. Acceleration and Update Formulae of MGDSIR

##### 5.1 Fast MGDSIR (FMGDSIR)

To accelerate MGDSIR, we further propose to introduce the orthogonality to column vectors of each dictionary in (8) as

$$\boldsymbol{\mu}^{(m)} = \max \left\{ \mathbf{0}, \boldsymbol{\mu}^{(m-1)} - \left[ R^T W (R \boldsymbol{\mu}^{(m-1)} - \mathbf{I}) + \lambda \sum_{s=1}^S H_s^T (H_s \boldsymbol{\mu}^{(m-1)} - D \mathbf{c}_s) \right] \oslash \left[ \left( R^T W R + \lambda \sum_{s=1}^S H_s^T H_s \right) \mathbf{1} \right] \right\} \quad (7)$$

$$\underset{\mu, C}{\text{minimize}} \|R\mu - I\|_w^2 + \sum_{q=1}^Q \lambda_q \sum_{s \in \mathcal{S}_q} (\|H_s \mu - \widehat{D}_q c_s\|_2^2 + \nu_s \|c_s\|_0), \quad (9)$$

where  $\widehat{D}_q \in \mathbb{R}^{P \times P}$  is an orthogonal dictionary s.t.  $\widehat{D}_q^T \widehat{D}_q = E$ , and  $E$  denotes the identity matrix of order  $P$ . In general, if we replace an overcomplete dictionary with an orthogonal dictionary, then dictionary learning can be accelerated while the accuracy of the sparse representation deteriorates. On the other hand, in the proposed method, since each dictionary has only to deal with similar patches thanks to clustering, the accuracy hardly deteriorates even if we use the orthogonal dictionaries. We call this method *Fast MGDSIR (FMGDSIR)*.

MGDSIR and FMGDSIR reconstruct  $\mu$  from the initial estimate  $\mu^{(0)}$  by alternately updating the sparse coefficients and the image  $M$  times after generating the global dictionaries. Because of the orthogonality, both dictionary generation and image reconstruction can be quickly done in FMGDSIR.

## 5.2 Global Dictionary Generation

First, patches are extracted from training images and divided into  $Q$  classes by the K-means algorithm in the same manner as in [16]. Note that the clustering centers are also used in the reconstruction step to classify target patches. In FMGDSIR, an orthogonal dictionary  $\widehat{D}_q$  is generated<sup>†</sup> for each class as an approximate solution to the optimization problem

$$\underset{\widehat{D}_q, C_q^{\text{tr}}}{\text{minimize}} \sum_{s \in \mathcal{S}_q^{\text{tr}}} (\|H_s \mu^{\text{tr}} - \widehat{D}_q c_s^{\text{tr}}\|_2^2 + \nu_s \|c_s^{\text{tr}}\|_0) \quad \text{s.t.} \quad \widehat{D}_q^T \widehat{D}_q = E, \quad (10)$$

where  $\mu^{\text{tr}}$  is a standard-dose training image,  $c_s^{\text{tr}}$  is a sparse coefficient vector for the  $s$ th training patch  $H_s \mu^{\text{tr}}$ ,  $\mathcal{S}_q^{\text{tr}}$  is a patch index set for training patches of class  $q$ , and all  $c_s^{\text{tr}}$  for  $s \in \mathcal{S}_q^{\text{tr}}$  are combined to a matrix  $C_q^{\text{tr}}$ . We solve the problem in (10) according to [8], and thus while fixing  $C_q^{\text{tr}}$ ,  $\widehat{D}_q$  is updated by

$$\widehat{D}_q^* = \underset{\widehat{D}_q}{\text{argmin}} \|\Phi_q^{\text{tr}} - \widehat{D}_q C_q^{\text{tr}}\|_F^2 \quad \text{s.t.} \quad \widehat{D}_q^T \widehat{D}_q = E, \quad (11)$$

where  $\Phi_q^{\text{tr}} \in \mathbb{R}^{P \times |\mathcal{S}_q^{\text{tr}}|}$  is a matrix whose column vectors are  $H_s \mu^{\text{tr}}$  s.t.  $s \in \mathcal{S}_q^{\text{tr}}$ , and  $\|\cdot\|_F$  denotes the Frobenius norm.

The cost function in (11) is expressed as

$$\begin{aligned} \|\Phi_q^{\text{tr}} - \widehat{D}_q C_q^{\text{tr}}\|_F^2 &= \|\Phi_q^{\text{tr}}\|_F^2 + \|\widehat{D}_q C_q^{\text{tr}}\|_F^2 - 2 \text{Trace}(C_q^{\text{tr}T} \widehat{D}_q^T \Phi_q^{\text{tr}}) \\ &= \|\Phi_q^{\text{tr}}\|_F^2 + \|C_q^{\text{tr}}\|_F^2 - 2 \text{Trace}(\Phi_q^{\text{tr}} C_q^{\text{tr}T} \widehat{D}_q^T), \end{aligned}$$

where Trace returns the sum of all diagonal components of a square matrix. In this equation, the first and second terms on the right side are constant for  $\widehat{D}_q$ . Thus, (11) is expressed as

$$\widehat{D}_q^* = \underset{\widehat{D}_q}{\text{argmax}} \text{Trace}(\Phi_q^{\text{tr}} C_q^{\text{tr}T} \widehat{D}_q^T) \quad \text{s.t.} \quad \widehat{D}_q^T \widehat{D}_q = E. \quad (12)$$

<sup>†</sup>In MGDSIR, an over-complete dictionary  $D_q$  is generated by solving the problem in (10) without the orthogonality constraint.

This problem is known as the *orthogonal Procrustes problem* [6] and can be solved as follows. By using singular value decomposition (SVD)  $\Phi_q^{\text{tr}} C_q^{\text{tr}T} = U \Sigma V^T$  with one diagonal matrix  $\Sigma \in \mathbb{R}_+^{P \times P}$  and two orthogonal matrices  $U \in \mathbb{R}^{P \times P}$  and  $V \in \mathbb{R}^{P \times P}$ , the cost function in (12) yields

$$\text{Trace}(\Phi_q^{\text{tr}} C_q^{\text{tr}T} \widehat{D}_q^T) = \text{Trace}(U \Sigma V^T \widehat{D}_q^T) = \text{Trace}(\Sigma V^T \widehat{D}_q^T U).$$

Since  $V^T \widehat{D}_q^T U$  is an orthogonal matrix and every component is smaller than 1, the cost function in (12) is maximized when  $V^T \widehat{D}_q^T U = E$ . As a result, the optimal solution is given by

$$\widehat{D}_q^* = UV^T. \quad (13)$$

After the dictionary update by (13), we update a coefficient matrix  $C^{\text{tr}}$  while fixing  $\widehat{D}_q = \widehat{D}_q^*$  in (10). In FMGDSIR, we utilize the property that the  $\ell_2$ -norm is *unitarily invariant*, and the coefficient matrix is updated<sup>††</sup> by

$$\begin{aligned} C_q^{\text{tr}*} &= \underset{C_q^{\text{tr}}}{\text{argmin}} \sum_{s \in \mathcal{S}_q^{\text{tr}}} (\|H_s \mu^{\text{tr}} - \widehat{D}_q c_s^{\text{tr}}\|_2^2 + \nu_s \|c_s^{\text{tr}}\|_0) \\ &= \underset{C_q^{\text{tr}}}{\text{argmin}} \sum_{s \in \mathcal{S}_q^{\text{tr}}} (\|\widehat{D}_q^T H_s \mu^{\text{tr}} - c_s^{\text{tr}}\|_2^2 + \nu_s \|c_s^{\text{tr}}\|_0). \end{aligned} \quad (14)$$

In the orthogonal dictionary learning, Type 1 update-rule is often used [8] and (14) is solved for each column vector  $c_s$  as

$$c_s^{\text{tr}*} = \underset{c_s^{\text{tr}}}{\text{argmin}} \|\widehat{D}_q^T H_s \mu^{\text{tr}} - c_s^{\text{tr}}\|_2^2 + \nu \|c_s^{\text{tr}}\|_0, \quad (15)$$

where the Lagrangian multipliers are fixed as  $\nu_s = \nu$  for all  $s \in \mathcal{S}_q^{\text{tr}}$ . The optimal solution to (15) is easily given by

$$c_s^{\text{tr}*} = \text{hard\_thresh}_{\sqrt{\nu}}(\widehat{D}_q^T H_s \mu^{\text{tr}}) \quad (16)$$

with the hard thresholding operator:

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

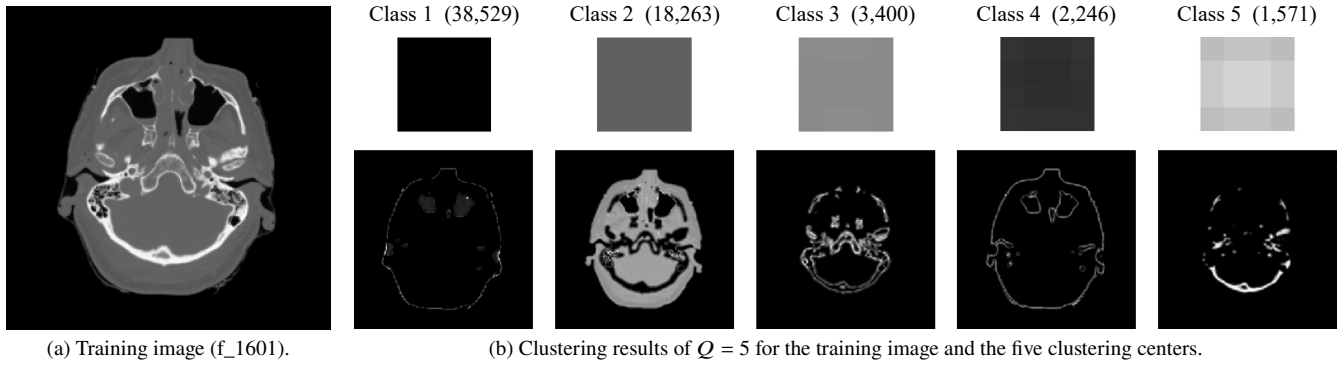
where  $[i]$  stands for the  $i$ th component of a vector. Note that Type 2 and 3 update-rules are also easily given. By repeating (13) and (16) for each class, the dictionaries  $\widehat{D}_q$  are obtained.

## 5.3 Image Reconstruction

To solve (9) with the multiclass orthogonal dictionaries  $\widehat{D}_q$ , first, the initial estimate  $\mu^{(0)}$  is obtained by applying the FBP method to the low-dose projection data  $I$ . Then, by using the stored clustering centers in the dictionary generation step, the initial patches  $H_s \mu^{(0)}$  are classified into one of  $Q$  classes, and hereafter the patch index sets  $\mathcal{S}_q$  are fixed. We reconstruct an image by alternately updating  $C$  and  $\mu$  for  $m = 1, 2, \dots, M$ .

At the  $m$ th update of  $C$ , if we choose Type 1 update-rule, then we solve, as shown in Sect. 4.2, the following problem

<sup>††</sup>In MGDSIR, coefficients are updated by OMP as in Sect. 4.2. Note that  $D_q$  can be updated by the same rule for any OMP type.



**Fig. 2** Training image used for dictionary learning and the clustering results of the K-means algorithm in the proposed method.

$$\text{minimize}_C \sum_{s=1}^S (\|H_s \boldsymbol{\mu}^{(m-1)} - \widehat{D}_{q(s)} \mathbf{c}_s\|_2^2 + \nu \|\mathbf{c}_s\|_0), \quad (17)$$

where  $\boldsymbol{\mu}$  is fixed to the current estimate  $\boldsymbol{\mu}^{(m-1)}$ . In the same manner as (14), the problem in (17) is solved for each  $\mathbf{c}_s$  as

$$\mathbf{c}_s^* = \text{hard\_thresh}_{\sqrt{\nu}} (\widehat{D}_{q(s)}^T H_s \boldsymbol{\mu}^{(m-1)}). \quad (18)$$

Differently from the problem in (5), the orthogonal dictionaries allow us to quickly compute the exact closed-form solution to (18) without using a greedy algorithm such as OMP. At the  $m$ th update of  $\boldsymbol{\mu}$ , we solve a least squares problem

$$\text{minimize}_{\boldsymbol{\mu} \in \mathbb{R}_+^J} \|R\boldsymbol{\mu} - \mathbf{I}\|_w^2 + \sum_{q=1}^Q \lambda_q \sum_{s \in \mathcal{S}_q} \|H_s \boldsymbol{\mu} - \widehat{D}_q \mathbf{c}_s\|_2^2 \quad (19)$$

for a fixed  $C$ . For the problem in (19), to avoid the computation of a large inverse matrix, we use the same MM technique as in [12], and  $\boldsymbol{\mu}$  is updated from  $\boldsymbol{\mu}^{(m-1)}$  by (20). Note that in MGDSIR,  $\widehat{D}_q$  is replaced with  $D_q$  in the update by (20).

## 6. Numerical Simulations and Discussions

In this section, we compare the proposed methods, MGDSIR and FMGDSIR, with the conventional methods, ADSIR and GDSIR, by simulations based on actual CT images in [18].

### 6.1 Simulation Data and Environment

We used a female head image ‘f\_1601’ as training data, and female head images ‘f\_1610’ and ‘f\_1615’ and male head images ‘m\_1114’ and ‘m\_1132’ were targets. The original sizes of these images were  $512 \times 512$ . By cropping the outer black pixels, we made training and target images composed of the central  $256 \times 256$  pixels ( $J = 65,536$ ). Thus, the pixel length in the training and target images was the same as the

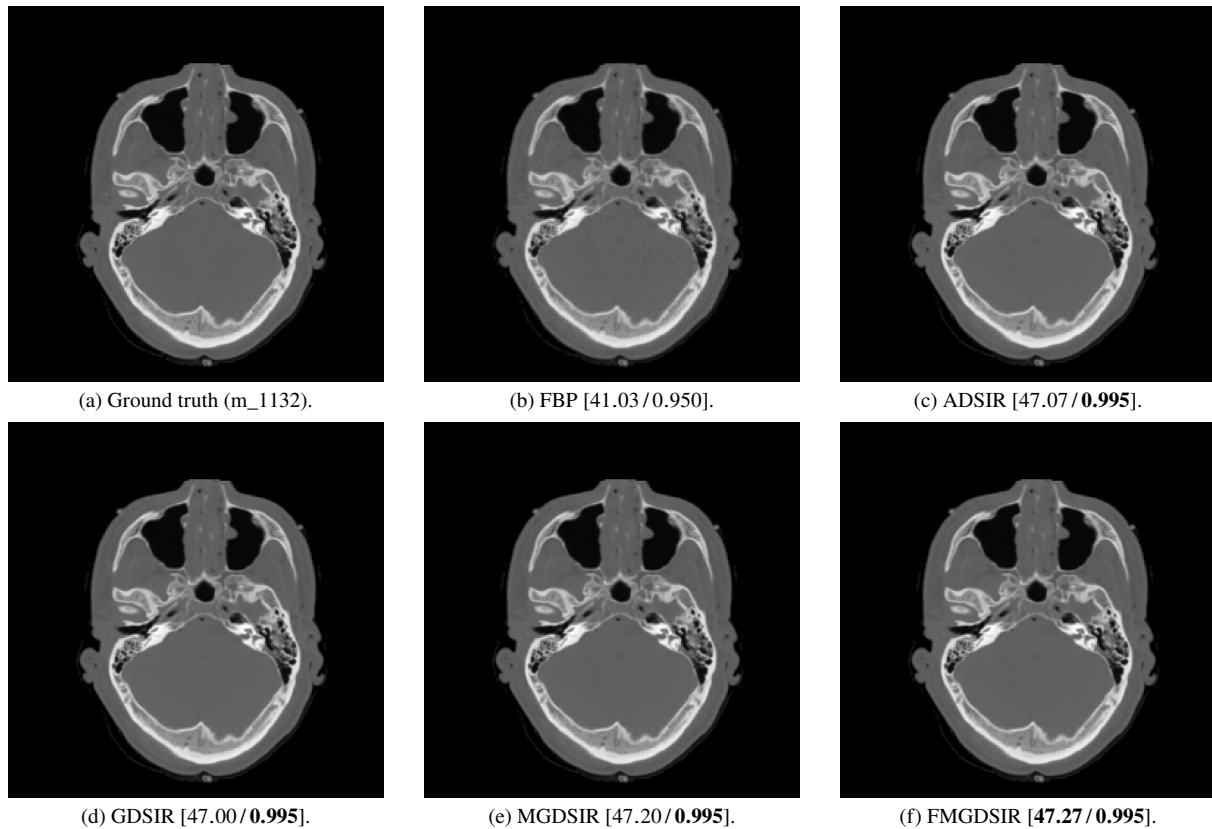
original pixel length, i.e., 1 mm. CT values  $v_{CT}$  in the images were transformed to the X-ray attenuation coefficients by  $\boldsymbol{\mu} = \mu_{\text{water}}(\mathbf{1} + v_{CT}/1000)$ , where  $\mu_{\text{water}} = 0.2059 \text{ cm}^{-1}$ . In (2), the Radon transform  $R$  was created on the basis of parallel beams to 579 X-ray detectors ( $N = 579$ ) of width 0.625 mm. Initial images  $\boldsymbol{\mu}^{(0)}$  were obtained by FBP with the Ram-Lak filter.

We reconstructed the target images by ADSIR, GDSIR, MGDSIR, and FMGDSIR with  $M = 1,000$  iterations. We set  $K_q = K = 256$  and  $\nu_s = \nu = 0.001$  for ADSIR, GDSIR and MGDSIR, and set  $\nu_s = \nu = 0.0007$  for FMGDSIR. In over-complete dictionary learning and OMP, we used the program in [17] for fast implementation. Dictionaries were generated with 100 iterations in ADSIR, with 2,000 iterations in GDSIR and MGDSIR, and with 1,000 iterations in FMGDSIR. Numerical simulations were conducted with MATLAB R2018a 64-bit on a MacBook Pro, OS Mojave ver. 10.14.6, CPU 3.1 GHz Intel Core i5, and memory 8 GB 2133 MHz LPDDR3.

### 6.2 Clustering and Regularization Parameters

We extracted  $S = 64,009$  image patches of size  $4 \times 4$  ( $P = 16$ ) by shifting the areas pixel by pixel. After clustering by the K-means algorithm, the mean value, called a DC component, of each patch is subtracted right before dictionary learning [17]. The dictionaries of GDSIR, MGDSIR, and FMGDSIR were generated from the training image in Fig. 2(a), where we divided the training patches into  $Q = 5$  classes in the proposed methods. The clustering results of the K-means algorithm and each clustering center are shown in Fig. 2(b). The cluster indices were ordered by the number of patches in descending order. The class 1 is mainly composed of dark patches, e.g., the background and the maxillary sinus. The class 2 is mainly composed of gray patches of soft tissues. The class 4 is mainly composed of middle-level patches in the boundary between the classes 1 and 2. The classes 3 and 5 are composed of bright patches of bones and detailed structures.

$$\boldsymbol{\mu}^{(m)} = \max \left\{ \mathbf{0}, \boldsymbol{\mu}^{(m-1)} - \left[ R^T W (R\boldsymbol{\mu}^{(m-1)} - \mathbf{I}) + \sum_{q=1}^Q \lambda_q \sum_{s \in \mathcal{S}_q} H_s^T (H_s \boldsymbol{\mu}^{(m-1)} - \widehat{D}_q \mathbf{c}_s) \right] \oslash \left[ \left( R^T W R + \sum_{q=1}^Q \lambda_q \sum_{s \in \mathcal{S}_q} H_s^T H_s \right) \mathbf{1} \right] \right\} \quad (20)$$



**Fig. 3** Reconstruction result of m\_1132 from its standard-dose projection data by each method [PSNR/SSIM].

**Table 1** PSNR [dB] and calculation time [sec] in 60/300 views.

SIR \ Image	f_1610	f_1615	m_1114	m_1132	calc. time
ADSIR	37.43	37.81	38.16	36.76	0.240/iter.
GDSIR	37.27	37.66	38.04	36.67	0.146/iter.
MGDSIR	38.13	38.34	38.78	<b>37.24</b>	0.168/iter.
FMGDSIR	<b>38.14</b>	<b>38.41</b>	<b>38.92</b>	37.20	<b>0.090/iter.</b>

**Table 2** PSNR [dB] and calculation time [sec] in 1/40 X-ray intensity.

SIR \ Image	f_1610	f_1615	m_1114	m_1132	calc. time
ADSIR	38.15	39.07	39.00	38.00	0.345/iter.
GDSIR	38.46	39.36	39.33	38.20	0.244/iter.
MGDSIR	<b>39.12</b>	<b>39.97</b>	<b>39.73</b>	<b>38.68</b>	0.266/iter.
FMGDSIR	38.50	39.36	39.24	38.18	<b>0.185/iter.</b>

To determine the values of the multiple regularization parameters  $\lambda_q$  in MGDSIR and FMGDSIR, we use that of  $\lambda$  in GDSIR as a reference. For patches in the classes 1 and 2, DC components are dominant, and the patches are well approximated by both the overcomplete dictionaries  $D_q$  and the orthogonal dictionaries  $\hat{D}_q$ . Hence,  $\lambda_1$  and  $\lambda_2$  can be larger than  $\lambda$ . On the other hand, patches in the classes 3 and 5 are very complex and not well approximated by the dictionaries, and hence  $\lambda_3$  and  $\lambda_5$  can be smaller than  $\lambda$ . The class 4 is difficult to handle. Patches in the class 4 mainly correspond to the boundary between the classes 1 and 2, which are better approximated by the overcomplete dictionary than by the orthogonal dictionary. Based on these considerations, we first adjusted  $\lambda$  in GDSIR, and then adjusted  $\lambda_q$  in MGDSIR and FMGDSIR manually by the simple change-and-test process.

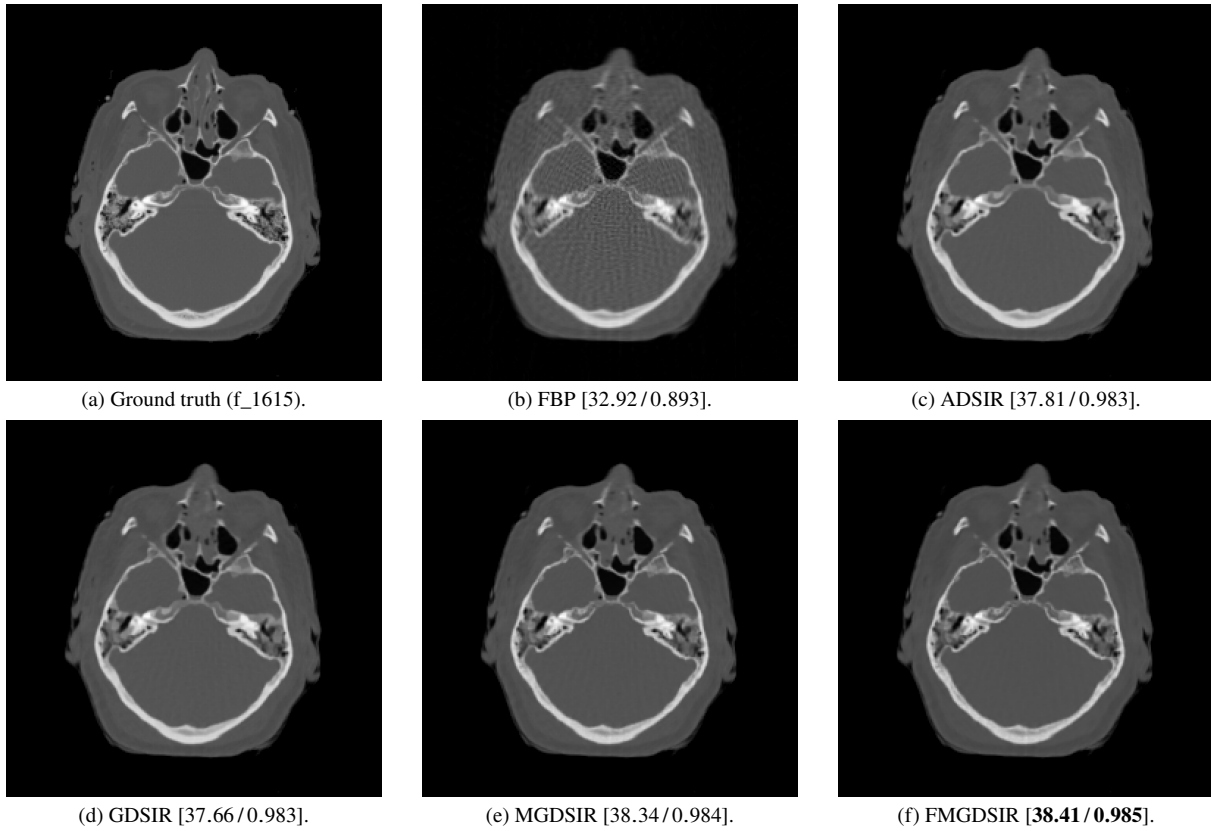
### 6.3 Standard-Dose Simulation Results

We first reconstructed ‘m\_1132’ from standard-dose projection data in  $G = 300$  views, where we set  $\theta_g := 0.6(g-1)$  deg and the X-ray radiation intensity to  $b = 10^6$ . We set the regu-

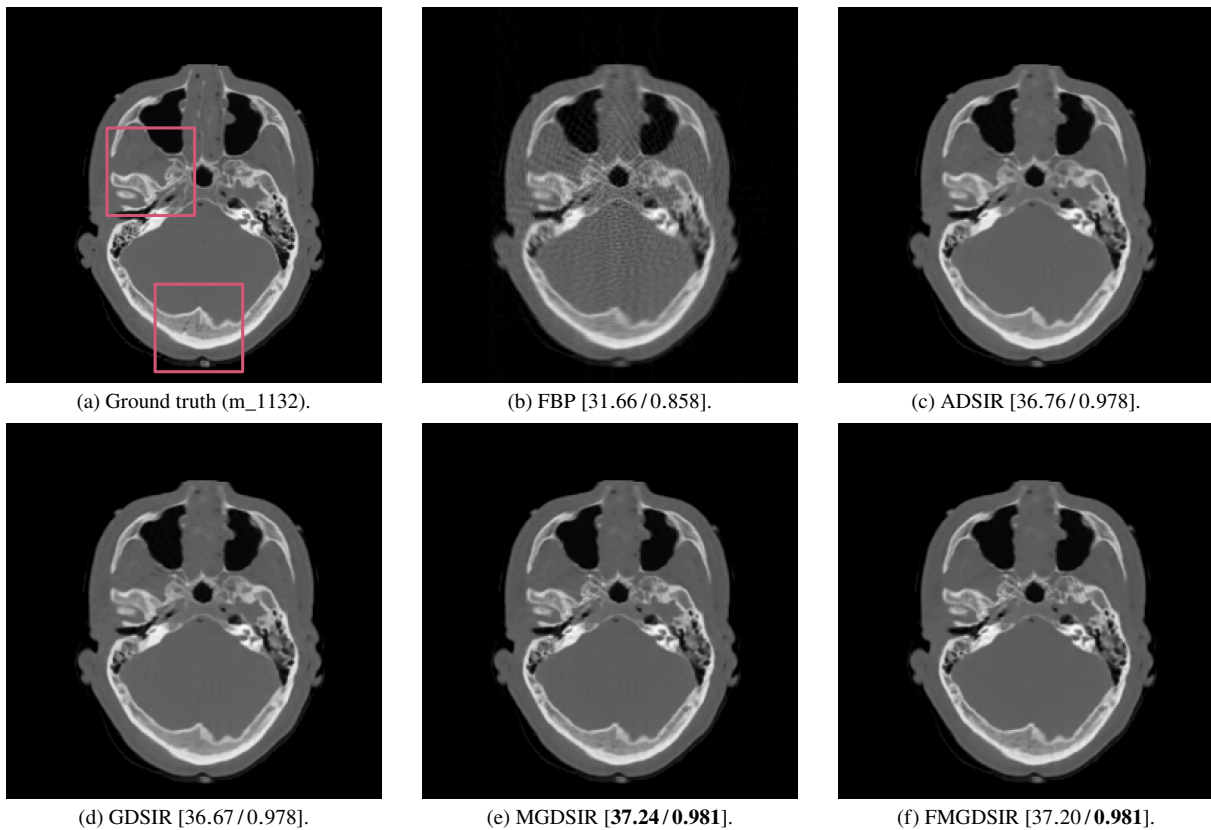
larization parameters to  $\lambda = 2,000$  in ADSIR and GDSIR, to  $\lambda_1 = 2,560$ ,  $\lambda_2 = 2,560$ ,  $\lambda_3 = 1,280$ ,  $\lambda_4 = 2,880$ ,  $\lambda_5 = 1,280$  in MGDSIR, and to  $\lambda_1 = 4,000$ ,  $\lambda_2 = 2,600$ ,  $\lambda_3 = 1,600$ ,  $\lambda_4 = 2,200$ ,  $\lambda_5 = 1,600$  in FMGDSIR. Figures 3(a), (b), (c), (d), (e), and (f) show the original image, the initial image  $\mu^{(0)}$  given by FBP, the reconstructed images by ADSIR, GDSIR, MGDSIR, and FMGDSIR, respectively. In this case, since a lot of projection data ( $I = 173,700 \gg J$ ) is available with small noise, we can see that FBP reconstructs a high-quality image. Therefore, we can confirm that  $G = 300$  and  $b = 10^6$  are proper values as the setting of a standard-dose situation. We can also see that all SIR methods reconstruct even better results than that by FBP with mostly the same quality.

### 6.4 Low-Dose Simulation Results

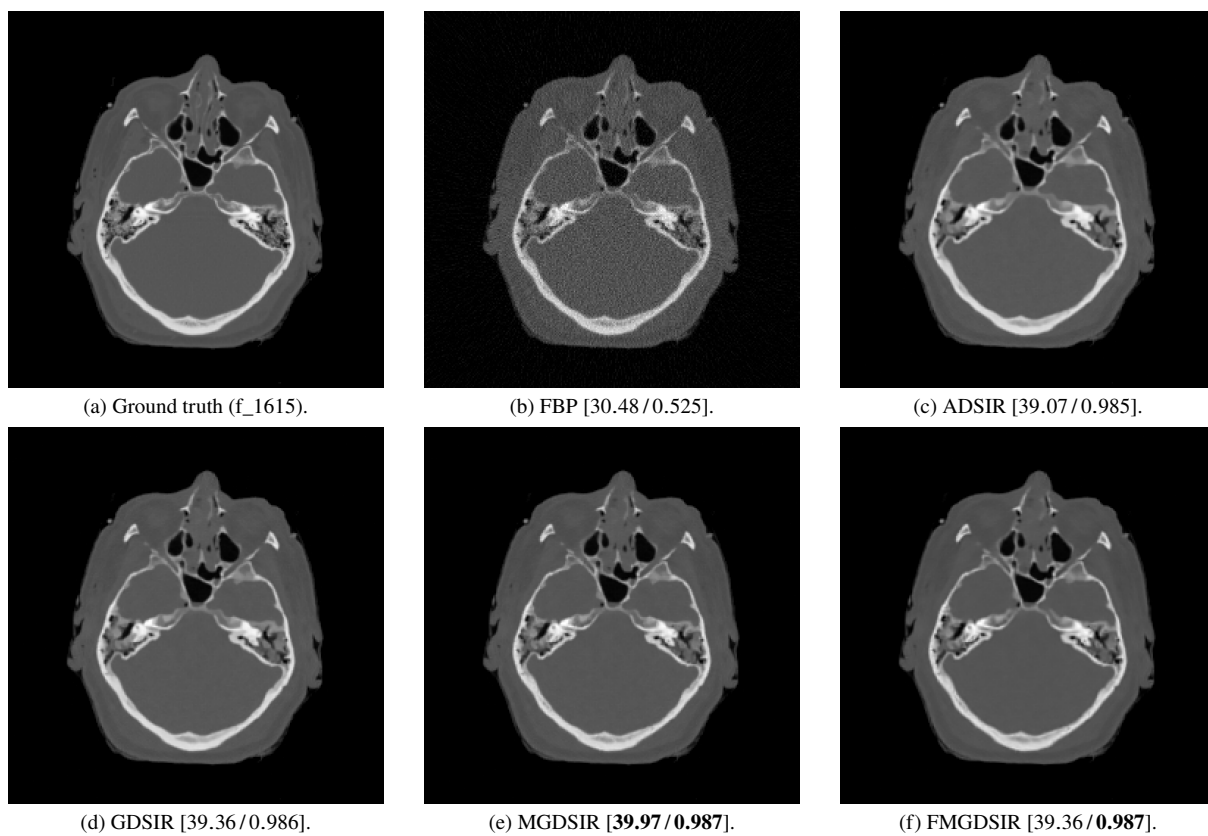
Low-dose projection data were generated in two situations: (i) reducing projection views from  $G = 300$  to  $G = 60$ , and (ii) lowering X-ray intensity from  $b = 10^6$  to  $b = 2.5 \times 10^4$ . Note that there is a significant difference between these two approaches. The first approach reduces the number of obser-



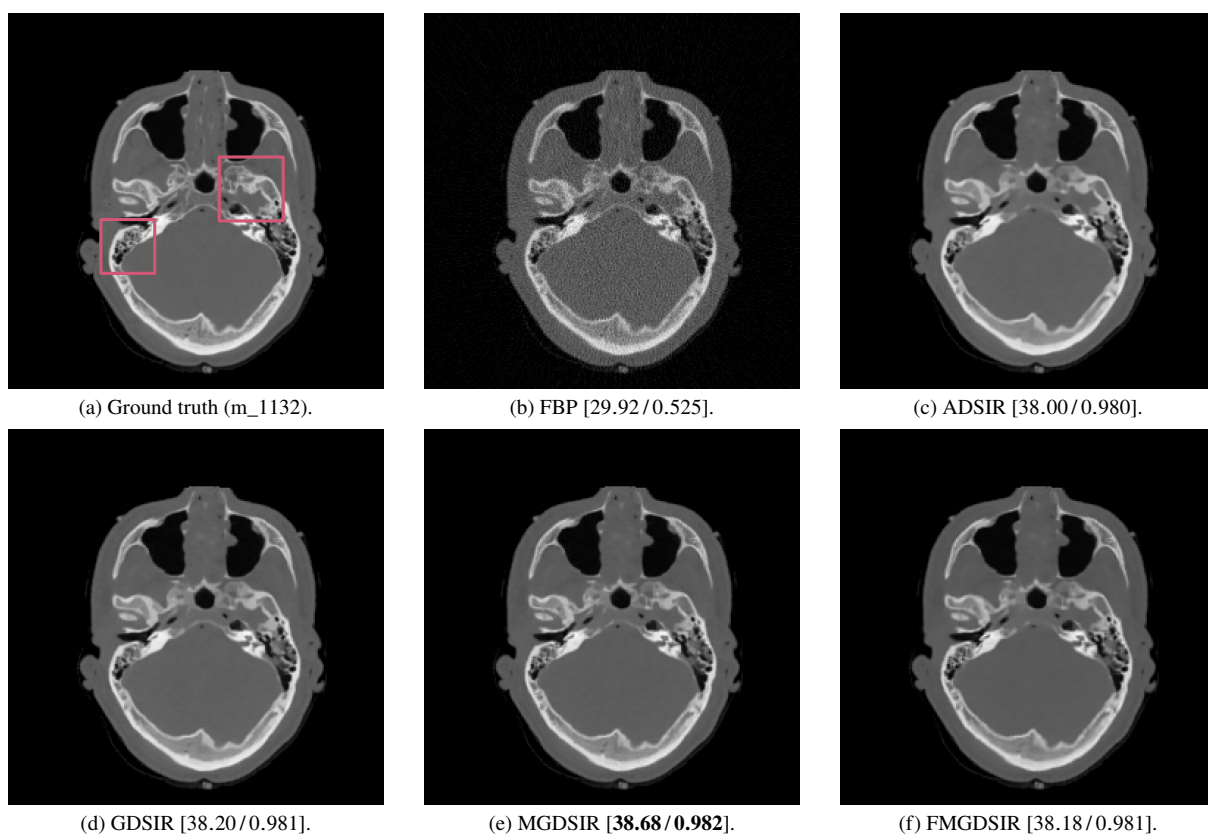
**Fig. 4** Reconstruction result of f\_1615 from its low-dose projection data (60/300 views) by each method [PSNR/SSIM].



**Fig. 5** Reconstruction result of m\_1132 from its low-dose projection data (60/300 views) by each method [PSNR/SSIM].

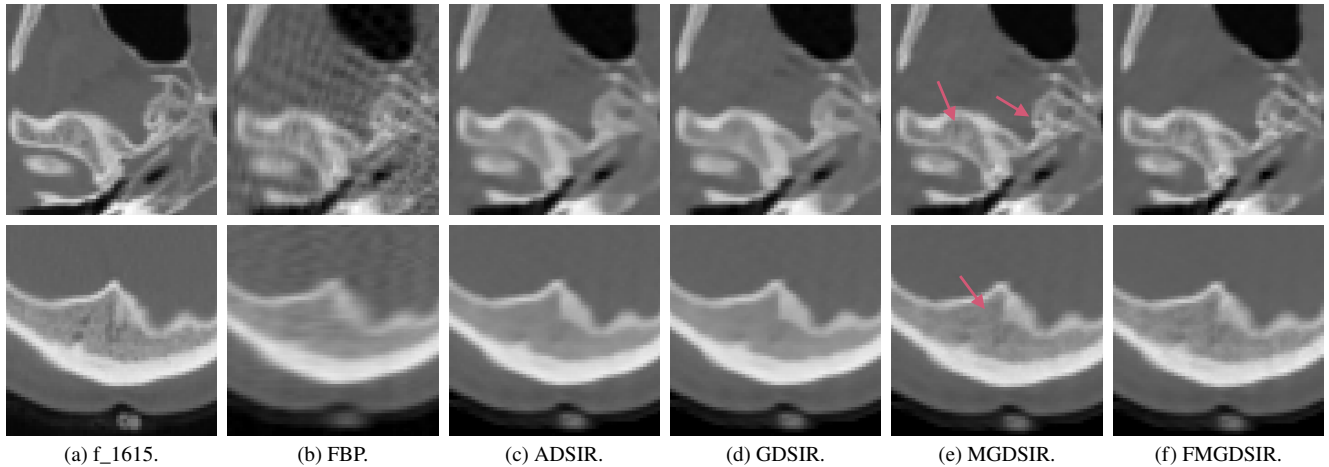


**Fig. 6** Reconstruction result of f\_1615 from its low-dose projection data (1/40 intensity) by each method [PSNR/SSIM].

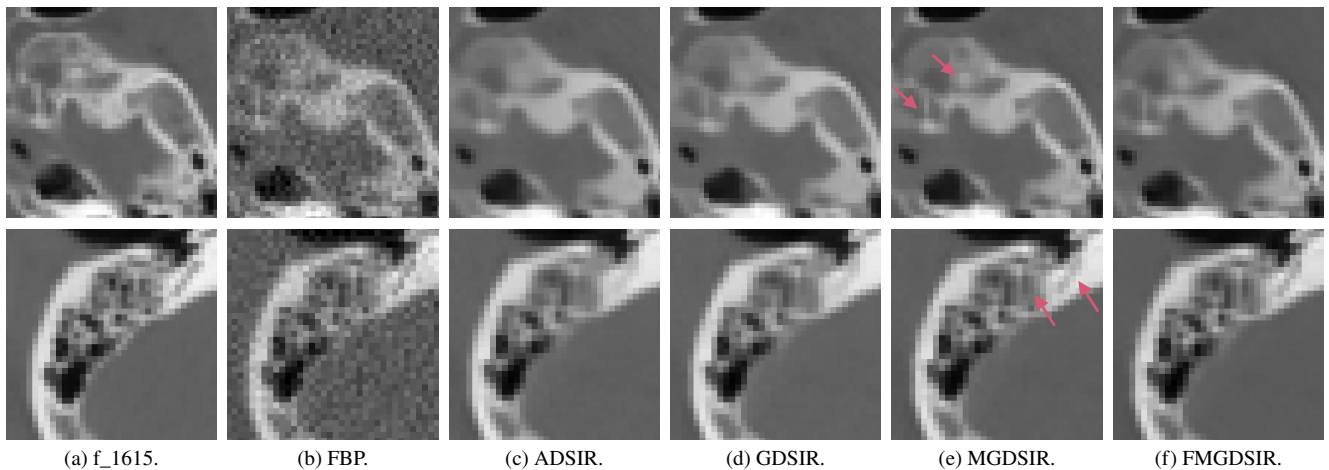


**Fig. 7** Reconstruction result of m\_1132 from its low-dose projection data (1/40 intensity) by each method [PSNR/SSIM].





**Fig. 8** Enlarged images of each reconstruction result in Fig. 5.



**Fig. 9** Enlarged images of each reconstruction result in Fig. 7.

vations ( $I = 34,740 < J$ ), while the second approach does not reduce the number of observations ( $I = 173,700 \gg J$ ) but enlarges the variance, 40 times, of  $z_{\theta_g, n}/b = \exp(-l_{\theta_g, n})$ . The reduction of the number  $I$  of the observations in the first approach makes the reconstruction problem in (2) more difficult than the increase of the noise level  $\|\xi\|_2$  in the second approach. If we set a same reduction ratio of the X-ray dose for both approaches, then the reconstruction quality is low in the first approach or difference of the quality is not visible in the second one. Hence, we used the different reduction ratios 1/5 and 1/40 in the first and second approaches, respectively.

1)  $G = 60$  and  $b = 10^6$ : We set  $\theta_g := 3(g-1)$  deg and the regularization parameters to  $\lambda = 3,800$  in ADSIR and GDSIR, to  $\lambda_1 = 7,500$ ,  $\lambda_2 = 3,800$ ,  $\lambda_3 = 1,000$ ,  $\lambda_4 = 2,500$ ,  $\lambda_5 = 1,000$  in MGDSIR, and to  $\lambda_1 = 7,500$ ,  $\lambda_2 = 6,000$ ,  $\lambda_3 = 1,000$ ,  $\lambda_4 = 1,500$ ,  $\lambda_5 = 1,000$  in FMGDSIR. When  $\mu^{(0)}$  was created by FBP, the projection data were linearly interpolated to 300 views. Table 1 summarizes PSNR of the reconstructed images and the calculation time per iteration for each method. From Table 1, we see that MGDSIR and FMGDSIR achieved higher reconstruction accuracy than by ADSIR and GDSIR with the same regularization parameters for all the images.

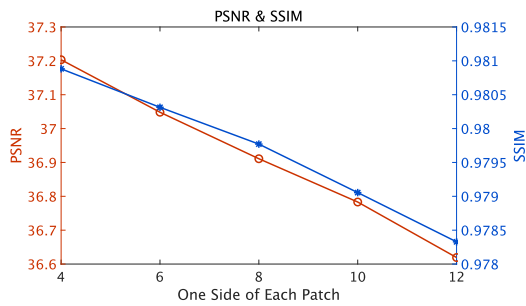
In addition, FMGDSIR reduced the calculation time approximately 46% from MGDSIR without accuracy degradation.

2)  $G = 300$  and  $b = 2.5 \times 10^4$ : We set the regularization parameters to  $\lambda = 800$  in ADSIR and GDSIR, to  $\lambda_1 = 800$ ,  $\lambda_2 = 800$ ,  $\lambda_3 = 400$ ,  $\lambda_4 = 900$ ,  $\lambda_5 = 400$  in MGDSIR, and to  $\lambda_1 = 2,000$ ,  $\lambda_2 = 1,300$ ,  $\lambda_3 = 800$ ,  $\lambda_4 = 1,100$ ,  $\lambda_5 = 800$  in FMGDSIR, which are constant multiples of  $\lambda_g$  in Sect. 6.3 since the matrix  $R$  is the same and the noise level is different in (2). Table 2 shows PSNR and the calculation time for each method. From Table 2, we see that MGDSIR achieved the highest reconstruction accuracy. FMGDSIR reduced the calculation time by about 30% from MGDSIR with almost the same accuracy as with the conventional GDSIR. In our simulations, FMGDSIR performed more effectively in the view reduced situation than in the weakened intensity situation.

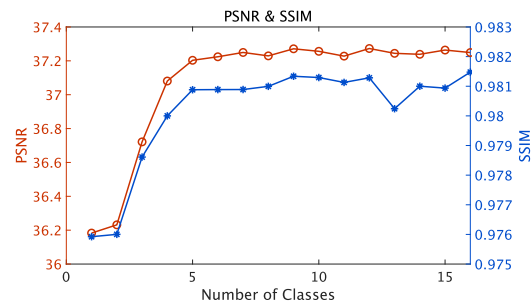
In Figs. 4–7, (a), (b), (c), (d), (e), and (f) show the original image, the initial image  $\mu^{(0)}$  by FBP, the reconstructed images by ADSIR, GDSIR, MGDSIR, and FMGDSIR, respectively, in the two low-dose simulations for ‘f\_1615’ and ‘m\_1132’, where Figs. 4 and 5 show the results in the view reduced situation, and Figs. 6 and 7 show ones in the weakened intensity situation. From these figures, we see that each ini-

**Table 3** Settings of  $P$ ,  $Q$  and  $\lambda_q$  in FMGDSIR and the reconstruction accuracy for m\_1132 in the view reduced situation.

$P$	$Q$	$(\lambda_1, \lambda_2, \dots, \lambda_Q)$	PSNR/SSIM
$4 \times 4$	5	$5000 \times (1.5, 1.2, 0.2, 0.3, 0.2)$	<b>37.20/0.981</b>
$6 \times 6$		$2500 \times (1.5, 1.2, 0.2, 0.3, 0.2)$	37.05/0.980
$8 \times 8$		$1500 \times (1.5, 1.2, 0.2, 0.3, 0.2)$	36.91/0.980
$10 \times 10$		$800 \times (1.5, 1.2, 0.3, 0.2, 0.2)$	36.78/0.979
$12 \times 12$		$600 \times (1.5, 1.2, 0.3, 0.2, 0.2)$	36.62/0.978
$4 \times 4$	1	4200	36.18/0.976
	2	$5000 \times (1.5, 0.8)$	36.23/0.976
	3	$5000 \times (1.5, 1.0, 0.2)$	36.72/0.979
	4	$5000 \times (1.5, 1.2, 0.2, 0.3)$	37.08/0.980
	5	$5000 \times (1.5, 1.2, 0.2, 0.3, 0.2)$	<b>37.20/0.981</b>
	6	$5000 \times (1.5, 1.2, 0.2, 0.2, 0.3, 0.3)$	<b>37.22/0.981</b>
	7	$5000 \times (1.5, 1.2, 0.3, 0.2, 0.2, 0.2, 0.2)$	<b>37.25/0.981</b>
	8	$5000 \times (1.5, 1.2, 0.3, 0.2, 0.2, 0.2, 0.2, 0.2)$	<b>37.23/0.981</b>
	9	$5000 \times (1.5, 1.2, 0.3, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2)$	<b>37.27/0.981</b>
	10	$5000 \times (1.5, 1.2, 0.3, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2)$	<b>37.26/0.981</b>
	11	$5000 \times (1.5, 1.2, 0.3, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2)$	<b>37.23/0.981</b>
	12	$5000 \times (1.5, 1.2, 0.3, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2)$	<b>37.27/0.981</b>
	13	$5000 \times (1.5, 1.2, 0.2, 0.2, 0.2, 0.2, 0.3, 0.2, 0.3, 0.3, 0.2, 0.3, 0.2)$	37.24/0.980
	14	$5000 \times (1.5, 1.2, 0.2, 0.2, 0.2, 0.3, 0.2, 0.3, 0.3, 0.2, 0.3, 0.2, 0.3, 0.2)$	<b>37.24/0.981</b>
	15	$5000 \times (1.5, 1.2, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2, 0.3, 0.3, 0.3, 0.3, 0.3, 0.3, 0.3)$	<b>37.26/0.981</b>
	16	$5000 \times (1.5, 1.2, 0.2, 0.2, 0.2, 0.2, 0.3, 0.3, 0.3, 0.2, 0.2, 0.3, 0.3, 0.2, 0.2, 0.2)$	<b>37.25/0.981</b>



(a) Relation between the accuracy and the patch size.



(b) Relation between the accuracy and the number of classes.

**Fig. 10** Values of PSNR (○) and SSIM (\*) in terms of the patch size  $P$  and the number  $Q$  of classes.

tial estimate has serious artifacts or noise while all SIR methods reconstruct high-quality images. Figures 8 and 9 show enlarged images of Figs. 5 and 7, and the red arrows indicate that detailed structures are lost in ADSIR and GDSIR while the proposed methods can reconstruct them more precisely. For example, a light-gray vertical line is visible clearly at the tip of the left arrow in the upper image of Fig. 9(e) while the lines in Figs. 9(c) and 9(d) are blurred. Such detailed structures are often used for medical diagnosis, and a radiographer in a hospital supported the results of the proposed methods.

### 6.5 On the Patch Size $P$ and the Number $Q$ of Classes

To analyze the effects of  $P$  and  $Q$ , we reconstructed ‘m\_1132’ by FMGDSIR in the view reduced situation, using several  $P$  and  $Q$  with a rough parameter tuning for  $\lambda_q$  as in Table 3.

1) Patch size: We used  $\sqrt{P} = 4, 6, \dots, 12$ . Figure 10(a) shows PSNR and SSIM in terms of the patch size. We found that the reconstruction accuracy decreased as the patch size increased. Hence, we have chosen  $P = 16$  as the patch size. Note that the width and height of each pixel were 1 mm and  $P = 16$  was optimal for this pixel length. We expect that if

the pixel length is smaller than 1 mm, the optimal  $P$  will be larger. If the target is not a head, we have to adjust  $P$  again.

2) Number of classes: We used  $Q = 1, 2, \dots, 16$ . Figure 10(b) shows that PSNR and SSIM largely increased until  $Q = 5$ . After that, the accuracy vibrated and the image quality hardly changed. Hence, we have chosen  $Q = 5$  considering the difficulty in tuning of the regularization parameters  $\lambda_q$ , although we can see that, in Table 3, FMGDSIR keeps the reconstruction accuracy with the rough parameter tuning.

## 7. Conclusion

In this paper, we proposed low-dose CT reconstruction methods exploiting multiclass dictionary learning. We used multiclass dictionaries and multiclass regularization parameters. The proposed methods, MGDSIR and FMGDSIR, achieved higher reconstruction accuracy than the conventional single-class dictionary-based methods, ADSIR and GDSIR. In addition, since MGDSIR performs clustering for target image patches only once, computational cost hardly increases from GDSIR. Moreover, by introducing orthogonal dictionaries, FMGDSIR achieved fast reconstruction. Simulation results

showed that the proposed methods reconstruct detailed structures of the targets while suppressing artifacts and noise.

## Acknowledgments

This work was supported by JSPS KAKENHI (17H02069). The authors are very grateful to Mr. Hiroki Okada in Kusatsu General Hospital for his valuable advice as a radiographer.

## References

- [1] H. Kamoshita, T. Shibata, D. Kitahara, K. Fujimoto, and A. Hirabayashi, "Low-dose CT reconstruction with multiclass orthogonal dictionaries," Proc. IEEE Int. Conf. Image Process. (ICIP), Taipei, Taiwan, pp.2055–2059, Sept. 2019.
- [2] A.C. Kak and M. Slaney, Principles of Computerized Tomographic Imaging, SIAM, Philadelphia, PA, 2001.
- [3] J. Hsieh, Computed Tomography: Principles, Design, Artifacts, and Recent Advances, 3rd ed., SPIE Press, Bellingham, WA, 2015.
- [4] Q. Xu, H. Yu, X. Mou, L. Zhang, J. Hsieh, and G. Wang, "Low-dose X-ray CT reconstruction via dictionary learning," IEEE Trans. Med. Imag., vol.31, no.9, pp.1682–1697, Sept. 2012.
- [5] Y. Chen, L. Shi, Q. Feng, J. Yang, H. Shu, L. Luo, J.L. Coatrieux, and W. Chen, "Artifact suppressed dictionary learning for low-dose CT image processing," IEEE Trans. Med. Imag., vol.33, no.12, pp.2271–2292, Dec. 2014.
- [6] M. Elad, Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing, Springer, New York, NY, 2010.
- [7] Y. Chen, X. Yin, L. Shi, H. Shu, L. Luo, J.L. Coatrieux, and C. Toumoulin, "Improving abdomen tumor low-dose CT images using a fast dictionary learning based processing," Phys. Med. Biol., vol.58, no.16, pp.5803–5820, Aug. 2013.
- [8] C. Bao, J.F. Cai, and H. Ji, "Fast sparsity-based orthogonal dictionary learning for image restoration," Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Sydney, Australia, pp.3384–3391, Dec. 2013.
- [9] C. Zhang, T. Zhang, M. Li, C. Peng, Z. Liu, and J. Zheng, "Low-dose CT reconstruction via  $\ell_1$  dictionary learning regularization using iteratively reweighted least-squares," Biomed. Eng. Online, vol.15, no.66, 21 pages, June 2016.
- [10] K. Sauer and C. Bouman, "A local update strategy for iterative reconstruction from projections," IEEE Trans. Signal Process., vol.41, no.2, pp.534–548, Feb. 1993.
- [11] J.A. Fessler, "Statistical image reconstruction methods for transmission tomography," Handbook of Medical Imaging, Volume 2: Medical Image Processing and Analysis, eds. M. Sonka and J.M. Fitzpatrick, pp.1–70, SPIE Press, Bellingham, WA, 2000.
- [12] I.A. Elbakri and J.A. Fessler, "Statistical image reconstruction for polyenergetic X-ray computed tomography," IEEE Trans. Med. Imag., vol.21, no.2, pp.89–99, Feb. 2002.
- [13] Y.C. Pati, R. Rezaifar, and P.S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," Proc. Asilomar Conf. Signal Syst. Comput. (ACSSC), Pacific Grove, CA, pp.40–44, Nov. 1993.
- [14] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding," J. Mach. Learn. Res., vol.11, pp.19–60, Jan. 2010.
- [15] Z. Zhan, J.F. Cai, D. Guo, Y. Liu, Z. Chen, and X. Qu, "Fast multiclass dictionaries learning with geometrical directions in MRI reconstruction," IEEE Trans. Biomed. Eng., vol.63, no.9, pp.1850–1861, Sept. 2016.
- [16] N. Inamuro, M. Shibata, C. Tang, T. Ijiri, and A. Hirabayashi, "High-quality MR imaging based on dictionary learning using training images and observed signals," IEICE Technical Report, vol.116, no.395, pp.123–128, Jan. 2017 (in Japanese).
- [17] J. Mairal, J.P. Chieze, Y. Chen, G. Durif, R. Jenatton, F. Bach, J. Ponce, G. Obozinski, B. Yu, G. Sapiro, and Z. Harchaoui, Spams: a SParse Modeling Software ver. 2.6, <http://spams-devel.gforge.inria.fr/doc/html>, accessed Jan. 20, 2020.
- [18] Magnetic Resonance Research Facility, Visible Human Project CT Datasets, <https://medicine.uiowa.edu/mri/facility-resources/images/visible-human-project-ct-datasets>, accessed Jan. 20, 2020.



**Hiryu Kamoshita** received his B.E. degree and M.E. degree in computer science from Ritsumeikan University, Shiga, Japan, in 2018 and 2020, respectively. His research interest includes compressed sensing and its applications. He received the Outstanding Research Award from the Graduate School of Information Science and Engineering, Ritsumeikan University in 2020.



**Daichi Kitahara** received his B.E. degree in computer science in 2012, and his M.E. and Ph.D. degrees in communications and computer engineering in 2014 and 2017 from the Tokyo Institute of Technology, Tokyo, Japan, respectively. Since 2017, he has been an Assistant Professor with the College of Information Science and Engineering, Ritsumeikan University, Shiga, Japan. His current research interest includes signal processing and its applications, inverse problem, optimization theory, and multivariate spline theory.



**Ken'ichi Fujimoto** received his M.E. degree from the Tokyo Institute of Technology in 1997 and his Ph.D. degree from the University of Tokushima in 2000. He is currently an Associate Professor at Kagawa University. His research interest spans interdisciplinary basic research on the dynamics of nonlinear dynamical systems and its applications to medical engineering as tomographic image reconstruction. He is a member of IEEE, IEICE, IEEJ, SICE, and ISCIE.



**Laurent Condat** received his Ph.D. degree in applied mathematics from Grenoble Institute of Technology in 2006. From 2006 to 2008, he was a postdoc in the Helmholtz Zentrum München in Munich, Germany. He was hired as a permanent researcher by the French National Center for Scientific Research (CNRS) in 2008. From 2016 to 2019, he was a member of the French National Committee for Scientific Research. Since Nov. 2019, he is a Research Scientist in the Visual Computing Center of King Abdullah University of Science and Technology (KAUST), Saudi Arabia. His area of interest includes convex optimization models and algorithms, and signal and image processing. He received a best student paper award at the IEEE ICIP 2005 and a best Ph.D. award from Grenoble Institute of Technology in 2007.



**Akira Hirabayashi** received his D.E. degree in computer science in 1999, from the Tokyo Institute of Technology, Tokyo, Japan. Currently, he is a Professor of the College of Information Science and Engineering, Ritsumeikan University, Shiga, Japan. His research interests include compressed sensing, sampling theory, and signal and image processing. Prof. Hirabayashi is also a member of the Institute of Electrical and Electronics Engineers (IEEE) and the Acoustical Society of Japan (ASJ).