Optimization-based image reconstruction in x-ray computed tomography by sparsity exploitation of local continuity and nonlocal spatial self-similarity*

Han-Ming Zhang(张瀚铭), Lin-Yuan Wang(王林元), Lei Li(李磊), Bin Yan(闫镔)†, Ai-Long Cai(蔡爱龙), and Guo-En Hu(胡国恩)

National Digital Switching System Engineering and Technological Research Center, Zhengzhou 450002, *China*

(Received 29 December 2015; revised manuscript received 26 February 2016; published online 1 June 2016)

The additional sparse prior of images has been the subject of much research in problems of sparse-view computed tomography (CT) reconstruction. A method employing the image gradient sparsity is often used to reduce the sampling rate and is shown to remove the unwanted artifacts while preserve sharp edges, but may cause blocky or patchy artifacts. To eliminate this drawback, we propose a novel sparsity exploitation-based model for CT image reconstruction. In the presented model, the sparse representation and sparsity exploitation of both gradient and nonlocal gradient are investigated. The new model is shown to offer the potential for better results by introducing a similarity prior information of the image structure. Then, an effective alternating direction minimization algorithm is developed to optimize the objective function with a robust convergence result. Qualitative and quantitative evaluations have been carried out both on the simulation and real data in terms of accuracy and resolution properties. The results indicate that the proposed method can be applied for achieving better image-quality potential with the theoretically expected detailed feature preservation.

Keywords: computed tomography, image reconstruction, sparsity exploitation, nonlocal gradient

PACS: 87.59.–e, 07.85.–m **DOI:** 10.1088/1674-1056/25/7/078701

1. Introduction

X-ray computed tomography (CT) has had a revolutionary effect in fields ranging from biology to medicine. Given that excessive x-ray radiation exposure may cause genetic disease, the reduction of radiation dose during x-ray CT inspection has been a focus of recent investigations.[1–3] Consequently, strategies that lower the required number of projections have received great attention.[4] Owing to insufficient angular sampling, the reconstruction results usually suffers aliasing artifacts, and conventional reconstruction methods such as the well-known filtered back projection (FBP), the algebraic reconstruction technique (ART),[5] simultaneous ART (SART),[6] and expectation maximization (EM) methods[7] do not reliably converge on the correct solution or afford satisfactory image quality. Reconstructing a target image from few-view data has become a research goal in medical imaging.

Inspired by the compressive sensing (CS) theory, methods involving total variation (TV) minimization have been developed and widely studied for reconstruction from few-view data.[4,8–13] The TV norm is the ℓ_1 norm of the image gradient magnitude, and minimizing the ℓ_1 norm is known to exploit sparsity in its argument.[4] However, owing to the piecewise constant assumption, TV regularization reconstruction methods may lead to a staircasing effect, which may suffer from over-smoothness in fine structures and produce undesirable images.[14,15] To achieve more accurate results, improved edge-preserving methods that involve the introduction of a penalty weight on the original TV norm with more local information have been proposed.[16–20]

The conventional TV norm uses only the sparsity information of the local continuity of an image, an approach that may encounter limitations in some richly detailed or piecewise polynomial situations. In recent years, a nonlocal (NL) gradient characterizing the structural similarity has been studied and has shown good performance for detail preservation in many image-processing applications.[21–23] A method using the ℓ_1 norm of the NL gradient, termed nonlocal total variation (NLTV) as a regularization term, has been proposed to improve image quality.[24–26]

Though NLTV shows potential advantages, unfortunately, it often suffers problems in the application of CT image reconstruction. The calculation of an NL gradient is based on the statistics of similarity in a reference image which reflect the similarity interaction of the reconstructed object. However, this reference is difficult to obtain in sparse-view CT image reconstruction as the observed CT projection data do not directly reflect the similarity information of the object. Thus, in CT image iterative reconstruction, the available reference used to calculate similarity weight is an intermediate result that struggles to approximate the actual result. Compared to the true image, it always contains many fakes and blurs. When calculating the NL operator from these references, it may in-

*Project supported by the National Natural Science Foundation of China (Grant No. 61372172).

†Corresponding author. E-mail: ybspace@hotmail.com

© 2016 Chinese Physical Society and IOP Publishing Ltd

troduce much false information and lead to an incorrect result. Reference [25] used the FBP image as a reference to build the weight. This method requires all angular projections to guarantee the accuracy of similarity estimation. Reference [26] updated the similarity weight in each iteration to generate a more reliable result. However, because the calculation of the similarity weight needs lots of time and memory, this method will cost a great amount of computation.

This paper intends to develop a robust and efficient reconstruction algorithm to make the most of the sparsity brought from the similarity prior. A hybrid optimization model utilizing both TV and NLTV regularizations is proposed. To solve the optimization problem, we then present an efficient iterative algorithm based on the alternating direction method (ADM) optimization approach.[27–29] The new method shows a good performance in sparsity exploitation of both local continuity and nonlocal spatial self-similarity. In particular, it only needs to take several times of calculation of the similarity weight. Thus, the proposed method could preserve a good balance between performance and computation. Finally, we performed qualitative and quantitative evaluations to verify the feasibility of artifact reduction and detail recovery for sparse-view CT image reconstruction. For simplicity, the present method is termed "HTV-ADM" in this paper.

2. Methods

2.1. Nonlocal gradient operator

The motivation of NL methods is to exploit the structural similarity property. In this case, the NL gradient is calculated with pixels belonging to the whole or semilocal region of an image, instead of only the nearest neighboring pixels as used in the conventional gradient model. As described in Ref. [23], the NL gradient of an image $u : \Omega \to R$ is defined as

$$\nabla_{\mathrm{NL}} u(x,y) = (u(y) - u(x))\sqrt{\omega(x,y)}, \quad x,y \in \Omega, \tag{1}$$

with $\omega(x,y)$ being a similarity weight between points x and y.

Let $U_\omega(x)$ denote the gray-level vector of a patch P_x centered at point x and let $\Omega_\omega(x)$ denote a neighborhood around point x. The similarity weight $\omega(x,y)$ can be written as follows:[22]

$$\omega(x,y) = \begin{cases} \varpi(U_\omega(x), U_\omega(y)), & y \in \Omega_\omega(x), \\ 0, & \text{otherwise}, \end{cases} \tag{2}$$

where $\varpi(s_i,s_j)$ is a similarity function with the symmetric property. Some examples of similarity functions are formulated in Ref. [22] for different applications.

In this work, for a general discussion of sparsity exploring and performance evaluation of the NL gradient, we used a typical weight expression that may show basic similarity affinity for most applications. The following similarity function is examined:

$$\varpi(s_i,s_j) = \exp\left(\frac{-d^2(s_i,s_j)}{h^2}\right), \tag{3}$$

where h is a scale parameter and $d(s_i,s_j)$ are some distance functions (here the Euclidean distance between sets s_i and s_j).

Figure 1 illustrates how the NL gradient is calculated: for a stable pixel x, traverse all pixels in its search window, measure the similarity between the target patch P_x and the others, and then construct the neighborhood $\Omega_\omega(x)$ by taking the m most similar neighbors of the pixel.

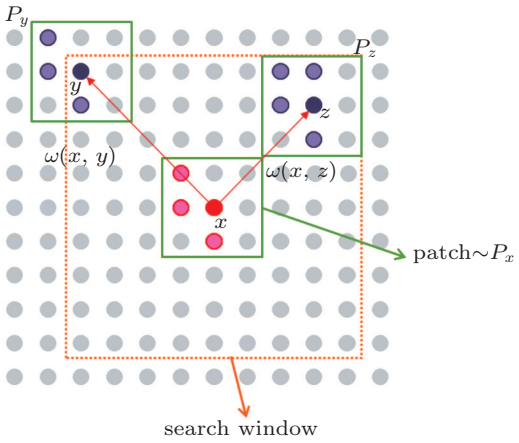


Fig. 1. (color online) Calculation of the image NL gradient. For a target pixel x in red, all pixels in its search window (orange square) are matched with x by calculating similarity weights. For example, the similarity between pixel y and x is measured by the weight $\omega(x,y)$, and the weight is calculated and depends on an ℓ_2 distance between the patches (green squares) centered on the pixels in the comparison.

In most medical and other tomographic imaging applications, the object image generally has the property of spatial self-similarity. There are often many similar structures in different regions of the image, no matter whether the structure is smooth or complex. An NL gradient penalizes the differences between these similar regions, and thus can exploit the considerable information of the structural similarity prior. This approach will obviously lead to a sparse representation of an image u in the basis ∇_{NL}.

For CT image reconstruction, the similarity weight function ϖ should be calculated based on a reference image which reflects the similarity interaction of the desired object. Nevertheless, the observed CT projection data do not directly reflect the similarity information of the image. In particular, the image is recovered gradually in the iterative reconstruction procedure. A suitable reference image is difficult to obtain and a direct measurement of NL gradient may result in some false information of similarity. Thus, we consider a hybrid method to exploit the sparsity in both local difference and nonlocal spatial self-similarity for CT image reconstruction.

2.2. Imaging model with HTV minimization

CT projection and desired image can be mathematically assumed to satisfy the following discrete linear equations:

$$\boldsymbol{A}\boldsymbol{u} = \boldsymbol{p}, \tag{4}$$

where vector $\boldsymbol{p}$ represents log-transformed projection data, $\boldsymbol{A}$ stands for the projection matrix in system geometry, and $\boldsymbol{u}$ is the image whose entries correspond to the x-ray linear attenuation coefficients at different voxels. A CT reconstruction problem is formulated to retrieve the unknown vector $\boldsymbol{u}$ based on the system matrix $\boldsymbol{A}$ and the estimated projection $\boldsymbol{p}$. When it comes to an under-sampled problem, the linear equation will become under-determined, and this problem will be difficult to solve directly. To mitigate the poor condition, the constrained TV minimization method has been used for few-view image reconstruction.[8–13]

Inspired by the studies of standard TV in image reconstruction, we specify an HTV model that implements the following optimization problem:

$$\min_{\boldsymbol{u}} \alpha_1 ||\boldsymbol{u}||_{\mathrm{TV}} + \alpha_2 ||\boldsymbol{u}||_{\mathrm{NLTV}}, \quad \text{s.t.} \quad \boldsymbol{A}\boldsymbol{u} = \boldsymbol{p}. \tag{5}$$

The introduction of the TV and NLTV terms in this optimization process differentiates infinitely many solutions to Eq. (4) and picks out the object by utilizing the gradient sparsity and spatial self-similarity property.

2.3. Optimization approach

By introducing vectors $\boldsymbol{y}$ and $\boldsymbol{z}$, we consider the following constrained minimization problem, which is equivalent to Eq. (5):

$$\begin{aligned} &\min_{u,y,z} \alpha_1 ||\boldsymbol{y}||_1 + \alpha_2 ||\boldsymbol{z}||_1, \quad \text{s.t.} \ \boldsymbol{\nabla}\boldsymbol{u} = \boldsymbol{y}, \\ &\boldsymbol{\nabla}_{\mathrm{NL}}\boldsymbol{u} = \boldsymbol{z} \ \text{ and } \ \boldsymbol{A}\boldsymbol{u} = \boldsymbol{p}. \end{aligned} \tag{6}$$

Next, we reformulate the constrained minimization problem as an unconstrained optimization task by an augmented Lagrangian method (ALM).[30] The augmented Lagrangian function associated to Eq. (6) is defined as

$$\begin{aligned} \mathcal{L}_{\mathrm{A}} :=\ & \alpha_1 ||\boldsymbol{y}||_1 + \boldsymbol{r}_y^{\mathrm{T}}(y - \boldsymbol{\nabla} u) + \frac{\lambda_1}{2}||y - \boldsymbol{\nabla} u||_2^2 \\ & + \alpha_2 ||\boldsymbol{z}||_1 + \boldsymbol{r}_z^{\mathrm{T}}(z - \boldsymbol{\nabla}_{\mathrm{NL}} u) + \frac{\lambda_2}{2}||z - \boldsymbol{\nabla}_{\mathrm{NL}} u||_2^2 \\ & + \boldsymbol{r}_{\mathrm{A}}^{\mathrm{T}}(\boldsymbol{p} - \boldsymbol{A}\boldsymbol{u}) + \frac{\mu}{2}||\boldsymbol{p} - \boldsymbol{A}\boldsymbol{u}||_2^2, \end{aligned} \tag{7}$$

where $\boldsymbol{r}_{\mathrm{A}}$, $\boldsymbol{r}_y$, and $\boldsymbol{r}_z$ are the Lagrange multipliers, μ, λ_1, and λ_2 are positive constants used to balance the terms. Thus, the constraint minimization problem (6) is converted to find the saddle point of the augmented Lagrangian energy $\mathcal{L}_{\mathrm{A}}$.

The solution to minimizing $\mathcal{L}_{\mathrm{A}}$ can be approximated efficiently by the alternating direction method: at each iteration find an approximate minimizer with respect to the variables $\boldsymbol{u}$, $\boldsymbol{y}$, $\boldsymbol{z}$ and update the Lagrange multipliers alternately.

(i) The sub-minimization problem with respect to $\boldsymbol{y}$ can be written as follows:

$$\min_y \alpha_1 ||\boldsymbol{y}||_1 + \frac{\lambda_1}{2}||\boldsymbol{y} + \frac{\boldsymbol{r}_y^k}{\lambda_1} - \boldsymbol{\nabla}\boldsymbol{u}^k||_2^2. \tag{8}$$

This minimization problem corresponds to an $\ell_1 - \ell_2$ norm and can be given directly by

$$\boldsymbol{y}^{k+1} = \mathrm{shrink}_2\left(\boldsymbol{\nabla}\boldsymbol{u}^k - \frac{\boldsymbol{r}_y^k}{\lambda_1}, \alpha_1/\lambda_1\right), \tag{9}$$

where $\mathrm{shrink}_2(\cdot, 1/\beta)$, known as the two-dimensional shrinkage operator,[27] is defined as

$$\mathrm{shrink}_2(\boldsymbol{\xi}, 1/\beta) = \max\{||\boldsymbol{\xi}|| - 1/\beta, 0\} \cdot \frac{\boldsymbol{\xi}}{||\boldsymbol{\xi}||}. \tag{10}$$

(ii) The sub-minimization problem with respect to z can be written as follows:

$$\min_z \alpha_2 ||\boldsymbol{z}||_1 + \frac{\lambda_2}{2}\left\|\boldsymbol{z} + \frac{\boldsymbol{r}_z^k}{\lambda_2} - \boldsymbol{\nabla}_{\mathrm{NL}}\boldsymbol{u}^k\right\|_2^2. \tag{11}$$

This problem is solved with the same adaptation of the shrinkage operator as follows:

$$\boldsymbol{z}^{k+1} = \mathrm{shrink}_2\left(\boldsymbol{\nabla}_{\mathrm{NL}}\boldsymbol{u}^k - \frac{\boldsymbol{r}_z^k}{\lambda_2},\ \alpha_2/\lambda_2\right). \tag{12}$$

(iii) The sub-minimization problem w.r.t $\boldsymbol{u}$ corresponds to the following quadratic positive definite problem:

$$\begin{aligned} &\min_u \frac{\mu}{2}\|\boldsymbol{p} - \boldsymbol{A}\boldsymbol{u}\|_2^2 + \boldsymbol{r}_{\mathrm{A}}^{\mathrm{T}}(\boldsymbol{p} - \boldsymbol{A}\boldsymbol{u}) \\ &+ \frac{\lambda_1}{2}\|\boldsymbol{y} + \frac{\boldsymbol{r}_y^k}{\lambda_1} - \boldsymbol{\nabla}\boldsymbol{u}^k\|_2^2 + \frac{\lambda_2}{2}\|\boldsymbol{z} + \frac{\boldsymbol{r}_z^k}{\lambda_2} - \boldsymbol{\nabla}_{\mathrm{NL}}\boldsymbol{u}^k\|_2^2. \end{aligned} \tag{13}$$

The first-order necessary conditions for optimality are

$$\begin{aligned} &(\mu\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} + \lambda_1\boldsymbol{\nabla}^{\mathrm{T}}\boldsymbol{\nabla} + \lambda_2\boldsymbol{\nabla}_{\mathrm{NL}}^{\mathrm{T}}\boldsymbol{\nabla}_{\mathrm{NL}})\boldsymbol{u} \\ &= \mu\boldsymbol{A}^{\mathrm{T}}\boldsymbol{p} - \boldsymbol{A}^{\mathrm{T}}\boldsymbol{r}_{\mathrm{A}} + \lambda_1\boldsymbol{\nabla}^{\mathrm{T}}\left(\boldsymbol{y}^{k+1} + \frac{\boldsymbol{r}_y^k}{\lambda_1}\right) \\ &\quad + \lambda_2\boldsymbol{\nabla}_{\mathrm{NL}}^{\mathrm{T}}\left(\boldsymbol{z}^{k+1} + \frac{\boldsymbol{r}_z^k}{\lambda_2}\right). \end{aligned} \tag{14}$$

The exact minimizer of Eq. (14) is formulated as

$$\begin{aligned} \boldsymbol{u}^{k+1} =\ & \left(\mu\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} + \lambda_1\boldsymbol{\nabla}^{\mathrm{T}}\boldsymbol{\nabla} + \lambda_2\boldsymbol{\nabla}_{\mathrm{NL}}^{\mathrm{T}}\boldsymbol{\nabla}_{\mathrm{NL}}\right)^{+} \\ & \times \left(\mu\boldsymbol{A}^{\mathrm{T}}\boldsymbol{p} - \boldsymbol{A}^{\mathrm{T}}\boldsymbol{r}_{\mathrm{A}} + \lambda_1\boldsymbol{\nabla}^{\mathrm{T}}\left(\boldsymbol{y}^{k+1} + \frac{\boldsymbol{r}_y^k}{\lambda_1}\right)\right. \\ & \left. + \lambda_2\boldsymbol{\nabla}_{\mathrm{NL}}^{\mathrm{T}}\left(\boldsymbol{z}^{k+1} + \frac{\boldsymbol{r}_z^k}{\lambda_2}\right)\right), \end{aligned} \tag{15}$$

where $\boldsymbol{M}^{+}$ stands for the Moore–Penrose pseudo inverse of matrix $\boldsymbol{M}$. This subproblem can be solved efficiently by nonmonotone alternating direction algorithm (NADA).[30]

(iv) Finally, the Lagrange multipliers are updated as follows:

$$\begin{aligned} \boldsymbol{r}_y^{k+1} &= \boldsymbol{r}_y^k + \lambda_1(\boldsymbol{y}^{k+1} - \boldsymbol{\nabla}\boldsymbol{u}^{k+1}), \\ \boldsymbol{r}_z^{k+1} &= \boldsymbol{r}_z^k + \lambda_2(\boldsymbol{z}^{k+1} - \boldsymbol{\nabla}_{\mathrm{NL}}\boldsymbol{u}^{k+1}), \\ \boldsymbol{r}_{\mathrm{A}}^{k+1} &= \boldsymbol{r}_{\mathrm{A}}^k + \mu(\boldsymbol{p} - \boldsymbol{A}\boldsymbol{u}^{k+1}). \end{aligned} \tag{16}$$

2.4. Presentation of the HTV-ADM algorithm

The calculation of similarity weight of NL operator usually takes a huge amount of computation and is time-consuming. Based on our experience, the weight does not need to be calculated and updated in every iteration. In our implementation, the similarity weight was set to be updated at regular intervals with a number of sub-iterations to approximate the true results, and it only needs to update two or three times of the weight for most image reconstruction procedure. As the available information in iterative procedure does not yield a good estimation for the similarity weight ω at the start, we consider a weight updating strategy: in the beginning of the iterative procedure, z and r_z are not updated and set to zero, and the algorithm is executed the same as the TV-ADM algorithm; when the changes of images between two adjacent iterations become small, then the current image is set as the initial reference, and the weight ω is fixed and updated during reconstruction iterations.

In summary, the implementation of the present HTV-ADM method for x-ray CT image reconstruction can be described as follows:

Algorithm 1: Constrained HTV minimization in the ADM (HTV-ADM) framework

Input $\boldsymbol{A}$, $\boldsymbol{p}$, μ, λ_1, λ_2, α_1, α_2, Ω_{NL}, initialize y^0, z^0, u^0, and $k = 0$.

While "not converged," **Do**

(1) Update $\boldsymbol{y}^{k+1}$ by

$$\boldsymbol{y}^{k+1} \leftarrow \mathrm{shrink}_2\left(\boldsymbol{\nabla}\boldsymbol{u}^k - \frac{\boldsymbol{r}_y^k}{\lambda_1}, \alpha_1/\lambda_1\right);$$

(2) Update $\boldsymbol{z}^{k+1}$ by

$$\boldsymbol{z}^{k+1} \leftarrow \mathrm{shrink}_2(\boldsymbol{\nabla}_{\mathrm{NL}}\boldsymbol{u}^k - \frac{\boldsymbol{r}_z^k}{\lambda_2}, \alpha_2/\lambda_2);$$

(3) Update $\boldsymbol{u}^{k+1}$ by

$$\boldsymbol{u}^{k+1} \leftarrow \left(\mu\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} + \lambda_1\boldsymbol{\nabla}^{\mathrm{T}}\boldsymbol{\nabla} + \lambda_2\boldsymbol{\nabla}_{\mathrm{NL}}^{\mathrm{T}}\boldsymbol{\nabla}_{\mathrm{NL}}\right)^{+} \left(\mu\boldsymbol{A}^{\mathrm{T}}\boldsymbol{p} - A^{\mathrm{T}}\boldsymbol{r}_{\mathrm{A}} + \lambda_1\boldsymbol{\nabla}^{\mathrm{T}}(\boldsymbol{y}^{k+1} + \frac{\boldsymbol{r}_y^k}{\lambda_1}) + \lambda_2\boldsymbol{\nabla}_{\mathrm{NL}}^{\mathrm{T}}(z^{k+1} + \frac{\boldsymbol{r}_z^k}{\lambda_2})\right);$$

(4) Update multipliers by

$$\begin{aligned} \boldsymbol{r}_y^{k+1} &\leftarrow \boldsymbol{r}_y^k + \lambda_1(\boldsymbol{y}^{k+1} - \boldsymbol{\nabla}u^{k+1}), \\ \boldsymbol{r}_z^{k+1} &\leftarrow \boldsymbol{r}_z^k + \lambda_2(z^{k+1} - \boldsymbol{\nabla}_{\mathrm{NL}}\boldsymbol{u}^{k+1}), \\ \boldsymbol{r}_{\mathrm{A}}^{k+1} &\leftarrow \boldsymbol{r}_{\mathrm{A}}^k + \mu(\boldsymbol{p} - \boldsymbol{A}\boldsymbol{u}^{k+1}); \end{aligned}$$

(5) Update similarity weight by

$$\omega \leftarrow \begin{cases} W_{\mathrm{NL}}(\boldsymbol{u}^{k+1}), & k \in \Omega_{\mathrm{NL}}, \\ \omega, & \text{else}; \end{cases}$$

(6) $k \leftarrow k + 1$.

End Do

Obtain reconstruction result: $\boldsymbol{u}$.

In the above implementation scheme, Ω_{NL} denotes the set of the sub-iteration number for updating similarity weight and $W_{\mathrm{NL}}(\boldsymbol{u}^k)$ denotes a similarity weight calculated from the reference image $\boldsymbol{u}^k$.

2.5. Parameter selection

For any reconstruction design, multiple parameter settings are likely to be involved, which can have a significant impact on reconstruction results. This section discusses the parameter settings of the NL gradient for guiding adequate adaptation of the specific CT imaging task.

As expressed in Section 2.1, it is easy to see that the similarity weight ω is the kernel of the NL gradient and is affected by four important parameters: i) the size of the patch, ii) the size of the search window, iii) the number of neighbors in $\Omega_\omega(x)$, and iv) the similarity scaling parameter h. The details of these parameter settings are described below.

2.5.1. Patch size

The patch size controls the areas considered to match with each other. In general, a greater patch size means more pixels in each patch. As it is hard to match every pixel between two large patches, a large size will introduce limited information on similarity redundancy. In contrast, a small patch size may be preferred. The complexity of the object image is also an important factor. If the image contains too many intricate details, it will be impossible to obtain sufficient information to discern subtle structures without a small patch size. Based on our experience, a value between 3 and 13 is adequate for most images.

2.5.2. Search window

In theory, when the patch size and neighbor number are fixed, a search window of increased size can explore more regions and may update pixels considered as neighbors in the set $\Omega_\omega(x)$, as they have higher similarity with the target pixel x. If the size of the search window is too small, the performance of the NL gradient will be similar to that of an anisotropic gradient.

However, increasing the size of the search window also results in a heavy computational demand. To preserve a balance between performance and computation, a realistic approach is to make an appropriate selection according to image content. Based on our experience, a value which is twice as large as the patch size is appropriate.

2.5.3. Neighbors used per pixel

Neighbor number m plays a key role in specifying the level of similarity. Employing more neighbors could directly increase similarity information, but at the same time, may increase some unsuitable connections that introduce an unfavorable prior. Thus, it appears to seek a balance point beyond which similarity information used increases little, while the influence of unsuitable information continues to increase. Based on our experience, a value between 7 and 13 seems to be an appropriate choice for most cases. Another feasible choice is to use a moderate threshold to limit the weight value.

2.5.4. Similarity scaling parameter

The scale parameter h acts as a soft threshold controlling the decay of the exponential function and determines the norm values considered similar. In general, h often corresponds to the noise level, and is fixed as the standard deviation of the noise.[24] Thus, it is necessary to build a noise estimation model for choosing h.

As a single scale parameter h may not be optimal for all the patches in the image,[31] in this paper we use an adaptive local estimation inspired by local scaling model[32]

$$h = \sqrt{\sigma_i \sigma_j}, \tag{17}$$

where σ_i denotes a distance function form x_i to its $[(m+1)/2]^{\text{th}}$ nearest neighbor,

$$\sigma_i = \|s_i - s_{(m+1)/2}\|_2. \tag{18}$$

The advantage of localizing the parameter h is that it can control the difference penalty according to the region's content by studying the local statistics of the neighborhoods.

3. Results

To demonstrate and validate our new method for the CT image reconstruction, two types of data (computer-simulated digital CS-phantom projection data, and experimental rabbit head projection data) were used. All of the experiments were performed under Matlab 2012a running on an HP-Z820 workstation with Intel Xeon E5-2650 dual-core CPU 2.60 GHz. To bring forth the advantages of the presented HTV-ADM algorithm, the TV-ADM[28] algorithm was also performed for comparison.

3.1. Digital CS-phantom study

In the simulation study, a modified CS-phantom[33] designed for compressed sensing reconstruction was used to simulate the few-view projection data. For reference, the used CS-phantom was shown in Fig. 2. For the CT projection simulation, we chose a geometry that was representative of a fan-beam CT scanner setup. Each projection datum along an x-ray through the sectional image was calculated as the intersection length of an x-ray with a pixel.[34] To check the capability of our algorithm, the noisy case was also considered. Noise was generated using a Poisson model with mean equal to the mean of the number of transmitted photons. In the simulation of noisy study, the incident x-ray intensity was set to 66000. The scanning and reconstruction parameters for both noise-free and noisy cases were listed in Table 1.

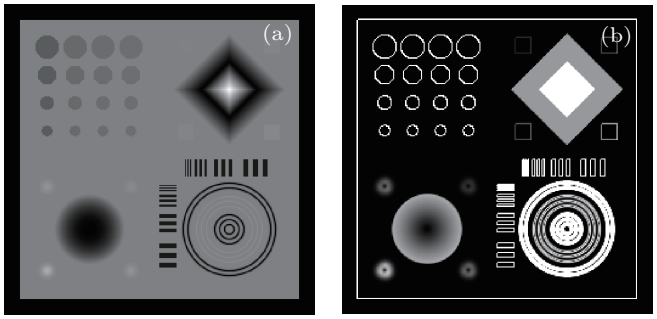


Fig. 2. (a) A modified CS-phantom shown in the grayscale window [0,1]. (b) A gradient magnitude image of the phantom shown in the grayscale window [0,0.05]. The phantom array comprises 52780 nonzero pixels, and there are 22779 nonzero values in its gradient image.

Table 1. Parameters in simulation studies.

Parameters	Noise-free cases	Noisy cases
detector elements	716	716
detector bin size	0.127 mm	0.127 mm
source to axis distance	300 mm	300 mm
source to detector distance	600 mm	600 mm
views of projection data	20	120
reconstruction size	256×256 pixels	256×256 pixels
pixel size	0.127×0.127 mm^2	0.127×0.127 mm^2

3.1.1. Noise-free cases

In the first group of simulation, the number of iterations for all methods was 2500. For TV-ADM and HTV-ADM algorithms, the parameters of ADM framework were the same in the experiments: μ and λ_1 were set to 1024 and 32, respectively. For the HTV-ADM algorithm, α_1, α_2, and λ_2 were set to 1, 1, and 32, respectively. The NL gradient was updated at iterations 500 and 1000, and the patch size and searching window were set to 9×9 and 31×31, respectively.

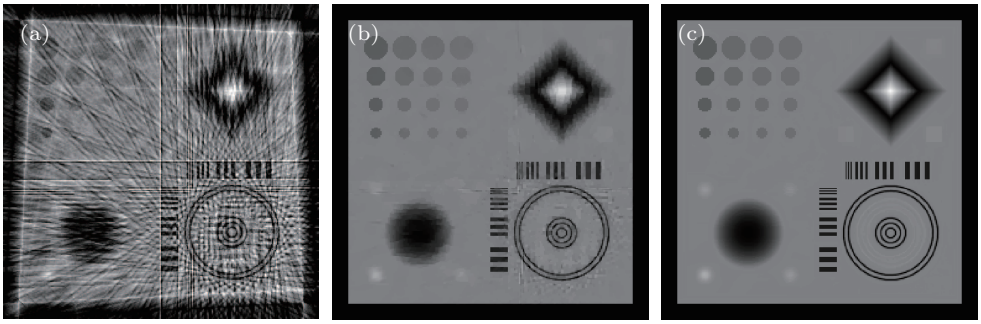


Fig. 3. Image reconstruction of the CS-phantom from noise-free projection dataset. Results obtained from (a) FBP, (b) TV-ADM, and (c) HTV-ADM methods. Display window is [0, 1].

The images reconstructed from ideal data by FBP, TV-ADM, and HTV-ADM methods are shown in Fig. 3. To reveal texture details, the zoomed region of interest (ROI) images are shown in Fig. 4. TV-ADM reconstructions have many blocky artifacts in smoothly varying image regions. Compared with the TV-ADM method, the HTV-ADM method can efficiently avoid the staircase effect and obtain better recovery of details.

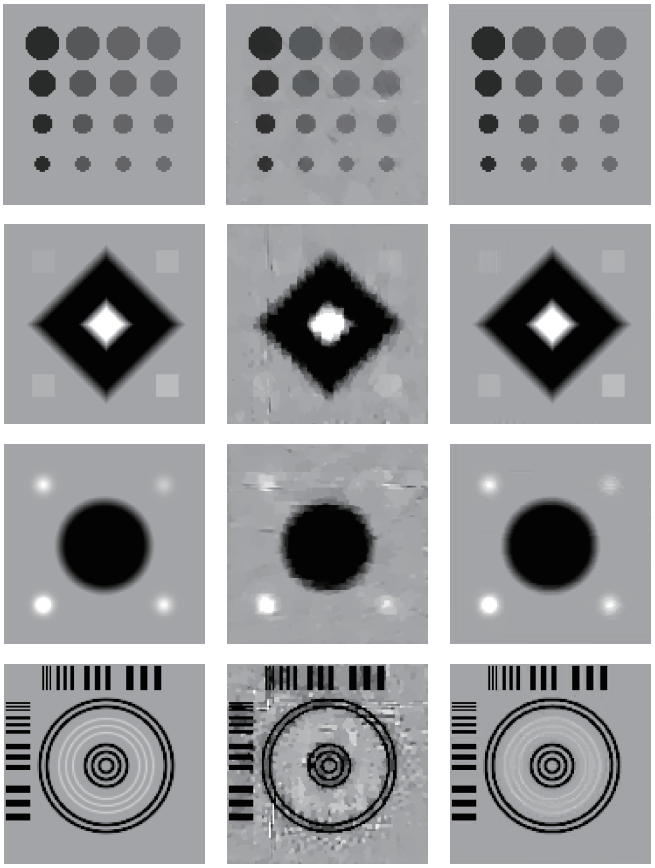


Fig. 4. Reconstructed ROIs of the CS-phantom from noise-free projection dataset. Results obtained from (a) original image, results of the (b) TV-ADM, and (c) HTV-ADM methods. Display window is [0.3, 0.6].

To evaluate the reconstruction accuracy quantitatively, the root mean squared error (RMSE) is used as a measure of the reconstruction deviation between the reconstructed and ideal images. The RMSE is defined as

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} |f_i - u_i|^2}{N}}, \tag{19}$$

where f and u denote the ideal phantom and the reconstruction, respectively, and N is the total pixels in the image. An RMSE value close to zero suggests high similarity to the ideal phantom image.

The corresponding RMSE of the reconstructions is calculated and plotted in Fig. 5. The recovery accuracy and running time of each reconstruction method are listed in Table 1. The quantitative results from the proposed HTV-ADM method showed obviously better results than that from the TV-ADM method. In iteration 500, the weight of NL operator is updated by estimating the reconstructed image, and then the HTV-ADM algorithm works differently from the TV-ADM algorithm. Interestingly, the HTV-ADM method shows a faster convergence when the similarity weight is updating. Furthermore, it shows that the time consumption of the HTV-ADM method is slightly larger than that of the TV-ADM method, and the present algorithm can preserve a good balance between performance and computation.

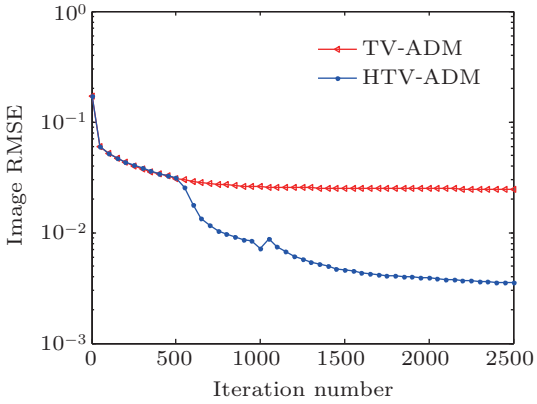


Fig. 5. (color online) Root mean squared errors as a function of iterations for different methods in noise-free case.

Table 2. RMSE and running time of different methods in the noise-free case.

Iteration steps		500	1000	2500
TV-ADM	RMSE	0.0314	0.0251	0.0247
	running time/s	14.54	32.87	88.60
HTV-ADM	RMSE	0.0314	0.00716	0.00351
	running time/s	14.54	39.57	106.53

3.1.2. Noisy cases

In the simulation of a noisy case, admissible reconstructions need more projection views than the noiseless case because of inconsistencies in data acquisition. Thus, simulations with noisy data were conducted using 120 views. For the TV-ADM and HTV-ADM algorithms, μ and λ_1 were set to 128 and 32, respectively. For the HTV-ADM algorithm, α_1, α_2, and λ_2 were set to 1, 1, and 32, respectively. As the noise would easily be introduced into the reconstructed image with the increase in iteration number, the iteration number for TV-ADM was selected as 200. The NL gradient was updated at iterations 200, and the total number of iterations for HTV-ADM was 500. The patch size and searching window were set to 11×11 and 31×31, respectively.

The images reconstructed by the FBP, TV-ADM, and HTV-ADM methods are shown in Fig. 6 and the corresponding zoomed-in ROIs are shown in Fig. 7. The corresponding RMSEs of the reconstructions are calculated and plotted in Fig. 8, and the recovery accuracy and running time of different reconstruction methods are shown in Table 3. The reconstruction results indicate that HTV-ADM produces relatively better results with projections contaminated by noise than the TV-ADM method.

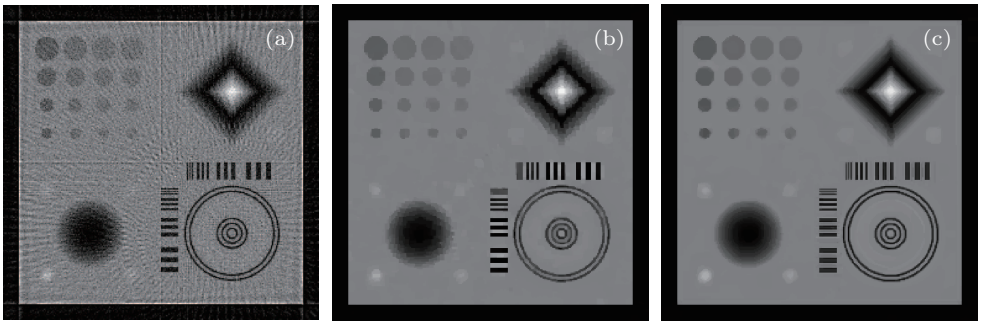


Fig. 6. Image reconstruction of the CS-phantom from noisy projection dataset. Results obtained from (a) FBP, (b) TV-ADM (iteration number: 200), and (c) HTV-ADM (iteration number: 500) methods. Display window is [0, 1].

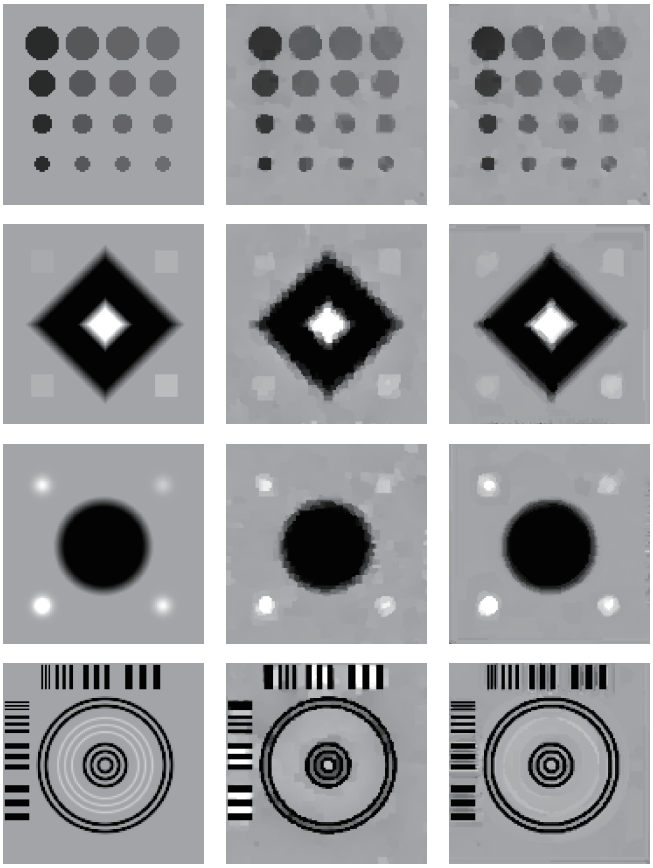


Fig. 7. Reconstructed ROIs of the CS-phantom from noisy projection dataset. Results obtained from (a) original image, (b) TV-ADM (iteration number: 200), and (c) HTV-ADM (iteration number: 500) methods. Display window is [0.3, 0.6].

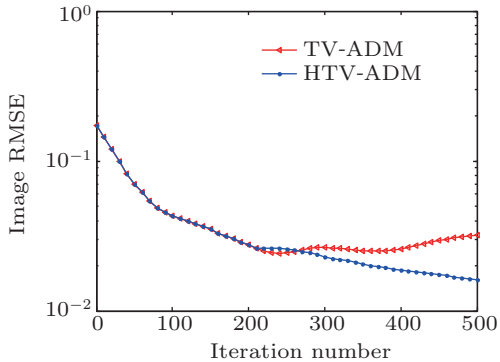


Fig. 8. (color online) Root mean squared errors as a function of iterations for different methods in the noisy case.

Table 3. RMSE and running time of different methods in the noisy case.

Iteration steps		100	200	400	500
TV-ADM	RMSE	0.0429	0.0275	0.0260	0.0321
	Running time /s	12.41	24.47	49.84	61.57
HTV-ADM	RMSE	0.0429	0.0275	0.0188	0.0161
	Running time /s	12.41	24.47	56.96	69.86

3.2. Real data study

To further demonstrate the potential capability of the proposed method, we performed an experimental rabbit head study from real data for clinical use. In this study, CT projection data were acquired using a CT scanner primarily comprising an x-ray source (FXE-225, YXLON, Germany) and a flat-panel detector (Varian 4030E, USA). The central slice of the projection data was extracted for 2D investigation and was modeled with 1600 bins on a 1D detector for 2D image reconstruction. The associated imaging parameters of the CT scanner were as follows: (i) 360 projection views were acquired evenly for a 360° rotation on a circular orbit, (ii) the distance from the detector to the x-ray source was 1434.73 mm, (iii) the distance from the rotation center to the source was 502.808 mm, and (iv) the space of each detector bin was 0.254 mm. All of the reconstructed images comprised a 720×720 square of pixels. The size of each pixel was 0.089 mm × 0.089 mm. We evenly extracted a 72-, 120-, and 180-view projection from the projection data for illustration purposes.

In the experiment, the algorithm parameters of all ADM frameworks were set to $\mu = 128$ and $\lambda_1 = 64$. For the HTV-ADM algorithm, α_1, α_2, and λ_2 were set to 1, 1, and 64, respectively. Similar to Section 3.1.2, the numbers of iterations for the TV-ADM and HTV-ADM were 200 and 600, respectively. The NL gradient was updated at iterations 200, and the patch size and searching window were set to 11×11 and 31×31, respectively.

The reconstructed images of the rabbit head from the real CT scan are presented in Fig. 9. Both TV-ADM and HTV-ADM method can effectively restrain the artifacts and improve the image quality from few-view projection data. To further demonstrate the superiority of our algorithm, the zoomed-in

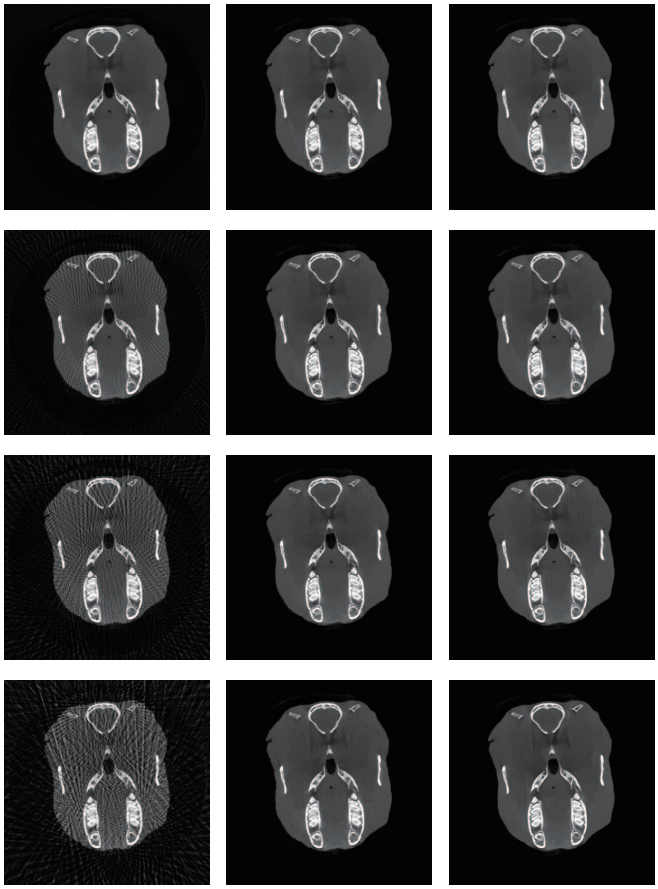


Fig. 9. (color online) Images reconstructed using the (a) FBP, (b) TV-ADM, and (c) HTV-ADM methods from 360- (1st row), 180- (2nd row), 120- (3rd row), and 72-view (4th row) projections, respectively. Display window is [0.0047, 0.08] mm^{-1}.

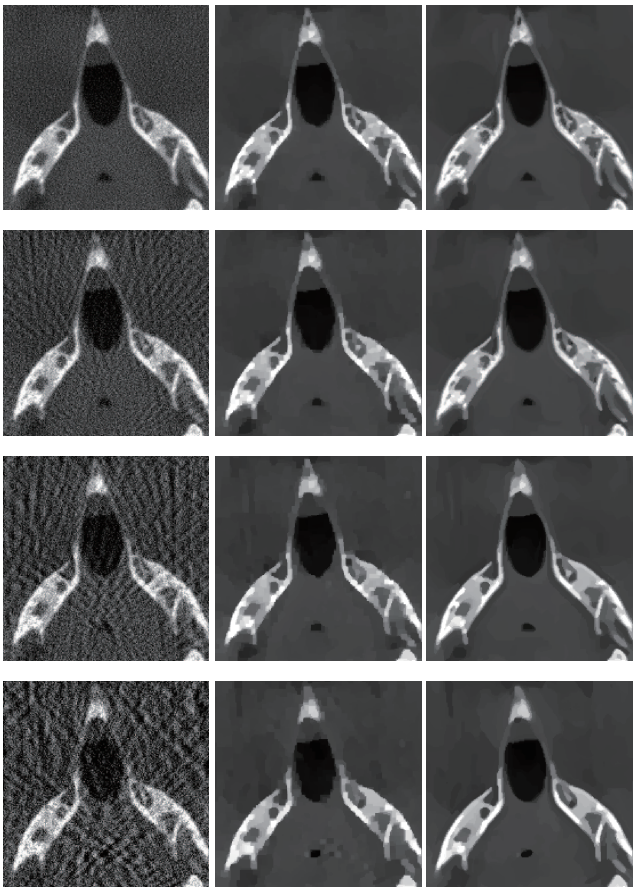


Fig. 10. (color online) Zoomed-in views of images reconstructed using the (a) FBP, (b) TV-ADM, and (c) HTV-ADM methods from 360- (1st row), 180- (2nd row), 120- (3rd row), and 72-view (4th row) projections, respectively. Display window is [0.0047, 0.08] mm^{-1}.

images corresponding to the selected ROIs in the middle are shown in Fig. 10. It can be seen that the edges and details are more accurate and better preserved by the HTV-ADM method than other methods.

4. Discussion and conclusion

Sparsity in the image gradient is exploited widely for reducing the x-ray CT sampling rate in the projection view angle, but the sparsity in the NL gradient is embedded and often ignored in state-of-the-art CT image reconstruction. To extract this information, we investigate the sparse representation of the to-be-estimated image by an NL gradient, and then improve the optimization-based reconstruction by exploring the local continuity and nonlocal similarity prior. Experiments have shown a clear improvement over standard TV minimization algorithms owing to the introduction of the self-similarity information of spatial interactions into the image reconstruction process.

As the sparsity of the NL gradient is related to the exploitation of image similarity redundancy, the setting of the similarity description plays a vital role and has considerable value to be explored and discussed. For different parameter settings, the reconstruction is likely to yield different levels of image quality. In this study, we focus on the sparsity-guided metrics of parameter setting determinations to use most of both similarity and sparsity potential for a CT imaging model. Although we cannot provide a "best" selection strategy for this HTV minimization procedure, the results show a similar performance in accordance with our analysis. The finding that the suggested method in this study employing HTV minimization allows accurate image recovery with sparse projection data suggests a clinically useful potential for radiation dose reduction.

The framework and metrics are only considered for the typical l_2 distance-based similarity weight. Other affinity measures between regions and pixels may provide a more reasonable performance. The complexity of the algorithm is also an important issue. As the computational complexity of weight computing is linear, it will increase rapidly with increasing image size. We hope to extend the theoretical foundations and also to investigate effective affinity expressions for gaining much improvement at very low computational cost. Addressing this question is one of our future research focuses.

References

[1] Fazel R, Krumholz H M, Wang Y, *et al.* 2009 *New Engl. J. Med.* **361** 849

[2] Brenner D J and Hall E J 2007 *New Engl. J. Med.* **357** 2277

[3] Wang J, Liang Z, Lu H and Xing L 2010 *Curr. Med. Imaging Rev.* **6** 72

[4] Sidky E Y, Kao C M and Pan X C 2006 *J. X-Ray Sci. Technol.* **14** 119

[5] Gordon R, Bender R and Herman G T 1970 *J. Theor. Biol.* **29** 471

[6] Andersen A and Kak A 1984 *Ultrasonic Imaging* **6** 81

[7] Lange K and Carson R 1984 *J. Comput. Assist. Tomo.* **8** 306

[8] Sidky E Y and Pan X C 2008 *Phys. Med. Biol.* **17** 4777

[9] Han X, Bian J G, Eaker D R, Kline T L, Sidky E Y, Ritman E L and Pan X C 2011 *IEEE Trans. Med. Imag.* **30** 606

[10] Han X, Bian J G, Ritman E L, Sidky E Y and Pan X C 2012 *Phys. Med. Biol.* **57** 5245

[11] Duan X H, Cheng J P, Zhang L, Xing Y X, Chen Z Q and Zhao Z R 2009 *IEEE Trans. Nucl. Sci.* **56** 1377

[12] Chen Z Q, Jin X, Li L and Wang G 2013 *Phys. Med. Biol.* **58** 2119

[13] Wang L Y, Zhang H M, Cai A L, Yan B, Li L and Hu G E 2013 *Acta Phys. Sin.* **62** 198701 (in Chinese)

[14] Bian J G, Siewerdsen J H, Han X, Sidky E Y, Prince J L, Pelizzari C A and Pan X C 2010 *Phys. Med. Biol.* **22** 6575

[15] Tang J, Nett B E and Chen G 2009 *Phys. Med. Biol.* **54** 5781

[16] Tian Z, Jia X, Yuan K H, Pan T and Jiang S B 2011 *Phys. Med. Biol.* **56** 5949

[17] Liu Y, Ma J, Fan Y and Liang Z 2012 *Phys. Med. Biol.* **57** 7923

[18] Liu Y, Liang Z L, Ma J H, Lu H B, Wang K, Zhang H and Moore W 2014 *IEEE Trans. Med. Imag.* **33** 749

[19] Cai A L, Wang L Y, Zhang H M, Yan B, Li L, Xi X Q and Li J X 2014 *J. X-ray Sci. Technol.* **22** 335

[20] Niu S Z, Gao Y, Bian Z Y, Huang J, Chen W F, Yu G H, Liang Z R and Ma J H 2014 *Phys. Med. Biol.* **59** 2997

[21] Buades A, Coll B and Morel J M 2005 *Multiscale Model Simul.* **4** 490

[22] Gilboa G and Osher S 2007 *Multiscale Model Simul.* **6** 595

[23] Gilboa G and Osher S 2007 *Multiscale Model Simul.* **7** 1005

[24] Zhang X Q, Burger M, Bresson X and Osher S 2010 *SIAM J. Imaging Sci.* **3** 253

[25] Lou Y F, Zhang X Q, Osher S and Bertozzi A 2010 *Journal of Scientific Computing* **42** 185

[26] Zhang Y, Zhang W H and Zhou J L 2014 *The Scientific World Journal* **2014** 458496

[27] Wang Y L, Yang J F, Yin W T and Zhang Y 2008 *SIAM J. Imaging Sci.* **1** 248

[28] Zhang H M, Wang L Y, Yan B, Li L, Xi X Q and Lu L Z 2013 *Chin. Phys. B* **22** 078701

[29] Zhang H M, Wang L Y, Yan B, Li L, Cai A L and Hu G E 2016 *PLoS ONE* **11** e0149899

[30] Li C, Yin W, Jiang H and Zhang Y 2013 *Comput. Optim. Appl.* **56** 507

[31] Xu H Y, Sun Q S, Luo N, Cao G and Xia D 2013 *PLoS ONE* **8** e65865

[32] Zelnik-Manor L and Perona P 2004 *Advances in Neural Information Processing Systems*, December 13–18, 2004, Vancouver, Canada, p. 1601

[33] Smith D S and Welch E B 2011 *Proceedings of* 19*th Annual Meeting of International Society for Magnetic Resonance in Medicine*, May 7–13, 2011, Montreal, Canada, p. 2845

[34] Chen J L, Li L, Wang L Y, Cai A L, Xi X Q, Zhang H M, Li J X and Yan B 2015 *Chin. Phys. B* **24** 028703