Article

Nonlocal Total Variation Using the First and Second Order Derivatives and Its Application to CT image Reconstruction

Yongchae Kim and Hiroyuki Kudo *

Department of Computer Science, Graduate School of Systems and Information Engineering, University of Tsukuba, Tennoudai 1-1-1, Tsukuba 305-8573, Japan; acevip12@hotmail.com

* Correspondence: kudo@cs.tsukuba.ac.jp

check for
updates

Abstract: We propose a new class of nonlocal Total Variation (TV), in which the first derivative and the second derivative are mixed. Since most existing TV considers only the first-order derivative, it suffers from problems such as staircase artifacts and loss in smooth intensity changes for textures and low-contrast objects, which is a major limitation in improving image quality. The proposed nonlocal TV combines the first and second order derivatives to preserve smooth intensity changes well. Furthermore, to accelerate the iterative algorithm to minimize the cost function using the proposed nonlocal TV, we propose a proximal splitting based on Passty's framework. We demonstrate that the proposed nonlocal TV method achieves adequate image quality both in sparse-view CT and low-dose CT, through simulation studies using a brain CT image with a very narrow contrast range for which it is rather difficult to preserve smooth intensity changes.

Keywords: image reconstruction; computed tomography; compressed sensing; nonlocal total variation; sparse-view CT; low-dose CT; proximal splitting; row-action; brain CT image

1. Introduction

Nonlocal Total Variation (TV) [1–6] was proposed as an improved version of ordinary TV. Nonlocal TV can use a weighting function (e.g., the weight of nonlocal means filter) by taking the intensity difference between the pixel pair into account, and can obtain higher image quality than the ordinary TV that uses only pairs of adjacent pixels.

Since G. Gilboa and S. Osher (2009) [5] proposed nonlocal operator, nonlocal TV has been widely applied to image reconstruction problems in sparse-view CT and low-dose CT [1–4].

H. Kim et al. (2016) [2] applied nonlocal TV to sparse-view CT and showed that nonlocal TV improves image quality over ordinary TV and incorporating the reweighted L1 norm into nonlocal TV further improves tissue contrast and structural details. Following this, K. Kim et al. (2017) [3] applied nonlocal TV to low-dose CT and showed nonlocal TV is effective for low-dose noise (Poisson noise).

D. Lv et al. (2019) [4] proposed a hybrid prior distribution (called NLTG prior) that combines nonlocal TV with Gaussian distribution. Additionally, they showed the possibility of applying NLTG prior to a large class of image reconstruction problems, especially when reference images were available.

However, the most existing nonlocal TV studies [2–4] are based on the first-order derivative, and still contain the staircase artifact problem as a potential drawback. Since the first-order derivative is too sensitive to the pixel values, even linear intensity changes are detected as false edges, which leads to staircase artifacts in the same way as local TV does [7–9].

On the other hand, higher-order derivatives (the second-order or more) possesses a potential risk that, as the order of differentiation is larger, its ability enhance image edges is smaller leading to an image blurring problem.

In previous studies, to overcome the disadvantage of using only the first-order or only the second-order, Bredies et al. (2010) [10] proposed Total Generalized Variation (TGV) that involves and balances higher-order derivatives and showed impressive results in the denoising problem. After that work, Ranftl et al. (2014) [11] proposed a nonlocal version of TGV.

Our proposed method is also a combined method of the first and second order derivatives. We define the regularization term as a weighted sum of two terms, where the first term is the ordinary nonlocal TV based on the first-order and the second term is based on the second-order. Additionally, in this paper, we introduce a newly discovered idea called nonlocal Total K-Split Variation, which is based on the second-order derivative using nonlocal regularization. This idea allows us to preserve smooth intensity changes well.

Furthermore, to accelerate the iterative algorithm associated with the proposed nonlocal TV, we propose a specially designed proximal splitting based on Passty's framework. In this proximal splitting, as the number of dividing the cost function into small subfunctions increases, the faster convergence is achieved [12–14]. The structure of final iterative algorithm is row-action type with respect to both the data-fidelity term and the regularization term [15–17], which converges to a minimizer very quickly [18–22].

Finally, in order to demonstrate the performances of combining the first and second order derivatives in reconstructed images, we use a brain CT image, in which a very narrow contrast range is used to display the image, where it is very difficult to preserve smooth intensity changes. Also, simulation studies are performed for both the sparse-view CT and the low-dose CT. We demonstrate that the proposed nonlocal TV method achieves adequate image quality within a small number of iterations.

2. Methodology

2.1. Problem Definition

We define the following unconstrained cost function:

$$\underset{\vec{x} \geq 0}{argmin} J(\vec{x}) = f(\vec{x}) + u(\vec{x}) = \left\| A\vec{x} - \vec{b} \right\|_2^2 + \beta\omega \left\| W\vec{x} \right\|_1^1, \tag{1}$$

where $f(\vec{x}) = \|A\vec{x} - \vec{b}\|_2^2$ is the data-fidelity term, and $u(\vec{x}) = \beta\omega\|W\vec{x}\|_1^1$ is the regularization term, and $A = \{a_{ij}\}$ is the $I \times J$ system matrix, β is the hyper-parameter to control regularization strength, and ω is the weight of regularization term, and W is the sparsifying transform to make $W\vec{x}$ sparse. Image reconstruction is an inverse problem to recover the image of attenuation coefficients $\vec{x} = (x_1, x_2, \ldots, x_J)^T$ from the measured projection data $\vec{b} = (b_1, b_2, \ldots, b_I)^T$.

In the sparse-view CT [23,24], by using the projection data corresponding to less than 100 directions (the conventional CT uses 1000–2000 projection data), the equation $A\vec{x} = \vec{b}$ becomes severely underdetermined, i.e., the dimension J of unknowns $\vec{x}$ is larger than the dimension I of measurements $\vec{b}$ ($I < J$). In this case, the regularization term acts to avoid the ill-posed problem by introducing the prior knowledge that most components of the vector $W\vec{x}$ are close to 0.

On the other hand, in the low-dose CT, the equation $A\vec{x} = \vec{b}$ becomes inconsistent due to Poisson noise $\vec{n}$ ($A\vec{x} - \vec{b} = \vec{n}$). In this case, the regularization term helps to reduce the effect of Poisson noise $\vec{n}$ by a smoothing.

First, we begin by the modified anisotropic nonlocal TV based on the first-order derivative expressed as

$$Nonlocal\ TV = u^{TV}(\vec{x}) = \beta t \sum_{j=1}^{J} \sum_{j' \in \Omega} \omega_{jj'} |x_j - x_{j'}|, \tag{2}$$

where x_j is the intensity in the pixel j, and $x_{j'}$ is the intensity in a distant pixel j', and $\omega_{jj'}$ is the weight of smoothing assigned for each pixel pair $(x_j, x_{j'})$.

Next, we describe a newly discovered regularization term based on the second-order derivative called nonlocal Total K-split Variation (TKV). The main idea is to consider more derivatives around x_j and $x_{j'}$ as

$$Nonlocal\ TKV = u^{TKV}(\vec{x}) = \frac{\beta(1-t)}{8} \sum_{j=1}^{J} \sum_{j' \in \Omega} \sum_{k=1}^{8} \omega_{jj'} \left| (x_j - x_{j_k}) - (x_{j'} - x_{j'_k}) \right|, \tag{3}$$

where x_{j_k} is the adjacent pixel value of x_j, $x_{j'_k}$ is the adjacent pixel value of $x_{j'}$. We remark that the TKV term is divided into a sum of eight terms dependent of the direction to take the pixel difference as shown in Figure 1.
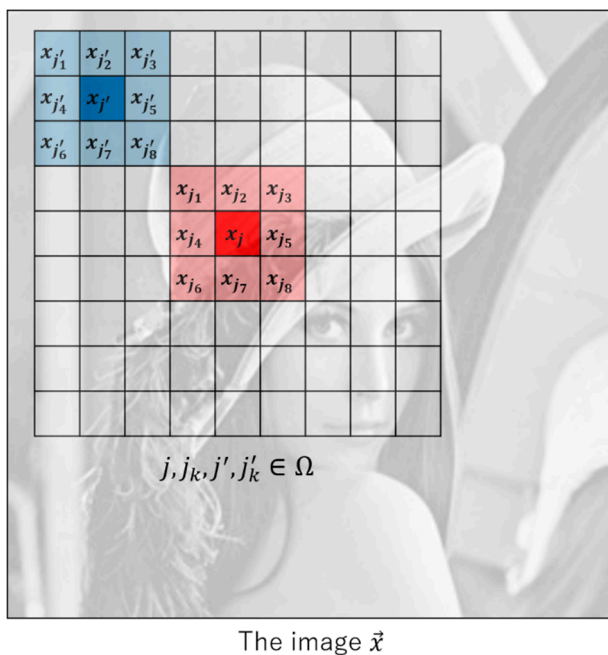

Figure 1. Definition of the pixel location in the proposed regularization term corresponding to $k = 1, 2, 3, \cdots, 7, 8$.

In the proposed method, we assume that the reconstructed image $\vec{x}$ is close to piecewise-polynomial of first order. Under this assumption, our proposed regularization term is designed to include both the first and second order derivatives as

$$u(\vec{x}) = \beta \sum_{j=1}^{J} \sum_{j' \in \Omega} \omega_{jj'} \left[t|x_j - x_{j'}| + \frac{(1-t)}{8} \sum_{k=1}^{8} \left| (x_j - x_{j_k}) - (x_{j'} - x_{j'_k}) \right| \right], (0 < t < 1), \tag{4}$$

where the proposed regularization term is a combination of two terms $(u(\vec{x}) = u^{TV}(\vec{x}) + u^{TKV}(\vec{x}))$. Additionally t is the trade-off parameter between nonlocal TV and TKV. If t is large, the reconstructed image becomes closer to that of nonlocal TV. If t is small, the reconstructed image becomes closer to that of nonlocal TKV. To match the strength of the first term into that of the second term, we divide the

second-order derivative based nonlocal TKV into 8 directions. Figure 1 shows the location of pixels appearing in $u^{TV}(\vec{x})$ and $u^{TKV}(\vec{x})$ for $k = 1, 2, 3, \cdots, 7, 8$.

2.2. Accelerated Algorithm Using the Proximal Splitting with Passty's Framework

To accelerate the iterative algorithm to minimize the cost function, we propose a specially designed proximal splitting with Passty's framework. The ordinary proximal splitting is able to minimize a cost function consisting of a sum of only two component terms, where the proximity operator corresponding to each subfunction is applied alternately [12,13]. On the other hand, by using Passty's framework which is not well-known in the image reconstruction community, a cost function can be divided into a number of much simpler subfunctions [14]. This results in considerable benefits for optimization problems appearing in CT reconstruction. By applying the proximity operator corresponding to each subfunction alternately, the proposed algorithm possesses a form of row-action type [15–17], which converges to a minimizer very quickly [18–22].

We begin by a brief review the proximal splitting with Passty's framework. Let us consider a convex minimization problem formulated as

$$\underset{\vec{x}}{\operatorname{argmin}} J(\vec{x}), \tag{5}$$

where $J(\vec{x})$ is a lower semi-continuous (lsc) convex function.

The proximity operator corresponding to the function $J(\vec{x})$ is defined as

$$\vec{x} = prox_{\alpha J}(\vec{z}) \equiv \underset{\vec{x}}{\operatorname{argmin}} (J(\vec{x}) + \frac{1}{2\alpha} \|\vec{x} - \vec{z}\|_2^2), \tag{6}$$

where α is the parameter called step-size. We note that $J(\vec{x})$ can be a non-differentiable function such as component terms of TV or nonlocal TV. The proximity operator is a non-expansive mapping such that its fixed-points $\vec{x}$ satisfying $\vec{x} = prox_{\alpha J}(\vec{x})$ coincides with a minimizer of $J(\vec{x})$ for any $\alpha > 0$. Therefore, the minimization problem of $J(\vec{x})$ can be solved by the iterative formula expressed as $\vec{x}^{(n+1)} = prox_{\alpha J}(\vec{x}^{(n)})$ (i.e., proximal algorithm). Next, we are going to explain the proximal splitting with Passty's framework.

[Passty's framework] Let us consider the case where $J(\vec{x})$ can be divided into a sum of subfunctions as:

$$J(\vec{x}) = \sum_{i=1}^{I} J_i(\vec{x}). \tag{7}$$

The iterative algorithm can be constructed by applying the proximity operator corresponding to each subfunction $J_i(\vec{x})$ $(i = 1, 2, \ldots, I)$ as below

$$\vec{x}^{(n+1)} = prox_{\alpha^{(n)} J_I} \cdots \cdots prox_{\alpha^{(n)} J_2} \cdot prox_{\alpha^{(n)} J_1}(\vec{x}^{(n)}). \tag{8}$$

Furthermore, let us consider the case where $J_i(\vec{x})$ is a sum of two subfunctions, like our cost function including data-fidelity term and regularization term as

$$J_i(\vec{x}) = f_i(\vec{x}) + u(\vec{x}). \tag{9}$$

The update can be constructed by applying two operators corresponding to each subfunction alternately as below

$$\vec{x}^{(n+1)} = prox_{\alpha^{(n)}u} \cdot prox_{\alpha^{(n)}f_i}\left(\vec{x}^{(n)}\right). \tag{10}$$

Finally, we show in Algorithm 1 the optimization model applied to this paper:

Algorithm 1: Proximal splitting with Passty's framework

Give an initial vector $\vec{x}^{(0,1)}$. Execute the following.

$\begin{aligned} &\lceil \text{For } n = 0, 1, 2, ..., \quad (n \text{ is the main iteration})\\ &\quad \Big\lceil \text{For } i = 1, 2, ..., I\\ &\quad \Big\lfloor \vec{x}^{(n,i+1)} = prox_{\alpha^{(n)}J_i}\left(\vec{x}^{(n,i)}\right)\\ &\lfloor \vec{x}^{(n+1,1)} = \vec{x}^{(n,I+1)} \end{aligned}$

In Passty's framework, by increasing the number to divide the cost function into smaller subfunctions, the better convergence can be expected. In this paper, this division is performed as follows. First, the data-fidelity term is divided as shown in Equation (7) in such a way that each subfunction $f_i(\vec{x})$ contains only a single term corresponding to projection data b_i. Therefore, the final algorithm can be designed in the form of a row-action type algorithm such as ART method.

We mention that $u(\vec{x})$ can also be divided into a sum of many subfunctions. In this paper, we divide $u(\vec{x})$ as finely as possible similarly to the case of the data-fidelity term. This idea leads to a significant benefit to simplify the processing of nonlocal TV+TKV term as well as improving the convergence speed. With respect to the regularization term $u(\vec{x})$, we perform a division described in Section 2.3.2.

2.3. Optimization

In this section, we focus on how to divide the cost function and how to derive the resulting iterative algorithm.

2.3.1. Update the Data-Fidelity Term

The data-fidelity term can be divided into I subfunctions as below

$$f(\vec{x}) = \left\| A\vec{x} - \vec{b} \right\|_2^2 = \sum_{i=1}^{I} f_i(\vec{x}) = \sum_{i=1}^{I} (\vec{a}_i^T \vec{x} - b_i)^2, \tag{11}$$

where I is the number of projection data, and $\vec{a}_i$ is i-th row vector of the system matrix A, and b_i is i-th component of projection data. Furthermore, we note that $f_i(\vec{x})$ is a subfunction corresponding to the data-fidelity term.

The minimization problem for the data-fidelity term can be defined as

$$\vec{x}^{(n,i+1)} = prox_{\alpha^{(n)}f_i}\left(\vec{x}^{(n,i)}\right) = \underset{\vec{x}}{\text{argmin}}\left\{\left(\vec{a}_i^T \vec{x} - b_i\right)^2 + \frac{1}{2\alpha^{(n)}}\|\vec{x} - \vec{x}^{(n,i)}\|_2^2\right\}. \tag{12}$$

By introducing a slack variable $z = \vec{a}_i^T \vec{x}$, the above minimization problem for each subfunction $f_i(\vec{x})$ can be converted into the constrained minimization below

$$\underset{x,z}{\min} \ (z - b_i)^2 + \frac{1}{2\alpha^{(n)}}\left\|\vec{x} - \vec{x}^{(n,i)}\right\|_2^2 \quad s.t. \ z = \vec{a}_i^T \vec{x}. \tag{13}$$

The Lagrange function can be defined by

$$L_{f_i}(\vec{x}, z, \lambda) = (z - b_i)^2 + \frac{1}{2\alpha^{(n)}}\|\vec{x} - \vec{x}^{(n,i)}\|_2^2 + \lambda\left(z - \vec{a}_i^T\vec{x}\right), \tag{14}$$

where λ is the Lagrange multiplier called the dual variable.

Finally, by optimizing the Lagrange function, we obtain the following expression of row-action type iteration.

$$\vec{x}^{(n,i+1)} = \vec{x}^{(n,i)} + \alpha^{(n)}\frac{b_i - \vec{a}_i^T\vec{x}^{(n,i)}}{1/2 + \alpha^{(n)}\|\vec{a}_i\|_2^2}\vec{a}_i, \ \alpha^{(n)} = \frac{\alpha_0}{1 + \varepsilon n}, \left(i \in R_{[I]}\right), \tag{15}$$

$$\left(\alpha_0 : \text{ initial value of step size}, \ \varepsilon : \text{ deceleration rate of step-size}, \ R_{[I]} : \text{ ordered subsets}\right)$$

where $\alpha^{(n)}$ is the step-size parameter to control the convergence, and n is the number of main iterations. After updating all elements of projection data, n is increased by 1. The step-size $\alpha^{(n)}$ is diminished gradually to zero as the iteration proceeds (i.e. diminishing step-size rule). In Passty's framework, it is known that this diminishing contributes to ensuring an exact convergence to a minimizer, thereby, avoiding the so-called limit cycle problem. Furthermore, we introduce a random-access order of projection data $R_{[I]}$ to enable a fast convergence within 20–30 iterations [19,20]. The mathematical detail to derive the update formula in Equation (15) has been already described in our previous studies [19–22].

2.3.2. Update the Regularization Term

We describe how to divide the regularization term. The modified anisotropic nonlocal TV possesses the following structure and is called L1 based group LASSO here. The L1 based group LASSO possesses a very simple structure, where each absolute value term can be considered a group element.

$$u^{TV}(\vec{x}) = \beta t \sum_{j=1}^{J}\sum_{j'\in\Omega}\omega_{jj'}|x_j - x_{j'}| = \beta t \sum_{g=1}^{G}\left[\omega_{jj'}\|\vec{x}_j - \vec{x}_{j'}\|_1^1\right]_g = \beta t \sum_{g=1}^{G}u_g^{TV}(\vec{x}), \quad (j = g). \tag{16}$$

where g is the group index, and the group index g corresponds to the pixel index j, and the TV term can be divided into G groups (G subfunctions). Furthermore, the group itself becomes a subfunction $u_g^{TV}(\vec{x})$.

In the case of the TKV term, it can be divided into $(G \times 8)$ subfunctions as below

$$u^{TKV}(\vec{x}) = \frac{\beta(1-t)}{8}\sum_{k=1}^{8}\sum_{g=1}^{G}\left[\omega_{jj'}\|(\vec{x}_j - \vec{x}_{j_k}) - (\vec{x}_{j'} - \vec{x}_{j'_k})\|_1^1\right]_g = \frac{\beta(1-t)}{8}\sum_{k=1}^{8}\sum_{g=1}^{G}u_g^{TKV}(\vec{x}). \tag{17}$$

The detailed structure of the TV term is shown in Figure 2. Among the elements of the group, the pixel j is common and distant pixel j' has different values for each other.
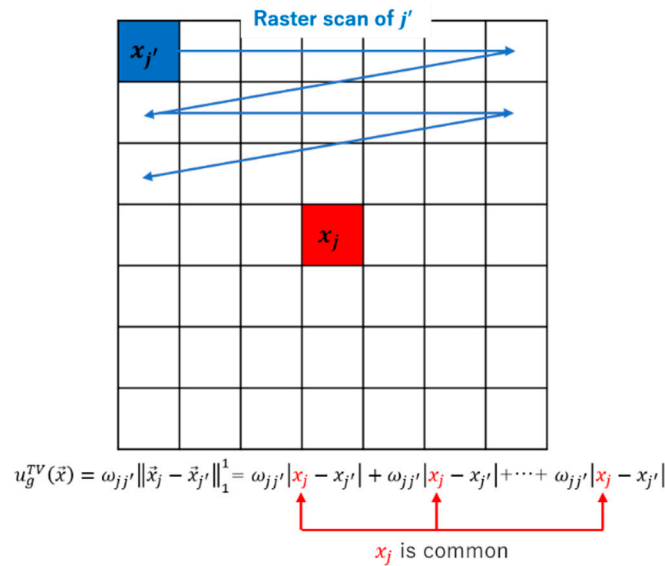
Figure 2. Raster scanning during the update ($j' = 1, 2, 3, \dots$).

The sequential update is related to raster scanning.

In the TV term, when assuming that x_j and $x_{j'}$ are updated simultaneously, x_j is updated sequentially J' times, and $x_{j'}$ is updated once respectively.

In the case of TKV term, when assuming that x_j, x_{j_k}, $x_{j'}$, $x_{j'_k}$ are updated simultaneously, x_j is updated sequentially $(J' \times 8)$ times, and $x_{j'}$ is updated sequentially eight times. Following this, x_{j_k} and $x_{j'_k}$ (the adjacent pixels) are updated once respectively.

[Update the TV term] First, we consider updating the pixel j.

The proximity operator for a subfunction can be defined as follows

$$\underset{\vec{x}_j}{\operatorname{argmin}}(\omega_{jj'}\|\vec{x}_j - \vec{x}_{j'}^{(n)}\|_1^1 + \frac{1}{2\alpha^{(n)}}\|\vec{x}_j - \vec{x}_j^{(n,j')}\|_2^2), \tag{18}$$

where $x_{j'}^{(n)}$ is the current updated solution as a constant approximation, which is the image updated from the data-fidelity term. The above L1 norm minimization problems can be solved with the following soft-thresholding.

$$S_\nabla\left(x_j^{(n,j')}\right) = x_j^{(n,j')} - \begin{cases} \nabla, & \tau > \nabla \\ -\nabla, & \tau < -\nabla \\ \tau_{TV} & (otherwise) \end{cases},$$
$$\left(\tau_{TV} = (x_j^{(n,j')} - x_{j'}^{(n)}), \nabla = \alpha^{(n)}\beta t\omega_{jj'}\right), \tag{19}$$

where $S_\nabla\left(x_j^{(n,j')}\right)$ is the soft-thresholding function. The solution $(S_\nabla\left(x_j^{(n,j')}\right))$ corresponding to otherwise is simply $x_{j'}^{(n)}$.

For better convergence, although it is possible to update only pixel j, we update the pixel j and j' simultaneously.

For updating the pixel j and j', we further divide a subfunction $u_g^{TV}(\vec{x})$ into two subfunctions as below

$$
\begin{aligned}
u^{TV}(\vec{x}) &= \beta t \sum_{g=1}^{G} u_g^{TV}(\vec{x}) = \beta t \sum_{g=1}^{G} \left[\omega_{jj'} \| \vec{x}_j - \vec{x}_{j'} \|_1^1 \right]_g \\
&= \beta t \sum_{g=1}^{G} \left[\left[\omega_{jj'} \left\| \frac{\vec{x}_j - \vec{x}_{j'}}{2} \right\|_1^1 \right]_{x_j} + \left[\omega_{jj'} \left\| \frac{\vec{x}_j - \vec{x}_{j'}}{2} \right\|_1^1 \right]_{x_{j'}} \right]_g \\
&= \beta t \sum_{g=1}^{G} \left[p(\vec{x}_j) + p(\vec{x}_{j'}) \right]_g .
\end{aligned}
\tag{20}
$$

where $p(\vec{x}_j) = [\,]_{x_j}$, $p(\vec{x}_{j'}) = [\,]_{x_{j'}}$.

The proximity operator for each subfunction $\left(prox_{\alpha^{(n)}p}\!\left(\vec{x}_j^{(n,j')}\right), prox_{\alpha^{(n)}p}\!\left(\vec{x}_{j'}^{(n)}\right) \right)$ can be defined as follows

$$
\begin{aligned}
\underset{\vec{x}_j}{\text{argmin}} \left(\omega_{jj'} \left\| \vec{x}_j - \frac{\vec{x}_j^{(n,j')} + \vec{x}_{j'}^{(n)}}{2} \right\|_1^1 + \frac{1}{2\alpha^{(n)}} \left\| \vec{x}_j - \vec{x}_j^{(n,j')} \right\|_2^2 \right), \\
\underset{\vec{x}_{j'}}{\text{argmin}} \left(\omega_{jj'} \left\| - \vec{x}_{j'} + \frac{\vec{x}_j^{(n,j')} + \vec{x}_{j'}^{(n)}}{2} \right\|_1^1 + \frac{1}{2\alpha^{(n)}} \left\| \vec{x}_{j'} - \vec{x}_{j'}^{(n)} \right\|_2^2 \right),
\end{aligned}
\tag{21}
$$

where $x_j^{(n,j')}$ and $x_{j'}^{(n)}$ are the current updated solution as a constant approximation, which is the image updated from the data-fidelity term.

Finally, the above L1 norm minimization problems can be solved with the following soft-thresholding.

$$
\begin{aligned}
S_\nabla\!\left(x_j^{(n,j')}\right) &= x_j^{(n,j')} - \begin{cases} \nabla, & \tau > \nabla \\ -\nabla, & \tau < -\nabla \\ \tau_{TV} & (otherwise) \end{cases}, \\
S_\nabla\!\left(x_{j'}^{(n)}\right) &= x_{j'}^{(n)} + \begin{cases} \nabla, & \tau > \nabla \\ -\nabla, & \tau < -\nabla \\ \tau_{TV} & (otherwise) \end{cases}, \\
\left(\tau_{TV} \right. &= \left. \left(x_j^{(n,j')} - x_{j'}^{(n)}\right)/2, \quad \nabla = \alpha^{(n)}\beta t \omega_{jj'} \right),
\end{aligned}
\tag{22}
$$

where $S_\nabla\!\left(x_j^{(n,j')}\right)$ and $S_\nabla\!\left(x_{j'}^{(n)}\right)$ are soft-thresholding functions. By dividing a subfunction and updating for each variable as shown in Equations (20)–(22), the solution $\left(S_\nabla\!\left(x_j^{(n,j')}\right), S_\nabla\!\left(x_{j'}^{(n)}\right)\right)$ corresponding to otherwise becomes the average of $x_j^{(n,j')}$ and $x_{j'}^{(n)}$ $\left((x_j^{(n,j')} + x_{j'}^{(n)})/2\right)$. Compared to Equation (19), this weak average that occurs otherwise can reduce the error in convergence and improve the stability of convergence.

[Update the TKV term] We update the pixel j, j_k, j', j'_k simultaneously. For updating the pixel j, j_k, j', j'_k, we further divide a subfunction $u_g^{TKV}(\vec{x})$ into four subfunctions as below

$$
\begin{aligned}
u^{TKV}(\vec{x}) &= \frac{\beta(1-t)}{8} \sum_{k=1}^{8} \sum_{g=1}^{G} u_g^{TKV}(\vec{x}) = \frac{\beta(1-t)}{8} \sum_{k=1}^{8} \sum_{g=1}^{G} \left[\omega_{jj'} \left\| (\vec{x}_j - \vec{x}_{j_k}) - (\vec{x}_{j'} - \vec{x}_{j'_k}) \right\|_1^1 \right]_g \\
&= \frac{\beta(1-t)}{8} \sum_{k=1}^{8} \sum_{g=1}^{G} \left[\left[\omega_{jj'} \left\| \frac{(\vec{x}_j - \vec{x}_{j_k}) - (\vec{x}_{j'} - \vec{x}_{j'_k})}{4} \right\|_1^1 \right]_{x_j} + \left[\omega_{jj'} \left\| \frac{(\vec{x}_j - \vec{x}_{j_k}) - (\vec{x}_{j'} - \vec{x}_{j'_k})}{4} \right\|_1^1 \right]_{x_{j_k}} \right. \\
&\quad \left. + \left[\omega_{jj'} \left\| \frac{(\vec{x}_j - \vec{x}_{j_k}) - (\vec{x}_{j'} - \vec{x}_{j'_k})}{4} \right\|_1^1 \right]_{x_{j'}} + \left[\omega_{jj'} \left\| \frac{(\vec{x}_j - \vec{x}_{j_k}) - (\vec{x}_{j'} - \vec{x}_{j'_k})}{4} \right\|_1^1 \right]_{x_{j'_k}} \right]_g \\
&= \frac{\beta(1-t)}{8} \sum_{k=1}^{8} \sum_{g=1}^{G} \left[q(\vec{x}_j) + q(\vec{x}_{j_k}) + q(\vec{x}_{j'}) + q(\vec{x}_{j'_k}) \right]_g,
\end{aligned}
$$

(23)

where $q(\vec{x}_j) = []_{x_j}$, $q(\vec{x}_{j_k}) = []_{x_{j_k}}$, $q(\vec{x}_{j'}) = []_{x_{j'}}$, $q(\vec{x}_{j'_k}) = []_{x_{j'_k}}$.

The proximity operator for each subfunction $(prox_{\alpha^{(n)}q}(\vec{x}_j^{(n,j',k)})$, $prox_{\alpha^{(n)}q}(\vec{x}_{j_k}^{(n)})$, $prox_{\alpha^{(n)}q}(\vec{x}_{j'}^{(n,k)})$, $prox_{\alpha^{(n)}q}(\vec{x}_{j'_k}^{(n)}))$ can be defined as follows

$$
\begin{aligned}
&\underset{\vec{x}_j}{\operatorname{argmin}} (\omega_{jj'} \left\| \vec{x}_j - \frac{3\vec{x}_j^{(n,j',k)} + \vec{x}_{j_k}^{(n)} + \vec{x}_{j'}^{(n,k)} - \vec{x}_{j'_k}^{(n)}}{4} \right\|_1^1 + \frac{1}{2\alpha^{(n)}} \| \vec{x}_j - \vec{x}_j^{(n,j',k)} \|_2^2), \\
&\underset{\vec{x}_{j_k}}{\operatorname{argmin}} (\omega_{jj'} \left\| - \vec{x}_{j_k} + \frac{\vec{x}_j^{(n,j',k)} + 3\vec{x}_{j_k}^{(n)} - \vec{x}_{j'}^{(n,k)} + \vec{x}_{j'_k}^{(n)}}{4} \right\|_1^1 + \frac{1}{2\alpha^{(n)}} \| \vec{x}_{j_k} - \vec{x}_{j_k}^{(n)} \|_2^2), \\
&\underset{\vec{x}_{j'}}{\operatorname{argmin}} (\omega_{jj'} \left\| - \vec{x}_{j'} + \frac{\vec{x}_j^{(n,j',k)} - \vec{x}_{j_k}^{(n)} + 3\vec{x}_{j'}^{(n,k)} + \vec{x}_{j'_k}^{(n)}}{4} \right\|_1^1 + \frac{1}{2\alpha^{(n)}} \| \vec{x}_{j'} - \vec{x}_{j'}^{(n,k)} \|_2^2), \\
&\underset{\vec{x}_{j'_k}}{\operatorname{argmin}} (\omega_{jj'} \left\| \vec{x}_{j'_k} - \frac{-\vec{x}_j^{(n,j',k)} + \vec{x}_{j_k}^{(n)} + \vec{x}_{j'}^{(n,k)} + 3\vec{x}_{j'_k}^{(n)}}{4} \right\|_1^1 + \frac{1}{2\alpha^{(n)}} \| \vec{x}_{j'_k} - \vec{x}_{j'_k}^{(n)} \|_2^2),
\end{aligned}
$$

(24)

where $\vec{x}_j^{(n,j',k)}$, $\vec{x}_{j_k}^{(n)}$, $\vec{x}_{j'}^{(n,k)}$, $\vec{x}_{j'_k}^{(n)}$ are the current updated solution as a constant approximation, which is the image updated from the TV term.

Finally, the above L1 norm minimization problems can be solved with the following soft-thresholding.

$$
\begin{aligned}
S_\nabla\!\left(x_j^{(n,j',k)}\right) &= x_j^{(n,j',k)} - \begin{cases} \nabla, & \tau > \nabla \\ -\nabla, & \tau < -\nabla \\ \tau_{TKV} & (otherwise) \end{cases}, \\
S_\nabla\!\left(x_{jk}^{(n)}\right) &= x_{jk}^{(n)} + \begin{cases} \nabla, & \tau > \nabla \\ -\nabla, & \tau < -\nabla \\ \tau_{TKV} & (otherwise) \end{cases}, \\
S_\nabla\!\left(x_{j'}^{(n,k)}\right) &= x_{j'}^{(n,k)} + \begin{cases} \nabla, & \tau > \nabla \\ -\nabla, & \tau < -\nabla \\ \tau_{TKV} & (otherwise) \end{cases}, \\
S_\nabla\!\left(x_{j'_k}^{(n)}\right) &= x_{j'_k}^{(n)} - \begin{cases} \nabla, & \tau > \nabla \\ -\nabla, & \tau < -\nabla \\ \tau_{TKV} & (otherwise) \end{cases}, \\
\left(\tau_{TKV} = \left(x_j^{(n,j',k)} - x_{jk}^{(n)} - x_{j'}^{(n,k)} + x_{j'_k}^{(n)}\right)\!\Big/4, \right. & \left. \quad \nabla = \alpha^{(n)}\beta(1-t)\omega_{jj'}/8\right),
\end{aligned}
\tag{25}
$$

where $S_\nabla\!\left(x_j^{(n,j',k)}\right)$, $S_\nabla\!\left(x_{jk}^{(n)}\right)$, $S_\nabla\!\left(x_{j'}^{(n,k)}\right)$, $S_\nabla\!\left(x_{j'_k}^{(n)}\right)$ are soft-thresholding function.

2.3.3. The Weight

In this paper, we used the weight of nonlocal means filter [25] as

$$
\omega_{jj'} = \frac{exp(-\max(\|B(x_j) - B(x_{j'})\|_2^2 - 2\sigma^2, 0)/h^2)}{\sum_{j'\in\Omega} exp(-\max(\|B(x_j) - B(x_{j'})\|_2^2 - 2\sigma^2, 0)/h^2)},
\tag{26}
$$

where $\|B(x_j) - B(x_{j'})\|_2^2$ means the average Euclidean distance between patches $(B(x_j), B(x_{j'}))$ centered in an interest pixel x_j and a distant pixel $x_{j'}$.

Theoretically, the weight must be a fixed value as a hyper-parameter of the regularization term. However, there have been previous studies showing that reweighting at each iteration contributes to better image quality [2,9]. Additionally, the larger the size of weight (Ω), the better the performance of removing artifacts or noise. As long as the computer is capable of processing, we recommend increasing the size of the weight. However, if the size of the weight is too large, the calculation cost will increase enormously as compared with the image quality improvement. Therefore, it is important to decide the size of the weight while paying attention to cost performance considering the level of artifacts or noise. Figure 3 shows the cost performance of changing the size of weight.
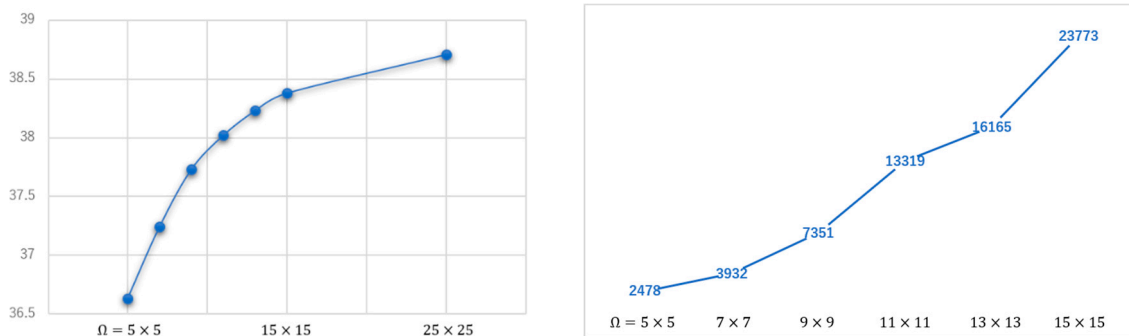
(**a**) Image quality improvement (PSNR)　　　　　(**b**) Computation time [ms]

Figure 3. Cost performance of changing the size of weight. (**a**) Image quality improvement (PSNR), (**b**) Computation time [ms].

Next, we describe the proper estimation of the weight. If the ground truth image is known, the weight can be calculated from the ground truth image. However, in the image reconstruction, only the projection data is given and the reconstructed image and the weight must be estimated simultaneously from the projection data. Therefore, we construct the optimization by alternating the estimation of the reconstructed image and the weight.

We show the reweighting process of optimization including the data-fidelity term and regularization term:

$$
\begin{cases}
For\ i = 1,2,3,\dots,I\ (I\ \text{is the number of projection data}) \\
\quad 1)\ \text{Update the data term by Equation (15)}: \vec{x}^{(n,i+1)} = prox_{\alpha^{(n)} f_i}\left(\vec{x}^{(n,i)}\right); \\
\quad\ if\left((i\ mod\ S) == 0\right) \\
\quad\quad\left|\begin{matrix}
2)\ \text{Calculate the weight}\ \omega_{jj'}\ \text{from}\ \vec{x}^{(n,i+1)}; \\
3)\ \text{Update the TV term by Equation (22)}; \\
4)\ \text{Update the TKV term by Equation (25)};
\end{matrix}\right. \\
\vec{x}^{(n+1,1)} \leftarrow \vec{x}^{(n,I+1)}
\end{cases} \tag{27}
$$

where the weight is calculated once as common to nonlocal TV and TKV and the weight is calculated from the image updated from the data-fidelity term. The span parameter S determines how often regularization is performed in the main iteration n. If S is small, many regularization updates are performed in one iteration. Theoretically, the smaller the S, the more accurate the convergence. However, since nonlocal TV has a large amount of computational complexity, it is desirable to determine an appropriate value of S.

Figure 4 shows how the size of weight (Ω) influences image quality and computation time.
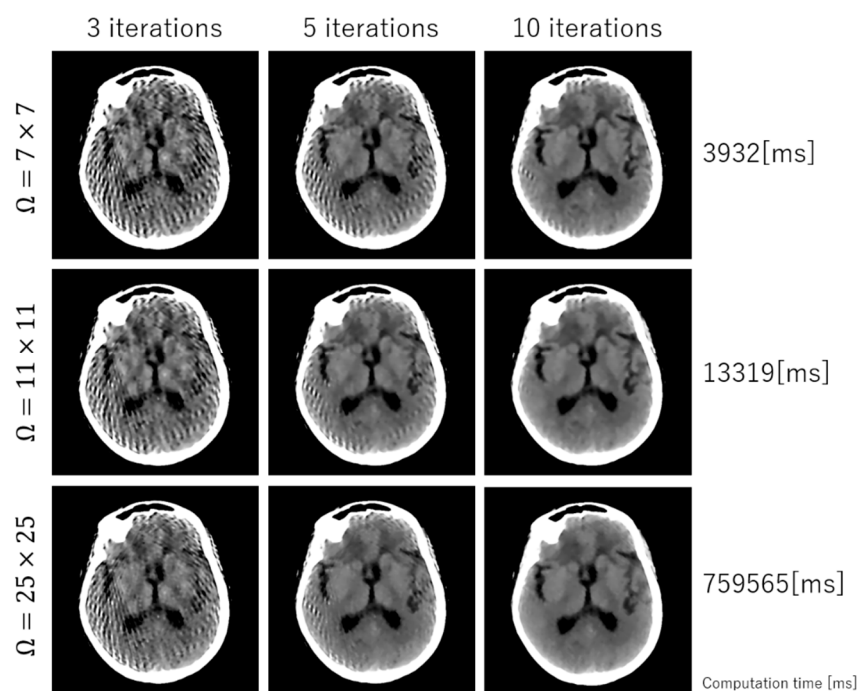
Figure 4. Demonstration of how the size of weight (Ω) influences image quality and computation time.

3. Experimental Results

We performed simulation studies using a brain CT image. The reason behind using the brain image is as follows. In the brain CT imaging, the reconstructed images are shown with a compressed gray scale range much larger compared to the other CT imaging like chest imaging or abdominal CT imaging. Therefore, preserving smooth intensity changes and avoiding the staircase artifacts are

much important in the brain case. Additionally, simulation studies were performed for both the sparse-view CT (the number of projection data was 64) and the low-dose CT (the number of photons was 3×10^6). The reconstructed image consisted of 512×512 pixels, where the pixel size was 0.0585 (cm^2). We compressed the range showing the reconstructed images to [7.82, 62.30] HU, where this contrast range was determined based on the contrast range used in clinical brain CT imaging. To evaluate image quality, standard RMSE, PSNR, SSIM values were used as metrics. The number of iterations in image reconstructions was 20 for nonlocal TV, TKV, and TV+TKV, which was determined by the fact that changes in image were small enough with this iteration number. We also showed the reconstructed images by the standard Filtered Back-Projection (FBP), and differences in image quality by changing values of the hyper-parameter *t* (i.e., the trade-off parameter between nonlocal TV (first derivative) term and the TKV (second derivative) term). The ground truth image and the FBP reconstructions are shown in Figure 5. The reconstructed images in the case of sparse-view CT are shown in Figure 6. In Figure 7, we show the used brain CT image with three display gray-scale ranges, from which we observe that the staircase artifacts are severe when the range of display gray-scale range is small. The reconstructed images in the case of low-dose CT are shown in Figure 8. In Figures 9–11, we show convergence properties of our iterative algorithm based on Passty's proximal splitting framework. In Figures 12 and 13, to show the effect of acceleration by Passty's proximal splitting, we incorporated the TV+TKV term into SIRT (simultaneous iterative reconstruction technique) which is a non-row-action method (a type of the standard iterative algorithm) and compared this with row-action based on our proposed nonlocal TV+TKV. From these figures, it can be observed that our algorithm converged very quickly. It is well-known that the standard iterative algorithms such as Chambolle–Pock [26] and proximal gradient algorithms require several hundreds of iteration up to the convergence. The benefit of our iterative algorithm mainly originates from the fact that our algorithm is of row-action type.
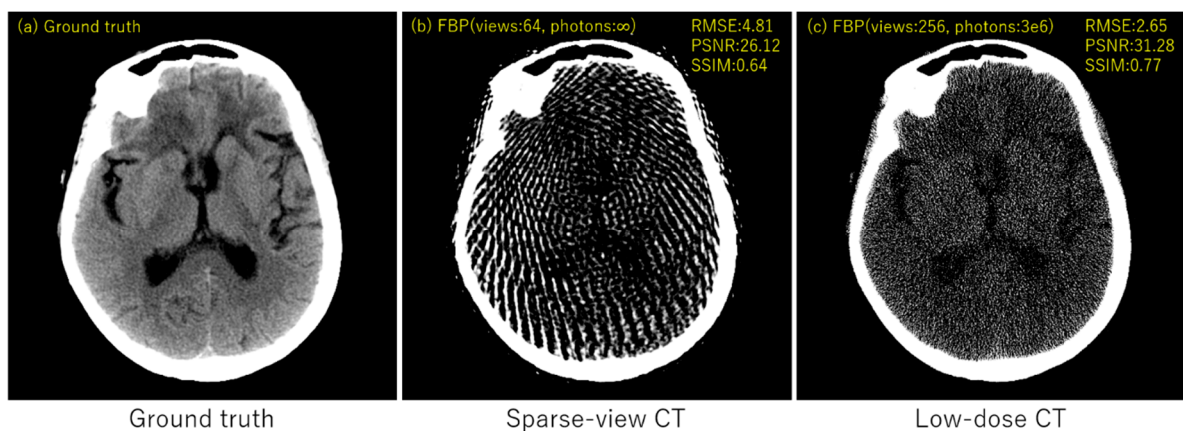

Figure 5. (**a**) Ground truth, (**b**) FBP with 64 projection data with no noise, (**c**) FBP with 256 projection data with the number of photon counts 3×10^6. All images are displayed with the same window of [7.82, 62.30] HU.
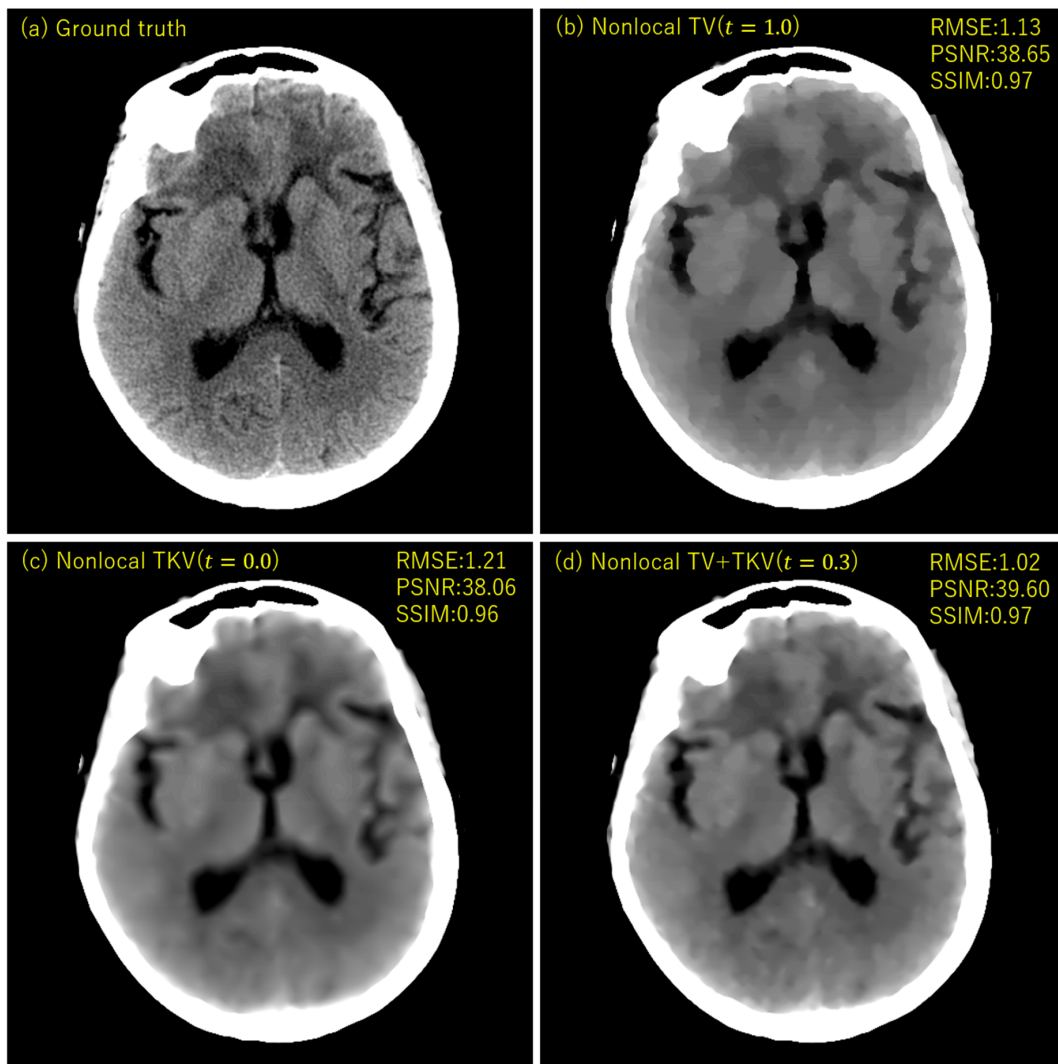
Figure 6. The reconstructed images of sparse-view CT (64 projection data with no noise). (**a**) Ground truth, (**b**) Nonlocal TV ($t = 1.0$), (**c**) Nonlocal TKV ($t = 0.0$), (**d**) Nonlocal TV+TKV ($t = 0.3$) were compared. All images are displayed with the same window of [7.82, 62.30] HU.
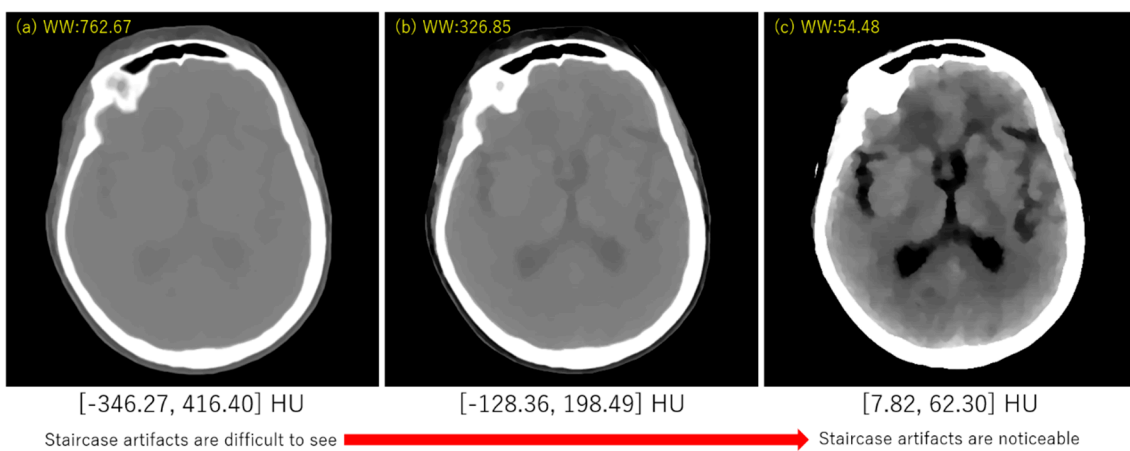

Figure 7. Demonstration of appearance of the staircase artifacts with various gray-scale ranges in displaying the brain CT image. (**a**) Window Width [−346.27, 416.40] HU, (**b**) Window Width [−128.36, 198.49] HU, (**c**) Window Width [7.82, 62.30] HU.
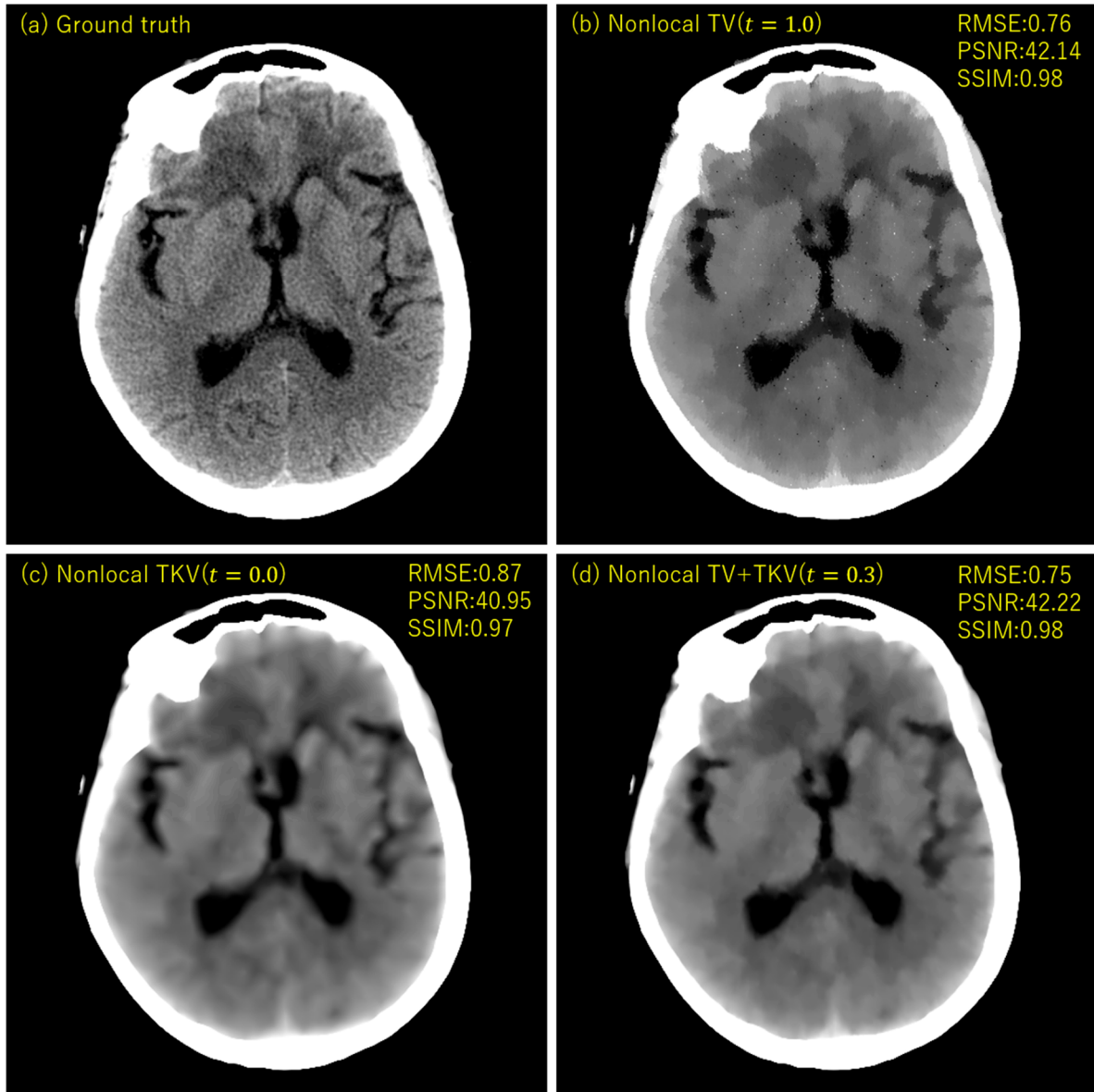
Figure 8. The reconstructed images of low-dose CT (256 projection data and the number of photon counts 3×10^6). (**a**) Ground truth, (**b**) Nonlocal TV ($t = 1.0$), (**c**) Nonlocal TKV ($t = 0.0$), (**d**) Nonlocal TV+TKV ($t = 0.3$) were compared. All images are displayed with the same window of [7.82, 62.30] HU.
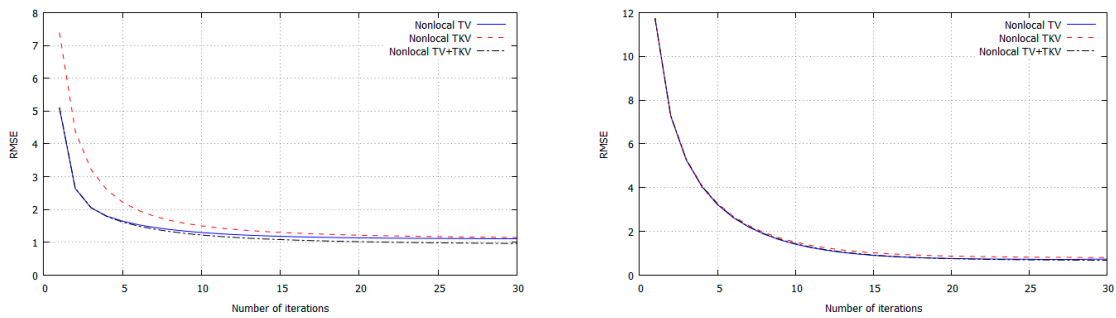
(**a**) Sparse-view CT

(**b**) Low-dose CT

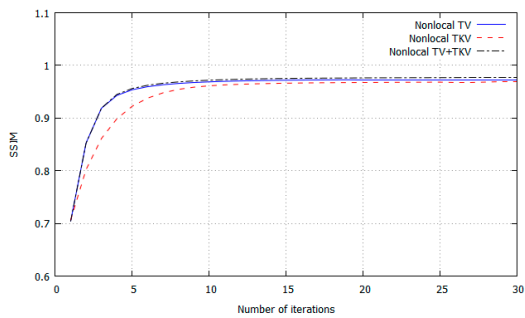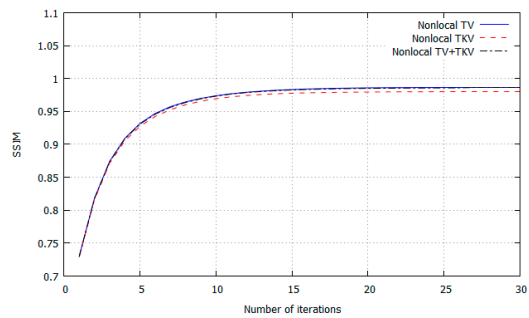**Figure 9.** RMSE for each iteration. (**a**) Sparse-view CT, (**b**) Low-dose CT.
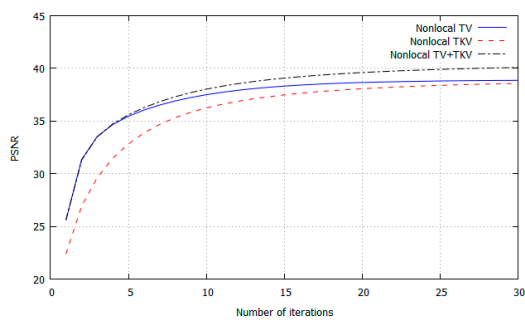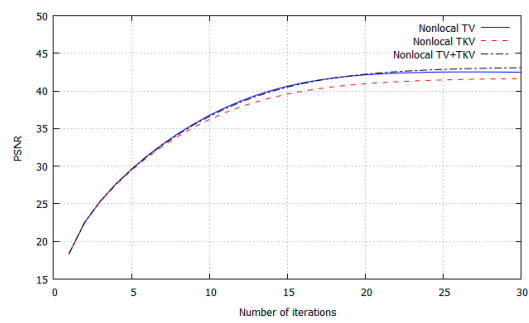
(**a**) Sparse-view CT　　　　　　　　　　　　　　(**b**) Low-dose CT

Figure 10. PSNR for each iteration. (**a**) Sparse-view CT, (**b**) Low-dose CT.

(**a**) Sparse-view CT　　　　　　　　　　　　　　(**b**) Low-dose CT

Figure 11. SSIM for each iteration. (**a**) Sparse-view CT, (**b**) Low-dose CT.
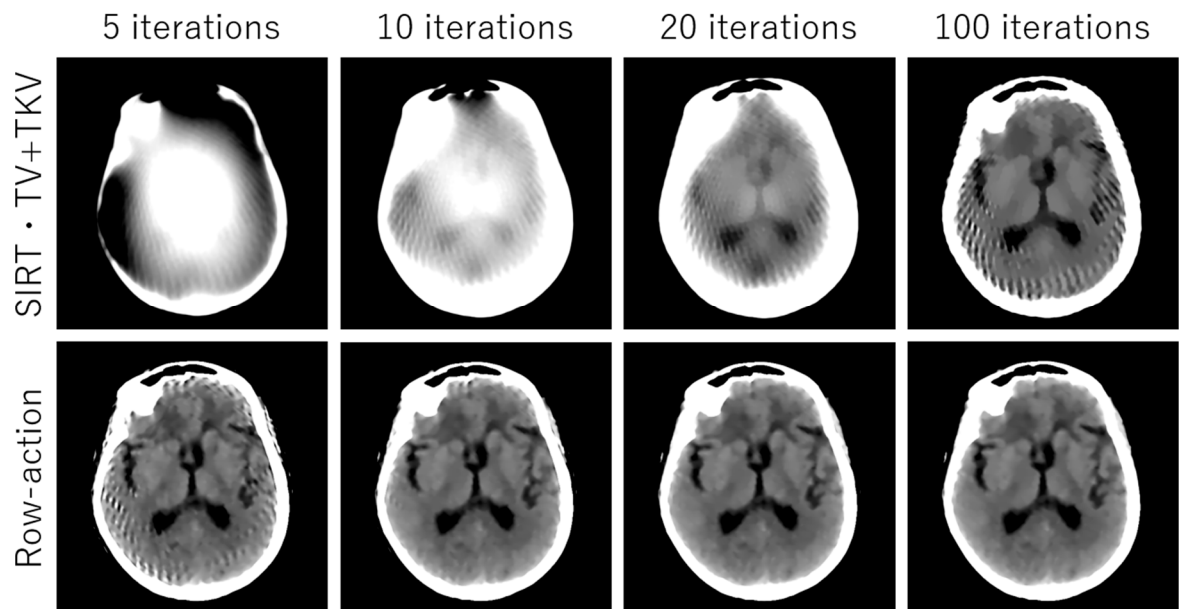

Figure 12. The reconstructed images of sparse-view CT (64 projection data with no noise). SIRT nonlocal TV+TKV and row-action accelerated nonlocal TV+TKV (our proposed method) were compared. All images are displayed with the same window of [7.82, 62.30] HU.
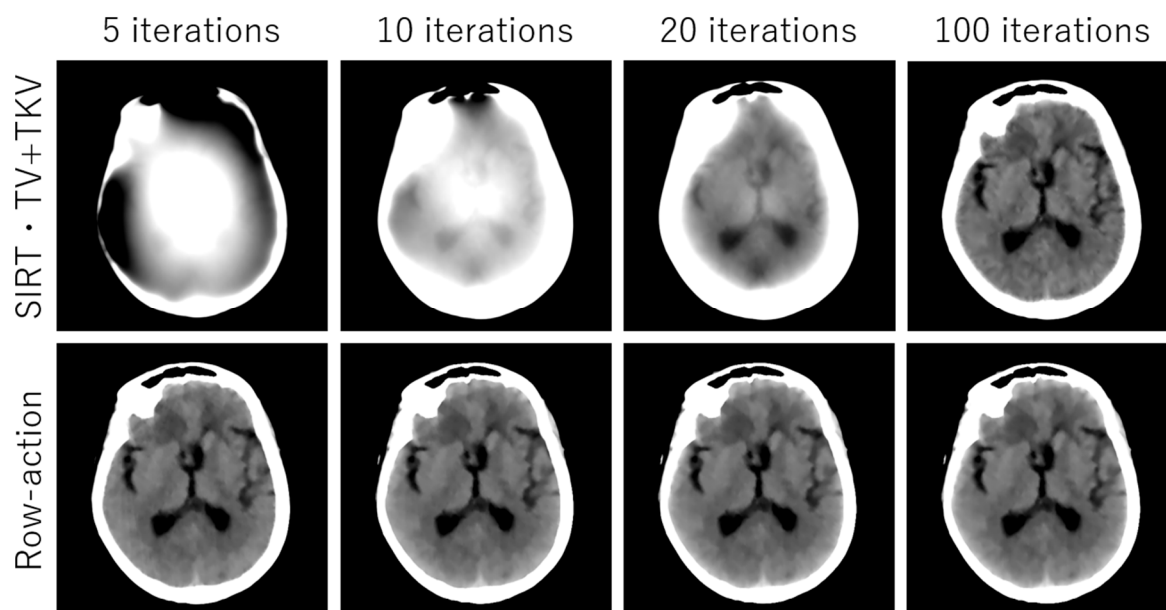
Figure 13. The reconstructed images of low-dose CT (256 projection data and the number of photon counts 3×10^6). SIRT nonlocal TV+TKV and row-action accelerated nonlocal TV+TKV (our proposed method) were compared. All images are displayed with the same window of [7.82, 62.30] HU.

4. Discussion

The experimental results Table 1 showed that the reconstructed image by nonlocal TV+TKV was closest to the ground truth image, with good RMSE, PSNR and SSIM values. Furthermore, no isolated points caused by outliers of the soft-thresholding, which often appear in the TV-class reconstruction methods, were visible. In the sparse-view CT, the result of nonlocal TV ($t = 1.0$) using only the first-order derivative was of very high-contrast, but the staircase artifacts appeared in the smooth intensity changes. In other words, the region with small intensity changes was over-smoother by the regularization as if the oil-painting. In the low-dose CT, the result of nonlocal TV showed isolated points in the region closer to the center of the image. In both the sparse-view CT and the low-dose CT, nonlocal TKV ($t = 0.0$) using only the second-order derivative was able to preserve fine soft tissues well including textures and low-contrast objects, however, it suffered from the blurring in the edge parts. This is because, by using a large weight in the second-order derivative, the threshold value (τ_{TKV}) of the soft-thresholding operation becomes very small (i.e., $\tau_{TV} \gg \tau_{TKV}$).

Table 1. Summary of each method.

	Nonlocal TV	Nonlocal TKV	Nonlocal TV + TKV
Convergence	**Good**	Not bad	**Good**
High contrast	**Yes**	No	**Yes**
Smooth intensity change	No	**Yes**	Yes

5. Conclusions

In this paper, we proposed a new concept in nonlocal TV, in which the first and second order derivatives are combined in the regularization term. By combining the two terms, we were able to compensate for each other's weaknesses, i.e., staircase artifact and loss in smooth intensity changes in the first derivative and image blurring in the second derivative. Furthermore, we proposed a specially designed proximal splitting algorithm that is based on Passty's framework. The key idea is to split the original cost function to minimize as finely as possible to accelerate convergence and simplify necessary computations. This allows as to make the final iterative algorithm into a form of row-action type, which

is known to converge very quickly compared to other standards such as Chambolle–Pock algorithm and proximal gradient. In our experiments, we experimentally confirmed that the proposed algorithm converges within 20 iterations even for the case of brain CT imaging in which the requirement of image contrast is very severe. The simulation results with the brain CT image were performed for both the sparse-view CT and the low-dose CT. We showed that our proposed algorithm works well in practice.

As future work, our proposed nonlocal TV can be compared with the latest technology e.g., deep leaning [27,28] or other applied methods such as low-rank minimization [29,30].

Recently, image reconstruction methods using deep learning have been actively investigated. Our proposed method can be compared with existing deep leaning [27,28] as advanced compressed sensing. Additionally, our proposed method can be applied to low-rank TV, which can improve image quality by combing low-rank minimization and Total Variation [29,30].

References

1. Kim, Y.; Kudo, H.; Chigita, K.; Lian, S. Image reconstruction in sparse-view CT using improved nonlocal total variation regularization. In Proceedings of the SPIE Optical Engineering + Applications, San Diego, CA, USA, 10 September 2019.
2. Kim, H.; Chen, J.; Wang, A.; Chuang, C.; Held, M.; Pouliot, J. Non-local total-variation (NLTV) minimization combined with reweighted L1-norm for compressed sensing CT reconstruction. *Phys. Med. Biol.* **2016**, *61*, 6878–6891. [CrossRef]
3. Kim, K.; El Fakhri, G.; Li, Q. Low-dose CT reconstruction using spatially encoded nonlocal penalty. *Med. Phys.* **2017**, *44*, e376–e390. [CrossRef] [PubMed]
4. Lv, D.; Zhou, Q.; Choi, J.K.; Li, J.; Zhang, X. NLTG Priors in Medical Image: Nonlocal TV-Gaussian (NLTG) prior for Bayesian inverse problems with applications to Limited CT Reconstruction. *Inverse Prob. Imaging* **2019**, *14*, 117–132. [CrossRef]
5. Gilboa, G.; Osher, S. Nonlocal Operators with Applications to Image Processing. *Multiscale Model. Simul.* **2009**, *7*, 1005–1028. [CrossRef]
6. Bresson, X. A Short Note for Nonlocal TV Minimization. Technical Report. Available online: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.210.471&rep=rep1&type=pdf (accessed on 15 June 2020).
7. Chambolle, A. An algorithm for Total variation Regularization and Denoising. *J. Math. Imaging.* **2004**, *20*, 89–97.
8. Rudin, L.; Osher, S.; Fatemi, E. Nonlinear total variation based noise removal algorithms. *Physica D* **1992**, *60*, 259–268. [CrossRef]
9. Zhang, X.; Xing, L. Sequentially reweighted TV minimization for CT metal artifact reduction. *Med. Phys.* **2013**, *40*, 1–12. [CrossRef]
10. Bredies, K.; Kunisch, K.; Pock, T. Total generalized variation. *SIAM J. Imag. Sci.* **2010**, *3*, 492–526. [CrossRef]
11. Ranftl, R.; Bredies, K.; Pock, T. Non-local total generalized variation for optical flow estimation. *Lect. Notes Comput. Sci.* **2014**, *8689*, 439–454.
12. Parikh, N.; Boyd, S. Proximal Algorithms. Available online: https://web.stanford.edu/~{}boyd/papers/pdf/prox_algs.pdf (accessed on 15 June 2020).
13. Combettes, P.L.; Pesquet, J.C. Proximal splitting methods in signal processing. In *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*; Springer: New York, NY, USA, 2011.
14. Passty, G.B. Ergodic convergence to a zero of the sum of monotone operators in Hilbert space. *J. Math. Anal. Appl.* **1979**, *72*, 383–390. [CrossRef]

15. Herman, G.T.; Meyer, L.B. Algebraic reconstruction techniques can be made computationally efficient (positron emission tomography application). *IEEE Trans. Med. Imaging* **1993**, *12*, 600–609. [CrossRef]

16. Tanaka, E.; Kudo, H. Subset-dependent relaxation in block-iterative algorithms for image reconstruction in emission tomography. *Phys. Med. Biol.* **2003**, *48*, 1405–1422. [CrossRef] [PubMed]

17. Wang, G.; Jiang, M. Ordered subset simultaneous algebraic reconstruction techniques (OS SART). *J. of X Ray Sci. Technol.* **2004**, *12*, 169–177.

18. Dong, J.; Kudo, H. Proposal of Compressed Sensing Using Nonlinear Sparsifying Transform for CT Image Reconstruction. *Med. Imaging Technol.* **2016**, *34*, 235–244.

19. Dong, J.; Kudo, H. Accelerated Algorithm for Compressed Sensing Using Nonlinear Sparsifying Transform in CT Image Reconstruction. *Med. Imaging Technol.* **2017**, *35*, 63–73.

20. Dong, J.; Kudo, H.; Kim, Y. Accelerated Algorithm for the Classical SIRT Method in CT Image Reconstruction. In Proceedings of the 5th International Conference on Multimedia and Image Processing, Nanjing, China, 10–12 January 2020; pp. 49–55.

21. Kudo, H.; Takaki, K.; Yamazaki, F.; Nemoto, T. Proposal of fault-tolerant tomographic image reconstruction. In Proceedings of the SPIE Optical Engineering + Applications, San Diego, CA, USA, 3 October 2016.

22. Kudo, H.; Dong, J.; Chigita, K.; Kim, Y. Metal artifact reduction in CT using fault-tolerant image reconstruction. In Proceedings of the SPIE Optical Engineering + Applications, San Diego, CA, USA, 10 September 2019.

23. Li, M.; Yang, H.; Kudo, H. An accurate iterative reconstruction algorithm for sparse objects: Application to 3 D blood vessel reconstruction from a limited number of projections. *Phys. Med. Biol.* **2002**, *47*, 2599–2609. [CrossRef] [PubMed]

24. Kudo, H.; Suzuki, T.; Rashed, E.A. Image reconstruction for sparse-view CT and interior CT-introduction to compressed sensing and differentiated backprojection. *Quant. Imaging Med. Surg.* **2013**, *3*. [CrossRef]

25. Buades, A.; Coll, B.; Morel, J.M. On image denoising methods. Available online: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.100.81&rep=rep1&type=pdf (accessed on 15 June 2020).

26. Sidky, E.Y.; Jørgensen, J.H.; Pan, X. Convex optimization problem prototyping for image reconstruction in computed tomography with the Chambolle–Pock algorithm. *Phys. Med. Biol.* **2012**, *57*, 3065–3091. [CrossRef]

27. Han, Y.; Ye, J.C. Framing U-Net via Deep Convolutional Framelets: Application to Sparse-View CT. *IEEE Trans. Med. Imaging* **2018**, *37*, 1418–1429. [CrossRef]

28. Jin, K.H.; McCann, M.T.; Froustey, E.; Unser, M. Deep Convolutional Neural Network for Inverse Problems in Imaging. *IEEE Trans. Image Process.* **2017**, *26*, 4509–4522. [CrossRef]

29. Shi, F.; Cheng, J.; Wang, L.; Yap, P.-T.; Shen, D. Low-Rank Total Variation for Image Super-Resolution. *Med. Image Comput. Comput. Assist. Interv.* **2013**, *16*, 155–162. [PubMed]

30. Niu, S.; Yu, G.; Ma, J.; Wang, J. Nonlocal low-rank and sparse matrix decomposition for spectral CT reconstruction. *Inverse Probl.* **2018**, *34*. [CrossRef] [PubMed]