Article

Impulse Noise Denoising Using Total Variation with Overlapping Group Sparsity and Lp-Pseudo-Norm Shrinkage

Lingzhi Wang, Yingpin Chen *, Fan Lin, Yuqun Chen, Fei Yu and Zongfu Cai

School of Physics and Information Engineering, Minnan Normal University, Zhangzhou 363000, China; wanglingzhi@mnnu.edu.cn (L.W.); linfan@mnnu.edu.cn (F.L.); cyq@mnnu.edu.cn (Y.C.); phoenix@mnnu.edu.cn (F.Y.); ypc@mnnu.edu.cn (Z.C.)

* Correspondence: 201511050301@std.uestc.edu.cn; Tel.: +86-139-0696-8591

check for updates

Featured Application: This paper proposes a new model for restoring polluted images under impulse noise, which makes a contribution to the research in image processing and image reconstruction.

Abstract: Models based on total variation (TV) regularization are proven to be effective in removing random noise. However, the serious staircase effect also exists in the denoised images. In this study, two-dimensional total variation with overlapping group sparsity (OGS-TV) is applied to images with impulse noise, to suppress the staircase effect of the TV model and enhance the dissimilarity between smooth and edge regions. In the traditional TV model, the L1-norm is always used to describe the statistics characteristic of impulse noise. In this paper, the Lp-pseudo-norm regularization term is employed here to replace the L1-norm. The new model introduces another degree of freedom, which better describes the sparsity of the image and improves the denoising result. Under the accelerated alternating direction method of multipliers (ADMM) framework, Fourier transform technology is introduced to transform the matrix operation from the spatial domain to the frequency domain, which improves the efficiency of the algorithm. Our model concerns the sparsity of the difference domain in the image: the neighborhood difference of each point is fully utilized to augment the difference between the smooth and edge regions. Experimental results show that the peak signal-to-noise ratio, the structural similarity, the visual effect, and the computational efficiency of this new model are improved compared with state-of-the-art denoising methods.

1. Introduction

Image denoising is one of the most important research areas in the field of image processing, and it has great value in both theoretical studies and engineering applications. Its usage spans the broad fields of image restoration [1], detection [2], photoelectric detection [3], geological exploration [4], remote sensing [5], and medical image analysis [6], among others [7,8]. With the development of compressed sensing theory, image processing algorithms based on sparse representation and constrained regularization have evolved into promising methods of image restoration [9]. Models based on total variation (TV) regularization [10–12] are found to be effective in removing random noise. The TV model has been successfully used in image restoration tasks such as denoising [13], deblurring [14], and super-resolution [15]. Although TV regularization can recover sharp edges of a degraded image, it also leads to some undesired effects and transforms the smooth signal into

piecewise constants, the so-called staircase effect. Several models have been proposed by scholars to make an improvement on the TV model [16–21]. One usual method is to replace the original TV norm by a high-order TV norm. The high-order TV overcomes the staircase effect while preserving the edges in the restored image. However, the high-order TV based methods may transform the smooth signal to over-smoothing and take more time to compute. More details can be referred to in References [22]. In 2010, Bredies et al. proposed the total generalized variation (TGV) model [19]. The TGV model puts constraints on both the first- and second-order gradients of an image, thus effectively attenuating the staircase effect of the TV model. Still, it is difficult to both preserve the image details and suppress the noise simultaneously in TGV. Furthermore, some scholars pay attention to fractional-order gradients replacing integer-order gradients [23]. Their research shows that using a fractional differential operator with $0 < v < 1$ can appropriately process the noise and edge information, but it also has un-denoised "spots" in the image.

Although these improved methods can alleviate the staircase artifacts, they might lead to "spots" effects on the processed image. How to choose a good regularization functional is a key point in imaging science, in order to balance the staircase artifacts and "spots" effects. Recently, Selesnick and Chen proposed the total variation with overlapping group sparsity (OGS-TV) [24–28], which introduces the concept of the group gradient into the TV model and takes into full consideration the dissimilarity between smooth and edge regions. The OGS-TV model can distinguish the individual noise point and image edge point, so it greatly alleviates the staircase effect. Based on this work, Liu et al. applied this method in the removal of speckle noise. Wu and Du applied the OGS model in the field of Magnetic Resonance (MR) image reconstruction [27]. In this paper, we introduce the OGS model into the denoising of impulse noises.

In the typical denoising method, the L1-norm is commonly used as the fidelity term of impulse noise. However, the solution to the L1-norm usually involves the soft-thresholding function, which reduces large values by a constant amount. As a result, noise signals are estimated systematically underestimated for large signal values [26]. To improve this shortcoming, many non-convex reconstruction methods are proposed. Non-convex regularizers have also shown to exhibit a sparser solution than the L1 regularizer [24,29,30]. Inspired by their research, we propose a total variation model based on overlapping group sparsity and Lp-pseudo-norm shrinkage (called OGS-Lp for short). Compared with the L1-norm, the Lp-pseudo-norm adds another degree of freedom to the model, which better characterizes the sparsity features of the image [31].

To solve the problem, the alternating multiplier iterative method (ADMM) [32] and the majorization-minimization (MM) algorithm [33] were used to split the complex problem into several subproblems. Furthermore, an accelerated ADMM with a restart [34] is used to solve the new model (OGS-Lp-FAST for short). In this way, a large amount of spatial-domain calculations are transferred to the frequency domain, which significantly reduces the complexity of the algorithm and speeds up its convergence.

The anisotropic total variation (ATV), isotropic total variation (ITV), total generalized variation (TGV), overlapping group sparsity with L1-norm (OGS-L1), overlapping group sparsity with pseudo-norm (OGS-Lp), and our method are compared experimentally using criteria such as peak signal-to-noise ratio (PSNR), structural similarity (SSIM), and runtime. The model and algorithm proposed here could further improve the image denoising performance.

The contributions of this study are as follows: (1) By introducing OGS into TV, a new regularization was proposed, which incorporated the advantages of TV and OGS models. In the OGS-TV model, the neighborhood difference of each point is fully utilized to augment the difference between the smooth and edge regions. It could balance the staircase artifacts and "spots" effects well. (2) We adopt the Lp-pseudo-norm instead of the L1-norm to describe the fidelity term of impulse noise, extending the L1-norm-based OGS-TV to the OGS-Lp model. (3) The ADMM framework was employed to solve the proposed model. In the ADMM framework, the complex multi-constraints optimization problem is changed to several decoupled subproblems, which are easier to solve. Fourier transform

technology is introduced to transform the matrix operation from the spatial domain to the frequency domain, which avoids large-scale matrix calculations. (4) In order to achieve faster convergence speed, the rapid ADMM with a restart process is adopted to improve the speed of the proposed algorithm. This improved model is named OGS-Lp-FAST. The rate of convergence of the model increases from $O(\frac{1}{k})$ to $O(\frac{1}{k^2})$.

This paper is organized as follows: Section 2 gives a review of the traditional TV model; Section 3 describes the incorporation of overlapping group sparsity and Lp-pseudo-norm shrinkage into the TV model, and uses accelerated ADMM with a restart to solve the new model; Section 4 seeks to validate the proposed algorithm with standard images, and compares it with four other models; and Section 5 summarizes this paper and proposes future work.

2. Traditional TV Model

An image could contain many types of noise. According to the distribution condition of probability density function (PDF) of their amplitude, noises are classified into Gaussian noise, Rayleigh noise, uniform noise, exponential noise, impulse noise, gamma noise, etc.

In this paper, the discussion focuses on impulse noise denoising of images. Impulse noise is additive and is mainly caused by black-and-white bright and dark spots produced by image sensors, transmission channels, decoding processes, etc. In 2004, Nikolova [11] proposed the use of a L1 data-fidelity term for impulse noise related problems. Since then, many research papers adopted this model in their characterization of this type of noise [35–37]. The ATV model of impulse noise based on this model is:

$$F = \underset{F}{\mathrm{argmin}} \|F - G\|_1 + \mu R_{ATV}(F). \tag{1}$$

where $G \in \mathbb{R}^{M \times M}$ is the image with noise and $F \in \mathbb{R}^{M \times M}$ is the denoised image. With $\|\bullet\|_1$ being the L1-norm, the first term in Equation (1), i.e., $\underset{F}{\mathrm{argmin}} \|F - G\|_1$ is called the fidelity term, and the second term $\mu R_{ATV}(F)$ is the sparsity regularization term, which includes the prior sparsity information of the image. μ is a regularization parameter for weighing between the fidelity and regularization terms. The image restoration problem can then be solved by finding the smallest value of F satisfying the conditions, so that Equation (1) is valid. Since the regularization term of the anisotropic total variation model needs to ensure the minimization of both horizontal and vertical gradients, $R_{ATV}(F)$ can be defined as

$$R_{ATV}(F) = \|K_h * F\|_1 + \|K_v * F\|_1 \tag{2}$$

where * represents the convolution, and $K_h = [-1, 1]$ with $K_v = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ are the differential operators used for convolution operations in horizontal and vertical directions, respectively.

3. Proposed Method

3.1. Overlapping Group Sparsity with L1 Norm (OGS-L1) Model

To reduce the staircase effect of the ATV model, Selesnick and Chen proposed the overlapping group sparsity regularization term [5–7] in 2006, which expands the vertical and horizontal gradient of pixels to the group gradient of N adjacent points (N is the size of the group). Setting a reasonable threshold, the individual noise points and image edge points can be distinguished. This model preserves the edge information of the image and mitigates the disadvantages of the staircase effect. With reference to the work of Selesnick and Chen, Liu et al. extended the overlapping group sparsity regularizer from one-dimensional to two-dimensional cases, then substituted it in the anisotropic total variation model for the deconvolution and removal of salt and pepper noise [8].

This model is centered at the pixel $x_{i,j}$ and extends into all directions, forming multiple staggered and overlapped squares. The variable $\widetilde{X}_{i,j,N,N} \in \mathbb{R}^{N \times N}$ is the $N \times N$ pixel matrix, centered at the coordinates (i, j), as shown in Equation (3):

$$\widetilde{X}_{i,j,N,N} = \begin{bmatrix} x_{i-N_l,j-N_l} & x_{i-N_l,j-N_l+1} & \cdots & x_{i-N_l,j+N_r} \\ x_{i-N_l+1,j-N_l} & x_{i-N_l+1,j-N_l+1} & \cdots & x_{i-N_l+1,j+N_r} \\ \vdots & \vdots & \ddots & \vdots \\ x_{i+N_r,j-N_l} & x_{i+N_r,j-N_l+1} & \cdots & x_{i+N_r,j+N_r} \end{bmatrix}, \tag{3}$$

where $N_l = \left\lfloor \frac{N-1}{2} \right\rfloor$, $N_r = \left\lfloor \frac{N}{2} \right\rfloor$. and $\lfloor \ \rfloor$ is the rounding-down operators. We define $\varphi(X)$ to represent the overlapping group sparsity functional of the two-dimensional array:

$$\varphi(X) = \sum_{i=1}^{N} \sum_{j=1}^{N} \left\| \widetilde{X}_{i,j,N,N} \right\|_2. \tag{4}$$

The ATV model can then be extended to embody overlapping group sparsity regularization (called OGS-L1 model for short), as shown in Equation (5):

$$F = \underset{F}{\mathrm{argmin}} \|F - G\|_1 + \mu [\varphi(K_h * F) + \varphi(K_v * F)], \tag{5}$$

where the regularization term $\varphi(X) = \sum\limits_{i=1}^{N} \sum\limits_{j=1}^{N} \left\| \widetilde{X}_{i,j,N,N} \right\|_2$. is the group gradient. Equation (4) shows that the OGS-TV model takes into full consideration the gradient information close to a pixel, so it strengthens the dissimilarity between the smooth and edge regions of the image.

3.2. Overlapping Group Sparsity with Lp-Pseudo-Norm (OSG-Lp) Model

The L1-norm is commonly used as the fidelity term of impulse noise. However, the L1-norm is only the convex relaxation of the L0-norm. The pth power of the Lp-norm ($0 \leq p \leq 1$, for simplicity, we name it Lp-pseudo-norm) is another relaxation of the L0-norm. In fact, the L1-norm constraint is a particular case of the Lp-pseudo-norm.

Recently, the Lp-pseudo-norm [38–42] has attracted much attention in academia. Woodworth and Chartrand pointed out that the Lp-quasinorm is better for approximating the original L0-norm than the L1-norm and developed an iterative Lp-quasinorm shrinkage (LpS) solving the problem [43].

The Lp-pseudo-norm makes an improvement on the sparsity-based shrinkage operator by introducing another degree of freedom, thus giving the model a better ability to depict the sparsity of an image in the gradient domain, as shown in Figure 1.

Lp-pseudo-norm contour lines are given in Figure 1, where $p = 2$ and $p = 1$ represent L2- and L1-norms, respectively. Assuming that the image is contaminated by impulse noises with an absolute difference of τ, Figure 2 shows the schematic plots of anisotropic total variation contour lines $R_{\mathrm{ApTV}}(F) = \|K_h * F\|_p^p + \|K_v * F\|_p^p (0 < p \leq 1)$ intersecting with the fidelity term. As shown in Figure 2, the intersections of the contour lines with the fidelity term are more sparse for $0 < p < 1$, as shown in Figure 2b, than for $p = 1$, as shown in Figure 2a, therefore the robustness of the model against noise is better.

Based on the above analysis, the advantages of Lp-quasinorm regularization are listed as follows: (1) The LpS operator may converge to an accurate solution. (2) The Lp-quasinorm is more flexible than the L1-norm. This might be useful to adapt the degree of sparsity to the signal being processed. (3) The Lp-quasinorm feasible domain makes the solution robust to noise.
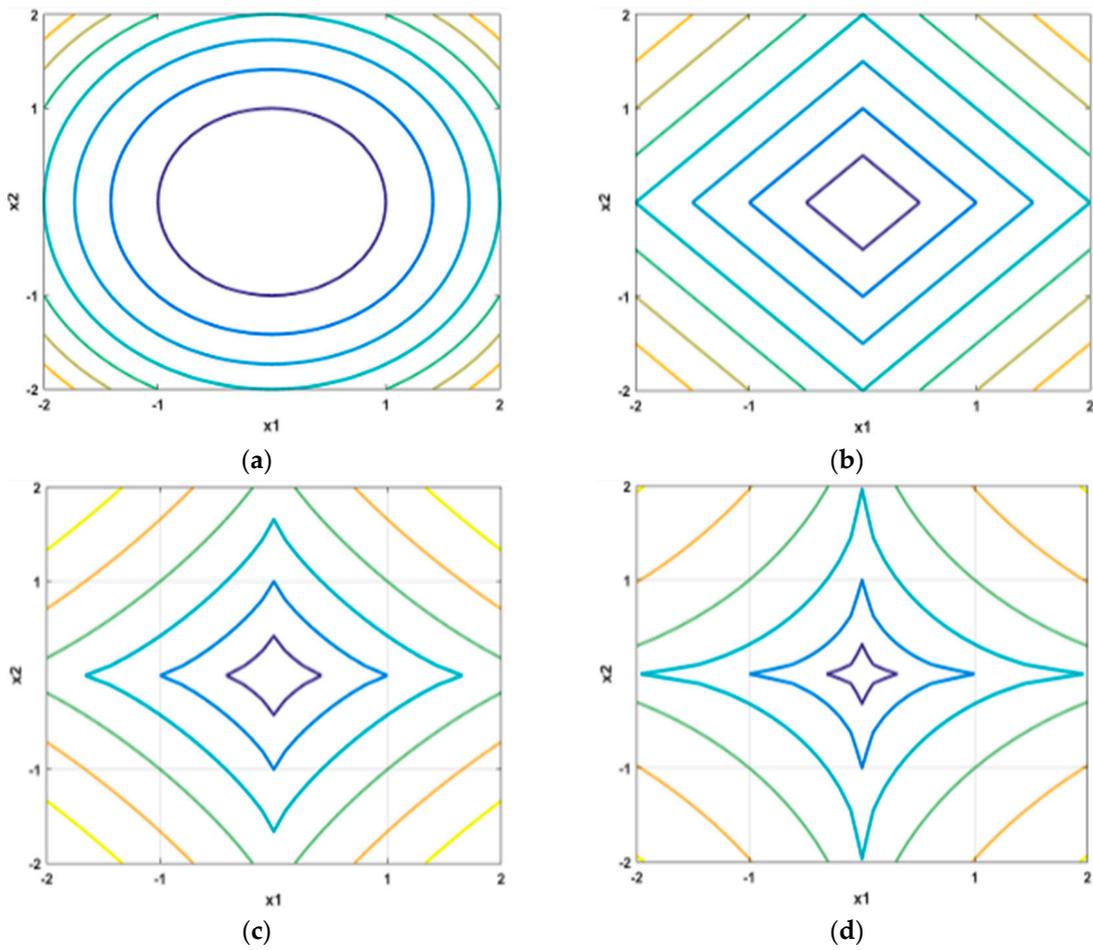
Figure 1. Contour lines of Lp-pseudo-norm: (**a**) $p = 2$; (**b**) $p = 1$; (**c**) $p = 0.8$; (**d**) $p = 0.6$.
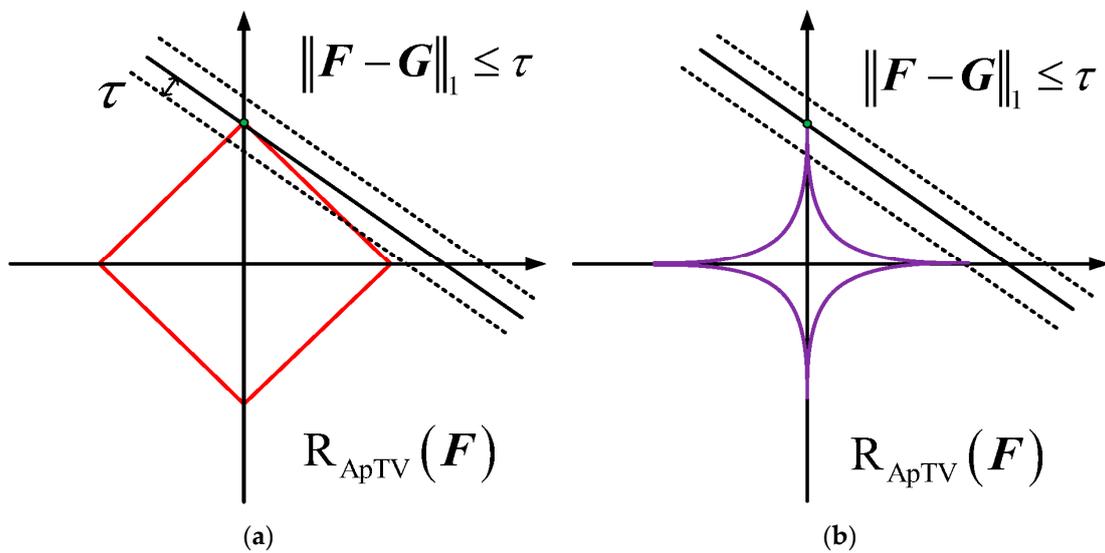
Figure 2. Feasible regions of $\mathrm{R}_{\mathrm{ApTV}}(F)$: (**a**) $p = 1$; (**b**) $0 < p < 1$.

Thus, the L1-norm-based OGS-TV could be extended to the Lp-pseudo-norm (abbreviated as OGS-Lp) [38,39,44] and is expressed as follows:

$$F = \operatorname*{argmin}_{F} \|F - G\|_p^p + \mu[\varphi(K_h * F) + \varphi(K_v * F)]. \tag{6}$$

4. Solution

4.1. Solving the OSG-Lp Model

The OSG-Lp model is treated as a minimization problem whose computation is given below. In real-life images, the values of all pixels usually fall in a limited interval $[a, b]$. For calculation and verification convenience, the image data are normalized so that they are all within $[0, 1]$. The operator P_Ω is first defined with $\Omega = \left\{ F \in \mathbb{R}^{M \times M} \middle| 0 \le F_{i,j} \le 1 \right\}$.

$$\Omega\left(F_{i,j}\right) = \left\{ \begin{array}{ll} 0 \, , & F_{i,j} < 0, \\ F_{i,j} \, , & F_{i,j} \in [0,1] \\ 1 \, , & F_{i,j} > 1, \end{array} \right. . \tag{7}$$

To solve the proposed model, we employed the ADMM framework and changed the complex problem into several subproblems. First, some intermediate variables were introduced to decouple the subproblems. That is, $Z_1 = K_h * F, Z_2 = K_v * F, Z_3 = F - G, Z_4 = F$. then Equation (6) can be reformulated into the following constrained optimization problem:

$$F = \arg\min_{F_{i,j} \in \Omega} \|Z_3\|_p^p + \mu[\varphi(Z_1) + \varphi(Z_2)]$$
$$s.t. \quad Z_1 = K_h * F, Z_2 = K_v * F, Z_3 = F - G, Z_4 = F. \tag{8}$$

According to the principle of the ADMM framework, the Lagrange multipliers and the quadratic penalty term are needed to establish the augmented function. Then we have the following:

$$
\begin{aligned}
J(F, Z_1, Z_2, Z_3, Z_4, \beta_1, \beta_2, \beta_3, \beta_4) &= \|Z_3\|_p^p + \mu[\varphi(Z_1) + \varphi(Z_2)] \\
&- \langle \Lambda_1, (Z_1 - K_h * F) \rangle + \frac{\beta_1}{2}\|Z_1 - K_h * F\|_2^2 \\
&- \langle \Lambda_2, (Z_2 - K_v * F) \rangle + \frac{\beta_2}{2}\|Z_2 - K_v * F\|_2^2, \\
&- \langle \Lambda_3, (Z_3 - F + G) \rangle + \frac{\beta_3}{2}\|Z_3 - F + G\|_2^2 \\
&- \langle \Lambda_4, (Z_4 - F) \rangle + \frac{\beta_4}{2}\|Z_4 - F\|_2^2
\end{aligned}
\tag{9}
$$

where $\Lambda_1, \Lambda_2, \Lambda_3, \Lambda_4$ are the Lagrange multipliers, and β_1, β_2, β_3, $\beta_4 > 0$ are the penalty coefficients. $\langle A, B \rangle$ is the inner-production operator of the matrixes A and B.

Introducing the scaled Lagrange multipliers $\widetilde{Z}_1, \widetilde{Z}_2, \widetilde{Z}_3, \widetilde{Z}_4$, which are also called dual variables. Defined $\widetilde{Z}_i (i = 1, 2, 3, 4)$ as following $\widetilde{Z}_1 = \frac{1}{\beta_1}\Lambda_1, \widetilde{Z}_2 = \frac{1}{\beta_2}\Lambda_2, \widetilde{Z}_3 = \frac{1}{\beta_3}\Lambda_3, \widetilde{Z}_4 = \frac{1}{\beta_4}\Lambda_4$. Add the term $\frac{\beta_i}{2}\left(\widetilde{Z}_i\right)^2 - \frac{\beta_i}{2}\left(\widetilde{Z}_i\right)^2 = 0 (i = 1, 2, 3, 4)$ to Equation (9) to complete the formula item. Equation (10) is obtained after rearrangement.

$$
\begin{aligned}
&J(F, Z_1, Z_2, Z_3, Z_4, \beta_1, \beta_2, \beta_3, \beta_4) \\
&= \|Z_3\|_p^p + \mu[\varphi(Z_1) + \varphi(Z_2)] \\
&+ \frac{\beta_1}{2}\|Z_1 - K_h * F\|_2^2 - \beta_1\left\langle \widetilde{Z}_1, Z_1 - K_h * F \right\rangle + \frac{\beta_1}{2}\left(\widetilde{Z}_1\right)^2 - \frac{\beta_1}{2}\left(\widetilde{Z}_1\right)^2 \\
&+ \frac{\beta_2}{2}\|Z_2 - K_v * F\|_2^2 - \beta_2\left\langle \widetilde{Z}_2, Z_2 - K_v * F \right\rangle + \frac{\beta_2}{2}\left(\widetilde{Z}_2\right)^2 - \frac{\beta_2}{2}\left(\widetilde{Z}_2\right)^2 \\
&+ \frac{\beta_3}{2}\|Z_3 - F + G\|_2^2 - \beta_3\left\langle \widetilde{Z}_3, Z_3 - F + G \right\rangle + \frac{\beta_3}{2}\left(\widetilde{Z}_3\right)^2 - \frac{\beta_3}{2}\left(\widetilde{Z}_3\right)^2 \\
&+ \frac{\beta_4}{2}\|Z_4 - F\|_2^2 - \beta_4\left\langle \widetilde{Z}_4, Z_4 - F \right\rangle + \frac{\beta_4}{2}\left(\widetilde{Z}_4\right)^2 - \frac{\beta_4}{2}\left(\widetilde{Z}_4\right)^2 .
\end{aligned}
\tag{10}
$$

In Equation (10), the expressions

$$\frac{\beta_i}{2}\|Z_i - A\|_2 - \beta_i\left\langle \widetilde{Z}_i, Z_i - A \right\rangle + \frac{\beta_i}{2}\left(\widetilde{Z}_i\right)^2 \quad (i = 1, 2, 3, 4) \ , \tag{11}$$

satisfy the form of $a^2 - 2ab + b^2 = (a-b)^2$, so Equation (10) can be written as Equation (12):

$$
\begin{aligned}
& J(F, Z_1, Z_2, Z_3, Z_4, \beta_1, \beta_2, \beta_3, \beta_4) \\
& = \|Z_3\|_p^p + \mu\varphi(Z_1) + \mu\varphi(Z_2) + \frac{\beta_1}{2}\|Z_1 - K_h * F - \widetilde{Z}_1\|_2^2 \\
& + \frac{\beta_2}{2}\|Z_2 - K_v * F - \widetilde{Z}_2\|_2^2 + \frac{\beta_3}{2}\|Z_3 - F + G - \widetilde{Z}_3\|_2^2 \\
& + \frac{\beta_4}{2}\|Z_4 - F - \widetilde{Z}_4\|_2^2 - \frac{\beta_1}{2}\left(\widetilde{Z}_1\right)^2 - \frac{\beta_2}{2}\left(\widetilde{Z}_2\right)^2 - \frac{\beta_3}{2}\left(\widetilde{Z}_3\right)^2 - \frac{\beta_4}{2}\left(\widetilde{Z}_4\right)^2.
\end{aligned}
\tag{12}
$$

Because $Z_1, Z_2, Z_3, Z_4, \widetilde{Z}_1, \widetilde{Z}_2, \widetilde{Z}_3, \widetilde{Z}_4$ are mutually independent, so, they can be solved as some independent subproblems according to the principle of the ADMM algorithm. These problems can be solved by the iterative algorithm for minimizing Z_i.

Defining $Z_i^{(k)}$ $(i = 1, 2, 3, 4)$ represents the value of Z_i after $(k) - th$ iterations. For a given $Z_i^{(k)}$, the next iteration $Z_i^{(k+1)}$ is generated as follows:

$$
Z_1^{(k+1)} = \underset{Z_1}{\operatorname{argmin}}\mu\varphi(Z_1) + \frac{\beta_1}{2}\|Z_1 - K_h * F^{(k)} - \widetilde{Z}_1^{(k)}\|_2^2.
\tag{13}
$$

$$
Z_2^{(k+1)} = \underset{Z_2}{\operatorname{argmin}}\mu\varphi(Z_2) + \frac{\beta_2}{2}\|Z_2 - K_v * F^{(k)} - \widetilde{Z}_2^{(k)}\|_2^2.
\tag{14}
$$

$$
Z_3^{(k+1)} = \underset{Z_3}{\operatorname{argmin}}\|Z_3\|_p^p + \frac{\beta_3}{2}\|Z_3 - F^{(k)} + G - \widetilde{Z}_3^{(k)}\|_2^2.
\tag{15}
$$

$$
Z_4^{(k+1)} = \underset{Z_4}{\operatorname{argmin}}\frac{\beta_4}{2}\|Z_4 - F^{(k)} - \widetilde{Z}_4^{(k)}\|_2^2.
\tag{16}
$$

Each of these four equations are solved below:

(1) The $Z_1^{(k+1)}$ and the $Z_2^{(k+1)}$ are solved by the majorization-minimization (MM) algorithm [33], which approximates the solution of the target problem by finding a well-behaving multi-variable auxiliary function and constructing an iterative sequence.

First, suppose a minimization optimization problem with the following form:

$$
\underset{v}{\min}P(v) = \left\{\frac{\alpha}{2}\|v - v_0\|_2^2 + \varphi(v)\right\}, v \in \mathbb{R}^{M^2 \times 1},
\tag{17}
$$

where $\alpha > 0$ and $\varphi(v)$ satisfies $\varphi(v) = \sum_{i=1}^{N}\sum_{j=1}^{N}\|\widetilde{v}_{i,j,N,N}\|_2$. To avoid having to solve the complex minimization problem $P(v)$ directly, a function $Q(v, u)$ for which $Q(v, u) \geq P(v)$ for all v, u could be constructed. The equality sign holds if and only if $u = v$. Thus, the optimal solution of $P(v)$ is the minimum value of $Q(v, u)$. Generally, an MM iterative algorithm for minimizing $P(v)$ has the form:

$$
v^{(n+1)} = \underset{v}{\operatorname{argmin}}Q\left(v, v^{(n)}\right),
\tag{18}
$$

It can be solved step-by-step in the following way.

The properties of the function $\varphi(v) = \sum_{i=1}^{N}\sum_{j=1}^{N}\|\widetilde{v}_{i,j,N,N}\|_2$ are first observed. It is known that the equality sign holds when $u = v$, as shown in Equation (19):

$$
\frac{1}{2}\left(\frac{1}{\|u\|_2}\|v\|_2^2 + \|u\|_2\right) \geq \|v\|_2,
\tag{19}
$$

which means $\varphi(v)$ can be solved by constructing the function $S(v)$ of Equation (20):

$$S(v,u) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left[\frac{1}{\|\widetilde{u}_{i,j,N,N}\|_2}\|\widetilde{v}_{i,j,N,N}\|_2^2 + \|\widetilde{u}_{i,j,N,N}\|_2\right] \geq \varphi(v) = \sum_{i=1}^{N}\sum_{j=1}^{N}\|\widetilde{v}_{i,j,N,N}\|_2 , \qquad (20)$$

After a simple calculation [31], $S(v,u)$ is rewritten as

$$S(v,u) = \frac{1}{2}\|\mathbf{D}(u)v\|_2^2 + C(u), \qquad (21)$$

to facilitate future calculations.

$C(u)$ in the above equation is independent of v, and $\mathbf{D}(u) \in \mathbb{R}^{M^2 \times M^2}$ is a diagonal matrix with its elements defined as

$$[\mathbf{D}(u)]_{m,m} = \sqrt{\sum_{i=-N_l}^{N_r}\sum_{j=-N_l}^{N_r}\left[\sum_{k_1=-N_l}^{N_r}\sum_{k_2=-N_l}^{N_r}|u_{m-i+k_1,m-i+k_2}|^2\right]^{-\frac{1}{2}}} \qquad \left(m = 1,2,\cdots,M^2\right), \qquad (22)$$

The entries of $\mathbf{D}$ can be easily computed by using MATLAB built-in function "conv2". Putting Equations (17), (21), and (22) together, the optimization problem $P(v)$ can be written as

$$Q(v,u) = \frac{\alpha}{2}\|v - v_0\|_2^2 + S(v,u) = \frac{\alpha}{2}\|v - v_0\|_2^2 + \frac{1}{2}\|\mathbf{D}(u)v\|_2^2 + C(u). \qquad (23)$$

When $v = u$, $Q(u,u) = P(u)$, to minimize $P(v)$, the MM aims to iteratively solve

$$v^{(n+1)} = \underset{v}{\arg\min}\,\frac{\alpha}{2}\|v - v_0\|_2^2 + \frac{1}{2}\left\|\mathbf{D}\left(v^{(n)}\right)v\right\|_2^2, n = 1,2,3\ldots \qquad (24)$$

with the solution

$$v^{(n+1)} = \left(\mathbf{I} + \frac{1}{\alpha}\mathbf{D}^2\left(v^{(n)}\right)\right)^{-1}v_0, n = 1,2,3\ldots, \qquad (25)$$

where $\mathbf{I} \in \mathbb{R}^{M^2 \times M^2}$ is an identity matrix with the same size of $\mathbf{D}^2\left(v^{(n)}\right)$, $\mathbf{D}^2\left(v^{(n)}\right)$ is also a diagonal matrix which has the same form of Equation (22).

Observing functions of Equations (13) and (14), the subproblem $\mathbf{Z}_1, \mathbf{Z}_2$ conforms to the function in Equation (17) and can be solved iteratively using Equation (26). $\mathbf{Z}_{i(n+1)}^{(k+1)}$ $(i = 1,2)$ represents the (n)-th iteration of the MM algorithm in the $(k+1)$-th outer loop.

$$\mathbf{Z}_{i(n+1)}^{(k+1)} = mat\left\{\left[\mathbf{I} + \frac{\mu}{\beta_i}\mathbf{D}^2\left(\mathbf{Z}_{i(n)}^{(k+1)}\right)\right]^{-1}\mathfrak{z}_{i(0)}^{(k+1)}\right\}, (i = 1,2), \qquad (26)$$

where ***mat*** plays the role of reshaping a vector to a matrix and $\mathfrak{z}_{i(0)}^{(k+1)}$ is the vector form of $\mathbf{Z}_{i(0)}^{(k+1)}$.

The initial values of $\mathbf{Z}_{1(0)}^{(k+1)}, \mathbf{Z}_{2(0)}^{(k+1)}$ in the above equation are:

$$\begin{cases} \mathbf{Z}_{1(0)}^{(k+1)} = \mathbf{K}_h * \mathbf{F}^{(k)} + \widetilde{\mathbf{Z}}_1^{(k)} \\ \mathbf{Z}_{2(0)}^{(k+1)} = \mathbf{K}_v * \mathbf{F}^{(k)} + \widetilde{\mathbf{Z}}_2^{(k)} \end{cases}. \qquad (27)$$

(2) $Z_3^{(k+1)}$ could be solved by the famous soft thresholding shrinkage method [45] as shown below:

$$
\begin{aligned}
Z_3^{(k+1)} &= \operatorname*{argmin}_{Z_3} \|Z_3\|_p^p + \tfrac{\beta_3}{2}\|Z_3 - F^{(k)} + G - \widetilde{Z}_3^{(k)}\|_2^2 \\
&= shrinkage(F^{(k)} - G + \widetilde{Z}_3^{(k)}, \tfrac{1}{\beta_3}, p) \\
&= \operatorname{sign}\!\left(F^{(k)} - G + \widetilde{Z}_3^{(k)}\right) \\
&. * \max\!\left(\operatorname{abs}\!\left(F^{(k)} - G + \widetilde{Z}_3^{(k)}\right) - \left(\tfrac{1}{\beta_3}\right).\hat{\ }(p-2).*\operatorname{abs}\!\left(F^{(k)} - G + \widetilde{Z}_3^{(k)}\right).\hat{\ }(p-1), 0\right).
\end{aligned}
\tag{28}
$$

(3) The $Z_4^{(k+1)}$ subproblem is computed as

$$
Z_4^{(k+1)} = \operatorname*{argmin}_{Z_4} \frac{\beta_4}{2}\|Z_4 - F^{(k)} - \widetilde{Z}_4^{(k)}\|_2^2 = \min\!\left(1, \max\!\left(F^{(k)} + \widetilde{Z}_4^{(k)}, 0\right)\right),.
\tag{29}
$$

(4) The $F^{(k+1)}$ subproblem is solved by substituting $Z_1^{(k+1)}, Z_2^{(k+1)}, Z_3^{(k+1)}, Z_4^{(k+1)}$ into Equation (12) and calculating for the variable $F^{(k+1)}$. Under the assumption of periodic boundary conditions, fast Fourier transform is applied to both sides of the equation to perform the computation in the frequency domain instead of a spatial domain, in order to reduce the computational complexity caused by matrix multiplication. Matrix multiplication is converted to dot product operation, in other words, solving the following normal equation:

$$
\begin{aligned}
&\left(\beta_1 \overline{K}_h^*.*\overline{K}_h + \beta_2 \overline{K}_v^*.*\overline{K}_v + \beta_3 \mathbf{1} + \beta_4 \mathbf{1}\right).*\overline{F}^{(k+1)} \\
&= \beta_1 \overline{K}_h^*.*\left(\overline{Z}_1^{(k+1)} - \overline{\widetilde{Z}}_1^{(k)}\right) + \beta_2 \overline{K}_v^*.*\left(\overline{Z}_2^{(k+1)} - \overline{\widetilde{Z}}_2^{(k)}\right) \\
&+ \beta_3 \left(\overline{Z}_3^{(k+1)} + \overline{G} - \overline{\widetilde{Z}}_3^{(k)}\right) + \beta_4 \left(\overline{Z}_4^{(k+1)} - \overline{\widetilde{Z}}_4^{(k)}\right),
\end{aligned}
\tag{30}
$$

where $\overline{x}$ is the frequency-domain representation of x, ".$*$" stands for dot product, "$*$" is the conjugate, $\mathbf{1}$ is the matrix whose entries are all 1, and $\mathcal{F}$ represents the two-dimensional fast Fourier transform (FFT). Rearranging of the equation gives $F^{(k+1)}$ as

$$
\begin{aligned}
&F^{(k+1)} = \\
&\mathcal{F}^{-1}\!\left(\frac{\beta_1 \overline{K}_h^*.*\left(\overline{Z}_1^{(k+1)} - \overline{\widetilde{Z}}_1^{(k)}\right) + \beta_2 \overline{K}_v^*.*\left(\overline{Z}_2^{(k+1)} - \overline{\widetilde{Z}}_2^{(k)}\right) + \beta_3\left(\overline{Z}_3^{(k+1)} + \overline{G} - \overline{\widetilde{Z}}_3^{(k)}\right) + \beta_4\left(\overline{Z}_4^{(k+1)} - \overline{\widetilde{Z}}_4^{(k)}\right)}{\beta_1 \overline{K}_h^* \circ \overline{K}_h + \beta_2 \overline{K}_v^* \circ \overline{K}_v + \beta_3 \mathbf{1} + \beta_4 \mathbf{1}}\right).
\end{aligned}
\tag{31}
$$

(5) The dual variables $\widetilde{Z}_1, \widetilde{Z}_2, \widetilde{Z}_3, \widetilde{Z}_4$ could be updated via the gradient ascent method.

$$
\begin{cases}
\widetilde{Z}_1^{(k+1)} = \widetilde{Z}_1^{(k)} + \gamma\beta_1(K_h * F^{(k+1)} - Z_1^{(k+1)}) \\
\widetilde{Z}_2^{(k+1)} = \widetilde{Z}_2^{(k)} + \gamma\beta_2(K_v * F^{(k+1)} - Z_2^{(k+1)}) \\
\widetilde{Z}_3^{(k+1)} = \widetilde{Z}_3^{(k)} + \gamma\beta_3(F^{(k+1)} - G - Z_3^{(k+1)}) \\
\widetilde{Z}_4^{(k+1)} = \widetilde{Z}_4^{(k)} + \gamma\beta_4(F^{(k+1)} - Z_4^{(k+1)})
\end{cases}.
\tag{32}
$$

4.2. OGS-Lp-FAST Model

The denoising models based on OGS is more time consuming than the TV-based model. This is mainly because that OGS model considers the gradient information of the neighborhood in a reconstructed image, thus making the computation more complex. Thus, this is a shortcoming.

Goldstein et al. [12] proposed an accelerated ADMM algorithm with a restart that improves the convergence rate of the ADMM algorithm from $O(\frac{1}{k})$ to $O(\frac{1}{k^2})$. Inspired by them, we adopted this algorithm to improve the OGS-Lp model. This modified model is named OGS-Lp-FAST ("Ours" for

short). The auxiliary variables $U_i (i = 1, 2)$ and $\widetilde{U}_i (i = 1, 2)$ are first adopted. Under this framework, $Z_i (i = 1, 2, 3, 4)$ is updated in the following way:

$$
\begin{cases}
Z_{1(n+1)}^{(k+1)} = mat\left\{ \left[I + \frac{\mu}{\beta_1} D^2 \left(v^{(n)} \right)^2 \left(Z_{1(n+1)}^{(k+1)} \right) \right]^{-1} \mathfrak{z}_{1(0)}^{(k+1)} \right\} \\
Z_{2(n+1)}^{(k+1)} = mat\left\{ \left[I + \frac{\mu}{\beta_2} D^2 \left(v^{(n)} \right) \left(Z_{2(n+1)}^{(k+1)} \right) \right]^{-1} \mathfrak{z}_{2(0)}^{(k+1)} \right\} \\
Z_3^{(k+1)} = shrinkage(F^{(k)} - G + \widetilde{Z}_3^{(k)}, \frac{1}{\beta_3}, p) \\
Z_4^{(k+1)} = min\left(1, max\left(F^{(k)} + \widetilde{Z}_4^{(k)}, 0 \right) \right),
\end{cases}
\tag{33}
$$

The initial values of $Z_{1(0)}^{(k+1)}, Z_{2(0)}^{(k+1)}$ in the above equation are:

$$
\begin{cases}
Z_{1(0)}^{(k+1)} = K_h * F^{(k)} + \widetilde{U}_1^{(k)} \\
Z_{2(0)}^{(k+1)} = K_v * F^{(k)} + \widetilde{U}_2^{(k)}
\end{cases}
\tag{34}
$$

The dual variable $\widetilde{Z}_i (i = 1, 2)$ can be updated as follows:

$$
\begin{cases}
\widetilde{Z}_1^{(k+1)} = \widetilde{U}_1^{(k)} + \gamma \beta_1 \left(K_h * F^{(k)} - Z_1^{(k+1)} \right) \\
\widetilde{Z}_2^{(k+1)} = \widetilde{U}_2^{(k)} + \gamma \beta_2 \left(K_v * F^{(k)} - Z_2^{(k+1)} \right) \\
\widetilde{Z}_3^{(k+1)} = \widetilde{Z}_3^{(k)} + \gamma \beta_3 \left(F^{(k)} - G - Z_3^{(k+1)} \right) \\
\widetilde{Z}_4^{(k+1)} = \widetilde{Z}_4^{(k)} + \gamma \beta_4 \left(F^{(k)} - Z_4^{(k+1)} \right)
\end{cases}
\tag{35}
$$

As image denoising is not a strong convex problem, the iteration needs to be restarted to ensure the convergence of the algorithm. When Equation (33) is not satisfied, the algorithm is restarted.

$$
c_i^{(k)} < \eta c_i^{(k-1)} (i = 1, 2),
\tag{36}
$$

where $c_i^{(k)} = \beta^{-1} \| \widetilde{Z}_i^{(k)} - \widetilde{U}_i^{(k)} \|_2^2 + \beta \| Z_i^{(k)} - U_i^{(k)} \|_2^2$ is the sum of the $(k) - th$ primal residuals $\beta \| Z_i^{(k)} - U_i^{(k)} \|_2^2$ and the dual residuals $\beta^{-1} \| \widetilde{Z}_i^{(k)} - \widetilde{U}_i^{(k)} \|_2^2$, and η is a number close to 1. To prevent frequent restarts, $\eta = 0.97$ is set.

When $c_i^{(k)} < \eta c_i^{(k-1)}$, the acceleration step size is set to ε_i, and the auxiliary variables $U_i (i = 1, 2)$ and $\widetilde{U}_i^{(k+1)} (i = 1, 2)$ are updated as follows:

$$
\begin{cases}
\varepsilon_i^{(k+1)} = \frac{1 + \sqrt{1 + 4(\varepsilon_i^{(k)})^2}}{2} (i = 1, 2) \\
U_i^{(k+1)} = Z_i^{(k+1)} + \frac{\varepsilon_i^{(k)} - 1}{\varepsilon_i^{(k+1)}} (Z_i^{(k+1)} - Z_i^{(k)})(i = 1, 2) \\
\widetilde{U}_i^{(k+1)} = \widetilde{Z}_i^{(k+1)} + \frac{\varepsilon_i^{(k)} - 1}{\varepsilon_i^{(k+1)}} (\widetilde{Z}_i^{(k+1)} - \widetilde{Z}_i^{(k)})(i = 1, 2)
\end{cases}
\tag{37}
$$

which is updated according to the equations below upon restart:

$$
\begin{cases}
\varepsilon_i^{(k+1)} = 1 (i = 1, 2) \\
U_i^{(k+1)} = Z_i^{(k+1)} (i = 1, 2) \\
\widetilde{U}_i^{(k+1)} = \widetilde{Z}_i^{(k+1)} (i = 1, 2) \\
c_i^{(k+1)} = \eta^{-1} c_i^{(k)} (i = 1, 2)
\end{cases}
\tag{38}
$$

Up to this point, all subproblems of the proposed model are solved. The OGS-Lp-FAST algorithm just described is summarized as Algorithm 1.

The *F* sub-problem should be updated as

$$F^{(k+1)} =$$
$$\mathcal{F}^{-1}\left(\frac{\beta_1\overline{K}_h^*.*\left(\overline{U}_1^{(k+1)}-\overline{\widetilde{U}}_1^{(k+1)}\right)+\beta_2\overline{K}_v^*.*\left(\overline{U}_2^{(k+1)}-\overline{\widetilde{U}}_2^{(k+1)}\right)+\beta_3\left(\overline{Z}_3^{(k+1)}+\overline{G}-\overline{\widetilde{Z}}_3^{(k+1)}\right)+\beta_4\left(\overline{Z}_4^{(k+1)}-\overline{\widetilde{Z}}_4^{(k+1)}\right)}{\beta_1\overline{K}_h^*\circ\overline{K}_h+\beta_2\overline{K}_v^*\circ\overline{K}_v+\beta_3\mathbf{1}+\beta_4\mathbf{1}}\right). \tag{39}$$

Algorithm 1 OGS-Lp-FAST pseudo-code

Input: image **G** with noise
Output: denoised image **F**
Initialize:
　　$k = 1, n = 0, Z_i^{(k)} = 0, \widetilde{Z}_i^{(k)} = 0 (i = 1, 2, \cdots, 4), \beta_1, \beta_2, \beta_3, \beta_4, \mu, p, \eta,$
　　　　$U_j^{(k)} = 0, \widetilde{U}_j^{(k)} = 0 (j = 1, 2), \gamma, tol, c^1 = inf, F^{(k)} = G, Max$
1: for $k = 1 : Max$
2: If : $c_i^{(k+1)} < \eta c_i^{(k)}$:
3: $Z_i^{(k+1)}, (i = 1, 2, 3, 4)$, are updated with Equations **(33)** and **(34)**
4: $\widetilde{Z}_i^{(k+1)} (i = 1, 2, 3, 4)$ are updated with Equation **(35)**
5: $\varepsilon_i^{(k+1)}, U_j^{(k+1)}, \widetilde{U}_j^{(k+1)} (j = 1, 2)$ is updated with Equation **(37)**
6: $F^{(k+1)}$ are updated with Equation **(39)**
7: $E = \|F^{(k+1)} - F^{(k)}\|_2 / \|F^{(k)}\|_2$
8 Else
9: Restart as in Equation **(38)**
10: End If
11: If $E < tol$ Break;
12: End For
13: Return $F^{(k)}$ *as* **F**

---

5. Experimental Results and Analyses

In this section, eight typical grayscale images with a size of 256×256 pixels are chosen to validate the denoising performance of the Lp-OSG-TV-FAST method. The test images are as shown in Figure 3. The image "House" is downloaded from http://sipi.usc.edu/database/database.php? volume%92=%92misc&image=5top. The images "Lena" and "Pepper" are from http://decsai.ugr. es/cvg/dbimagenes/. The images "Woman", "Girl", and "Reagan" are from http://www.hlevkin. com/default.html#testimages. The images "Milk drop" and "Shoulder" are from http://www.cs.cmu. edu/~cil/v-images.html. The versions of the images in our paper are other special formats which are converted by Photoshop from the sources above.

The method proposed here is compared with ATV, ITV, TGV, OSG-L1, and OGS-Lp methods, and is evaluated objectively in terms of the peak signal-to-noise ratio (PSNR), structural similarity (SSIM), runtime, and other experimental indicators [45]. Simulations are performed on MATLAB R2014a platform running in a hardware environment of Inter(R) Core (TM) i7-6700@3.4GHz CPU and 16 GB memory.
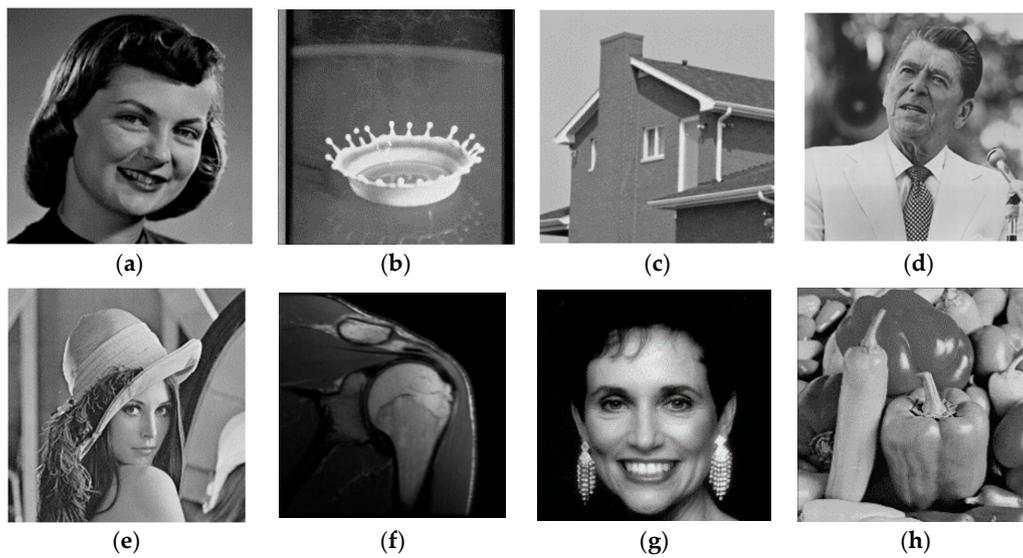
Figure 3. Original images. Top row: from left to right, (**a**) Girl.bmp, (**b**) Milk drop.tiff, (**c**) House.png, (**d**) Reagan.bmp, (**e**) Lena.png, (**f**) Shoulder.jpg, (**g**) Woman.tif, and (**h**) Pepper.tiff.

5.1. Evaluation Method

In the denoising field, the common evaluating criteria including PSNR, SSIM, and runtime. The PSNR and SSIM [46] are defined in Equations (40) and (41):

$$PSNR(\boldsymbol{X}, \boldsymbol{Y}) = 10\lg \frac{(MAX(\boldsymbol{X}))^2}{\frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} (\boldsymbol{X}_{ij} - \boldsymbol{Y}_{ij})^2}, \tag{40}$$

where $\boldsymbol{X}$ denotes the original image, $\boldsymbol{Y}$ is the reconstructed image, and $MAX(\boldsymbol{X})$ represents the largest gray value in the original image.

$$SSIM(\boldsymbol{X}, \boldsymbol{Y}) = \frac{(2u_X u_Y + (Lk_1)^2)(2\sigma_{XY} + (Lk_2)^2)}{(u_X^2 + u_Y^2 + (Lk_1)^2)(\sigma_X^2 + \sigma_Y^2 + (Lk_2)^2)}, \tag{41}$$

where u_X is the mean of $\boldsymbol{X}$; u_Y is the mean of $\boldsymbol{Y}$; σ_X^2 is the variance of $\boldsymbol{X}$; σ_Y^2 is the variance of $\boldsymbol{Y}$; σ_{XY} is the covariance between $\boldsymbol{X}$ and $\boldsymbol{Y}$; and $L = 512$, $k_1 = 0.05$, $k_2 = 0.05$. The parameter $L = 255$.

5.2. Sensitivity of the Parameters

In this section, an important parameter of the proposed algorithm, the group size K, is tested and compared to evaluate its overall effect on the algorithm. PSNR and SSIM are used as criteria to evaluate the algorithm objectively. Three images ("Girl", "House", and "Lena") with 30% noise level are selected, on which K is made to vary continuously from 1 to 10. The other parameters are adjusted to the optimum. PSNR and SSIM values are recorded and plotted into graphs, as shown in Figures 4 and 5. In Figure 4, PSNR and SSIM increase with the increase of K and reach their maximum at K = 5. Further increases in K leads to decreased PSNR values. Thus, the neighborhood information of an image has a positive impact on the performance of the algorithm. With K set to appropriate values, the edge information of the image is better preserved, and the noise resistance is improved. However, K should not be too large either, or nearby regions with drastic pixel changes could be taken, which will result in decreased PSNR and SSIM.

Then, we tested how to select a good regularization parameter μ for different images. We started with the value of μ from a low value then increased the parameters empirically as the level of noise

improved to get the best visual effect. For example, for the "Girl" image corrupted by impulse noise from 20% to 50%, μ = 0.14, 0.15, 0.15, 0.18, respectively.
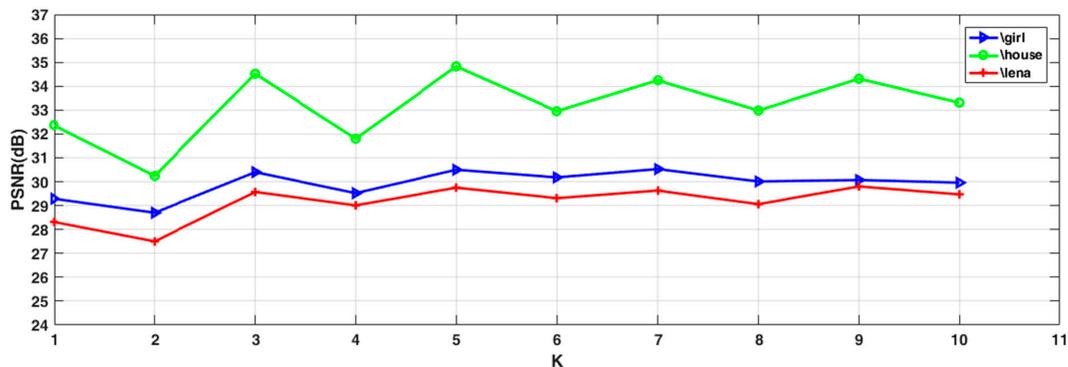

Figure 4. Comparison of peak signal-to-noise ratio (PSNR) under different K values (Noise Level = 30%). The test images ("Girl", "House", "Lena") are corrupted by 30% impulse noise.
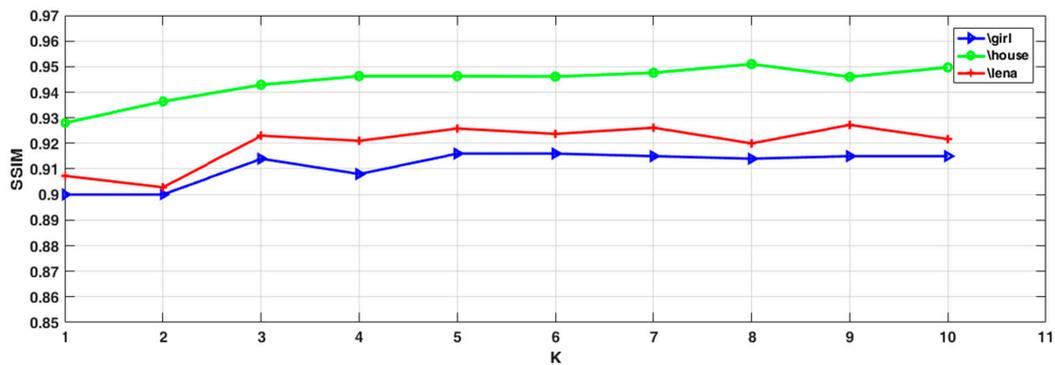

Figure 5. Comparison of structural similarity (SSIM) under different K values (Noise Level = 30%). The test images ("Girl", "House", "Lena") are corrupted by 30% impulse noise.

In the selection of the parameter p, the p value is set between 0 and 1. On the premise of fixing other parameters, we increase p by 0.1 step. After several rounds of experiments, we select the optimal when the image gets the best visual effect.

The optimal parameters for different images with the noise level from 20% to 50% are given in Table 1.

Table 1. The optimal parameters for different images with the noise level from 20% to 50%.

Image	House	Lena	Woman	Milk Drop	Girl	Shoulder
Parameter	μ/p	μ/p	μ/p	μ/p	μ/p	μ/p
Level						
20%	0.14/0.45	0.15/0.45	0.15/0.65	0.14/0.65	0.14/0.65	0.14/0.65
30%	0.15/0.45	0.15/0.65	0.15/0.7	0.16/0.65	0.15/0.65	0.15/0.65
40%	0.15/0.45	0.16/0.65	0.15/0.7	0.19/0.65	0.15/0.65	0.15/0.65
50%	0.18/0.55	0.17/0.65	0.2/0.7	0.2/0.65	0.18/0.55	0.18/0.55

5.3. Testing and Comparing the Denoising Performance of Different Algorithms

Six images are selected from the original images of Figure 3 for testing, on which impulse noise at levels from 20% to 50%, to compare the denoising effects of ATV, ITV, TGV, OGS-L1, OGS-Lp, and Ours algorithms (six in total). To ensure the objectiveness and fairness of the evaluation, the above algorithms all adopt the following iterative condition:

$$\|\boldsymbol{F}^{(k+1)} - \boldsymbol{F}^{(k)}\|_2 \cdot \|\boldsymbol{F}^{(k)}\|_2^{-1} < 10^{-4}. \tag{42}$$

Regularization parameters of these algorithms are adjusted to ensure the best denoising effect of each, which ensured the fairness of the test. For methods based on the OGS model, the group size is set to K = 5. The test results on different images are given in Tables 2–5. The best indicator values are labeled as black and bold. By observing the data in each table, the following conclusions could be made:

1. With the introduction of different levels of noise to the images, our model generates higher PSNR and SSIM values for the reconstructed images than other methods, indicating its superior denoising effect. The recovered images also resemble the original ones more.
2. The proposed model works better at lower noise levels. For example, at a 20% noise level, as shown in Table 2, the PSNR value of the "House" image (37.72 dB) given by our model is 5.91 dB higher than that given by the ITV model (31.81 dB) and 5.4 dB higher than that of the TGV model (32.32 dB). Even at high noise levels, our model still performs better than the others, which shows the clear advantages that total variation with overlapping group sparsity has over the classic anisotropic TV model.
3. Compared to OGS-L1, our proposed method incorporates the Lp-pseudo-norm shrinkage, which adds another degree of freedom to the algorithm and improves the depiction of the gradient-domain sparsity of the images, achieving a better denoising effect. For example, at a 20% noise level, as shown in Table 2, the PSNR value of the "Girl" image (32.34 dB) given by our model is 1.67 dB higher than that given by the OGS-L1 model (30.67 dB). Even at a noise level of 50%, as shown in Table 5, the PSNR value of the "Girl" image (27.35 dB) given by our model is still 0.90 dB higher than that given by the OGS-L1 model (26.45 dB). This proves that the Lp-pseudo-norm is more suitable as a regularizer for describing the sparsity of images than the L1-norm.
4. In terms of the runtime of the six models, the OGS-based method is more time consuming than ATV, ITV, and TGV. This is mainly because the OGS model considers the gradient information of the neighborhood in an image undergoing reconstruction, thus making the computation more complex.
5. Comparing the values of PSNR and SSIM in Tables 2–5, OGS-Lp-FAST and OGS-Lp have the same denoising effect. However, by observing the value of runtime of all testing images, we find that convergence is sped up in the OGS-LP-FAST method with the use of accelerated ADMM with a restart. For example, at a 20% noise level, as shown in Table 2, the time value of the "Woman" image (8.69 s) given by the OGS-Lp-FAST model is 7.53 s less than that given by the OGS-L1 model (16.22 s).

Table 2. Numerical comparison of our proposed method and other models (images are corrupted by impulse noise of 20%). ATV: anisotropic total variation; ITV: isotropic total variation; TGV: total generalized variation; OGS-L1: overlapping group sparsity with L1-norm; OGS-Lp: overlapping group sparsity with pseudo-norm.

Level	Image	Method	The Output Seismic Signal		
			PSNR (dB)	SSIM	Time (s)
20%	Lena	ATV	28.71	0.8854	4.81
		ITV	28.83	0.8936	**2.45**
		TGV	28.63	0.8966	9.59
		OGS-L1	29.79	0.9115	16.34
		OGS-Lp	31.47	0.9474	13.94
		OGS-Lp-FAST	**31.55**	**0.9482**	8.64

Table 2. *Cont.*

Level	Image	Method	The Output Seismic Signal		
			PSNR (dB)	SSIM	Time (s)
	House	ATV	32.31	0.8871	3.09
		ITV	31.81	0.8888	**1.84**
		TGV	32.32	0.9135	8.38
		OGS-L1	33.04	0.9127	15.27
		OGS-Lp	37.47	0.9667	12.45
		OGS-Lp-FAST	**37.72**	**0.9679**	10.48
	Shoulder	ATV	35.32	0.9636	5.42
		ITV	35.33	0.9649	**3.56**
		TGV	35.29	0.9256	14.98
		OGS-L1	37.00	0.9719	17.77
		OGS-Lp	38.89	0.9829	16.06
		OGS-Lp-FAST	**38.92**	**0.983**	14.61
	Girl	ATV	29.45	0.8868	4.53
		ITV	30.05	0.894	**3.22**
		TGV	30.14	0.8907	9.58
		OGS-L1	30.67	0.8995	13.48
		OGS-Lp	32.29	0.9365	14.17
		OGS-Lp-FAST	**32.34**	**0.9371**	11.69
	Milk Drop	ATV	32.32	0.8973	4.83
		ITV	31.02	0.9039	**3.63**
		TGV	30.48	0.894	8.59
		OGS-L1	33.35	0.911	16.39
		OGS-Lp	35.76	0.9533	13.42
		OGS-Lp-FAST	**35.87**	**0.9538**	8.58
	Woman	ATV	29.45	0.8868	4.53
		ITV	29.65	0.9015	**3.73**
		TGV	29.84	0.8853	8.98
		OGS-L1	30.35	0.908	16.03
		OGS-Lp	**31.71**	**0.9395**	16.22
		OGS-Lp-FAST	31.7	0.9398	8.69

Table 3. Numerical comparison of our proposed method and other models (images are corrupted by impulse noise of 30%).

Level	Image	Method	The Output Seismic Signal		
			PSNR (dB)	SSIM	Time (s)
30%	Lena	ATV	27.08	0.829	5
		ITV	27.06	0.837	**2.88**
		TGV	27.23	0.8319	8.16
		OGS-L1	27.59	0.8543	9.56
		OGS-Lp	29.19	0.9035	15.72
		OGS-Lp-FAST	**29.21**	**0.9039**	7.48
	House	ATV	30.4	0.8717	3.8
		ITV	30.16	0.8545	**2.17**
		TGV	30.82	0.8862	11.61
		OGS-L1	31.5	0.8807	7.77
		OGS-Lp	35.03	0.9432	13.8
		OGS-Lp-FAST	**35.13**	**0.9444**	10.77
	Shoulder	ATV	34.48	0.9551	4.95
		ITV	34.33	0.9556	**3.94**
		TGV	34.74	0.9493	22
		OGS-L1	35.34	0.9599	16.97
		OGS-Lp	**36.5**	**0.9633**	18.88
		OGS-Lp-FAST	36.47	0.9628	5.88

Table 3. *Cont.*

Level	Image	Method	The Output Seismic Signal		
			PSNR (dB)	SSIM	Time (s)
	Girl	ATV	28.17	0.8586	5.17
		ITV	28.53	0.8737	**3.31**
		TGV	28.82	0.8422	8.84
		OGS-L1	29.11	0.8818	8.88
		OGS-Lp	**30.42**	**0.9155**	15.64
		OGS-Lp-FAST	30.41	0.9148	12.84
	Milk Drop	ATV	30.24	0.8788	4.95
		ITV	30.1	0.8681	**2.33**
		TGV	29.42	0.8878	11.45
		OGS-L1	31.08	0.8836	9.33
		OGS-Lp	32.7	0.9261	16.88
		OGS-Lp-FAST	**33.19**	**0.9274**	10.64
	Woman	ATV	27.86	0.8534	4.27
		ITV	28.43	0.866	**2.83**
		TGV	28.32	0.844	9.53
		OGS-L1	29.05	0.8725	11.59
		OGS-Lp	**30.15**	**0.9063**	17.25
		OGS-Lp-FAST	30.13	0.9047	10.89

Table 4. Numerical comparison of our proposed method and other models (images are corrupted by impulse noise of 40%).

Level	Image	Method	The Output Seismic Signal		
			PSNR (dB)	SSIM	Time (s)
	Lena	ATV	25.85	0.796	5
		ITV	25.8	0.8009	**3.11**
		TGV	26.2	0.8041	10.27
		OGS-L1	26.22	0.8151	11.25
		OGS-Lp	27.64	0.8683	18.88
		OGS-Lp-FAST	**27.67**	**0.8675**	9.81
	House	ATV	28.5	0.8433	4.94
		ITV	28.69	0.8356	**2.31**
		TGV	29.21	0.8284	9.48
		OGS-L1	29.31	0.8566	11.95
		OGS-Lp	32.84	0.9182	13.77
		OGS-Lp-FAST	**32.92**	**0.9189**	12.8
	Shoulder	ATV	32.4	0.9328	4.86
		ITV	32.54	0.9357	**3.86**
		TGV	32.8	0.9345	24.81
		OGS-L1	32.62	0.9310	16.36
		OGS-Lp	**33.25**	**0.9510**	20.66
		OGS-Lp-FAST	33.24	0.9509	17.61
40%	Girl	ATV	27.19	0.8348	4.94
		ITV	27.33	0.8447	**3.09**
		TGV	27.86	0.8135	12.11
		OGS-L1	27.88	0.8526	11.78
		OGS-Lp	**28.87**	**0.8861**	18.05
		OGS-Lp-FAST	28.86	0.8847	12.69
	Milk Drop	ATV	27.51	0.8307	3.97
		ITV	28.1	0.8424	**3.19**
		TGV	28.09	0.8521	12.97
		OGS-L1	29.34	0.8569	10.27
		OGS-Lp	**30.56**	**0.8938**	16.94
		OGS-Lp-FAST	30.46	0.8933	13
	Woman	ATV	26.97	0.8338	4.67
		ITV	27.13	0.8366	**3.63**
		TGV	27.34	0.7708	9.58
		OGS-L1	27.7	0.8483	13.13
		OGS-Lp	28.27	0.8722	18.45
		OGS-Lp-FAST	**28.29**	**0.8716**	13.11

Table 5. Numerical comparison of our proposed method and other models (images are corrupted by impulse noise of 50%).

Level	Image	Method	The Output Seismic Signal		
			PSNR (dB)	SSIM	Time (s)
50%	Lena	ATV	23.44	0.7353	6
		ITV	23.56	0.7457	**4.14**
		TGV	25.05	0.7631	13.44
		OGS-L1	25.08	0.7612	15.86
		OGS-Lp	25.74	0.8264	19.06
		OGS-Lp-FAST	25.72	0.8262	11.67
	House	ATV	26.48	0.8046	3.61
		ITV	27.09	0.8139	**2.88**
		TGV	27.88	0.8264	15.14
		OGS-L1	27.85	0.8245	10.05
		OGS-Lp	**31.24**	**0.8932**	16.41
		OGS-Lp-FAST	31.15	0.8923	14.61
	Shoulder	ATV	30.99	0.9137	4.91
		ITV	31.07	0.9168	3.59
		TGV	31.93	0.8954	24.31
		OGS-L1	30.99	0.9137	16.5
		OGS-Lp	31.94	0.9161	19.45
		OGS-Lp-FAST	**32.05**	**0.9277**	17.72
	Girl	ATV	25.1	0.7761	5.58
		ITV	25.73	0.7977	**3.3**
		TGV	26.75	0.814	15.67
		OGS-L1	26.45	0.811	14.16
		OGS-Lp	27.32	0.846	19.28
		OGS-Lp-FAST	**27.35**	**0.8463**	15.91
	Milk Drop	ATV	26.19	0.8082	4.92
		ITV	26.51	0.8188	**3.42**
		TGV	27.01	0.8349	19.05
		OGS-L1	27.61	0.8351	14.67
		OGS-Lp	**28.65**	0.868	17.8
		OGS-Lp-FAST	28.59	**0.8682**	14.11
	Woman	ATV	25.84	0.8049	5.11
		ITV	25.24	0.7939	**3.67**
		TGV	25.99	0.685	10.84
		OGS-L1	26.29	0.8143	13.95
		OGS-Lp	26.56	0.8165	18.19
		OGS-Lp-FAST	**26.61**	**0.8169**	14.02

To verify our proposed method further, we compared the image details restored by different algorithms. Figure 6 shows four images ("Woman", "Pepper", "Girl", and "House") with impulse noise from 20% to 50%. The enlarged details of the images restored by ATV, ITV, TGV, OGS-L1, and OGS-Lp-FAST algorithms are displayed for comparison.

In terms of the visual effect of the restored images, ATV denoising produces apparent blocking artifacts in the images. In the "Pepper" image ATV recovers, we can easily find two heavy noise pixels.

The ITV method also shows significant staircase effect. In the "House" image it restores, the edge of the image is not well preserved when the noise pollution is high.

For the denoising results of TGV, the blocking artifacts in the images are sufficiently suppressed, but local heavy noise spots are still observable.

Finally, by comparing the four images, we can easily see the visual improvement in the images by using our method. Even at high noise pollution, our method protects the edge information of the image very well, and at the same time, avoids the staircase effect.
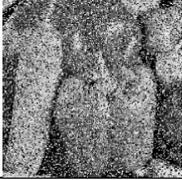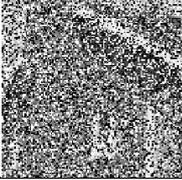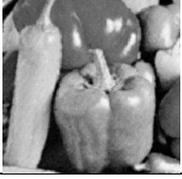
MODEL	Woman Level = 20%	Pepper Level = 30%	Girl Level = 40%	House Level = 50%
Noise Image				
ATV	(PSNR = 29.45 dB)	(PSNR = 26.15 dB)	(PSNR = 27.19 dB)	(PSNR = 26.48 dB)
ITV	(PSNR = 29.65 dB)	(PSNR = 26.05 dB)	(PSNR = 27.33 dB)	(PSNR = 27.29 dB)
TGV	(PSNR = 29.84 dB)	(PSNR = 26.13 dB)	(PSNR = 27.86 dB)	(PSNR = 27.88 dB)
OGS-L1	(PSNR = 30.35 dB)	(PSNR = 27.19 dB)	(PSNR = 27.88 dB)	(PSNR = 27.85 dB)
OGS-Lp-FAST	(PSNR = 29.45 dB)	(PSNR = 26.15 dB)	(PSNR = 27.19 dB)	(PSNR = 26.48 dB)

Figure 6. Comparison of image details recovered by our proposed models and other models. The test images ("Woman", "Pepper", "Girl", and "House") are corrupted by 20~50% impulse noise, respectively.

6. Discussion and Conclusions

In this work, we study a new regularization model by applying TV with OGS and Lp-pseudo-norm shrinkage for the image polluting under impulse noise. We provided the efficient algorithm OGS-Lp-FAST under the ADMM framework. This algorithm is rooted in overlapping group sparsity-based regularization, and incorporated the comparisons made with ATV, ITV, TGV, OGS-L1, and OGS-Lp models for validation of our proposed method. The following conclusions are drawn from the experimental results:

1. An overlapping group sparsity (OGS)-based regularizer is used to replace the anisotropic total variation (ATV), to describe the prior conditions of the image. OGS makes full use of the similarity among image neighborhoods and the dissimilarity in the surroundings of each point. It promotes the distinction between the smooth and edge regions of an image, thus enhancing the robustness of the proposed model.

2. Lp-pseudo-norm shrinkage is used in place of the L1-norm regularization to describe the fidelity term of images with salt and pepper noise. With the inclusion of another degree of freedom, Lp-pseudo-norm shrinkage reflects the sparsity of the image better and greatly improves the denoising performance of the algorithm.

3. The difference operator is used for convolution. Under the ADMM framework, the complex model is transformed into a series of simpler mathematical problems to solve.

4. Appropriate K values could effectively improve the overall denoising performance of the model. In practice, this parameter needs to be adjusted. If it is too small, the neighborhood information is not utilized completely. If the value is too big, too many dissimilar pixel blocks will be included, impairing the denoising result.

5. The adoption of accelerated ADMM with a restart accelerates the convergence of the algorithm. The running time is reduced.

6. In this paper, we focus on impulse noise removal, but the model is also applicable to other types of noise removal that we will further study in future work.

Author Contributions: Funding acquisition, F.Y.; Methodology, L.W., Y.C. (Yingpin Chen) and F.L.; Software, L.W., Y.C. (Yingpin Chen) and Y.C. (Yuqun Chen) and Z.C.; Writing—original draft, L.W.; Writing—review & editing, Y.C. (Yingpin Chen).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. He, W.; Zhang, H.; Zhang, L.; Shen, H. Total-Variation-Regularized Low-Rank Matrix Factorization for Hyperspectral Image Restoration. *IEEE Trans. Geosci. Remote Sens.* **2015**, *54*, 178–188. [CrossRef]

2. Cevher, V.; Sankaranarayanan, A.; Duarte, M.F.; Reddy, D.; Baraniuk, R.G.; Chellappa, R. *Compressive Sensing for Background Subtraction*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 155–168.

3. Huang, G.; Jiang, H.; Matthews, K.; Wilford, P. Lensless imaging by compressive sensing. In Proceedings of the IEEE International Conference on Image Processing, Melbourne, VIC, Australia, 15–18 September 2014; pp. 2101–2105.

4. Chen, Y.; Peng, Z.; Cheng, Z.; Tian, L. Seismic signal time-frequency analysis basedon multi-directional window using greedy strategy. *J. Appl. Geophys.* **2017**, *143*, 116–128. [CrossRef]

5. Lu, S.L.T.; Fang, L. Spectral–spatial adaptive sparse representation for hyperspectral image denoising. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 373–385. [CrossRef]

6. Zhao, W.; Lu, H. Medical Image Fusion and Denoising with Alternating Sequential Filter and Adaptive Fractional Order Total Variation. *IEEE Trans. Instrum. Meas.* **2017**, *66*, 1–12. [CrossRef]

7. Knoll, F.; Bredies, K.; Pock, T.; Stollberger, R. Second order total generalized variation (TGV) for MRI. *Magn. Reson. Med.* **2011**, *65*, 480–491. [CrossRef] [PubMed]

8. Kong, D.; Peng, Z. Seismic random noise attenuation using shearlet and total generalized variation. *J. Geophys. Eng.* **2015**, *12*, 1024–1035. [CrossRef]

9.   Zhu, Z.; Yin, H.; Chai, Y.; Li, Y.; Qi, G. A novel multi-modality image fusion method based on image decomposition and sparse representation. *Inform. Sci.* **2018**, *432*, 516–529. [CrossRef]

10. Marquina, A.; Osher, S.J. Image super-resolution by TV-regularization and Bregman iteration. *J. Sci. Comput.* **2008**, *37*, 367–382. [CrossRef]

11. Nikolova, M. A variational approach to remove outliers and impulse noise. *J. Math. Imaging Vis.* **2004**, *20*, 99–120. [CrossRef]

12. Rudin, L.I.; Osher, S.; Fatemi, E. Nonlinear total variation based noise removal algorithms. *Phys. D Nonlinear Phenom.* **1992**, *60*, 259–268. [CrossRef]

13. Qin, Z. An Alternating Direction Method for Total Variation Denoising. *Optim. Methods Softw.* **2015**, *30*, 594–615. [CrossRef]

14. Tai, X.C.; Wu, C. Augmented Lagrangian Method, Dual Methods and Split Bregman Iteration for ROF Model. In *International Conference on Scale Space and Variational Methods in Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 502–513.

15. Ng, M.; Wang, F.; Yuan, X.M. Fast Minimization Methods for Solving Constrained Total-Variation Superresolution Image Reconstruction. *Multidimens. Syst. Signal Process.* **2011**, *22*, 259–286. [CrossRef]

16. Chan, T.; Marquina, A.; Mulet, P. High-order total variation-based image restoration. *SIAM J. Sci. Comput.* **2000**, *22*, 503–516. [CrossRef]

17. Chan, T.F.; Esedoglu, S.; Park, F. A fourth order dual method for staircase reduction in texture extraction and image restoration problems. In Proceedings of the 2010 17th IEEE International Conference on Image Processing (ICIP), Hong Kong, China, 26–29 September 2010; pp. 4137–4140.

18. Wu, L.; Chen, Y.; Jin, J.; Du, H.; Qiu, B. Four-directional fractional-order total variation regularization for image denoising. *J. Electron. Imaging* **2017**, *26*, 053003. [CrossRef]

19. Bredies, K.; Kunisch, K.; Pock, T. Total generalized variation. *SIAM J. Imaging Sci.* **2010**, *3*, 492–526. [CrossRef]

20. Feng, W.; Lei, H.; Gao, Y. Speckle reduction via higher order total variation approach. *IEEE Trans. Image Process.* **2014**, *23*, 1831–1843. [CrossRef] [PubMed]

21. Cheng, Z.; Chen, Y.; Wang, L.; Lin, F.; Wang, H.; Chen, Y. Four-Directional Total Variation Denoising Using Fast Fourier Transform and ADMM. In Proceedings of the IEEE 3rd International Conference on Image, Vision and Computing (ICIVC), Chongqing, China, 27–29 June 2018; pp. 379–383.

22. Hajiaboli, M.R. *An Anisotropic Fourth-Order Partial Differential Equation for Noise Removal*; Springer: Berlin/Heidelberg, Germany, 2009.

23. Zhang, J.; Chen, K. A Total Fractional-Order Variation Model for Image Restoration with Non-homogeneous Boundary Conditions and its Numerical Solution. *Siam J. Imaging Sci.* **2015**, *8*, 2487–2518. [CrossRef]

24. Chen, P.-Y.; Selesnick, I.W. Group-Sparse Signal Denoising: Non-Convex Regularization, Convex Optimization. *arXiv* **2013**. [CrossRef]

25. Liu, J.; Huang, T.-Z.; Selesnick, I.W.; Lv, X.-G.; Chen, P.-Y. Image restoration using total variation with overlapping group sparsity. *Inform. Sci.* **2015**, *295*, 232–246. [CrossRef]

26. Adam, T.; Paramesran, R. Image denoising using combined higher order non-convex total variation with overlapping group sparsity. *Multidimens. Syst. Signal Process.* **2018**, 1–25. [CrossRef]

27. Wu, Y.C.L.; Du, H. Efficient compressedsensing MR image reconstruction using anisotropic overlapping group sparsitytotal variation. In Proceedings of the 2017 7th International Workshop on Computer Science and Engineering, Beijing, China, 25–27 June 2017.

28. Chen, Y.; Wu, L.; Peng, Z.; Liu, X. Fast Overlapping Group Sparsity Total Variation Image Denoising Based on Fast Fourier Transform and Split Bregman Iterations. In Proceedings of the 7th International Workshop on Computer Science and Engineering, Beijing, China, 25–27 June 2017.

29. Chartrand, R. Exact Reconstruction of Sparse Signals via Nonconvex Minimization. *IEEE Signal Process. Lett.* **2007**, *14*, 707–710. [CrossRef]

30. Parekh, A.; Selesnick, I.W. Convex Denoising using Non-Convex Tight Frame Regularization. *IEEE Signal Process. Lett.* **2015**, *22*, 1786–1790. [CrossRef]

31. Li, S.; He, Y.; Chen, Y.; Liu, W.; Yang, X.; Peng, Z. Fast multi-trace impedance inversion using anisotropic total p-variation regularization in the frequency domain. *J. Geophys. Eng.* **2018**, 15. [CrossRef]

32. Boyd, S.; Parikh, N.; Chu, E.; Peleato, B.; Eckstein, J. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.* **2011**, *3*, 1–122. [CrossRef]

33. Liu, J.; Huang, T.-Z.; Liu, G.; Wang, S.; Lv, X.-G. Total variation with overlapping group sparsity for speckle noise reduction. *Neurocomputing* **2016**, *216*, 502–513. [CrossRef]

34. Goldstein, T.; O'donoghue, B.; Setzer, S.; Baraniuk, R. Fast alternating direction optimization methods. *SIAM J. Imaging Sci.* **2014**, *7*, 1588–1623. [CrossRef]

35. Chan, R.H.; Ho, C.-W.; Nikolova, M. Salt-and-pepper noise removal by median-type noise detectors and detail-preserving regularization. *IEEE Trans. Image Process.* **2005**, *14*, 1479–1485. [CrossRef] [PubMed]

36. Zhong, Q.; Wu, C.; Shu, Q.; Liu, R.W. Spatially adaptive total generalized variation-regularized image deblurring with impulse noise. *J. Electron. Imaging* **2018**, *27*, 053006. [CrossRef]

37. Trivedi, M.C.; Singh, V.K.; Kolhe, M.L.; Goyal, P.K.; Shrimali, M. Patch-Based Image Denoising Model for Mixed Gaussian Impulse Noise Using L1 Norm. In *Intelligent Communication and Computational Technologies*; Springer: Singapore, 2018; pp. 77–84.

38. Zheng, L.; Maleki, A.; Weng, H.; Wang, X.; Long, T. Does ℓp-minimization outperform ℓ1-minimization? *IEEE Trans. Inform. Theory* **2017**, *63*, 6896–6935. [CrossRef]

39. Xie, Y.; Gu, S.; Liu, Y.; Zuo, W.; Zhang, W.; Zhang, L. Weighted Schatten p -Norm Minimization for Image Denoising and Background Subtraction. *IEEE Trans. Image Process.* **2016**, *25*, 4842–4857. [CrossRef]

40. Zhou, X.; Molina, R.; Zhou, F.; Katsaggelos, A.K. Fast iteratively reweighted least squares for lp regularized image deconvolution and reconstruction. *IEEE Int. Conf. Image Process.* **2015**, *24*, 1783–1787.

41. Chen, F.; Zhang, Y. Sparse Hyperspectral Unmixing Based on Constrained lp—l2 Optimization. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 1142–1146. [CrossRef]

42. Chen, Y.; Peng, Z.; Gholami, A.; Yan, J.; Li, S. Seismic Signal Sparse Time-Frequency Analysis by Lp-Quasinorm Constraint. *arXiv* **2018**.

43. Woodworth, J.; Chartrand, R. Compressed Sensing Recovery via Nonconvex Shrinkage Penalties. *Inverse Probl.* **2016**, *32*, 075004. [CrossRef]

44. Sidky, E.Y.; Chartrand, R.; Boone, J.M.; Pan, X. Constrained TpV Minimization for Enhanced Exploitation of Gradient Sparsity: Application to CT Image Reconstruction. *IEEE J. Transl. Eng. Health Med.* **2014**, *2*, 1–18. [CrossRef] [PubMed]

45. Wu, C.; Tai, X.C. Augmented Lagrangian Method, Dual Methods, and Split Bregman Iteration for ROF, Vectorial TV, and High Order Models. *Siam J. Imaging Sci.* **2012**, *3*, 300–339. [CrossRef]

46. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef] [PubMed]