Article

Mixed Fractional-Order and High-Order Adaptive Image Denoising Algorithm Based on Weight Selection Function

Shaojiu Bi, Minmin Li and Guangcheng Cai *

Department of Mathematics, Kunming University of Science and Technology, Kunming 650500, China;
bishaojiu@stu.kust.edu.cn (S.B.)

* Correspondence: guangchengcai@sina.com

Abstract: In this paper, a mixed-order image denoising algorithm containing fractional-order and high-order regularization terms is proposed, which effectively suppresses the staircase effect generated by the TV model and its variants while better preserving the edges and details of the image. Adding different regularization penalties in different regions is fundamental to improving the denoising performance of the model. Therefore, a weight selection function is designed using the structure tensor to achieve a more effective selection of regularization terms in different regions. In each iteration, the regularization parameters are adaptively adjusted according to the Morozov discrepancy principle to promote the performance of the algorithm. Based on the primal–dual theory, the original algorithm is improved by using the predictor–corrector scheme to obtain a more accurate approximate solution while ensuring the convergence of the algorithm. The effectiveness of the proposed algorithm is demonstrated through simulation experiments.

Keywords: image denoising; fractional-order; mixed-order; weight selection function; primal–dual algorithm; adaptive regularization parameter

1. Introduction

The field of digital image processing requires high-resolution and clean images. However, in most cases, images are often degraded due to noise and other factors, resulting in the loss of valuable image information. Therefore, it poses a significant challenge to model the noise accurately, employ efficient algorithms to remove it from the image, and achieve algorithm convergence.

Image denoising can be regarded as a reversible problem, but it is ill-posed, which means that the solution to this problem does not exist or is not unique. To overcome this problem and achieve better denoising performance, several methods have been proposed, including regularization methods [1], wavelet transforms [2,3], non-local methods such as NLM [4,5] and BM3D [6,7], deep learning methods such as DnCNN [8], DIVA [9], and BM3D-net [10], and quantum-based methods such as DeQuIP [11] and QAB [12]. The role of these methods in image denoising is acknowledged, but certain limitations are also present in them. For instance, the appearance of new unrealistic structures or textures is prone to occur in denoised images using the wavelet transform. Non-local methods are characterized by high computational complexity and are easily influenced by outliers. Deep learning methods necessitate a substantial dataset for model training, leading to longer modeling times. Quantum-based methods are still in the early stages of research and have many technical challenges and limitations. Although there are some limitations to regularization methods, they are supported by a solid understanding of mathematical theory, and the logical reasoning of the computational derivation process is also strong. Being easy to code and having a shorter modeling time, they are still widely utilized at present.

In models associated with regularization techniques, there are two prevalent issues: how to suppress the staircase effect caused by the total variation (TV) regularization model
proposed by Rudin et al. [13], and how to adaptively adjust the regularization parameters based on image features. Currently, a considerable amount of research has been conducted on these problems. Regarding the first issue, researchers have presented using higher-order regularizers [14–16] or fractional-order regularizers [17–19] to modify the TV regularization term and suppress the staircase effect. However, studies have found that the higher-order TV model (HOTV) does not smooth noise as effectively as the TV model. Although the fractional-order TV model (FOTV) has better noise filtering and detail preservation effects, its solving process is challenging. For the second issue, researchers have designed different adaptive regularization parameters based on prior information about the image [20,21]. However, there are few algorithms that simultaneously address both of these problems. Therefore, the advantages of the HOTV and FOTV models are inherited in this paper, and a mixed-order TV (MOTV) image denoising algorithm is proposed. Moreover, adaptive regularization parameters are designed based on image features. To efficiently solve the model while minimizing the number of parameters, an improved primal–dual algorithm based on the predictor–corrector scheme is proposed, and the convergence of this algorithm is theoretically proven.

The main contributions of this paper are summarized below: Firstly, a new MOTV image denoising algorithm is proposed to suppress the staircase artifacts while filtering out more noise. The approximate optimal solution of the model is obtained by using the improved primal–dual algorithm. Secondly, an edge-weighted selection function is constructed by using the structure tensor, which preserves more edge details of the image. Finally, an adaptive regularization parameter is designed using the Morozov discrepancy principle to facilitate the performance of the algorithm.

The rest of this article is organized as follows: Section 2 briefly introduces the relevant knowledge background and model. In Section 3, a new model is proposed, and the derivation process of the model solution is given. Then an adaptive regularization parameter is set, and the convergence of the algorithm is evidenced. In Section 4, simulation experiments are carried out to prove the effectiveness of the proposed algorithm. Section 5 provides a simple summary.

2. Related Works

2.1. Fractional-Order Variation Model

The fractional-order derivative, as a generalization of the integer derivative, can preserve more image details. Therefore, it has been extensively studied by a quantity of scholars in recent decades [22,23]. Firstly, the definition of fractional-order derivatives and their related properties are reviewed here.

The Euclidean space $\mathbb{R}^{m \times n}$ of an image $u$ with size $m \times n$ is represented as $V$. To facilitate interpretation, the image domain is discretized into a matrix grid $\{(x_i, y_j) : 1 \leq i \leq m, 1 \leq j \leq n\}$, where $(i, j)$ represents the position of the pixel in the matrix. Consequently, the image $u \subseteq V$ can be defined as $u_{ij} = u(x_i, y_j)$.

Various definitions of fractional-order derivatives exist in the literature [24], including the Grünwald–Letnikov (G-L) definition, the Riemann–Liouville (R-L) definition, the Caputo definition, and others. In digital image processing, the G-L definition is the most popular. The G-L definition is theoretically proven. In Section 4, simulation experiments are carried out to prove the effectiveness of the proposed algorithm. Section 5 provides a simple summary.

$$\nabla^a u = (\nabla^a_1 u, \nabla^a_2 u)^T, a > 0,$$  \hfill (1)$$

where $\nabla^a_1 u, \nabla^a_2 u \in V$ denote the derivatives of the image $u$ in the horizontal and vertical directions, respectively, which can be discretized and approximated as

$$\begin{align*}
(\nabla^a_1 u)_{ij} &= \sum_{k=0}^{K-1} (-1)^k C_k^a u_{i-k,j}, \\
(\nabla^a_2 u)_{ij} &= \sum_{k=0}^{K-1} (-1)^k C_k^a u_{i,j-k}.
\end{align*}$$  \hfill (2)$$

Here $K \geq 3$ is the number of adjacent pixels used to approximate the fractional-order derivative at each pixel point. The coefficient $\{C_k^a\}_{k=0}^{K-1}$ is defined as $C_k^a = \frac{\Gamma(a+1)}{\Gamma(k+1)\Gamma(a+1-k)}$. 

where $\Gamma(\cdot)$ is the Gamma function, $U = \{u \in V : \|u\|_{\infty} = \max_{ij}|u_{ij}| \leq 1\}$. Then the FOTV model can be defined as

$$
\min_u \frac{\lambda}{2} \|Hu - f\|_2^2 + \|\nabla^a u\|_1,
$$

(3)

where

$$
\|\nabla^a u\|_1 := \sum_{ij} |(\nabla^a u)_{ij}|, \quad \|\nabla^a u\|_1 = \sqrt{(\nabla^1 u)^2_{ij} + (\nabla^2 u)^2_{ij}},
$$

(4)

Here, $H$ is a blur operator, $u$ is the original image, and $f$ is the noisy image. According to Ref. [25], an important relation can be derived as $\langle \nabla^a u, p \rangle = \langle u, (-1)^a \text{div}^a p \rangle$. Here, $\forall p = (p^1, p^2) \in Z$ where $Z = V \times V$, and $p$ represents a dual variable. Utilizing the aforementioned relationship, the adjoint divergence of the discrete gradient (1) can be obtained as follows:

$$
(\text{div}^a p)_{ij} = (-1)^a \sum_{k=0}^{K-1} (-1)^k C^a_k (p^1_{i+k,j} + p^2_{i,j+k}).
$$

(5)

2.2. High-Order Variation Model

If the TV model is defined in the bounded variation (BV) function space, then the reconstruction of discontinuous points can be achieved. Accordingly, the HOTV models, which can better model noise, have been proposed by plenty of scholars to suppress the staircase effect and achieve improved results [14–16,26]. The HOTV model is generally defined as follows:

$$
\min_u \|Hu - f\|_2^2 + \beta R(u),
$$

(6)

where $\beta$ is a predefined parameter and $R(u)$ is a high-order regularization term.

2.3. Structural Tensor

In mathematics, the structure tensor, also known as the second moment matrix, is a matrix derived from the gradient of a function. It can summarize the main directions of the gradients in a specified neighborhood of a given point, as well as the degree of coherence in those directions. As a result, the structure tensor finds extensive application in the fields of computer vision and image processing [27,28]. In image processing, it excels at distinguishing between edge regions, flat regions, and corner regions. The structure tensor for the image $u$ can be expressed as

$$
ST = \begin{bmatrix}
  u_x^2 & u_x u_y \\
  u_y u_x & u_y^2
\end{bmatrix}.
$$

(7)

Here, $u_x, u_y$ denote the gradients of the image $u$ in the $x$ and $y$ directions, respectively. If $SK$ and $SH$ are used to represent the determinant and trace of matrix $ST$, respectively, then the judgment conditions for each region are summarized as follows:

- flat regions: $SH = 0$;
- edge regions: $SH > 0$ and $SK = 0$;
- corner regions: $SH > 0$ and $SK > 0$.

In Figure 1, the white portions of subgraphs (a–c), respectively, represent the flat regions, edge regions, and corner regions of the Lena image. However, in practice, the values of $SK$ and $SH$ are not ideal, so approximate judgment is usually used.
2.4. Primal–Dual Algorithm with Guaranteed Convergence

The standard Euclidean inner product and norm of the finite dimensional vector spaces \( X \) and \( Y \) are expressed as \( \langle \cdot, \cdot \rangle \) and \( \| \cdot \| = \langle \cdot, \cdot \rangle^{1/2} \), respectively. If a continuous linear operator \( A \) is a map on \( X \to Y \), then its induced norm can be expressed as

\[
L = \|A\| = \max\{\|Ax\| : x \in X, \|x\| \leq 1\}. \tag{8}
\]

According to Fenchel duality theory [29], the generalized saddle point problem can be expressed as

\[
\min_{x \in X} \max_{y \in Y} \langle Ax, y \rangle + G(x) - F^*(y), \tag{9}
\]

where \( G \) and \( F \) are lower semi-continuous functions, and \( F^* \) is the conjugate function of \( F \).

It is well known that the primal–dual problem of the following primal optimization problem (10) corresponds to this saddle point problem (9) mentioned above. The primal problem can be expressed as follows:

\[
\min_{x \in X} F(Ax) + G(x). \tag{10}
\]

Then, according to the primal–dual problem, the dual problem related to the primal problem can be calculated, which is expressed as

\[
\max_{y \in Y} - (G^*(-A^*y)) + F^*(y). \tag{11}
\]

Chambler and Pock proposed a primal–dual algorithm for solving the saddle point structure problem (9), and the convergence of the algorithm was proven [30]. By fixing the primal variable \( x \) of the primal–dual problem and solving the first-order derivative function about the dual variable \( y \), the expression for \( y \) can be obtained:

\[
y = (I + \partial F^*)^{-1} (y + Ax). \tag{12}
\]

Similarly, it can be obtained that the expression about the primal variable \( x \) is:

\[
x = (I + \partial G)^{-1} (x - A^*y) \tag{13}
\]

where \( \partial F^* \) and \( \partial G \) represent the subgradients of \( F^* \) and \( G \), respectively. The operators \( (I + \partial F^*)^{-1} \) and \( (I + \partial G)^{-1} \) are known as resolvent operators. When the concept of minimizing an energy functional is applied to define the resolvent operator, it can be expressed as

\[
x = (I + \tau \partial F)^{-1} (y) = \arg \min_x \left\{ \frac{\|x - y\|^2}{2\tau} + F(x) \right\}. \tag{14}
\]

See [31] for details of the subgradient and resolvent operators. Then, the primal–dual algorithm proposed in the literature [30] can ultimately be summarized as
3. Model Proposal and Solution

3.1. Proposed Model

In this study, the following image denoising problems with additive noise are considered:

\[ f = Hu + \eta, \quad \text{(14)} \]

where \( \eta \) represents the Gaussian noise. TV regularization is a classical method for restoring high-quality images, guaranteeing the stability of the solution. Models (3) and (6) are classical FOTV and HOTV models, respectively, used for removing Gaussian noise. Building upon the strengths of both models, a mixed-order image denoising model is proposed in this paper, described as follows:

\[
\min_{u} \left\{ \frac{\lambda}{2} ||Hu - f||_2^2 + \xi_1||\nabla u||_1 + \xi_2||\nabla^2 u||_2 + \beta ||u||_1 \right\}.
\quad \text{(15)}
\]

Here, \( \lambda \) represents the regularization parameter, \( \beta \) is the balance parameter, and \( \xi_i(i = 1,2) \) is a weight selection parameter. The purpose of introducing this weight selection parameter is to select different regions of the image by the structure tensor and to process the image using different regularization terms in different regions. To achieve this purpose, \( \xi_i \) is defined as a function of pixel point characteristics. In other words, the intensity of regularization varies for pixel points in the flat regions, edge regions, and corner regions. At present, the weight function, which is the inverse function of the image gradient, is used by most researchers [29,32,33], but this paper takes a different approach.

Next, a matrix related to the structure tensor is defined:

\[
K_\sigma(u(x,y)) = G_\sigma \ast (\nabla u \nabla u^T) = \begin{bmatrix}
G_\sigma \ast u_x^2 & G_\sigma \ast u_x u_y \\
G_\sigma \ast u_x u_y & G_\sigma \ast u_y^2
\end{bmatrix} = \begin{bmatrix}
a_{11} & a_{12} \\
a_{21} & a_{22}
\end{bmatrix},
\]

where \( G_\sigma = (\sigma \sqrt{2\pi})^{-1} \exp(-|x|^2/2\sigma^2) \) is a Gaussian filter function with standard deviation \( \sigma \).

The matrix \( K \) contains two eigenvectors that are standard orthogonal. Different eigenvalues \( \lambda_1, \lambda_2 \) describe the average contrast in the corresponding eigenvector direction. Different regions exhibit different eigenvalue relationships. In the edge regions, \( \lambda_1 \gg \lambda_2 \approx 0 \); in the flat regions, \( \lambda_1, \lambda_2 \approx 0 \); and in the corner regions, \( \lambda_1 > \lambda_2 \gg 0 \) [34]. Hence, the characteristics of each pixel can be measured using the difference between the eigenvalues. The edge coherence [35] based on the eigenvalue is defined as

\[
C(u) = (\lambda_1 - \lambda_2)^2 = (a_{11} - a_{22})^2 + 4a_{12}^2.
\quad \text{(16)}
\]

Furthermore, the weight selection function \( \xi_i \) can be defined as

\[
\xi_i = \begin{cases} 
1 - \exp\left[-\frac{1}{(C(u)/\omega)^2}\right], & i = 1 \\
\exp\left[-\frac{1}{(C(u)/\omega)^2}\right], & i = 2
\end{cases}
\quad \text{(17)}
\]
where $\omega \in [10^{-4} \sim 10^{-2}]$ is the comparison parameter. The role of $\omega$ is equivalent to a threshold, which can be used to distinguish various regions. Note that in order to prevent the impact of abnormal values, $C(u)$ should be normalized before calculating $\xi_i$.

Obviously, $\xi_i$ depends on the edge consistency of pixel features. In the edge and corner regions, $C \rightarrow 1$, $\xi_i \rightarrow 0$, $\xi_2 \rightarrow 1$, at this time, $\xi_2 \parallel \nabla^2 u \parallel_2 + \beta \parallel u \parallel_1$ is selected as a regularization term of the model, which can better preserve the edge details of the image. In the flat regions, $C \rightarrow 0$, $\xi_2 \rightarrow 0$, $\xi_1 \rightarrow 1$, at this time, $\xi_1 \parallel \nabla^u u \parallel_1 + \beta \parallel u \parallel_1$ is selected as a regularization term, which can reduce noise and artifacts in the image while also reinforcing the regularization process for the flat regions. By designing the selection function $\xi_i$ and performing correspondence matching within the model, the image edges and fine details can be effectively preserved, ultimately resulting in the recovery of high-quality images.

Below is the corresponding matching model:

$$\min_u \Phi(u, \lambda) = \left\{ \frac{\lambda}{2} \|Hu - f\|_2^2 + \xi_1 \parallel \nabla^u u \parallel_1 + \xi_2 \parallel \nabla^2 u \parallel_2 + \beta \parallel u \parallel_1 \right\}. \quad (18)$$

Solving the MOTV model is equivalent to solving the saddle point of the primal–dual problem. The iterative steps and convergence analysis of the model are introduced in detail below.

3.2. Numerical Algorithm

The MOTV model is a variant of the TV model. The famous Euler–Lagrange method can be utilized to solve both the TV model and its variant. However, the model contains a non-differentiable derivative, making its calculation process complex. To overcome the difficulties, the primal–dual algorithm is employed. The conjugate sets of all the dual variables $p$, $y$, and $w$ are $P$, $Y$, and $W$, respectively. Let $\delta(u) = \|\nabla^u u\|_1$, then its conjugate function satisfies the following properties:

$$\delta^*_p(p) = \begin{cases} 0, & p \in P \\ +\infty, & p \notin P \end{cases} \quad (19)$$

where $P = \{p \in Z : \|p\|_\infty = \max |p| \leq 1 \}$ denotes the conjugate set. Similarly, let $\zeta(u) = \|\nabla^2 u\|_2$, $\xi(u) = \|u\|_1$, then the associated conjugate functions $\zeta^*_Y(y)$, $\xi^*_W(w)$ and conjugate sets $Y$, $W$ also have the same properties.

Applying Fenchel duality theory to the primal problem (18), a general primal–dual model of the MOTV model can be obtained, which is expressed as

$$\min_u \max_{y,w} \frac{\lambda}{2} \|Hu - f\|_2^2 + \xi_1 \langle p, \nabla^u u \rangle + \xi_2 \langle y, \nabla^2 u \rangle + \beta \langle w, u \rangle$$

$$- \xi_1 \delta^*_p(p) - \xi_2 \zeta^*_y(y) - \beta \xi^*_w(w). \quad (20)$$

Of these, $y$, $w$ has the same properties as $p$. Similarly, the corresponding dual problem can be obtained as follows:

$$\max_{y,w} \min_{p_{y,w}} \frac{\lambda}{2} \|\nabla^u u\|_1 \langle p, \nabla^u u \rangle + \xi_2 \langle y, \nabla^2 u \rangle + \beta \langle w, u \rangle$$

$$- \xi_1 \delta^*_p(p) - \xi_2 \zeta^*_y(y) - \beta \xi^*_w(w). \quad (21)$$

The discretized $MTV$ of the image $u$ is also the Legendre–Fenchel conjugate of $\delta$, $\zeta$, and $\xi$ [36,37]. That is:

$$MTV(u) = \max_{p_{y,w}} \left\{ \xi_1 \langle p, \nabla^u u \rangle + \xi_2 \langle y, \nabla^2 u \rangle + \beta \langle w, u \rangle$$

$$- \xi_1 \delta^*_p(p) - \xi_2 \zeta^*_y(y) - \beta \xi^*_w(w) \right\}$$

$$= \max_{p_{y,w}} \left\{ \xi_1 \langle p, \nabla^u u \rangle + \xi_2 \langle y, \nabla^2 u \rangle + \beta \langle w, u \rangle \right\}. \quad (22)$$
Legendre–Fenchel’s duality is used to represent the $MTV(u)$ norm, (18) can be represented as

$$\min_u \max_{p,y,w} L(u, p, y, w; \lambda).$$  \hspace{1cm} (22)

where $L(u, p, y, w; \lambda) = \frac{1}{2} \|Hu - f\|_2^2 + \xi_1 \langle p, \nabla^\alpha u \rangle + \xi_2 \langle y, \nabla^2 u \rangle + \beta \langle w, u \rangle$. Let $G(u) = \frac{1}{2} \|Hu - f\|_2^2$, then the following equality holds [38].

$$p^{k+1} = (s_2 \xi_1 \partial \xi^* + I)^{-1} \hat{p} = p^k + s_2 \xi_1 \nabla^\alpha u^k$$  \hspace{1cm} (23)

$$y^{k+1} = (s_2 \xi_2 \partial \xi^* + I)^{-1} \hat{y} = y^k + s_2 \xi_2 \nabla^2 u^k$$  \hspace{1cm} (24)

$$w^{k+1} = (s \beta \partial \xi^* + I)^{-1} \hat{w} = w^k + s \beta u^k$$  \hspace{1cm} (25)

$$u^{k+1} = (\partial G + I)^{-1} \tilde{u}, \tilde{u} = u^k - t_1 \xi_1 (\nabla^\alpha)^T p^{k+1} - t_2 \xi_2 (\nabla^2)^T y^{k+1} - t \beta w^{k+1}$$  \hspace{1cm} (26)

Among them, four resolvent operators are involved, namely $(\partial \delta^* + I)^{-1}$, $(\partial \xi^* + I)^{-1}$, $(\partial \xi^* + I)^{-1}$, $(\partial G + I)^{-1}$, the solution of the resolvent operators is very difficult. Therefore, the method of solving saddle points can be employed to equivalently solve the above equations and reduce computational complexity. According to the literature [39], the function $L(u, p, y, w; \lambda)$ has a saddle point, which can be obtained by Lemma 1 as $(u^*, p^*, y^*, w^*)$.

**Lemma 1.** If $(u^*, p^*, y^*, w^*)$ is a point of the function $L(u, p, y, w; \lambda)$, and the point satisfies the inequality

$$L(u^*, y^*, w^*; \lambda) \leq L(u, p^*, y^*, w^*; \lambda) \leq L(u^*, p^*, y^*, w^*; \lambda),$$

then, this point can be represented as a saddle point of the function $L(u, p, y, w; \lambda)$.

In connection with the literature [40], it is clear that the solution procedure for the maximum and minimum of (22) is interchangeable, i.e.,

$$\min_u \max_{p,y,w} L(u, p, y, w; \lambda) = L(u^*, p^*, y^*, w^*; \lambda) = \max_{p,y,w} \min_u L(u, p, y, w; \lambda).$$

This further reflects that the minimalized primal problem can be solved by finding the saddle point of the primal–dual problem, where the dual variable is updated alternately with the primal variable [41,42]. If $u^*$, $p^*$, $y^*$, $w^*$ is the optimal solution of (22), then $(u^*, p^*, y^*, w^*)$ is a saddle point of (22) [39], i.e.,

$$u^* = \arg \min_u L(u, p^*, y^*, w^*; \lambda),$$  \hspace{1cm} (27)

$$p^* = \arg \min_p L(u^*, p, y^*, w^*; \lambda),$$  \hspace{1cm} (28)

$$y^* = \arg \min_y L(u^*, p^*, y, w^*; \lambda),$$  \hspace{1cm} (29)

$$w^* = \arg \min_w L(u^*, p^*, y^*, w; \lambda).$$  \hspace{1cm} (30)

Combining (22) and (27) yields

$$\lambda H^T (Hu^* - f) + \xi_1 (\nabla^\alpha)^T p^* + \xi_2 (\nabla^2)^T y^* + \beta w^* = 0.$$  \hspace{1cm} (31)

According to the Karush–Kuhn–Tucker (KKT) necessary conditions of the dual optimality, the Lagrange multipliers $\mu^1_{i,j}, \mu^2_{i,j}, \mu^3_{i,j} \geq 0$ can be yielded, such that
where either \( \mu_{ij}^1 > 0 \) with \(|p_{ij}| = 1\) or \( \mu_{ij}^1 = 0 \) with \(|p_{ij}| < 1\). At this time, \( \mu_{ij}^1 = |(\nabla^a u^*)_{ij}| \). Similarly, the other two formulas can be obtained. Thus, if the solutions of (28)–(30) are \( p^*, y^*, w^* \), respectively, then

\[
\begin{align*}
\text{(32)} & \\
\begin{cases}
(\nabla^a u^*)_{ij} + |(\nabla^a u^*)_{ij}| p_{ij}^* = 0, \\
(\nabla^2 u^*)_{ij} + |(\nabla^2 u^*)_{ij}| y_{ij}^* = 0, \\
u_{ij}^* + |u_{ij}^*| w_{ij}^* = 0.
\end{cases}
\end{align*}
\]

Therefore, the following lemma can be obtained:

**Lemma 2.** Suppose \( p, y, w \) are dual variables, and if a saddle point of the function \( L(u, p, y, w; \lambda) \) is \( (u^*, p^*, y^*, w^*) \), then (31) and (32) hold.

There exist plenty of methods available for computing saddle points through maximization and minimization techniques [37, 43, 44]. However, the hybrid gradient primal–dual algorithm is utilized by this paper to solve saddle points, which improves the original primal–dual algorithm. By employing this method to calculate the primal and dual variables alternately, a solution can be obtained as follows:

\[
\begin{align*}
p^{k+1} &= \arg \max_p \left\{ L(u, p, y, w; \lambda) - \frac{1}{2s} \| p - p^k \|_2^2 \right\}, \\
y^{k+1} &= \arg \max_y \left\{ L(u, p, y, w; \lambda) - \frac{1}{2s} \| y - y^k \|_2^2 \right\}, \\
w^{k+1} &= \arg \max_w \left\{ L(u, p, y, w; \lambda) - \frac{1}{2s} \| w - w^k \|_2^2 \right\}, \\
u^{k+1} &= \arg \min_u \left\{ L(u, p, y, w; \lambda) + \frac{1}{2t} \| u - u^k \|_2^2 \right\}.
\end{align*}
\]

Here, \( s, t > 0 \) represent the update steps for the dual and primal variables, respectively, which satisfies \( s \cdot t \leq 1/16 \). Note that the regularization parameter \( \lambda \) is not involved in the solution of the dual variables \( p, y, w \). Before discussing the solution of the subproblem (33)–(36), the projection of \( q \) on \( P \) is first defined as follows:

\[
\rho_P(q) = \arg \min_{p \in P} \| p - q \|_2.
\]

The Lagrange method can be used to solve (37). The corresponding Lagrangian function is expressed as

\[
\| p - q \|_2^2 + \sum_{i,j} \gamma_{ij}(\| p_{ij} \|_2^2 - 1).
\]

Under the constraint \( |p_{ij}|^2 \leq 1 \), the Lagrange multiplier \( \gamma_{ij} \geq 0 \) can be obtained. It follows from the complementarity condition that either \( \gamma_{ij} = 0 \) with \( |p_{ij}| < 1 \) and \( |q_{ij}| < 1 \) or \( \gamma_{ij} > 0 \) with \( |p_{ij}| = 1 \) and \( |q_{ij}| \geq 1 \). In the first case, \( p_{ij} = q_{ij} \). In the other case, the KKT condition yields

\[
p_{ij} - q_{ij} + \gamma_{ij} p_{ij} = 0, \quad \forall i, j.
\]
Furthermore, \( \gamma_{i,j} = |q_{i,j}| - 1 \). Thus, \( p_{i,j} = q_{i,j} / |q_{i,j}| \). Then, the expression of \( \rho_p(q) \) is:
\[
(\rho_p(q)) = \frac{q}{\max(1, |q|)}.
\] (38)

In what follows, the solution to each subproblem is discussed in detail. For the \( p \)-subproblem in (33), the solution of
\[
p^{k+1} = \arg \max_p \xi_1(p - p^k, \nabla^a u) - \frac{1}{2\varepsilon} \|p - p^k\|_2^2
\]
\[
= \arg \max_p \frac{1}{2\varepsilon} \{2(p - p^k, s\xi_1 \nabla^a u) - \|p - p^k\|_2^2\}
\] (39)
\[
= \arg \min_p \|p - (p^k + s\xi_1 \nabla^a u)\|_2^2.
\]
The minimization problem is equivalent to computing the projection of \( (p^k + s\xi_1 \nabla^a u) \) onto \( P \) set. Therefore
\[
p^{k+1} = \rho_p(p^k + s\xi_1 \nabla^a u).
\] (40)

Similarly, \( y^{k+1} \) in (34) and \( w^{k+1} \) in (35) can be computed by
\[
y^{k+1} = \rho_y(y^k + s\xi_2 \nabla^2 u),
\] (41)
\[
w^{k+1} = \rho_w(w^k + s\beta u).
\] (42)

For the \( u \)-subproblem in (36), the subproblem (36) is a quadratic function of image \( u \), which has a closed-form solution. The solution process can be expressed as follows:
\[
u^{k+1} = \arg \min_u \xi_1(p, \nabla^a (u - u^k)) + \frac{\lambda}{2} \|Hu - f\|_2^2 + \xi_2(y, \nabla^2 (u - u^k))
\]
\[
+ \beta(\omega, u - u^k) + \frac{1}{2\tau} \|u - u^k\|_2^2,
\] (43)

hence, \( u^{k+1} \) can be easily computed by
\[
u^{k+1} = (t\lambda H^T H + I)^{-1} (t\lambda H^T f + \tilde{u}).
\] (44)

Here, \( \tilde{u} = (u^k - t\xi_1(\nabla^a)^T p - t\xi_2(\nabla^2)^T y - t\beta w) \). Then, the solutions to each subproblem are summarized as follows:
\[
p^{k+1} = (s\xi_1 \partial u + I)^{-1} p \iff p^{k+1} = \frac{p^k + s\xi_1 \nabla^a u^k}{\max(1, |p^k + s\xi_1 \nabla^a u^k|)},
\] (45)
\[
y^{k+1} = (s\xi_2 \partial u + I)^{-1} y \iff y^{k+1} = \frac{y^k + s\xi_2 \nabla^2 u^k}{\max(1, |y^k + s\xi_2 \nabla^2 u^k|)},
\] (46)
\[
u^{k+1} = (s\partial \omega + I)^{-1} \omega \iff w^{k+1} = \frac{w^k + s\beta u^k}{\max(1, |w^k + s\beta u^k|)},
\] (47)
\[
u^{k+1} = (t\partial E + I)^{-1} \tilde{u} \iff u^{k+1} = (t\lambda H^T H + I)^{-1} (t\lambda H^T f + \tilde{u}).
\] (48)

To obtain more accurate computational results, the primal–dual algorithm can be further enhanced using the predictive correction method [43]. This method also facilitates the convergence analysis of the algorithm.

In summary, the solution process can be summarized as the following Algorithm 1.
The error is defined as \( \lambda \), according to the Lemma 3, the updated rules for \( \lambda \) are:

\[
\hat{\lambda}^k = \begin{cases} 
0 & \text{if } \|H\hat{u} - f\|_2^2 \leq c^2 \\
\text{the root of } \|\hat{\lambda}^{k+1}\|_2^2 = c^2 & \text{if } \|H\hat{u} - f\|_2^2 > c^2
\end{cases}
\]

3.3. Adaptive Regularization Parameter

Algorithm 1 involves an important regularization parameter \( \lambda \) when solving the subproblem \( u \). In fact, \( \lambda \) has a tremendous influence on the denoising performance of the algorithm. In general, a multitude of researchers have manually set the regularization parameter \( \lambda \), and several researchers have adaptively selected the parameter \( \lambda \) according to the prior information of the image [45,46]. In order to simplify the solution process, this paper designs an adaptive parameter \( \lambda \) by using the structural features of the data fidelity term while combining the Morozov deviation principle [40], so that the denoised image is always in the feasible set \( D \). The feasible set \( D \) is defined as

\[
D = \left\{ u : \|Hu - f\|_2^2 \leq c^2 \right\}
\]  

(49)

Here \( c \) is the positive number that depends on the noise variance \( \sigma^2 \) [17,40]. If the noise variance \( \sigma^2 \) is known, then the upper bound of the feasible set \( D \) can be expressed as \( c^2 = \tau\min\sigma^2 \). If the noise variance \( \sigma^2 \) is unknown, then \( \sigma^2 \) can be estimated by referring to Reference [47]. \( \tau \) is a predefined parameter, and \( \tau = 1 \) is chosen empirically.

For the denoised image to retain more image detail, an error associated with the adaptive parameter \( \lambda \) is set, which in turn allows the algorithm to adaptively select the parameter \( \lambda \). The error is defined as

\[
e_{k+1} = Hu^{k+1}((\lambda^{k+1}) - f = (\lambda^{k+1}H^TH + I)^{-1}(H\hat{u} - f).
\]

Whether \( H^TH \) is a singular matrix or not, \( (\lambda^{k+1}H^TH + I) \) is always invertible. According to the Lemma 3, the updated rules for \( \lambda \) in the iterative process are as follows:

\[
\lambda^k = \begin{cases} 
0 & \text{if } \|H\hat{u} - f\|_2^2 \leq c^2 \\
\text{the root of } \|\hat{\lambda}^{k+1}\|_2^2 = c^2 & \text{if } \|H\hat{u} - f\|_2^2 > c^2
\end{cases}
\]

(51)
Lemma 3. Let
\[ \phi(u, \lambda) = \|e\|^2. \] (52)

Then \( \phi(u, \lambda) \) is a convex function and strictly monotone decreasing. This function makes the following equation have a unique solution.
\[ \phi(u, \lambda) = c^2 \]

**Proof.** The first-order derivative of the function \( \phi(u, \lambda) \) is \( \frac{\partial \phi(u, \lambda)}{\partial \lambda} = \frac{-2t(\tilde{H}u - f)}{(\lambda + I)^2} \leq 0 \). From this, it can be seen that the function \( \phi(u, \lambda) \) is a monotonically decreasing function concerning the regularization parameter \( \lambda \). The second-order derivative of the function \( \phi(u, \lambda) \) is \( \frac{\partial^2 \phi(u, \lambda)}{\partial \lambda^2} = \frac{6\hat{c}(H\tilde{u} - f)^2}{(\lambda + I)^4} \geq 0 \). It is evident that the function \( \phi(u, \lambda) \) is a strictly positive convex function about the regularization parameter \( \lambda \). Therefore, the above equation has a unique solution. \( \square \)

Contacting Algorithm 1, a new MOTV denoising algorithm can be obtained, see Algorithm 2.

**Algorithm 2:** Primal–Dual Algorithm for New MOTV Image Restoration

**Require:** \( f, H, \beta, c^2 \).

1. Initialize \( u, p, y \) and \( w \). Set the size of \( s \) and \( t \).
2. **while** stopping criterion is not satisfied **do**
3. \( C(u) = (\lambda_1 - \lambda_2)^2 = (k_{11} - k_{22})^2 + 4k_{12}^2; \)
4. \( \bar{c}_1(u) = 1 - \exp[-\frac{1}{(C(u)/\gamma)^2}]; \)
5. \( \bar{c}_2(u) = \exp[-\frac{1}{(C(u)/\gamma)^2}]; \)
6. \( p^{k+1} = \frac{p^k + \bar{c}_1 \nabla u^k}{\max(1, p^k + \bar{c}_1 \nabla u^k)}; \)
7. \( y^{k+1} = \frac{y^k + \bar{c}_2 \nabla u^k}{\max(1, y^k + \bar{c}_2 \nabla u^k)}; \)
8. \( u^{k+1} = \frac{u^k + \beta w^k}{\max(1, u^k + \beta w^k)}; \)
9. \( \tilde{u} = u^k - \bar{c}_1 \nabla u^k, \bar{c}_2 \nabla u^k - t\beta w^k; \)
10. **if** \( \|H\tilde{u} - f\|^2 < c^2; \) **then**
11. \( \lambda_{k+1} = 0; \)
12. **else**
13. \( \lambda_{k+1} \) is set to the root of \( \phi(\lambda^{k+1}, u^*) = c^2; \)
14. **end if**
15. \( u^{k+1} = \frac{(\lambda A H T + I)^{-1}(\lambda AH T f + \tilde{u})}; \)
16. \( p^{k+1} = \frac{p^k + \bar{c}_1 \nabla u^{k+1}}{\max(1, p^k + \bar{c}_1 \nabla u^{k+1})}; \)
17. \( y^{k+1} = \frac{y^k + \bar{c}_2 \nabla u^{k+1}}{\max(1, y^k + \bar{c}_2 \nabla u^{k+1})}; \)
18. \( w^{k+1} = \frac{w^k + \beta u^{k+1}}{\max(1, w^k + \beta u^{k+1})}; \)
19. **end while**
20. **return** \( u = u^{k+1}. \)

### 3.4. Convergence Analysis

The sequence \( u^{k+1} \) generated by Algorithm 2 converges to the minimum \( u^* \) of \( MTV(u) \).

When \( u \) belongs to the feasible set \( D \), the sequence \( \lambda^k \) converges to the Lagrange multiplier \( \lambda^* \). Through the primal problem (18) and the dual problem (22), \( \Phi(u; \lambda^*) = \max_{p \in P, y \in Y, w \in W} L(u, p, y, w; \lambda^*) \) can be obtained. Therefore, when the dual variables satisfy the constraints \( p \in P, y \in Y, w \in W \), it is only necessary to prove that the sequence \( (u^k, p^k, y^k, w^k) \) generated by Algorithm 2 converges to a saddle point of the func-
tion \(L(u, p, y, w; \lambda^*)\). Firstly, two lemmas are introduced to aid in proving the convergence of the algorithm.

**Lemma 4.** Let the sequence generated by Algorithm 2 be \((u^k, p^k, y^k, w^k, \lambda^k)\), then

\[
\|u^{k+1} - u^k\|_2^2 \geq \max\left( \frac{1}{2s_1^2k(v_0^2 + v_1^2 + \cdots + v_{k-1}^2)} (u^{k+1} - u^k, (-1)^k \text{div}^a(p^{k+1} - p^{k+\frac{1}{2}})), \frac{1}{24s_2^2} (u^{k+1} - u^k, \text{div}^2(y^{k+1} - y^{k+\frac{1}{2}})), \right) \frac{1}{s^2} (u^{k+1} - u^k, (w^{k+1} - w^{k+\frac{1}{2}})) \right).
\]

(53)

**Proof.** Let \(q_1 = p^k + s_1^2 \nabla^a u^{k+1}\), \(q_2 = p^k + s_1^2 \nabla^a u^k\). According to the properties of div and \(\nabla\), the following can be obtained:

\[
s_1^2 (u^{k+1} - u^k, (-1)^k \text{div}^a(p^{k+1} - p^{k+\frac{1}{2}})) = s_1^2 (\nabla^a(u^{k+1} - u^k), \rho_p(q_1) - \rho_p(q_2)) = \langle q_1 - q_2, \rho_p(q_1) - \rho_p(q_2) \rangle.
\]

Using the classical inequality for any vector: \(2a^T b \leq \|a\|_2^2 + \|b\|_2^2\), it can be obtained that

\[
\langle q_1 - q_2, \rho_p(q_1) - \rho_p(q_2) \rangle \leq \frac{1}{2}(\|q_1 - q_2\|_2^2 + \|\rho_p(q_1) - \rho_p(q_2)\|_2^2) \leq \|q_1 - q_2\|_2^2 = (s_1^2)^2 \|\nabla^a(u^{k+1} - u^k)\|_2^2.
\]

Let \(v_i = (-1)^i C_i^a\), it can be obtained that

\[
\|\nabla^a u\|^2 = \sum_{ij} [(\nabla^a u^1)^2_{ij} + (\nabla^a u^2)^2_{ij}]
\]

\[
= \sum_{ij} [(v_0 u^1_{ij} + \cdots + v_{k-1} u^1_{i+k-1,j})^2 + (v_0 u^2_{ij} + \cdots + v_{k-1} u^2_{i+k-1,j})^2]
\]

\[
\leq 2k \sum_{ij} [(v_0 u^1_{ij})^2 + (v_0 u^2_{ij})^2 + \cdots + (v_{k-1} u^1_{i+k-1,j})^2 + (v_{k-1} u^2_{i+k-1})^2]
\]

\[
\leq 2k(v_0^2 + v_1^2 + \cdots + v_{k-1}^2) \|u\|^2.
\]

Consequently,

\[
\|u^{k+1} - u^k\|_2^2 \geq \frac{1}{2s_1^2k(v_0^2 + v_1^2 + \cdots + v_{k-1}^2)} (u^{k+1} - u^k, (-1)^k \text{div}^a(p^{k+1} - p^{k+\frac{1}{2}})),
\]

where \(s_1 \neq 0\). Similarly, it follows that

\[
\|u^{k+1} - u^k\|_2^2 \geq \frac{1}{24s_2^2} (u^{k+1} - u^k, \text{div}^2(y^{k+1} - y^{k+\frac{1}{2}})), \quad \xi_2 \neq 0,
\]

\[
\|u^{k+1} - u^k\|_2^2 \geq \frac{1}{s^2} (u^{k+1} - u^k, (w^{k+1} - w^{k+\frac{1}{2}})).
\]

The \(\|\nabla^2 u\|_2^2\) appearing in the solution process is expressed as
\[ \| \nabla^2 u \|^2_2 = \sum_{ij} (\nabla^2 u_{ij})^2 + (\nabla^2 u_{ij})^2 \]
\[ \leq 2 \sum_{ij} (u_{i+1,j} - 2u_{ij} + u_{i-1,j})^2 + (u_{i,j+1} - 2u_{ij} + u_{i,j-1})^2 \]
\[ \leq 24 \| u \|^2_2. \]

In summary,

\[ \| u^{k+1} - u^k \|^2_2 \geq \max \left\{ \frac{1}{2s\xi_1 k(t_0^2 + t_1^2 + \cdots + t_{k-1}^2)} (u^{k+1} - u^k, (\nabla \phi^1(p^{k+1} - p^{k+1/2})), \frac{1}{24s\xi_2} (u^{k+1} - u^k, \text{div}^2(y^{k+1} - y^{k+1/2})), \frac{1}{s^2} \right\} \]
\[ \| u^{k+1} - u^k, w^{k+1} - w^{k+1/2} \|, \]

which is proved. \[ \square \]

**Lemma 5.** Let the sequence generated by Algorithm 2 be \((u^k, p^k, y^k, w^k, \lambda^k)\), then

\[ \frac{1}{l} \| u^* - u^k \|^2_2 + \frac{1}{s} \| p^* - p^k \|^2_2 + \frac{1}{s} \| y^* - y^k \|^2_2 + \frac{1}{s} \| w^* - w^k \|^2_2 \geq \frac{1}{l} \| u^* - u^{k+1} \|^2_2 + \frac{1}{s} \| p^* - p^{k+1} \|^2_2 + \frac{1}{s} \| y^* - y^{k+1} \|^2_2 + \frac{1}{s} \| w^* - w^{k+1} \|^2_2. \] (54)

In particular, the sequence \((u^k, p^k, y^k, w^k, \lambda^k)\) converges to the limit point \((u^*, p^*, y^*, w^*, \lambda^*)\), at which point there are two cases, if \(u^*\) belongs to the feasible set \(D\), then \(\lambda^* = 0\). Conversely, \(\lambda^*\) takes the solution of the equation \(\| Hu^*(\lambda) - f \|^2_2 = c^2\).

See [40] for a related proof.

Next, prove that a saddle point of \(L(u, p, y, w; \lambda^*)\) is \((u^*, p^*, y^*, w^*)\), when \(p^* \in P, y^* \in Y, w^* \in W\), it follows from Algorithm 2 that \((u^*, p^*, y^*, w^*)\) satisfies the following equation:

\[ u^* = (t\lambda^* H^T H + I)^{-1} (t\lambda^* H^T f + u^* - t\xi_1 (\nabla u)^T p^* - t\xi_2 (\nabla^2 u)^T y^* - t\beta w^*). \] (55)

Further collation yields,

\[ \lambda^* H^T (Hu^* - f) + \xi_1 (\nabla u)^T p^* + \xi_2 (\nabla^2 u)^T y^* + \beta w^* = 0. \] (56)

This process is also the process of replacing \(\lambda\) with \(\lambda^*\). Starting from Algorithm 2, it is known that \((u^*, p^*, y^*, w^*)\) also satisfies the following equation system:

\[
\begin{cases}
    p^* = \rho_p(p^* + s\xi_1 \nabla^2 u^*), \\
    y^* = \rho_y(y^* + s\xi_2 \nabla^2 u^*), \\
    w^* = \rho_w(w^* + s\beta u^*). 
\end{cases}
\] (57)
The solution of the projection operator is shown in (38), which thus yields

\[
\begin{align*}
    p_{i,j}^* &= \frac{\beta_{i,j} + s_1^2(\nabla u^*)_{i,j}}{\max(1, |p_{i,j}^*|)} \\
    y_{i,j}^* &= \frac{\lambda_{i,j} + s_2^2(\nabla^2 u^*)_{i,j}}{\max(1, |y_{i,j}^*|)} \\
    w_{i,j}^* &= \frac{\alpha_{i,j} + s_3^2u_{i,j}}{\max(1, |w_{i,j}^*|)}.
\end{align*}
\]

(58)

If \( |p_{i,j}^* + s_1^2(\nabla u^*)_{i,j}| \leq 1 \), then \((\nabla u^*)_{i,j} = 0\) and \( |p_{i,j}^*| \leq 1 \). If \( |p_{i,j}^* + s_1^2(\nabla u^*)_{i,j}| > 1 \), then \( p_{i,j}^* = 1 \) and \( s_1^2(\nabla u^*)_{i,j} = (1 - |p_{i,j}^* + s_1^2(\nabla u^*)_{i,j}|)p_{i,j}^* \). Similarly, \( y_{i,j}^* \), \( w_{i,j}^* \) have a similar property. Therefore, the \( p^* \in P \), \( y^* \in Y \), \( w^* \in W \) and the KKT optimality condition of (32) hold.

Thus, the following theorem can be obtained.

**Theorem 1.** When \( \lambda^* \) is the Lagrange multiplier on the constraint \( u \in D \), assuming that the saddle point of \( L(u, p, y, w; \lambda) \) is \((u^*, p^*, y^*, w^*)\), then the sequence \((u^k, p^k, y^k, w^k; \lambda^k)\) generated by Algorithm 2 converges to \((u^*, p^*, y^*, w^*; \lambda^*)\). At this time, \( s \cdot t \leq 1/16 \). In particular, \( u^k \) converges to the minimum, \( \lambda^k \) converges to the Lagrange multiplier satisfying the constraint \( u \in D \).

4. Simulation Results

4.1. Quantitative Evaluation Index Description

In this paper, the peak signal-to-noise ratio (PSNR) and structural self-similarity (SSIM) are used to assess the quality of denoised images. PSNR is used to evaluate the quality similarity between two images, while SSIM measures the structural similarity between two images. Here, the PSNR is defined as

\[
\text{PSNR} = 10\log_{10}\left(\frac{b^2}{\text{MSE}}\right).
\]

(59)

where \( b = \max(m, n) \) and \( m, n \) denotes the size of the image. MSE denotes the mean square error between the two images, which is defined as

\[
\text{MSE} = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} (u_0(i,j) - u(i,j))^2.
\]

(60)

Here, \( u_0, u \) represents the original image and the denoised image, respectively. The definition of SSIM is as follows:

\[
\text{SSIM}(u_0, u) = \frac{(2A_{u_0}A_u + d_1) (2\text{cov}(u_0, u) + d_2)}{(A_{u_0}^2 + A_u^2 + d_1) (B_{u_0}^2 + B_u^2 + d_2)}.
\]

(61)

where \( A \) denotes the mean, \( \text{cov}(u_0, u) \) denotes the standard deviation, and \( d_1 \) and \( d_2 \) are two constants, and it is commonly recommended based on experience to set \( d_1 = 0.01 \) and \( d_1 = 0.03 \). To verify the feasibility of the algorithm, a set of images is first introduced as test images, as depicted in Figure 2.
4.2. The Influence of Fractional-Order $\alpha$

The regularization parameter $\lambda$ and the fractional-order $\alpha$ are both indefinite parameters in image denoising algorithms, but they both play a significant role. The effect of the fractional-order $\alpha$ on the performance of the algorithm is investigated by adding Gaussian noise with the noise standard deviation of $\sigma = 10, 20$ to the test images Lena, Boats, and Pepper. From Figure 3, it can be observed that when $\sigma = 10$, with the increase of $\alpha$, PSNR shows a decreasing trend. When $\sigma = 20$, with the increase of $\alpha$, PSNR shows an increasing trend. The value near $\alpha = 1.5$ can be considered to ensure that the changing trend of PSNR has certain adaptability to different noise standard deviations. From Figure 4, it can be observed that if $\alpha$ is small, the image is prone to a blocky effect, and if $\alpha$ is large, the image is not well denoised. To effectively suppress the phenomenon observed in Figure 4, it is also possible to consider selecting values around $\alpha = 1.5$. In conclusion, it is preferable to select $\alpha$ from the range of $\alpha \in [1.4, 1.6]$. 

Figure 2. Original images. (a) Lena, (b) Parrots, (c) Pepper, (d) Barbara, (e) Boats, (f) Cameraman, (g) Foot, (h) Head.

Figure 3. (a–c) represent the PSNR curves of Boats, Lena, and Pepper under different $\alpha$ when $\sigma = 10$, respectively. (d–f) represent the PSNR curves of Boats, Lena, and Pepper under different $\alpha$ when $\sigma = 20$, respectively.
4.3. The Influence of $\lambda$ on Algorithm Performance

The regularization parameter $\lambda$ is also known to affect the denoising performance of the algorithm. Hence, Gaussian noise with $\sigma = 30, 40$ is added to the Lena, Boats, and Pepper images to verify the impact of the regularization parameter $\lambda$ on the performance of the algorithm. From Figure 5, it can be observed that when manually selecting the parameter $\lambda$, the PSNR and SSIM of each algorithm are not satisfactory. However, when using the adaptive parameter $\lambda$ selection method proposed in Section 3.3, the algorithms achieve higher PSNR and SSIM. Therefore, it can be concluded that the adaptive parameter $\lambda$ proposed can improve the performance of the algorithm to a certain extent.

4.4. Comparison with Other Methods

To further prove the denoising performance of the new algorithm, a first set of comparative experiments was conducted using test images Parrots, Lena, Barbara, Cameraman, Pepper, Foot, and Head, in comparison with some state-of-the-art methods [48–50]. Gaussian noise with $\sigma = 5, 10, 20, 30$ was added to the images, and the PSNR and SSIM of each algorithm after denoising are shown in Table 1. Figures 6 and 7 demonstrate the denoising effects of the Pepper and Foot images with $\sigma = 30$. Overall, the algorithm proposed can achieve better results.

Next, a second set of comparative experiments [51–53] was conducted by adding Gaussian noise with $\sigma = 10, 20, 30$ to Cameraman, Boats, Barbara, and Pepper. The experimental results are shown in Table 2. Based on the experimental results, it can be observed
that the results obtained in this study are slightly higher than those of other algorithms. In conclusion, the algorithm proposed is effective in a certain sense.

![Noise image](a) PSNR=22.5738, SSIM=0.8937
![Pepper denoising effect image of [48]](b) PSNR=24.4121, SSIM=0.8939
![Pepper denoising effect image of [49]](c) PSNR=25.4720, SSIM=0.8950
![Pepper denoising effect image of [50]](d) PSNR=29.0906, SSIM=0.9016
![Pepper denoising effect image of the proposed algorithm.](e)

Figure 6. (a) Noise image of $\sigma = 30$, (b) Pepper denoising effect image of [48], (c) Pepper denoising effect image of [49], (d) Pepper denoising effect image of [50], (e) Pepper denoising effect image of the proposed algorithm.

Table 1. The PSNR and SSIM of each algorithm under the first set of comparative experiments.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Parrots</td>
<td>5</td>
<td>32.8362/0.9071</td>
<td>32.3292/0.9050</td>
<td>34.3292/0.9058</td>
<td>34.8288/0.9080</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>27.7650/0.9085</td>
<td>31.4979/0.9005</td>
<td>33.4346/0.9089</td>
<td>33.6375/0.9101</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>23.9519/0.9146</td>
<td>28.2574/0.9162</td>
<td>28.5716/0.9218</td>
<td>28.5996/0.9241</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>22.7536/0.9201</td>
<td>23.1230/0.9204</td>
<td>24.2050/0.9277</td>
<td>26.3916/0.9279</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>33.1229/0.9023</td>
<td>32.2596/0.8938</td>
<td>33.1639/0.8992</td>
<td>33.4520/0.9000</td>
</tr>
<tr>
<td>Lena</td>
<td>10</td>
<td>27.7962/0.8963</td>
<td>31.4268/0.8974</td>
<td>32.4873/0.9023</td>
<td>32.5379/0.9011</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>24.0047/0.9144</td>
<td>28.0919/0.9050</td>
<td>28.1660/0.9141</td>
<td>28.4511/0.9153</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>23.9741/0.8973</td>
<td>24.2123/0.9075</td>
<td>24.8134/0.8896</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>33.1016/0.8908</td>
<td>29.2736/0.8904</td>
<td>31.0013/0.8909</td>
<td>31.0157/0.8924</td>
</tr>
<tr>
<td>Barbara</td>
<td>10</td>
<td>27.7793/0.8954</td>
<td>28.8140/0.8973</td>
<td>30.0985/0.8974</td>
<td>30.1247/0.8978</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>23.9799/0.9024</td>
<td>26.7437/0.9033</td>
<td>26.5292/0.9051</td>
<td>26.5692/0.9133</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>22.1423/0.8937</td>
<td>22.9751/0.9102</td>
<td>22.6868/0.9188</td>
<td>22.9827/0.9124</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>33.1716/0.9021</td>
<td>29.8097/0.8991</td>
<td>31.8976/0.9000</td>
<td>32.2383/0.9006</td>
</tr>
<tr>
<td>Cameraman</td>
<td>10</td>
<td>27.8417/0.9056</td>
<td>29.3846/0.9015</td>
<td>30.3292/0.9034</td>
<td>31.4454/0.9059</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>23.9418/0.9140</td>
<td>27.0952/0.9084</td>
<td>27.6781/0.9082</td>
<td>27.7247/0.9144</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>22.8149/0.8902</td>
<td>23.1143/0.8907</td>
<td>23.5803/0.8905</td>
<td>24.6137/0.8911</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>34.6851/0.8698</td>
<td>32.6693/0.8620</td>
<td>34.5614/0.8713</td>
<td>34.7247/0.8729</td>
</tr>
<tr>
<td>Pepper</td>
<td>10</td>
<td>28.8834/0.8764</td>
<td>32.0109/0.8689</td>
<td>32.9939/0.8786</td>
<td>33.8284/0.8767</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>24.2617/0.8853</td>
<td>28.6892/0.8808</td>
<td>29.1745/0.8902</td>
<td>29.3960/0.8913</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>22.5736/0.8937</td>
<td>24.4121/0.8938</td>
<td>25.4702/0.8950</td>
<td>29.0906/0.9016</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>37.5561/0.8856</td>
<td>37.4102/0.8524</td>
<td>35.4676/0.8713</td>
<td>35.7180/0.8561</td>
</tr>
<tr>
<td>Foot</td>
<td>10</td>
<td>33.8384/0.8340</td>
<td>33.9815/0.8312</td>
<td>34.0148/0.8311</td>
<td>34.6360/0.8350</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>27.1051/0.8286</td>
<td>28.2964/0.8279</td>
<td>28.7614/0.8253</td>
<td>29.8339/0.8272</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>25.4720/0.8050</td>
<td>25.6541/0.8144</td>
<td>25.9459/0.8193</td>
<td>27.5594/0.8074</td>
</tr>
</tbody>
</table>
The proposed MOTV model plays a significant role in improving the staircase effect and eliminating residual noise.

Table 1. Cont.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Head</td>
<td>5</td>
<td>33.5699/0.8953</td>
<td>33.4973/0.8874</td>
<td>34.6343/0.9030</td>
<td>33.6142/0.8423</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>32.3203/0.8470</td>
<td>31.0355/0.8309</td>
<td>32.1153/0.8576</td>
<td>32.3258/0.8310</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>22.9786/0.7933</td>
<td>22.8761/0.7823</td>
<td>27.6259/0.8117</td>
<td>28.1583/0.7446</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>19.4785/0.7403</td>
<td>19.4509/0.7371</td>
<td>23.3967/0.7649</td>
<td>24.5390/0.6468</td>
</tr>
</tbody>
</table>

Table 2. The PSNR and SSIM of each algorithm under the second set of comparative experiments.

<table>
<thead>
<tr>
<th>Image</th>
<th>$\sigma$</th>
<th>PSNR/SSIM [51]</th>
<th>PSNR/SSIM [52]</th>
<th>PSNR/SSIM [53]</th>
<th>PSNR/SSIM Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cameraman</td>
<td>10</td>
<td>31.3925/0.8948</td>
<td>33.1347/0.9049</td>
<td>25.0623/0.9089</td>
<td>33.6375/0.9101</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>28.2944/0.8272</td>
<td>28.5723/0.8309</td>
<td>24.9803/0.9218</td>
<td>28.5996/0.9241</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>27.1039/0.7700</td>
<td>27.2705/0.7315</td>
<td>24.8771/0.9277</td>
<td>26.3916/0.9279</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>32.3014/0.9017</td>
<td>32.3436/0.9065</td>
<td>25.6609/0.9032</td>
<td>32.4058/0.8889</td>
</tr>
<tr>
<td>Boats</td>
<td>20</td>
<td>28.6036/0.8802</td>
<td>28.7042/0.8840</td>
<td>25.5432/0.9019</td>
<td>28.3939/0.9011</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>27.5687/0.8204</td>
<td>27.5816/0.8253</td>
<td>25.3779/0.9087</td>
<td>26.0102/0.9112</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>30.0487/0.9186</td>
<td>31.8501/0.9340</td>
<td>22.0302/0.8974</td>
<td>30.1247/0.8978</td>
</tr>
<tr>
<td>Barbara</td>
<td>20</td>
<td>26.1845/0.9259</td>
<td>26.4952/0.9247</td>
<td>21.9057/0.9188</td>
<td>22.9827/0.9124</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>25.7061/0.8790</td>
<td>25.7384/0.8729</td>
<td>21.9057/0.9188</td>
<td>22.9827/0.9124</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>33.3949/0.8950</td>
<td>33.9108/0.9032</td>
<td>28.8618/0.9034</td>
<td>34.7247/0.8729</td>
</tr>
<tr>
<td>Pepper</td>
<td>20</td>
<td>29.2765/0.8593</td>
<td>29.5572/0.8897</td>
<td>28.6174/0.8996</td>
<td>29.3960/0.8913</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>28.2187/0.8746</td>
<td>28.2854/0.8731</td>
<td>28.4424/0.8905</td>
<td>29.0906/0.9016</td>
</tr>
</tbody>
</table>

5. Conclusions

In summary, the achievements obtained in this article are as follows:

* The proposed MOTV model plays a significant role in improving the staircase effect and eliminating residual noise.
The appropriate fractional-order $\alpha$ and the selection of $\lambda$ based on the deviation principle have a certain effect on the performance of the algorithm. According to the results of two groups of comparative experiments, it can be concluded that the algorithm proposed outperforms other algorithms and is effective.

In fact, the algorithm proposed also has certain limitations. Specifically, when the noise variance increases, the noise suppression effect of the algorithm may decrease. In future research, it will be considered by the authors to incorporate the ideas of this paper into other methods mentioned in the introduction section. Additionally, the application of regularizers will be considered to achieve better results. Furthermore, efforts will be made to reduce the computation time.

Author Contributions: Conceptualization, S.B. and M.L.; methodology, S.B.; software, S.B. and M.L.; validation, S.B.; formal analysis, S.B.; investigation, S.B.; resources, S.B.; data curation, S.B., G.C. and M.L.; writing—original draft, S.B.; writing—review and editing, S.B., visualization, S.B., supervision, S.B., G.C. and M.L.; project administration, S.B.; funding acquisition, G.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the National Natural Science Foundation of China under Grant No. 11461037, and the High-Quality Postgraduate Courses of Yunnan Province under Grant No. 109920210027.

Data Availability Statement: Not applicable.

Acknowledgments: The authors thank the National Natural Science Foundation of China under Grant No. 11461037, and the High-Quality Postgraduate Courses of Yunnan Province under Grant No. 109920210027.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References
32. Harris, C.; Stephens, M. A combined corner and edge detector. *Alvey Vis. Conf.* 1998, 15, 10–5244. [CrossRef]

49. Li, M.; Cai, G.; Bi, S.; Zhang, X. Improved TV image denoising over inverse gradient. *Symmetry* 2023, 15, 678. [CrossRef]


52. Zhang, X. Center pixel weight based on wiener filter for non-local means image denoising. *Optik* 2021, 244, 167557. [CrossRef]


**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.