Research Article

Fractional-Order Total Variation Image Restoration Based on Primal-Dual Algorithm

Dali Chen,1 YangQuan Chen,2 and Dingyu Xue1

1 College of Information Science and Engineering, Northeastern University, Shenyang, Liaoning 110006, China
2 MESA Lab, University of California, Merced, 5200 North Lake Road, Merced, CA 95343, USA

Correspondence should be addressed to Dali Chen; chendali@ise.neu.edu.cn

Received 7 August 2013; Accepted 27 September 2013

Academic Editor: Dumitru Baleanu

Copyright © 2013 Dali Chen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper proposes a fractional-order total variation image denoising algorithm based on the primal-dual method, which provides a much more elegant and effective way of treating problems of the algorithm implementation, ill-posed inverse, convergence rate, and blocky effect. The fractional-order total variation model is introduced by generalizing the first-order model, and the corresponding saddle-point and dual formulation are constructed in theory. In order to guarantee \( O(1/N^2) \) convergence rate, the primal-dual algorithm was used to solve the constructed saddle-point problem, and the final numerical procedure is given for image denoising. Finally, the experimental results demonstrate that the proposed methodology avoids the blocky effect, achieves state-of-the-art performance, and guarantees \( O(1/N^2) \) convergence rate.

1. Introduction

Since the work of Rudin et al. [1], total variation (TV) minimization problems arise in many image processing applications for regularizing inverse problems where one expects the recovered image or signal to be piecewise constant [2–5]. The typical total variation model, for example, ROF model, has been proved to be able to achieve a good tradeoff between edge preservation and noise removal [1]. However, it tends to produce the so-called blocky (staircase) effects on the images because it favors a piecewise constant solution in bounded variation (BV) space [6]. In order to deal with blocky effects, the modification of TV model, which generalizes the differential order in regularization term, has aroused the more and more attentions of numerous scholars.

The improved methods of TV modelization are divided into two kinds: the high-order derivative and the fractional-order derivative. The first one replaces the first-order derivative in regularization term by the high-order derivative. For example, a fourth-order partial differential equation-(PDE-based) denoising mode was proposed by [7], in which the regularized solution is obtained by solving the minimization of potential function of second-order derivative of the image. It has been proved that this model is able to deal with the blocky problem. However, it tends to cause the sign of uplifting effect and formation of artifacts around edges. For this problem, an improved fourth-order PDE model, replacing the Laplacian operator of diffusivity function by the gradient operator, was proposed in [8].

In this paper, our interest focuses on the second generalization which deals with fractional-order differentiation. Fractional calculus is a rapidly growing mathematical discipline, which provides an important tool for nonlocal field theories [9]. Recently, it has been greatly studied in computer vision [10–16], and the main reason for this development is the expectation that the use of this theory will lead to a much more elegant and effective way of treating problems of blocky effect and detailed information protection. Specially, the fractional-order total variation (TV) models play an important role for image denoising, inpainting, and motion estimation [17–21]. So far, the methods adopted to deal with the fractional-order total variation problem are divided into three kinds: (1) solving the associated Euler-Lagrange equation, which is a nonlinear partial differential equation (PDE) [17, 18]; (2) using the methods based on duality [6]; (3) using the method based on majorization-minimization...
(MM) algorithm [22]. Although the convergence of these methods has been studied in the corresponding literatures, the rate of convergence for them has not been discussed yet. For this problem, in this paper, a primal-dual algorithm is used to solve the fractional-order total variation problem, which is able to guarantee $O(1/N^2)$ convergence rate.

The primal-dual algorithm was first presented by [23] and named as the primal-dual hybrid gradient (PDHG) method in [24]. In this method, each iteration updates both a primal and a dual variable. It is thus able to avoid some of the difficulties that arise when working only on the primal or dual side [25, 26]. The convergence of the primal-dual algorithm has been studied in [27, 28]. Recently a unified approach was studied in [29, 30], which demonstrated that, in some cases, these algorithms can achieve the rate of convergence.

This paper is organized as follows. Section 1 introduces prior work, focusing on the main problems with existing methods that are addressed by our model. In Section 2, the fractional-order total variation denoising model is described, and the corresponding saddle-point and dual formulation are constructed in theory. Based on this, the primal-dual algorithm was used to solve the constructed saddle-point problem, and the final numerical implementation is presented for image denoising. Experimental evaluation is presented in Section 3 and the paper is concluded in Section 4.

2. Fractional-Order Total Variation Model and Primal-Dual Algorithm

2.1. Model Description. Let $f_{i,j} = u_{i,j} + v_{i,j}$ denote the observed noisy image, where $(i, j)^T$ denotes the location with a rectangular image domain $\Omega \in \mathbb{R}^{N \times N}$ and $v$ is white Gaussian noise. The typical total variation (TV) denoising model estimates the desired clean image $u_{i,j}$ by solving the following finite-dimensional optimization problem:

$$
\hat{u} = \arg\min_{u \in \Omega} \left\{ E(u) := TV(u) + \frac{\lambda}{2} \| f - u \|_2^2 \right\},
$$

where $TV(u)$ is the regularization term, $\| f - u \|_2^2$ is the data fidelity term, $\| u \|_2$ is the $\ell_2$-norm of $u$, and $\lambda$ is regularization parameter which controls the degree of smoothing. In this paper, we consider a fractional-order total variation model, defined as

$$
\hat{u} = \arg\min_{u \in \Omega} \left\{ E(u) := TV^\alpha (u) + \frac{\lambda}{2} \| f - u \|_2^2 \right\},
$$

where $TV^\alpha(u)$ is obtained by the following formula:

$$
TV^\alpha (u) = \sqrt{(D_1^\alpha u)^2 + (D_2^\alpha u)^2},
$$

where $\sqrt{(D_1^\alpha u)^2 + (D_2^\alpha u)^2}$ is the discretization of $|D^\alpha u|$, $D^\alpha$ is the fractional-order derivative operator, and $D_1^\alpha u := (D_1^\alpha u, D_2^\alpha u)^T$. $D_1^\alpha$ and $D_2^\alpha$ are linear operators corresponding to horizontal and vertical fractional-order derivative.

From Grünwald-Letnikov fractional derivative definition [31], the finite fractional-order forward difference can be obtained by

$$
D_1^\alpha u_{i,j} = \sum_{k=0}^{K-1} w_k^{(\alpha)} u_{i,j+k}, \quad D_2^\alpha u_{i,j} = \sum_{k=0}^{K-1} w_k^{(\alpha)} u_{i,j+k},
$$

where $w_k^{(\alpha)} = (-1)^{k+1} \frac{\Gamma(\alpha+1)}{\Gamma(k+1)\Gamma(\alpha-k+1)}$ denotes the generalized binomial coefficient, and $\Gamma(x)$ is the Gamma function. In addition, the coefficients $w^{(\alpha)}$ can also be obtained recursively from

$$
\begin{align*}
&\begin{cases} 
\frac{\Gamma(\alpha+1)}{\Gamma(k+1)\sqrt{\alpha}} & \text{if } m \leq N-\alpha k, \\
0 & \text{otherwise}
\end{cases} \\
&\sum_{k=0}^{K-1} w_k^{(\alpha)} x_{m+nk}^T,
\end{align*}
$$

for $m = 1, 2, \ldots, N$.

When $\alpha = 1$, $w_k^{(\alpha)} = 0$ for $k > 1$ and (4) is the first-order forward derivative as usual. To describe (2) in matrix algebra language, we reord the image matrix $u$ and $f$ row wisely into the vector $x$ and $z$, associating the $(i, j)$ element of the two-dimensional structure to the element $(m, n) = (i, j)$ of the vector structure, $x_{(m,n)} = u_{i,j}$, and $z_{(m,n)} = f_{i,j}$. We have $x \in \mathbb{R}^N$ and $z \in \mathbb{R}^N$, $N = n^2$. The $(i, j)$ component of the fractional-order derivative $D^\alpha u$ can thus be represented as a multiplication of the vector $x \in \mathbb{R}^N$ by a matrix $A_m \in \mathbb{R}^{2\times N}$, for $m = 1, 2, \ldots, N$:

$$
\begin{align*}
&\begin{cases} 
\frac{\Gamma(\alpha+1)}{\Gamma(k+1)\sqrt{\alpha}} & \text{if } m \leq N-\alpha k, \\
0 & \text{otherwise}
\end{cases} \\
&\sum_{k=0}^{K-1} w_k^{(\alpha)} x_{m+nk}^T,
\end{align*}
$$

From this definition, the discrete version of the prime fractional-order total variation (FOTV) model (2) can be written as

$$
\min_{x \in \mathbb{R}^N} \left\{ E(x) := \sum_{m=1}^{N} \| A_m x \|_2 + \frac{\lambda}{2} \| x - z \|_2^2 \right\}.
$$
Define a norm \( \|v\|_E = \sum_{m=1}^{2N} \sqrt{v_m^2 + v_{2N}^2} \) where \( v = (v_1, v_2)^T \), and (7) can be rewritten in terms of \( \| \cdot \|_E \) as
\[
\min_{x \in \mathbb{R}^{2N}} \left\{ E(x) := \| A x \|_E + \frac{\lambda}{2} \| x - z \|_2^2 \right\}, \tag{8}
\]
where \( \lambda = 1 \) is the corresponding discretization of the fractional-order divergence, and when \( \alpha = 1 \), it is the discrete divergence as usual.

For primal-dual feasible pair \((x, y)\), the partial primal-dual gap \( G(x, y) \) is defined by
\[
G(x, y) = \max_{y' \in \mathbb{R}^{2N}} \left\{ \langle y', Ax \rangle - \delta (y') + \frac{\lambda}{2} \| x - z \|_2^2 \right\} - \min_{x \in \mathbb{R}^{2N}} \left\{ \langle y, Ax \rangle - \delta (y) + \frac{\lambda}{2} \| x - z \|_2^2 \right\}, \tag{14}
\]
which vanishes only if \((x, y)\) is itself the saddle point [29]. The primal-dual gap \( G(x, y) \) is a measure of closeness of the primal-dual \((x, y)\) to the primal-dual solution, and we use it to design the stopping criterion for our numerical algorithm in this paper.

2.3. Primal-Dual Method for Solving Saddle-Point Problem. The first-order primal-dual method summarized in [29] for convex problem was applied in this paper to solve the saddle-point problem described by (11), which is described by Algorithm 1.

Algorithm 1

1. Choose \( r_0, \sigma_0 > 0 \) with \( r_0 \sigma_0 L^2 \leq 1 \), \( x_0 \in \mathbb{R}^N \), \( y_0 \in \mathbb{R}^{2N} \), and \( \bar{x}_0 = x_0 \).
2. Update \( x_t, y_t, \bar{x}_t, \theta_t, r_t, \sigma_t \), as follows:
\[
y_{t+1} = \arg \min_{y \in \mathbb{R}^{2N}} \langle y, -A \bar{x}_t \rangle + \delta (y) + \frac{1}{2 \sigma_t} \| y - y_t \|_2^2, \tag{A}
\]
\[
x_{t+1} = \arg \min_{x \in \mathbb{R}^{2N}} \frac{\lambda}{2} \| x - z \|_2^2 + \langle A, x_t \rangle + \frac{1}{2 \tau_t} \| x - x_t \|_2^2, \tag{B}
\]
\[
\theta_t = \frac{1}{\sqrt{1+2 \sigma_t}} \quad r_{t+1} = \theta_t r_t, \quad \alpha_{t+1} = \frac{\sigma_t}{\theta_t}.
\]
\[
\bar{x}_{t+1} = x_{t+1} + \alpha_t (x_{t+1} - x_t),
\]
where \(-A^T\) is the corresponding discretization of the fractional-order divergence, and when \( \alpha = 1 \), it is the discrete divergence as usual.

3. Let \( r_t, \sigma_t, L^2 \leq 1 \); then there exists a saddle-point \((\bar{x}, \bar{y})\) such that \( x_t \to \bar{x} \) and \( y_t \to \bar{y} \). Since \( \langle \lambda/2 \rangle \| x - z \|_2^2 \) is uniformly convex such that it has a Lipschitz continuous gradient, the output \((x_t, y_t)\) possesses a rate of convergence of \( O(1/N^2) \) [29]. More specifically, we assume that the constant stepsizes are used; that is, \( \tau_t = \tau > 0, \sigma_t = \sigma > 0 \), and \( \theta_t = \theta \) for all \( t \geq 0 \). If \( \tau \sigma L^2 < 1 \); then the \( O(1/N) \) convergence rate can be guaranteed for \( \theta = 1 \), and \( O(1/\sqrt{N}) \) convergence rate can be guaranteed for \( \theta = 0 \) [30].
To summarize, our entire noise removal algorithm in a form of a pseudocode is done in following:

(1) initialization:
   (i) for a given noisy image \( f \) of size \( n \times n \), we record the image \( f \) row-wise into the vector \( z \in \mathbb{R}^N \) with \( N = n \times n, z_{(j-1)n+i} = f_{i,j} \);
   (ii) compute the fractional-order discrete operator \( A \) according to (6);
   (iii) initialize the regularization parameter \( \lambda \), iteration number \( T \), and maximum permissible error \( E_{\text{mpc}} \);
   (iv) set \( t = 0, x_0 = z, L = \| A \|, \tau_0 = 1/L^2, \sigma_0 = 1/L^2 \), and \( y = 0.7\lambda \);

(2) iteration: compute \( x \) by the following steps.

Step 1. Solve the problem (A) by

\[
y_t^{m+1} = \frac{y_t^m + \sigma_t A_m \bar{x}^t}{\max \{ 1, \| y_t^m + \sigma_t A_m \bar{x}^t \|_2 \}}, \quad m = 1, 2, \ldots, N. \tag{16}
\]

Step 2. Substitute \( y_t^{m+1} \) into the (B) and solve the problem (B) by

\[
x_t^{m+1} = \frac{x_t^m - \tau_t A_T y_t^{m+1} + \tau_t \lambda z}{1 + \tau_t \lambda}. \tag{17}
\]

Step 3. Compute \( \theta_t = 1/\sqrt{1 + 2\tau_t \lambda}, \tau_{t+1} = \tau_t \theta_t, \) and \( \sigma_{t+1} = \sigma_t/\theta_t \).

Step 4. Update \( \bar{x} \) by \( \bar{x}^{t+1} = (1 + \theta_t) x_t^{m+1} - \theta_t x_t^m \).

Step 5. Compute the primal-dual gap \( g_t^{m+1} \) according to (15) as follows:

\[
g_t^{m+1} = \sum_{m} \| A_m x_t^{m+1} \|_2 - \left( y_t^{m+1} \right)^T A x_t^{m+1} + \frac{\lambda}{2} \| x_t^{m+1} - z + \frac{1}{\lambda} A^T y_t^{m+1} \|_2^2. \tag{18}
\]

Step 6. If \( (g_t^{m+1} < E_{\text{mpc}}) \) or \( (t > T) \), then we terminate the iteration and output \( x_t^{m+1} \); otherwise, go back to Step 1.

Due to the simplicity of our model, it is easy to be implemented and can be effectively accelerated on parallel hardware such as field programmable gate array (FPGA) and graphics processing unit (GPU). In order to employ the huge computational power and the parallel processing capabilities of FPGA to obtain a fully accelerated implementation of our denoising method, the numerical method should be working on the regular grids. For this purpose, the implementation of our algorithm on FPGA is described as the following strategy. First, an image frame is captured and read in the processor from a live camera. For the captured frame \( f \in \mathbb{R}^{m \times n} \), the \( k \)th iteration of our method performs the following updates on prime variable \( u \in \mathbb{R}^{m \times n} \) and dual variable \( p \in \mathbb{R}^{m \times n \times 2} \):

\[
p_t^{k+1} = \frac{p_t^k + \sigma_t D^a u_t^k}{\max \{ 1, \| p_t^k + \sigma_t D^a u_t^k \|_2 \}}, \quad \tag{19}
\]

\[
u_t^{k+1} = u_t^k - \tau_t \text{div}^a p_t^{k+1} + \tau_t \lambda f, \quad \tag{20}
\]

where the dual variable \( p = (p^1, p^2)^T \) corresponds with \( y \) and the discrete fractional divergence operator \( \text{div}^a \) is adjoint to the discrete fractional gradient operator \( D^a \). Note that (19) and (20), respectively, correspond with the solutions of problem (A) and (B). The updates of the other parameters are the same as the description in the foregoing pseudocode, and the number of iteration is fixed to ensure the correct timing sequence. After the iteration, the denoised image \( u \) is outputted, the next image frame is captured, and the same process is repeated until the assignment is over.

3.3. Analysis of Denoising Performance. The aim of this section is to analyze the denoising performance of the FOTV...
algorithm. For this purpose, three famous $512 \times 512$ test images, called Barbara, Lena, and Peppers, are used in the experiments, as shown in Figure 3. In order to quantify the denoised image, we consider the peak-signal-to-noise ratio (PSNR), which has been largely used in the literature and commonly applied to determine the quality of a processed image. It can be calculated by the following formula:

\[
\text{PSNR} = 10 \log_{10} \frac{V_L^2MN}{\sum_{i=1}^{N} \sum_{j=1}^{M} (u(i,j) - \hat{u}(i,j))^2},
\]

where $u$ is the original image, $\hat{u}$ is the denoised image, and $V_L$ is the maximal gray level of the image.

Firstly, in order to decide the value of the fractional-order $\alpha$, we study the relation between the PSNR and $\alpha$. Figure 4 shows the relation between the PSNR and $\alpha$ on Lena and Peppers image corrupted by the additive white Gaussian noise with standard deviation (SD) of 20. The left image is the result of Lena image and the right one is the result of Peppers image. From the figure, we can obtain the following conclusions. Firstly, the PSNR reaches a maximum between $\alpha = 1$ and $\alpha = 2$. Secondly, the PSNR at $\alpha = 1$ is lower than PSNR at $\alpha > 1$, which is owing to the blocky effect. Finally, the PSNR decreases rapidly as $\alpha$ tends to zero, and when $\alpha = 0$, the PSNR reaches a minimum. According to these facts, we can select the fractional-order $\alpha$ between $\alpha = 1.2$ and $\alpha = 2$.

In order to analyze the denoising performance, the additive white Gaussian noise (AWGN) with standard deviation (SD) of 10, 20, and 30 is, respectively, added in the three test images. Five denoising algorithms are used to process these given noisy images, which are the improved Perona and Malik (IP-M) model [33], fourth order (F-O) PDE model [7], improved fourth order (IF-O) PDE model [8], ROF model [1], and the proposed FOTV denoising algorithm. In this experiment, we set $\sigma = 5$ and $\Delta t = 0.25$ for IP-M, F-O-PDE, and IF-O-PDE models, and we set $\lambda = 0.07$, $\tau = 0.1$, and $K = 20$ for ROF model. These parameters are able to ensure

\[
\text{Figure 1: Signal processed by FOTV with } \alpha = 1 \text{ and } \alpha = 1.5.
\]
the best denoising performance of the corresponding denoising model. Based on the conclusion of the foregoing experiment, we set $\alpha = 1.8$ for FOTV denoising algorithm, and the denoising results are shown in Table 1. In the table, the first column lists the given image and the second column lists the standard deviation of noisy image. The PSNRs of the denoised images processed by the different denoising algorithms are listed under the corresponding denoising model. The bigger the PSNR is, the better the denoising performance is. For easy observation, the biggest PSNR values are shown in boldface. From the table, it is obvious that the PSNR of our proposed algorithm is bigger than that of the other four algorithms, so we can conclude that our FOTV denoising algorithm outperforms the other algorithms.

In order to further verify the denoising performance of our proposed algorithm, Figure 5 shows the denoised images
Figure 3: Original image: Barbara, Lena, and Peppers.

Figure 4: Relation between the fractional-order $\alpha$ and PSNR on Lena (a) and Peppers (b) image corrupted by the additive white Gaussian noise with standard deviation (SD) of 20.
processed by four denoising algorithms. The first figure is the famous image “Barbara” corrupted by the additive white Gaussian noise with standard deviation (SD) of 20. In order to show the comparison clearly, the partial enlarged view of the noisy image is shown in the second figure. The denoised result of IP-M algorithm is shown in the right-top figure. The second row shows the denoised images processed by IF-O PDE algorithm, ROF algorithm, and our FOTV algorithm. In Figures 5(c) and 5(d), although the detailed information is preserved, there are a lot of noises unremoved. The result in Figure 5(e) looks blocky. Only the result of our model, as shown in Figure 5(f), looks natural and does not produce false edges. So we conclude that our proposed FOTV algorithm is able to achieve a better tradeoff between edge preservation and noise removal.

3.3. The Convergence and the Rate of Convergence. The convergence and the rate of convergence are two important factors for evaluating the performance of the denoising method. In this section, some experiments are given to show the convergence and convergence rate of our proposed fractional-order total variation (FOTV) denoising algorithm.

For this purpose, the proposed FOTV algorithm is, respectively, used to process the image “Lena” corrupted by the additive white Gaussian noise with standard deviation (SD) of 10, 20, and 30, and the PSNR is recorded at each iteration. Figure 6(a) shows the relation between the iteration and PSNR. The horizontal axis of this figure is iteration number and the vertical axis is PSNR. The different colorful curve denotes the result obtained from the different noisy image. The red dotted line is the result of the noisy image with $\sigma = 30$, the green dashed line is the result of the noisy image with $\sigma = 20$, and the red real line is the result of the noisy image with $\sigma = 10$. From the figure, we can see that the PSNR values are stable after 50 iterations, which indicate that our proposed FOTV denoising algorithm is convergent.

In addition, it can be seen that the bigger the variance of white noise is, the longer the stopping times are and the lower the PSNRs are.

Figure 6 plots the convergence of the FOTV algorithm for the above experiment together with the theoretical $O(1/N^2)$ and $O(1/N^3)$ rate. For easy observation, the horizontal axis is the log curve of iteration number and the vertical axis is the log curve of primal-dual gap calculated in Step 5. The dashed line shows the theoretical $O(1/N^2)$ rate and the dotted line shows the theoretical $O(1/N^3)$ rate. The red line shows the convergence of FOTV algorithm in the noisy image with $\sigma = 30$, the green line shows the convergence of FOTV algorithm in the noisy image with $\sigma = 20$, and the blue line shows the convergence of FOTV algorithm in the noisy image with $\sigma = 10$. From this figure, it can be seen that the smaller the variance of white noise is, the faster convergence is. In
addition, we can see that the FOTV algorithm converges at $O(1/N^2)$ rate firstly and $O(1/N^4)$ rate in the end.

4. Conclusion

A fractional-order total variation image denoising algorithm based on primal-dual method was proposed in this paper. The main contributions are as follows: (1) the saddle point and dual formulation of the proposed fractional-order total variation model are constructed in theory; (2) the primal-dual algorithm was used to solve the constructed saddle-point problem, in which $O(1/N^2)$ convergence rate can be guaranteed; (3) the experiments demonstrate that the proposed methodology is able to avoid the blocky effect, achieve state-of-the-art performance, and guarantee $O(1/N^2)$ convergence rate.

Acknowledgments

This work was supported by National Natural Science Foundation of China (no. 61201378), Scientific Research Fund of Liaoning Provincial Education Department (L2012073), and Fundamental Research Funds for the Central Universities (N110304001).

References


