

# Image Denoising with Higher Order Total Variation and Fast Algorithms

Wenchao Zeng<sup>1</sup>, Xueying Zeng<sup>2(✉)</sup>, and Zhen Yue<sup>1</sup>

<sup>1</sup> Jining First People's Hospital, Jining 272011, People's Republic of China  
<sup>2</sup> School of Mathematical Sciences, Ocean University of China, Qingdao 266100, People's Republic of China  
auwizeng@gmail.com

**Abstract.** In this paper, we propose an efficient higher order total variation regularization scheme for image denoising problem. By relaxing the constraints appearing in the traditional infimal convolution regularization, the proposed higher order total variation can remove the staircasing effects caused by total variation as well as preserve sharp edges and finer details well in the restored image. We characterize the solution of the proposed model using fixed point equations (via the proximity operator) and develop convergent proximity algorithms for solving the model. Our numerical experiments demonstrate the efficiency of the proposed method.

**Keywords:** Image denoising · Total variation · Infimal convolution · Proximity operator

## 1 Introduction

Noise removal is a long standing problem in image processing. Many different methods have been proposed for image denoising problem, among which the well-known Rudin-Osher-Fatemi (ROF) total variation (TV) model is one of the most popular methods [8]. This model restores the image by solving a minimization problem

$$\arg \min_u \left\{ \frac{1}{2} \|u - f\|_2^2 + \lambda \text{TV}(u) \right\}, \quad (1)$$

where  $f$  is the observed noisy image,  $\text{TV}(u)$  denotes the total variation of  $u$  and  $\lambda$  is the regularization parameter to balance both terms in (1). The distinctive feature of the total variation is that it can preserve sharp edges well. However, natural images usually contain smooth part that can be destroyed by TV, and therefore, the denoised image by total variation regularization often suffers the undesired staircasing effects in the flat region of the image [10]. Addressing this drawback of the total variation, many improved ROF models have been proposed in the recent literature. These approaches can be roughly divided into two categories. In the first category, some local gradient information is added to the total variation to make it differentiate different features of the image [4, 5, 11]. Another

category to reduce the staircasing effect in the restored image is to incorporate higher order derivatives into the regularization term. One successful approach in this direction was give in [2] who suggested to use the infimal convolution of functionals with first and second order derivatives as regularizer. The resulting image denoising model has the form of

$$\arg \min_u \left\{ \frac{1}{2} \|u - f\|_2^2 + \text{ICTV}(u) \right\}, \tag{2}$$

where  $\text{ICTV}(u)$  is the infimal convolution regularization that will be defined in Sect. 2. Although the staircaing effects caused by total variation can be alleviated to some extent, the infimal convolution regularization often penalizes the smooth regions of the image too much and may blur the sharp edges and finer details of the image.

In this paper, a novel higher order total variation regularization is proposed to remove the staircasing effects while still preserve sharp edges and finer details in the restored image. We first give an equivalent formulation of the traditional infimal convolution regularizer and motivate the proposed higher order total variation by relaxing the constrains appearing in traditional infimal convolution regularizer. This modification will not increase any difficulties for the numerical treatments of the resulting image denoising model. We characterize the solution of the proposed model via fixed points of fixed point equations in which the proximity operators of convex functions are involved. Based on the characterization, the convergent proximity algorithm is developed to solve the model.

The paper is organized as follows. We shall briefly introduce the difference matrices that are used to define total variation and the regularization terms in this paper in Sect. 2. In Sect. 3, we present the higher order total variation regularization term and the proposed image denoising model. We characterize the solution of the model via fixed point equations and develop numerical algorithms for the proposed model in Sect. 4. Numerical experiments showing the performance of our method is given in Sect. 5. Finally, we conclude our paper in Sect. 6.

## 2 Difference Matrices and Total Variation

In this section, we present the precise definition of the regularization terms appearing in model (1) and (2). For the sake of simplicity, we assume the image is an  $n \times n$  square matrix for some integer  $n$ . To describe the problem in matrix algebra language, we treat the image as a vector in  $\mathbb{R}^N$  by sequentially concatenating the columns of the image. Here we assume that  $N = n^2$ . To define the total variation of the image  $u$ , we need a  $n \times n$  backward difference matrix

$$D := \begin{bmatrix} 0 & & & & \\ -1 & 1 & & & \\ & & \ddots & \ddots & \\ & & & & -1 & 1 \end{bmatrix},$$

then  $-D^\top D$  is the second order central difference matrix. Since we reshape the square image into a vector, we can use

$$D_x := I_n \otimes D, \quad D_y := D \otimes I_n$$

as the first order difference matrices for the image, where  $\otimes$  denotes the Kronecker product of the matrices and  $I_n$  denotes the  $n \times n$  identity matrix. Hence, by setting the gradient matrix

$$B_1 := [D_x; D_y]$$

which is formed by stacking two matrices into a single one, and  $\varphi : \mathbb{R}^{2N} \rightarrow \mathbb{R}$  defined for  $z \in \mathbb{R}^{2N}$  as

$$\varphi(z) := \sum_{i=1}^N \sqrt{z_i^2 + z_{i+N}^2}, \tag{3}$$

the composition  $\varphi(B_1 u)$  computes the total variation  $\text{TV}(u)$  appearing in the model (1).

The second order difference matrices for the image can be defined as

$$D_{xx} := I_n \otimes (-D^\top D), \quad D_{yy} := (-D^\top D) \otimes I_n, \quad D_{xy} := (-D^\top) \otimes D, \quad D_{yx} := D \otimes (-D^\top),$$

using which we can get the second order gradient matrix

$$B_2 := [D_{xx}; D_{xy}; D_{yx}; D_{yy}].$$

The first and second order gradient matrices  $B_1$  and  $B_2$  are related by

$$B_2 = R B_1$$

where

$$R = \begin{bmatrix} I_n \otimes -D^t & 0 \\ -D^t \otimes I_n & 0 \\ 0 & I_n \otimes -D^t \\ 0 & -D^t \otimes I_n \end{bmatrix}.$$

Based on our notation, the regularizer  $\text{ICTV}(u)$  in (2) which use the infimal convolution of functionals with first and second order derivatives has the form of

$$\text{ICTV}(u) := \min_{u=u_1+u_2} \{ \lambda_1 \varphi(B_1 u_1) + \lambda_2 \psi(B_2 u_2) \}, \tag{4}$$

where  $\lambda_1$  and  $\lambda_2$  are positive parameters balancing between the first and second order derivatives, and  $\psi : \mathbb{R}^{4N} \rightarrow \mathbb{R}$  is defined  $z \in \mathbb{R}^{4N}$  as

$$\psi(z) := \sum_{i=1}^N \sqrt{\sum_{j=0}^3 z_{i+jN}^2}. \tag{5}$$

### 3 The Proposed Model

In this section, we give the proposed higher order total variation regularization term and the resulting image denoising model. The proposed regularizer can be seen as the improved version of the infimal convolution regularizer in (4). The following lemma presents an equivalent form of the infimal convolution regularizer in (4) which facilitates us to motivate the proposed higher order total variation regularization term.

**Lemma 1.** *For any  $u$ , we have that*

$$\text{ICTV}(u) = \min_{\substack{B_1u = d + x \\ d, x \in \mathcal{R}(B_1)}} \{ \lambda_1\varphi(d) + \lambda_2\psi(Rx) \}, \tag{6}$$

where  $\mathcal{R}(\cdot)$  denotes the range of a matrix.

*Proof.* Clearly, we can conclude that

$$\min_{\substack{B_1u = d + x \\ d, x \in \mathcal{R}(B_1)}} \{ \lambda_1\varphi(d) + \lambda_2\psi(Rx) \} = \min_{B_1u = B_1u_1 + B_1u_2} \{ \lambda_1\varphi(B_1u_1) + \lambda_2\psi(B_2u_2) \}$$

by setting  $d = B_1u_1$  and  $x = B_1u_2$ .

Since  $u = u_1 + u_2$  implies that  $B_1u = B_1u_1 + B_1u_2$ , we can get

$$\min_{u = u_1 + u_2} \{ \lambda_1\varphi(B_1u_1) + \lambda_2\psi(B_2u_2) \} \geq \min_{B_1u = B_1u_1 + B_1u_2} \{ \lambda_1\varphi(B_1u_1) + \lambda_2\psi(B_2u_2) \} \tag{7}$$

On the other hand, for any  $u_1, u_2$  satisfying  $B_1u = B_1u_1 + B_1u_2$ , there exists  $v \in \mathcal{N}(B_1)$  such that  $u = u_1 + u_2 + v$ , and we can have

$$\begin{aligned} \lambda_1\varphi(B_1u_1) + \lambda_2\psi(B_2u_2) &= \lambda_1\varphi(B_1(u_1 + v)) + \lambda_2\psi(B_2u_2) \\ &\geq \min_{u = u_1 + u_2} \{ \lambda_1\varphi(B_1u_1) + \lambda_2\psi(B_2u_2) \} \end{aligned}$$

which implies that

$$\min_{B_1u = B_1u_1 + B_1u_2} \{ \lambda_1\varphi(B_1u_1) + \lambda_2\psi(B_2u_2) \} \geq \min_{u = u_1 + u_2} \{ \lambda_1\varphi(B_1u_1) + \lambda_2\psi(B_2u_2) \}$$

This together with (7) complete the proof.

The proposed higher order total variation regularization term can be seen as the improved version of the infimal convolution regularizer and has the form of

$$\text{HTV}(u) = \min_{B_1u = d + x} \{ \lambda_1\varphi(d) + \lambda_2\psi(Rx) \}. \tag{8}$$

As can be seen from Lemma 1, the difference between the original infimal convolution regularization and its modified version (8) consists in the relaxed conditions on the new regularizer. For  $\text{HTV}(u)$  we no longer have the restriction that

$d, x \in \mathcal{R}(B_1)$  and  $B_1u$  can be decomposed into any components  $d$  and  $x$ . The authors in [9] showed that this modification can generally lead to better numerical results than the original model (4). This modification was given a theoretical fundament in [1] based on tensor algebra and the corresponding new regularizer was called total generalized variation.

Using the higher order total variation in (8) as the regularizer, the proposed image denoising model is

$$\arg \min_u \left\{ \frac{1}{2} \|u - f\|_2^2 + \text{HTV}(u) \right\}, \tag{9}$$

which can be reformulated as an equivalent unconstrained two variables minimization problem

$$\arg \min_{u,d} \left\{ \frac{1}{2} \|u - f\|_2^2 + \lambda_1 \varphi(d) + \lambda_2 \psi(B_2u - Rd) \right\}. \tag{10}$$

Both above models are equivalent in the sense that if the pair  $(\hat{u}, \hat{d})$  is the solution of model (10) then  $\hat{u}$  is the solution of model (9). Therefore, we can solve model (10) as an alternative of directly solving model (9).

### 4 Minimization Algorithm

The main focus of this section is to give numerical proximity algorithm that solve the proposed model (10). Proximity algorithm was proposed in [6] for solving the ROF model and outperformed the benchmarking split Bregman iteration method. The application of the proximity algorithm to other image models can be found in [3, 7]. The main idea of this algorithm is firstly characterize the solution of the model via the fixed point of fixed point equations in which the proximity operators of the convex functions are included, and then develop efficient numerical algorithms via various fixed point iterations. The proximity operator of a real-valued function  $g$  is defined for  $x \in \mathbb{R}^d$  and  $\gamma > 0$  by

$$\text{prox}_{\gamma g}(x) := \arg \min_y \left\{ \frac{1}{2} \|y - x\|_2^2 + \gamma g(y) \right\}$$

For convenience of developing numerical algorithms, we can rewrite the proposed model into a general form of

$$\arg \min_w \{ F(w) + \lambda_2 \psi(Bw) \}, \tag{11}$$

by defining the matrix  $B$ , the vector  $w$  and the function  $F$  as

$$B = [B_2, -R], w = [u; d]$$

and

$$F(w) = \frac{1}{2} \|u - f\|_2^2 + \lambda_1 \varphi(d)$$

respectively. The following proposition characterizes a solution of model (11) in terms of fixed point equations.

**Proposition 1.** *If  $\hat{w}$  is a solution of model (11), then for any  $\sigma, \gamma > 0$ , there exist two vectors  $b, c$  such that*

$$\begin{aligned} \hat{w} &= \text{prox}_{\frac{1}{\sigma}F}((I - \frac{\gamma}{\sigma}B^tB)\hat{w} - \frac{\gamma}{\sigma}B^t(b - c)) \\ c &= \text{prox}_{\frac{1}{\gamma}\psi}(b + B\hat{w}) \\ b &= b + Bw - c \end{aligned} \tag{12}$$

*Conversely, if there exist positive parameters  $\sigma, \gamma$  and vectors  $b, c$  and  $\hat{w}$  satisfying equations in (12), then  $\hat{w}$  is a solution of model (11).*

*Proof.* See Proposition 5 in [3].

According to Proposition 1, the solution of model (10) is characterized by the fixed point equations in (12). Based on these fixed point equations, we can naturally propose an iterative algorithm that generates the sequence  $\{w^k, c^k, b^k\}$  from arbitrary initial vector  $(w^0, c^0, b^0)$ . This iterative procedure may be stated as follows:

$$\begin{cases} w^{k+1} = \text{prox}_{\frac{1}{\sigma}F}((I - \frac{\gamma}{\sigma}B^tB)w^k - \frac{\gamma}{\sigma}B^t(b^k - c^k)) \\ c^{k+1} = \text{prox}_{\frac{1}{\gamma}\psi}(b^k + Bw^{k+1}) \\ b^{k+1} = b^k + Bw^{k+1} - c^{k+1} \end{cases} . \tag{13}$$

The following theorem establishes a convergence result of the proposed iterative scheme (13), its proof can be found in Theorem 4.1 in [3].

**Theorem 1.** *If the positive numbers  $\gamma$  and  $\sigma$  are chosen to satisfy*

$$\frac{\sigma}{\gamma} < \frac{1}{\|B\|_2^2}, \tag{14}$$

*then the sequence  $\{w^k, b^k, c^k\}$  generated by the iterative scheme (13) converges to the solution of model (10).*

Implementing the iterative procedure (13) requires the availability of explicit forms of two proximity operators  $\text{prox}_{\frac{1}{\sigma}F}$  and  $\text{prox}_{\frac{1}{\gamma}\psi}$ . We next present explicit forms of these proximity operators. We first give the closed forms of the proximity operator of  $F$ . Remembering that  $w = [u; d]$  and  $F(w) = \frac{1}{2}\|u - f\|_2^2 + \lambda_1\varphi(d)$ , the corresponding proximity operator of  $F$  in (13) reads

$$\begin{aligned} \begin{bmatrix} u^{k+1} \\ d^{k+1} \end{bmatrix} &= \text{prox}_{\frac{1}{\sigma}F}((I - \frac{\gamma}{\sigma}B^tB)w^k - \frac{\gamma}{\sigma}B^t(b^k - c^k)) = \arg \min_{u,d} \{ \frac{1}{2} \left\| \begin{bmatrix} u \\ d \end{bmatrix} \right. \\ &\quad \left. - \begin{bmatrix} (I - \frac{\gamma}{\sigma}B_2^tB_2)u^k - \frac{\gamma}{\sigma}B_2^t(b^k - c^k) \\ (I - \frac{\gamma}{\sigma}R^tR)d^k + \frac{\gamma}{\sigma}R^t(b^k - c^k) \end{bmatrix} \right\|^2 + \frac{1}{2\sigma}\|u - f\|^2 + \frac{\lambda_1}{\sigma}\varphi(d) \} \end{aligned} \tag{15}$$

The minimization problem in (15) is very easy to compute since it can be solved separately with respect to  $u^{k+1}$  and  $d^{k+1}$ , that is we have

$$\begin{aligned} u^{k+1} &= \frac{1}{1 + \sigma}f + \frac{\sigma}{1 + \sigma}((I - \frac{\gamma}{\sigma}B_2^tB_2)u^k - \frac{\gamma}{\sigma}B_2^t(b^k - c^k)) \\ d^{k+1} &= \text{prox}_{\frac{\lambda_1}{\sigma}\varphi}((I - \frac{\gamma}{\sigma}R^tR)d^k + \frac{\gamma}{\sigma}R^t(b^k - c^k)). \end{aligned} \tag{16}$$

Both of the proximity operators  $\text{prox}_{\frac{1}{\gamma}\psi}$  and  $\text{prox}_{\frac{\lambda_1}{\sigma}\varphi}$  appearing in (13) and (16) have the closed-form solutions respectively. For  $t > 0$  and  $x \in \mathbb{R}^{2N}$  and  $\tilde{x} \in \mathbb{R}^{4N}$ , set  $y := \text{prox}_{t\varphi}(x)$  and  $z := \text{prox}_{t\psi}(\tilde{x})$ , writing  $\bar{x}_i := [x_i; x_{i+N}]$ , and  $\tilde{\bar{x}}_i := [\tilde{x}_i; \tilde{x}_{i+N}; \tilde{x}_{i+2N}; \tilde{x}_{i+3N}]$ , for  $i = 1, 2, \dots, N$ , we have

$$\begin{bmatrix} y_i \\ y_{i+N} \end{bmatrix} = \max \{ \|\bar{x}_i\|_2 - t, 0 \} \frac{\bar{x}_i}{\|\bar{x}_i\|_2} \quad (17)$$

and

$$\begin{bmatrix} z_i \\ z_{i+N} \\ z_{i+2N} \\ z_{i+3N} \end{bmatrix} = \max \left\{ \|\tilde{\bar{x}}_i\|_2 - t, 0 \right\} \frac{\tilde{\bar{x}}_i}{\|\tilde{\bar{x}}_i\|_2} \quad (18)$$

respectively.

In summary, we have the following convergent algorithm for solving problem (10):

---

**Algorithm 1.** Proximity algorithm to solve the proposed model (10).

---

1. Given:  $f$  and positive parameters  $\lambda_1, \lambda_2, \sigma, \gamma$  satisfies  $\frac{\sigma}{\gamma} < \frac{1}{\|B\|^2}$
  2. Initialization:  $u^0 = 0, d^0 = 0, b^0 = 0, c^0 = 0$
  3. For  $k \in \mathbb{N}$ ,
    - (a)  $u^{k+1} = \frac{1}{1+\sigma}f + \frac{\sigma}{1+\sigma}((I - \frac{\gamma}{\sigma}B_2^t B_2)u^k - \frac{\gamma}{\sigma}B_2^t(b^k - c^k))$   
 $d^{k+1} = \text{prox}_{\frac{\lambda_1}{\sigma}\varphi}((I - \frac{\gamma}{\sigma}R^t R)d^k + \frac{\gamma}{\sigma}R^t(b^k - c^k))$
    - (b)  $c^{k+1} = \text{prox}_{\frac{1}{\gamma}\psi}(b^k + B_2 u^{k+1} - R d^{k+1})$
    - (c)  $b^{k+1} = b^k + B_2 u^{k+1} - R d^{k+1} - c^{k+1}$
    - (d) Stop, if  $u^{k+1}$  converges or satisfies a stopping criteria; otherwise, set  $k \leftarrow k+1$ .
  4. Write the output from the above loop as:  $\hat{u}, \hat{d}$
- 

## 5 Experiments

In this section, we justify the proposed higher order total variation regularization term and image denoising model by its performance in removing Gaussian noise from images. The peak signal-to-noise ration(PSNR) is used to measure the quality of the restored image. We select the standard pepper and lena images and a CT image as our testing data. Both of the lena and pepper are with the size of  $256 \times 256$ , and the CT image is with the size of  $555 \times 555$ . All of the images are shown in Fig. 1. The regularization parameters are chosen to produce the highest PSNR values in all of the experiments. The stopping criterion of all of the algorithms is that the relative error between the successive iterates of the restored images should satisfy the following inequality

$$\frac{\|u^{i+1} - u^i\|_2^2}{\|u^i\|_2^2} \leq 10^{-4}$$

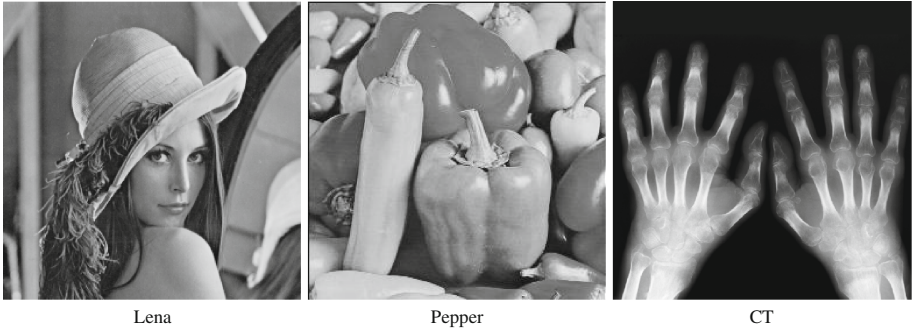


Fig. 1. The original images.

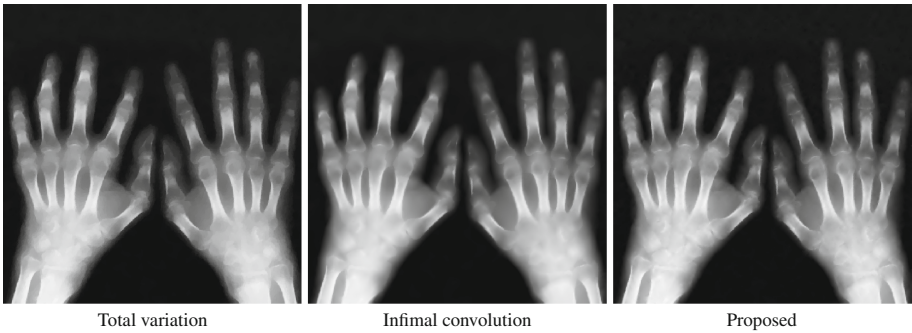


Fig. 2. The restored CT images in the case of Gaussian noise with  $\delta=10$ .

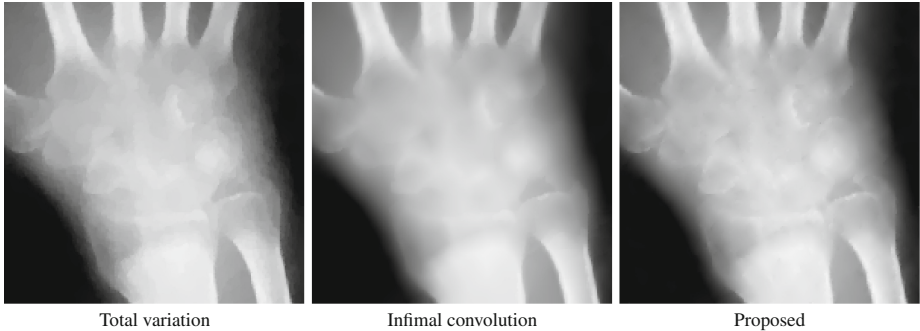
where  $u^i$  is the denoised image at the  $i$ -th iteration. Each PSNR value reported in this section is the averaged result of three runs.

We add additive white Gaussian noise with standard deviation  $\delta = 10$  and  $\delta = 20$  to the lena, pepper and CT images respectively. We use the ROF model (1), the infimal convolution model (2), and the model (9) which is based on our proposed higher order total variation to restore the images. The PSNR results of all restored images are shown in Table 1. We can see that our proposed model can produce the restored images with the higher PSNR values than those of the ROF model and infimal convolution model. To see the visual effects of the restored images, we show the CT images (the case of  $\delta = 10$ ) in Fig. 2. Figure 3 is the zoomed version of Fig. 2. It can be clearly seen the staircasing effects caused by the total variation when the image is zoomed in. The infimal convolution regularization can remove the staircasing effects caused by TV, however, the visible blurring artifacts at the edges are introduced. Our proposed higher order total variation regularization can produce more favorable result without visual artifacts while still preserve sharp edges and finer details well.



**Table 1.** The summary of the restoration results of Models (1), (2), and (9).

Model	$\delta = 10$			$\delta = 20$		
	Lena	Pepper	CT	Lena	Pepper	CT
Model (1)	32.30	32.47	36.84	29.07	29.67	33.67
Model (2)	33.07	33.28	37.55	29.55	30.21	34.51
Model (9)	<b>33.52</b>	<b>33.69</b>	<b>37.98</b>	<b>30.06</b>	<b>30.75</b>	<b>34.86</b>

**Fig. 3.** Zoomed results of the restored CT images in the case of Gaussian noise with  $\delta=10$ .

## 6 Concluding Remarks

We propose an higher order total variation regularization for image denoising problem. This regularization term can be seen as the improvement of the traditional infimal convolution type regularization and can alleviate the staircasing effects in the restored image caused by total variation. We use the proximity algorithm to solve the resulting model and establish the convergence result of the algorithm. Numerical results presented in this paper demonstrate the efficiency of the proposed method.

**Acknowledgment.** This work is supported by the Promotive Research Fund for Excellent Young Scientists of Shandong Province (No. BS2014DX003), by the Specialized Research Fund for the Doctoral Program of Higher Education of China (No. 20130132120022), and the Fundamental Research Funds of the Central Universities of Ocean University of China (No. 201313009).

## References

1. Bredies, K., Kunisch, K., Pock, T.: Total generalized variation. *SIAM J. Imaging Sci.* **3**, 492–526 (2010)
2. Chambolle, A., Lions, P.: Image recovery via total variation minimization and related problems. *Numer. Math.* **76**, 167–188 (1997)

3. Chen, F., Shen, L., Yuesheng, X., Zeng, X.: The moreau envelope approach for the  $l_1/tv$  image denoising model. *Inverse Probl. Imaging* **8**(1), 53–77 (2014)
4. Chen, Q., Montesinos, P., Sun, Q., Heng, P., Xia, D.: Adaptive total variation denoising based on difference curvature. *Image Vis. Comput.* **28**(3), 298–306 (2010)
5. Dong, Y., Hintermuller, M., Rincon-Camacho, M.: Automated regularization parameter selection in multiscale total variation methods for image restoration. *J. Math. Imaging Vis.* **40**(1), 82–104 (2011)
6. Micchelli, C.A., Shen, L., Xu, Y.: Proximity algorithms for image models: Denoising. *Inverse Probl.* **27**(4), 045009(30pp) (2011)
7. Micchelli, C.A., Shen, L., Yuesheng, X., Zeng, X.: Proximity algorithms for the  $l_1/tv$  image denoising models. *Adv. Comput. Math.* **38**(2), 401–426 (2013)
8. Rudin, L., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. *Phys. D* **60**, 259–268 (1992)
9. Setzer, S., Steidl, G., Teuber, T.: Infimal convolution regularizations with discrete  $l_1$ -type functionals. *Commun. Math. Sci.* **9**(3), 797–872 (2011)
10. Strong, D., Blomgren, P., Chan, T.: Spatially adaptive local feature-driven total variation minimizing image restoration. In: *Proceedings of SPIE*, vol. 3167, pp. 222–233 (1997)
11. Zeng, X., Li, S.: An efficient adaptive total variation regularization for image denoising. In: *2013 Seventh International Conference on Image and Graphics (ICIG)*, pp. 55–59. IEEE (2013)