Research Article

A Fast Alternating Minimization Algorithm for Nonlocal Vectorial Total Variational Multichannel Image Denoising

Rubing Xi, Zhengming Wang, Xia Zhao, and Meihua Xie

Department of Mathematics and Systems Science, National University of Defense Technology, Changsha 410073, China

Correspondence should be addressed to Rubing Xi; xirubing@gmail.com

Received 2 April 2014; Revised 20 July 2014; Accepted 5 August 2014; Published 26 August 2014

Academic Editor: Zhike Peng

Copyright © 2014 Rubing Xi 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.

The variational models with nonlocal regularization offer superior image restoration quality over traditional method. But the processing speed remains a bottleneck due to the calculation quantity brought by the recent iterative algorithms. In this paper, a fast algorithm is proposed to restore the multichannel image in the presence of additive Gaussian noise by minimizing an energy function consisting of an ℓ_2 -norm fidelity term and a nonlocal vectorial total variational regularization term. This algorithm is based on the variable splitting and penalty techniques in optimization. Following our previous work on the proof of the existence and the uniqueness of the solution of the model, we establish and prove the convergence properties of this algorithm, which are the finite convergence for some variables and the q -linear convergence for the rest. Experiments show that this model has a fabulous texture-preserving property in restoring color images. Both the theoretical derivation of the computation complexity analysis and the experimental results show that the proposed algorithm performs favorably in comparison to the widely used fixed point algorithm.

1. Introduction

The restoration of multichannel image is one of the most important inverse problems in image processing. Among all the methods for restoration, the vectorial total variational (VTV) regularization is most popular for preserving discontinuities in restored images. VTV has been introduced by [1] into multichannel image restoration as a vectorial extension of total variation (TV) in gray scale image processing.

The famous total variational (TV) methods have drawbacks due to its local property while calculating the gradient of the image. Thus the images that contain many small structures or textures cannot get satisfying restored results. As a result, TV method based on nonlocal (NL) operator is generally used in gray scale image processing because of its structure preserving property. This method is generalized from the Yaroslavsky denoising and the block method. The main idea is using the similar blocks in the whole image to estimate current pixel value. It was introduced by Efros and Leung [2] while studying texture synthesis. Then Buades et al. [3] generalized nonlocal method and proposed the well-known neighbor denoising filter, namely, nonlocal means (NLM). After that, Gilboa and Osher [4] united nonlocal

method into the regularization framework by defining the nonlocal formed TV model including NL-TV- ℓ_1 model, NL-TV- ℓ_2 model, and NL-TV-G model and proposed the algorithms separately. NL-TV model can both preserve the edges and the other small structures in images and has excellent application effect.

Our previous works extend NL-TV regularization to NL-VTV regularization for multichannel image restoration (see [5]). The corresponding algorithms are based on the variation calculus by solving the Euler-Lagrangian equations by fixed point iterations. This kind of methods is tested effectively but has a low convergence rate.

Recently, the optimization algorithms based on variable-splitting and penalty techniques have attracted much attention in image and signal processing, such as the dual methods, the Bregman methods, and the alternating direction method of multipliers (ADMM) methods (see [6–8]). Correspondingly, these methods have been extended for multichannel images. Bresson and Chan [6] has extended Chambolle's dual algorithm for TV problem to vectorial images by defining the standard dual definition of VTV. Wang et al. [9, 10] has applied the alternating minimization algorithm to VTV model for multichannel image deblurring.

In this paper, we propose a fast alternating algorithm for NL-VTV model to restore color images from noisy observations. The following sections are organized as follows. Section 2 is the brief introduction of the NL-VTV model and classical fixed point algorithm. In Section 3, we propose the alternating algorithm for NL-VTV model. In Section 4, we do the convergence analysis of this algorithm. In Section 5, we applied this algorithm to color image denoising to present the texture-preserving effect of the NL-VTV model and the high convergent rate of the proposed algorithm.

2. Nonlocal Vectorial Total Variation Model

In this section, we will briefly introduce the nonlocal vectorial total variational model for multichannel image processing. The usually used iterative algorithm will be introduced as well, which will be compared to our algorithm in Section 5.

2.1. Vectorial Total Variation Model. Consider a vector valued function $\mathbf{u}$, such as the multichannel images, defined on a bounded open domain $\Omega \subset \mathbb{R}^N$:

$$\begin{aligned} \mathbf{u} : \Omega &\longrightarrow \mathbb{R}^M \\ \mathbf{x} &\longmapsto \mathbf{u}(\mathbf{x}) = (u^1(\mathbf{x}), u^2(\mathbf{x}), \dots, u^M(\mathbf{x})), \end{aligned} \quad (1)$$

where $u^m : \Omega \rightarrow \mathbb{R}$ ($1 \leq m \leq M$) denotes the m th channel of $\mathbf{u}$.

The regularized VTV model for multichannel image reconstruction could be written as

$$\inf_{\mathbf{u} \in \mathbb{R}^M} E(\mathbf{u}) := \int_{\Omega} |\nabla \mathbf{u}| d\mathbf{x} + \frac{\lambda}{2} \int_{\Omega} |\mathbf{u} - \mathbf{u}_0|^2 d\mathbf{x}, \quad (2)$$

where $\mathbf{u}_0$ is the observed noisy multichannel image and $\lambda > 0$ is the regularized parameter. The vectorial gradient of $\mathbf{u}$ is defined as

$$\begin{aligned} \nabla \mathbf{u} = (\nabla u^1, \dots, \nabla u^M) &= \begin{bmatrix} \frac{\partial u^1}{\partial x} & \frac{\partial u^1}{\partial y} & \dots & \frac{\partial u^1}{\partial x} \\ \frac{\partial u^2}{\partial x} & \frac{\partial u^2}{\partial y} & \dots & \frac{\partial u^2}{\partial x} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial u^M}{\partial x} & \frac{\partial u^M}{\partial y} & \dots & \frac{\partial u^M}{\partial x} \end{bmatrix} \in \mathbb{R}^{2 \times M}, \\ |\nabla \mathbf{u}| = \|\nabla \mathbf{u}\|_F &= \sqrt{\sum_{m=1}^M \left(\frac{\partial u^m}{\partial x} \right)^2 + \sum_{m=1}^M \left(\frac{\partial u^m}{\partial y} \right)^2} = \sqrt{\sum_{m=1}^M |\nabla u^m|^2}. \end{aligned} \quad (3)$$

We refer to [6] for more information about the well-posedness analysis of the VTV model.

2.2. Nonlocal Functional. The neighborhood filter NLM proposed by Buades et al. developed to a famous image denoising method [3]. It uses the pixels similar to the current pixel to estimate the value of the current pixel. Let the reference image

be v . The similarity of the two pixels in the image x, y could be measured as the following weight function $w_v(x, y)$:

$$w_v(x, y) = \exp \left(- \frac{(G_a * |\gamma(x + \cdot) - \gamma(y + \cdot)|^2)(0)}{h^2} \right), \quad (4)$$

where G_a is the Gaussian kernel with the standard deviation of a :

$$\begin{aligned} &(G_a * |\gamma(x + \cdot) - \gamma(y + \cdot)|^2)(0) \\ &= \int_{\mathbb{R}^2} G_a(t) |\gamma(x + t) - \gamma(y + t)|^2 dt, \end{aligned} \quad (5)$$

where h is the filter parameter. Generally, h corresponds to the noise level, which is usually set as the standard deviation of the noise. It is shown that the weight function above could be calculated by the image blocks centered around x, y . The value of the weight function reflects the similarity of the two blocks. The estimation of the value at pixel x by the nonlocal means is

$$\text{NL}_v(x) := \frac{1}{C(x)} \int_{\Omega} w_v(x, y) v(y) dy, \quad (6)$$

where the normalized factor $C(x) = \int_{\Omega} w_v(x, y) dy$. It can be seen that this method is a kind of special filter algorithm, and the estimation of the value at pixel x is weighted by the values of all the pixels y ($y \in \Omega$) in the image.

Let the 2-dimensional image domain be $\Omega \subset \mathbb{R}^2$, pixels $x \in \Omega$, and $u(x)$ is a real function on $\Omega \rightarrow \mathbb{R}$. The derivation under the nonlocal frame is

$$\partial_y u(x) = (u(y) - u(x)) w(x, y), \quad (7)$$

where $w : \Omega \times \Omega \rightarrow \mathbb{R}$ is the nonnegative symmetrical weight function defined as (4), which is calculated by the reference image.

The nonlocal gradient $\nabla_w u(x, y)$ is defined as a vector composed by all the partial derivations at position x : $\nabla_w u(x, \cdot)$, where $\nabla_w u : \Omega \rightarrow \Omega \times \Omega$:

$$\nabla_w u(x, y) := (u(y) - u(x)) \sqrt{w(x, y)}, \quad \forall y \in \Omega. \quad (8)$$

Define the nonlocal divergence as

$$\text{div}_w \vec{v}(x) := \int_{\Omega} (v(x, y) - v(y, x)) \sqrt{w(x, y)} dy, \quad (9)$$

which satisfies the conjugation with the nonlocal gradient; that is,

$$\begin{aligned} \langle \nabla_w u, \vec{v} \rangle &= - \langle u, \text{div}_w \vec{v} \rangle, \quad \forall u : \Omega \rightarrow \mathbb{R}, \\ \forall \vec{v} : \Omega \times \Omega &\rightarrow \mathbb{R}. \end{aligned} \quad (10)$$

Thus the nonlocal Laplacian is defined as

$$\Delta_w u(x) := \frac{1}{2} \text{div}_w (\nabla_w u(x)) = \int_{\Omega} (u(y) - u(x)) w(x, y) dy. \quad (11)$$

The NL-TV norm of $u \in L^1(\Omega)$ based on nonlocal operators, that is, the weighted gradient $\nabla_w u$, is defined by

$$\begin{aligned} J_{\text{NL-TV}}(u) &= \int_{\Omega} |\nabla_w u| d\mathbf{x} \\ &= \int_{\Omega} \sqrt{\int_{\Omega} (u(x) - u(y))^2 w(x, y) dy} d\mathbf{x}, \end{aligned} \quad (12)$$

where function $w(x, y)$ is defined as (4). The NL-TV model can be written as

$$\|u\|_{L^1(\Omega)} + J_{\text{NL-TV}}(u). \quad (13)$$

The well-posedness of NL-TV model is guaranteed in [11].

2.3. Nonlocal Vectorial Total Variation Model and the Fixed Point Iterative Algorithm. In this section, the nonlocal gradient will be introduced to reconstruct the variation norm in the vectorial total variation model to deduce the NL-VTV model.

The vectorial nonlocal gradient could be derived by generalizing the nonlocal gradient defined previously:

$$\nabla_w \mathbf{u} = (\nabla_w u^1, \nabla_w u^2, \dots, \nabla_w u^M), \quad (14)$$

where $\nabla_w u^m = (u^m(y) - u^m(x))\sqrt{w^m(x, y)}$, $m = 1, \dots, M$, represents the gradient of each channel separately, which is calculated by the image of its own channel.

Use the vectorial nonlocal gradient $\nabla_w \mathbf{u}$ to replace the gradient $\nabla \mathbf{u}$ in VTV norm; we get the NL-VTV model:

$$\inf_{\mathbf{u} \in \mathbb{R}^M} E(\mathbf{u}) = \int_{\Omega} |\nabla_w \mathbf{u}| d\mathbf{x} + \frac{\lambda}{2} \int_{\Omega} |\mathbf{u} - \mathbf{u}_0|^2 d\mathbf{x}, \quad (15)$$

where $\int_{\Omega} |\nabla_w \mathbf{u}| d\mathbf{x}$ is the VTV regularization based on the nonlocal gradient, which could be denoted by

$$\begin{aligned} J(\mathbf{u}) &= \text{NL-VTV}(\mathbf{u}) = \int_{\Omega} |\nabla_w \mathbf{u}| d\mathbf{x} \\ &:= \int_{\Omega} \sqrt{\sum_{m=1}^M \int_{\Omega} (u^m(x) - u^m(y))^2 w^m(x, y) dy} d\mathbf{x}. \end{aligned} \quad (16)$$

The existence and the uniqueness of the solution to (15) have been proven in our previous work, in which we proposed a fixed point iterative algorithm to solve (15) as well. Here we will give out a brief introduction as follows.

The Euler-Lagrangian equation of the functional (15) is

$$-\text{div}_w \left(\frac{\nabla_w \mathbf{u}}{|\nabla_w \mathbf{u}|} \right) + \lambda (\mathbf{u} - \mathbf{u}_0) = 0. \quad (17)$$

Calculate the nonlocal curvature channel by channel; that is,

$$\begin{aligned} \kappa_w^m &:= \text{div}_w \left(\frac{\nabla_w u^m}{|\nabla_w \mathbf{u}|} \right) \\ &= \int_{\Omega} (u^m(y) - u^m(x)) w^m(x, y) \\ &\quad \times \left(\frac{1}{|\nabla_w \mathbf{u}|(x)} + \frac{1}{|\nabla_w \mathbf{u}|(y)} \right) dy, \end{aligned} \quad (18)$$

where $|\nabla_w \mathbf{u}|(q) := \sqrt{\sum_{m=1}^M \int_{\Omega} (u^m(z) - u^m(q))^2 w^m(q, z) dz}$, $m = 1, \dots, M$. Then the Euler-Lagrange equation becomes

$$\begin{aligned} - \int_{\Omega} (u^m(y) - u^m(x)) w^m(x, y) \left[\frac{1}{|\nabla_w \mathbf{u}|(x)} + \frac{1}{|\nabla_w \mathbf{u}|(y)} \right] dy \\ + \lambda (\mathbf{u} - \mathbf{u}_0) = 0, \quad m = 1, \dots, M. \end{aligned} \quad (19)$$

According to (19), when the weight function is fixed, the Euler-Lagrange equation is linear, which could be solved by the gradient descent kind of algorithm as follows.

First of all, we discretize (19). Denote u_i^m to be the pixel value at position i of the m th channel image, where $1 \leq m \leq M$, $1 \leq i \leq N$ (if the size of the image is $a \times b$, $N = a \times b$). $w_{i,j}^m$ is the discrete value of the weight function $w(x, y)$, that is, the weight function value with respect to pixel i and pixel j . The pixel j traverses the image channel where the pixel i is in. Denote the neighbor set to be $\mathcal{N}_i := \{j : w_{i,j}^m > \tau\}$, where τ denotes a threshold value as a judgment of the similarity between the two blocks centered at pixel i and pixel j , respectively. Let ∇_{wd} be the discrete ∇_w :

$$\nabla_{wd}(u_i^m) = (u_j^m - u_i^m) \sqrt{w_{i,j}^m}, \quad j \in \mathcal{N}_i. \quad (20)$$

And let div_{wd} be the discrete div_w : $\text{div}_{wd}(p_{i,j}) := \sum_{j \in \mathcal{N}_i} (p_{i,j} - p_{j,i}) \sqrt{w_{i,j}^m}$.

As for functions, the discrete inner product is $\langle u, v \rangle := \sum_i (u_i v_i)$, while, as for vectors, the discrete dot product is $\langle p, q \rangle := \sum_j (p_{i,j} q_{i,j})$, the discrete inner product is $\langle p, q \rangle := \sum_i \sum_j (p_{i,j} q_{i,j})$. The norm of vector is $|p_i| = \sqrt{\sum_j (p_{i,j})^2}$.

Then for channel m , where $1 \leq m \leq M$, the discrete nonlocal curvature is

$$\begin{aligned} \kappa_w^m &:= \text{div}_w^m \left(\frac{\nabla_w u^m}{|\nabla_w \mathbf{u}|} \right) \\ &= \sum_j (u_j^m - u_i^m) w_{i,j}^m \left[\frac{1}{|\nabla_w \mathbf{u}|(i)} + \frac{1}{|\nabla_w \mathbf{u}|(j)} \right], \end{aligned} \quad (21)$$

where $|\nabla_w \mathbf{u}|(i) = \sqrt{\sum_{m=1}^M \sum_j (u_j^m - u_i^m)^2 w_{i,j}^m}$ and $|\nabla_w \mathbf{u}|(j) = \sqrt{\sum_{m=1}^M \sum_i (u_j^m - u_i^m)^2 w_{i,j}^m}$. Then the Euler-Lagrange equation becomes

$$-\sum_j (u_j^m - u_i^m) w_{i,j}^m \left[\frac{1}{|\nabla_w \mathbf{u}|(i)} + \frac{1}{|\nabla_w \mathbf{u}|(j)} \right] + \lambda (u_i^m - u_{0i}^m) = 0, \quad m = 1, \dots, M. \quad (22)$$

Denote $w_{i,j}^m [(1/|\nabla_w \mathbf{u}|(i)) + (1/|\nabla_w \mathbf{u}|(j))] \triangleq W_{i,j}^m$. Then the Euler-Lagrange equation could be shortly written as

$$-\sum_j W_{i,j}^m (u_j^m - u_i^m) + \lambda (u_i^m - u_{0i}^m) = 0, \quad m = 1, \dots, M. \quad (23)$$

Furthermore, it could be written as

$$u_i^m = \sum_j \frac{W_{i,j}^m}{\sum_j W_{i,j}^m + \lambda} u_j^m + \frac{\lambda}{\sum_j W_{i,j}^m + \lambda} u_{0i}^m, \quad m = 1, \dots, M. \quad (24)$$

According to the fixed point iteration, we use the multi-channel image $\mathbf{u}(k)$ in step k to update image $\mathbf{u}(k+1)$ in step $k+1$. The weights are calculated by the original observation image. Then we get the iteration formula as

$$\mathbf{u}(0) = \mathbf{u}_0, \quad \mathbf{u}_i(k+1) = \sum_j \frac{W_{i,j}^m(k)}{\sum_j W_{i,j}^m(k) + \lambda} \mathbf{u}_j(k) + \frac{\lambda}{\sum_j W_{i,j}^m(k) + \lambda} \mathbf{u}_{0i}, \quad (25)$$

where $i \in \Omega$, $j \in \mathcal{N}_i$.

The convergence property of the fixed point algorithm has been proved in our in-press work. However, when $|\nabla_w \mathbf{u}| = 0$, the NLTV functional is nondifferentiable, which is similar to the classical TV model. Thus a regularization $\sqrt{|\nabla_w \mathbf{u}|^2 + \epsilon}$ is involved in the calculation of $W_{i,j}^m$ to escape the denominator to be zero. Consequently, this algorithm takes many iteration steps to obtain the best restoration result. Thus we consider designing a fast alternating minimization algorithm instead.

3. The Alternating Minimization Algorithm

For the sake of conquering the nondifferentiability of the VTV functional, we design an alternating minimization algorithm in this section to solve the problem (15). The main idea of the alternating minimization is introducing an auxiliary variable. Then the original problem could be split into two subproblems of which one is differential with respect to the original variable, while the other is convex with respect to the auxiliary variable, respectively.

Yang et al. [10] have already studied the alternating minimization algorithm for a general model:

$$\min_{\mathbf{u}} \sum_i \alpha_i \|G_i \mathbf{u}\| + \frac{\mu}{2} \|K\mathbf{u} - \mathbf{f}\|^2. \quad (26)$$

Model (26) reduces to (15) when $G_i = I_m \otimes \nabla_w$, $\alpha_i = 1$, and $K = I$, where I_m is the identity matrix of order m and $\otimes$ represents the Kronecker product. Yang et al. pointed out in [10] that if G_i is the finite difference operator D_i , then, under some boundary conditions for $\mathbf{u}$ and some other assumptions for the blurring kernel K , the problem (26) can be fast transformed by fast Fourier transform, which diagonalizes the block circulant matrices: D_i , K , and their transpose, and then can be solved by the fast algorithm such as the Gaussian elimination method.

However, the nonlocal gradient operator ∇_w in (15) is not block circulant. We will figure out another method to speed up the algorithm. We briefly propose the algorithm in this section. The convergence of this algorithm will be analysed in Section 4, as well as others properties of this algorithm.

3.1. Variable Splitting (Half-Quadratic Formulation of (15)). We introduce an auxiliary variable $\mathbf{d}$ to approximate the nonlocal gradient of $\mathbf{u}$, that is, $\nabla_w \mathbf{u}$. Then the minimization problem (15) is transformed into an equivalent constrained problem:

$$\min_{\mathbf{u} \in \mathbb{R}^M} E(\mathbf{u}) = \int_{\Omega} |\mathbf{d}| d\mathbf{x} + \frac{\lambda}{2} \int_{\Omega} |\mathbf{u} - \mathbf{u}_0|^2 d\mathbf{x}, \quad (27)$$

subject to $\mathbf{d} = \nabla_w \mathbf{u}$.

In order to solve problem (27), we use the penalty method to reformulate it as

$$\min_{\mathbf{u}, \mathbf{d}} E(\mathbf{u}, \mathbf{d}) = \int_{\Omega} |\mathbf{d}| d\mathbf{x} + \frac{\lambda}{2} \int_{\Omega} |\mathbf{u} - \mathbf{u}_0|^2 d\mathbf{x} + \frac{\beta}{2} \int_{\Omega} |\mathbf{d} - \nabla_w \mathbf{u}|^2 d\mathbf{x}, \quad (28)$$

where β is the penalty parameter. It could be easily proved that, as $\beta \rightarrow \infty$, the solution of problem (28) is convergent to the solution of problem (15).

3.2. Alternating Minimization Algorithm. The problem of minimizing (28) with respect to $\mathbf{d}$ for fixed $\mathbf{u}$ becomes

$$\min_{\mathbf{d}} \int_{\Omega} |\mathbf{d}| d\mathbf{x} + \frac{\beta}{2} \int_{\Omega} |\mathbf{d} - \nabla_w \mathbf{u}|^2 d\mathbf{x}. \quad (29)$$

The functional in (29) is convex with respect to $\mathbf{d}$. According to Lemma 3.3 in [10], the problem (29) has the unique solutions given by

$$\mathbf{d}^{k+1} = \text{shrink} \left(\nabla_w \mathbf{u}^{k+1}, \frac{1}{\beta} \right), \quad (30)$$

where the multidimensional shrinkage operator $\text{shrink}(\mathbf{x}, \lambda) = (\mathbf{x}/|\mathbf{x}|) \cdot \max\{|\mathbf{x}| - \lambda, 0\}$.

On the other hand, minimizing (28) with respect to $\mathbf{u}$ for fixed $\mathbf{d}$ becomes

$$\min_{\mathbf{u}} \frac{\lambda}{2} \int_{\Omega} |\mathbf{u} - \mathbf{u}_0|^2 d\mathbf{x} + \frac{\beta}{2} \int_{\Omega} |\mathbf{d} - \nabla_w \mathbf{u}|^2 d\mathbf{x}. \quad (31)$$

Initialization:
 for $k = 0$ to K , do
 Solve $\mathbf{d}^{k+1} = \text{shrink}(\nabla_w \mathbf{u}^{k+1}, 1/\beta)$.
 Solve $\mathbf{u}^{k+1} = (1 - (2/\beta))\mathbf{I}(\mathbf{u}_0 - (1/\beta)\text{div}_w \mathbf{d}^k) + (2/\beta)\mathbf{W}(\mathbf{u}_0 - (1/\beta)\text{div}_w \mathbf{d}^k)$.
 end for

ALGORITHM 1: Alternating minimization algorithm for nonlocal VTV denoising.

The functional in (31) is differential with respect to $\mathbf{u}$. Thus the solution to (31) satisfies the first order optimized condition; that is, the Euler-Lagrange equation of the functional (31) is

$$\lambda(\mathbf{u} - \mathbf{u}_0) + \beta \text{div}_w(\mathbf{d} - \nabla_w \mathbf{u}) = 0. \quad (32)$$

After some steps of simple calculation, we get

$$\mathbf{u} - \frac{\beta}{\lambda} \text{div}_w(\nabla_w \mathbf{u}) = \mathbf{u}_0 - \frac{\beta}{\lambda} \text{div}_w \mathbf{d}. \quad (33)$$

Then according to (11), we get

$$\left(\mathbf{I} - \frac{2\beta}{\lambda} \Delta_w\right) \mathbf{u} = \mathbf{u}_0 - \frac{\beta}{\lambda} \text{div}_w \mathbf{d}. \quad (34)$$

Finally, the multichannel image $\mathbf{u}$ is calculated as follows:

$$\left(\mathbf{I} - \frac{2\beta}{\lambda} \Delta_w\right) \mathbf{u}^{k+1} = \mathbf{u}_0 - \frac{\beta}{\lambda} \text{div}_w \mathbf{d}^k. \quad (35)$$

Since the graph Laplacian Δ_w is negative semidefinite, the operator $\mathbf{I} - (2\beta/\lambda)\Delta_w$ is diagonally dominant. Therefore we can solve $\mathbf{u}^{k+1}$ by a Gauss-Seidel algorithm. However, the subproblem (35) consists in solving the linear equations with the scale of $N \times N$ for each image channel. As a result, the calculation complexity of the Gauss-Seidel algorithm is heavy. We are inspired by Chen and Cheng [12] to simplify, and, without loss of generality as well, we chose $\lambda = \beta^2$. Furthermore, using Taylor expansion to approximate, there is

$$\begin{aligned} \left(\mathbf{I} - \frac{2\beta}{\lambda} \Delta_w\right)^{-1} &= \left(\mathbf{I} - \frac{2}{\beta} \Delta_w\right)^{-1} \\ &= \mathbf{I} + \frac{2}{\beta} \Delta_w + O(\beta^{-2}) \approx \mathbf{I} + \frac{2}{\beta} \Delta_w \\ &= \mathbf{I} + \frac{2}{\beta} (\mathbf{W} - \mathbf{I}) = \left(1 - \frac{2}{\beta}\right) \mathbf{I} + \frac{2}{\beta} \mathbf{W}, \end{aligned} \quad (36)$$

where $\mathbf{W} = (w_{ij})_{i,j=1}^{n^2}$ is the weight matrix of similar blocks of each channel, which is normalized to a symmetrical matrix with the sum of every row equal to 1. Thus the solution to (35) is

$$\mathbf{u}^{k+1} = \left(1 - \frac{2}{\beta}\right) \mathbf{I} \left(\mathbf{u}_0 - \frac{1}{\beta} \text{div}_w \mathbf{d}^k\right) + \frac{2}{\beta} \mathbf{W} \left(\mathbf{u}_0 - \frac{1}{\beta} \text{div}_w \mathbf{d}^k\right). \quad (37)$$

According to (9), the divergence at pixel i is $\text{div}_w d_i^m = \sum_j (d_{ij}^m - d_{ji}^m) \sqrt{w_{ij}}$, $m = 1, \dots, M$.

To conclude this section, we rewrite the alternating minimization algorithm as in Table 1.

TABLE 1: Comparison of the highest SNR obtained.

SNR	Noisy image	VTV	NLVTv
Barbara image	8.9013	13.1212	13.2873
Baboon image	9.2373	13.1618	12.5936

4. Convergence Analysis

Since it has been proved in our in-press work that the model (15) has the unique solution, in this section, we prove that the solution obtained by Algorithm 1 is the unique one. First, the convergence of the quadratic penalty method as the penalty parameter goes to infinity is well known [13]. That is, as $\beta \rightarrow \infty$, the solution of (31) converges to that of (15). We analyze the convergence properties of Algorithm 1 for a fixed $\beta > 0$. In this section, we show that, for any fixed β , the algorithm of minimizing $\mathbf{u}$ and $\mathbf{d}$ alternately has finite convergence for some variables and q -linear convergence for the rest.

Theorem 1. *The iterative sequence $\{(\mathbf{d}^k, \mathbf{u}^k)\}$ generated by (30) and (37) from any starting point $\{(\mathbf{d}^0, \mathbf{u}^0)\}$ converges to the solution of the minimization problem (31).*

Proof. According to (30) and (37), the functional $E(\mathbf{u}, \mathbf{d})$ in (28) satisfies

$$E(\mathbf{u}^{k+1}, \mathbf{d}^{k+1}) \leq E(\mathbf{u}^k, \mathbf{d}^{k+1}) \leq E(\mathbf{u}^k, \mathbf{d}^k). \quad (38)$$

Therefore, the sequence $\{E(\mathbf{u}^k, \mathbf{d}^k)\}$ is nonincreasing. Thus it is convergent in $\mathbb{R}^N$ because of its nonnegativity as well. We denote the limit of the sequence by $\mathbf{m}$. We prove that $\mathbf{m} = E(\mathbf{u}^*, \mathbf{d}^*)$ followed, where $\mathbf{u}^*$ and $\mathbf{d}^*$ are the solution of (31).

Because E is coercive and the sequence $\{E(\mathbf{u}^k, \mathbf{d}^k)\}$ is convergent, the sequence $\{(\mathbf{u}^k, \mathbf{d}^k)\}$ is bounded. Thus there exists a subsequence $\{(\mathbf{u}^k, \mathbf{d}^k)\}$, which converges to an accumulation point $(\tilde{\mathbf{u}}, \tilde{\mathbf{d}})$ when $n_k \rightarrow \infty$. According to (30) and (37), for all $n_k \in \mathbb{N}$, $\mathbf{d}$,

$$E(\mathbf{u}^{n_k}, \mathbf{d}^{n_k+1}) \leq E(\mathbf{u}^{n_k}, \mathbf{d}), \quad (39)$$

and, for all $n_k \in \mathbb{N}$, $\mathbf{u}$,

$$E(\mathbf{u}^{n_k+1}, \mathbf{d}^{n_k+1}) \leq E(\mathbf{u}, \mathbf{d}^{n_k+1}). \quad (40)$$

We denote $\tilde{\mathbf{d}}$ as the accumulation point of the sequence $\{\mathbf{d}^{n_k+1}\}$, because of that the functional E is continuous, according to (38),

$$\mathbf{m} = E(\tilde{\mathbf{u}}, \tilde{\mathbf{d}}) = E(\tilde{\mathbf{u}}, \hat{\mathbf{d}}). \quad (41)$$

Because of that the threshold function, shrink, in (30), is continuous, we compute the limits of both sides of the equation and get

$$\tilde{\mathbf{d}} = \text{shrink}\left(\nabla_w \tilde{\mathbf{u}}, \frac{1}{\beta}\right). \quad (42)$$

Thus $\tilde{\mathbf{d}} = \arg \min_{\mathbf{d}} E(\tilde{\mathbf{u}}, \mathbf{d})$. According to (41), $\hat{\mathbf{d}} = \arg \min_{\mathbf{d}} E(\hat{\mathbf{u}}, \mathbf{d})$. On the other hand, E is strictly convex with respect to $\mathbf{d}$, then we have $\tilde{\mathbf{d}} = \hat{\mathbf{d}}$ because of the uniqueness of the minimizer of $\min_{\mathbf{d}} E(\tilde{\mathbf{u}}, \mathbf{d})$. Similarly, we prove that $\tilde{\mathbf{u}} = \hat{\mathbf{u}}$.

We compute the limits of (39) and (40),

$$\begin{aligned} \inf_{\mathbf{d}} E(\tilde{\mathbf{u}}, \mathbf{d}) &= E(\tilde{\mathbf{u}}, \tilde{\mathbf{d}}) \leq E(\tilde{\mathbf{u}}, \mathbf{d}) \\ \inf_{\mathbf{u}} E(\mathbf{u}, \tilde{\mathbf{d}}) &= E(\tilde{\mathbf{u}}, \tilde{\mathbf{d}}) \leq E(\mathbf{u}, \tilde{\mathbf{d}}). \end{aligned} \quad (43)$$

According to (43), $(\mathbf{0}, \mathbf{0})^T \in \partial E(\tilde{\mathbf{u}}, \tilde{\mathbf{d}})$. Thus the limit points of all of the subsequence of $\{(\mathbf{u}^k, \mathbf{d}^k)\}$ are the solutions to (28). As a result, $\{(\mathbf{u}^k, \mathbf{d}^k)\}$ converges to $(\mathbf{u}^*, \mathbf{d}^*)$. $\square$

Let $\mathbf{S}_{1/\beta}(\nabla_w \mathbf{u}) = \text{shrink}(\nabla_w \mathbf{u}, 1/\beta)$ and let $\mathbf{P} = (\mathbf{I} - (2\beta/\lambda)\Delta_w)^{-1}$. We define set $\mathcal{H} = \{(\mathbf{y}, \mathbf{y}') \in \mathbb{R}^{m \times n^2} \otimes \mathbb{R}^{m \times n^2} \mid \|\mathbf{W} \cdot (\mathbf{y} - \mathbf{y}')\|_2 \leq \nu \|\mathbf{y} - \mathbf{y}'\|_2\}$, where $\nu < 1$ is a positive constant. Then we have the following conclusion.

Theorem 2 (*q-linear convergence*). Assume that the sequence $\{\mathbf{u}^k\} \subseteq \mathcal{H}$; then there is a constant $\beta_0 > 0$, such that for all $\beta > \beta_0$, the sequence $\{\mathbf{u}^k\}$ generated by (30) and (37) is *q-linear convergence*.

Proof. Combine (30) and (37); we have

$$\mathbf{u}^{k+1} = \mathbf{P} \left(\mathbf{u}_0 - \frac{1}{\beta} \text{div}_w (\mathbf{S}_{1/\beta}(\nabla_w \mathbf{u}^k)) \right). \quad (44)$$

We know that $\mathbf{u}^*$ is a fixed point of (44). Let $\mathbf{Q} = \mathbf{P}^{-1} = \mathbf{I} - (2\beta/\lambda)\Delta_w = (1 + (2/\beta))\mathbf{I} - (2/\beta)\mathbf{W}$. Then

$$\mathbf{Q}(\mathbf{u}^{k+1} - \mathbf{u}^*) = -\frac{1}{\beta} \text{div}_w (\mathbf{S}_{1/\beta}(\nabla_w \mathbf{u}^k) - \mathbf{S}_{1/\beta}(\nabla_w \mathbf{u}^*)). \quad (45)$$

Since the operator $\mathbf{S}_{1/\beta}$ is nonexpansive,

$$\begin{aligned} &\left\| -\frac{1}{\beta} \text{div}_w (\mathbf{S}_{1/\beta}(\nabla_w \mathbf{u}^k) - \mathbf{S}_{1/\beta}(\nabla_w \mathbf{u}^*)) \right\|_2 \\ &\leq \frac{\|\text{div}_w\|_2 \|\nabla_w\|_2}{\beta} \|\mathbf{u}^k - \mathbf{u}^*\|_2. \end{aligned} \quad (46)$$

Because $\{\mathbf{u}^k\} \subseteq \mathcal{H}$,

$$\begin{aligned} \|\mathbf{Q}(\mathbf{u}^{k+1} - \mathbf{u}^*)\|_2 &= \left\| \left(\left(1 + \frac{2}{\beta}\right)\mathbf{I} - \frac{2}{\beta}\mathbf{W} \right) (\mathbf{u}^{k+1} - \mathbf{u}^*) \right\|_2 \\ &\geq \left(1 + \frac{2}{\beta}\right) \|\mathbf{u}^{k+1} - \mathbf{u}^*\|_2 \\ &\quad - \frac{2}{\beta} \|\mathbf{W}(\mathbf{u}^{k+1} - \mathbf{u}^*)\|_2 \\ &\geq \left(1 + \frac{2}{\beta} - \frac{2}{\beta}\nu\right) \|\mathbf{u}^{k+1} - \mathbf{u}^*\|_2. \end{aligned} \quad (47)$$

FIGURE 1: Comparison of SNR.

We write $\tau = 1/(1 + (2/\beta) - (2/\beta)\nu) < 1$. Then by uniting (46) and (47), we have

$$\|\mathbf{u}^{k+1} - \mathbf{u}^*\|_2 \leq \tau \frac{\|\text{div}_w\|_2 \|\nabla_w\|_2}{\beta} \|\mathbf{u}^k - \mathbf{u}^*\|_2. \quad (48)$$

Let $\beta_0 = \tau \|\text{div}_w\|_2 \|\nabla_w\|_2 > 0$. Then for all $\beta > \beta_0$, the sequence $\{\mathbf{u}^k\}$ generated by (30) and (37) is *q-linear convergent* to $\mathbf{u}^*$. $\square$

5. Experiments

In this section, we apply Algorithm 1 to color image denoising to test its efficiency. We compare this method to that in [10] to show the detail-preserving property of the nonlocal VTV model. Then we compare the convergent rate of this algorithm with the fixed point algorithm presented in Section 2.

5.1. Texture-Preserving Property of the NLVTV Model. We choose the color Barbara picture size $300 \times 300 \times 3$ and the color Baboon picture size $256 \times 256 \times 3$ for testing, both of which contain many textures, such as the woman's scarf and the baboon's beard. We add the Gaussian white noise with standard deviation 20 to the two original pictures. Then we use the VTV model and the NLVTV model to restore them, separately. In this experiment, the signal to noise rate (SNR) is used to estimate the restoration effect of the methods, which is defined as follows:

$$\text{SNR} = 10 \cdot \log_{10} \left(\frac{\text{var}(\mathbf{u})}{\text{var}(\mathbf{u}^k - \mathbf{u})} \right), \quad (49)$$

where $\text{var}(x)$ represents the variance of all the elements of matrix x .

As is shown in Figure 1 and Table 1, the NLVTV model obtains similar SNR to VTV model. In the experiment of

FIGURE 2: Comparison between the NLVTV and the VTV in the Barbara image.

restoring the color Barbara picture, we choose the restored image by NLVTV model under the SNR (13.0678) lower than but close to that of the VTV model (SNR = 13.1212) instead of the restored image under the best SNR (13.2873). Figure 2 shows the results. Figure 3 shows the zoom part of the restored image, that is, the woman's scarf. We can see that the NVTV model preserves the details better than the VTV model. Figures 4 and 5 show the results of the experiment of restoring the color Baboon picture. Although the NLVTV model gets lower SNR than the VTV model, the textures in this picture are better preserved by the NLVTV model rather than the VTV model.

5.2. Computation Complexity Analysis and the Convergent Rate Comparison. In this section, we compare the computation complexity and the convergent rate of this algorithm to the fixed point algorithm.

First, we analyze the computation complexities of the two algorithms, respectively. Assume the size of the multichannel image is $n \times n \times M$. The size of the nonlocal searching window is $m \times m$. We do not consider the computation complexity

of the calculation of the nonlocal weight, because that is executed exactly the same in both algorithms.

At each iterative step of Algorithm 1, namely, (30) and (37), the dominant calculation of (30) contains the calculation of the nonlocal gradient and the threshold computation of $M(nm)^2$ times. Thus the computation complexity is $O(Mn^4) + O(M(nm)^2)$. The dominant calculation of (37) contains a step of nonlocal divergence computation with the computation complexity of $O(Mn^2)$ and a matrix multiplication with the computation complexity of $O(Mn^4)$. As a result, the total computation complexity for each iteration is $O(Mn^4) + O(M(nm)^2) + O(Mn^2) + O(Mn^4)$.

On the other hand, each iterative step of the fixed point iteration contains the nonlocal curvature with the computation complexity of $O(Mn^4)$ and three matrix multiplications with the computation complexity of $O(Mn^4) + O(Mn^4) + O(Mn^4)$. As a result, the total computation complexity for each iteration is $O(Mn^4) + O(Mn^4) + O(Mn^4) + O(Mn^4)$.

Consequently, the computation complexity of the proposed algorithm is much less than the fixed point algorithm.

FIGURE 3: Zoom in the Barbara image.

FIGURE 4: Comparison between the NLVTV and the VTV in the Baboon image.

FIGURE 5: Zoom in the Baboon image.

FIGURE 6: Comparison of the convergence rate.

The convergence rate can be represented by the relationship between the iterative step and the standard deviation of the restored image, intuitively. Figure 6 shows the comparison of the convergence rate between the fixed point iteration and the proposed alternating minimization algorithm. The horizontal axis represents the iterative step. And the vertical axis represents the standard deviation of the restored image in each step. The standard deviation is calculated as

$$\text{std} \left(\left| \mathbf{u}^k - \mathbf{u}^{k-1} \right| \right), \quad (50)$$

where $\mathbf{u}^k$ and $\mathbf{u}^{k-1}$ represent the restored images of step k and step $k-1$, respectively. When calculating $\text{std}(x)$, we reshape

matrix x into a vector x whose elements are taken column-wise from matrix x . Then

$$\text{std}(x) = \left(\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{1/2}. \quad (51)$$

Figure 6 shows that the alternating minimization algorithm converges faster.

6. Conclusion

In this paper, we propose a fast alternating minimization algorithm for multichannel image processing, based on the new nonlocal vectorial total variational model introduced recently. We use the variable splitting and penalty techniques to split the energy minimization problem into two optimization subproblems. One is a simple ℓ_1 minimization problem, which is solved by the threshold method. The other is a simple ℓ_2 minimization problem, which is solved by the variation calculus. As a result, this algorithm is faster than the classical algorithm because of avoiding dealing with the nondifferentiability of the nonlocal TV norm.

The main contribution of this paper is introducing the alternating minimization algorithm for the nonlocal vectorial total variational model, which is proved fast and can be applied to many multichannel image processing problems. In this paper, we apply this algorithm to the RGB color image denoising as an example. It can be easily extended to other vectorial problems, such as color image deblurring, and to other color image models, such as hue-saturation-value (HSV) model, as well.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This paper and the related works are supported by grants from the National Natural Science Foundation of China under Contracts 61072142, 61271437, and 61201337 and the Science Research Project of National University of Defense Technology under Contracts JC12-02-05 and JC13-02-03.

References

- [1] G. Sapiro and D. L. Ringach, "Anisotropic diffusion of multivalued images with applications to color filtering," *IEEE Transactions on Image Processing*, vol. 5, no. 11, pp. 1582–1586, 1996.
- [2] A. A. Efros and T. K. Leung, "Texture synthesis by non-parametric sampling," in *Proceedings of the 7th IEEE International Conference on Computer Vision (ICCV '99)*, pp. 1033–1038, Kerkyra, Greece, September 1999.
- [3] A. Buades, B. Coll, and J. M. Morel, "A review of image denoising algorithms, with a new one," *Multiscale Modeling & Simulation*, vol. 4, no. 2, pp. 490–530, 2005.
- [4] G. Gilboa and S. Osher, "Nonlocal operators with applications to image processing," cam report 07-23, UCLA, 2007.
- [5] R. Xi, Z. Wang, and X. Zhao, "A non-local means based vectorial total variational model for multi-channel SAR image denoising," in *Proceedings of the 6th International Congress on Image and Signal Processing (CISP '13)*, pp. 217–222, 2013.
- [6] X. Bresson and T. F. Chan, "Fast dual minimization of the vectorial total variation norm and applications to color image processing," *Inverse Problems and Imaging*, vol. 2, no. 4, pp. 455–484, 2008.
- [7] W. Yin, S. Osher, and D. Goldfarb, "Bregman iterative algorithms for ℓ_1 -minimization with applications to compressed sensing," *SIAM Journal on Imaging Sciences*, vol. 1, no. 1, pp. 143–168, 2008.
- [8] D. Boley, "Local linear convergence of the alternating direction method of multipliers on quadratic or linear programs," *SIAM Journal on Optimization*, vol. 23, no. 4, pp. 2183–2207, 2013.
- [9] Y. Wang, J. Yang, W. Yin, and Y. Zhang, "A new alternating minimization algorithm for total variation image reconstruction," *SIAM Journal on Imaging Sciences*, vol. 1, no. 3, pp. 248–272, 2008.
- [10] J. Yang, W. Yin, Y. Zhang, and Y. Wang, "A fast algorithm for edge-preserving variational multichannel image restoration," *SIAM Journal on Imaging Sciences*, vol. 2, no. 2, pp. 569–592, 2009.
- [11] X. Bresson and T. F. Chan, "Non-local unsupervised variational image segmentation models," UCLA cam report 08-67, 2008.
- [12] D. Chen and L. Cheng, "Alternative minimisation algorithm for non-local total variational image deblurring," *IET Image Processing*, vol. 4, no. 5, pp. 353–364, 2010.
- [13] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer, New York, NY, USA, 1999.