

# A Block Nonlocal TV Method for Image Restoration

Jun Liu\*

*School of Mathematical Sciences, Laboratory of Mathematics and Complex Systems, Beijing Normal University,  
Beijing 100875, P.R. China*

Xiaojun Zheng

*School of Mathematical Sciences, Laboratory of Mathematics and Complex Systems, Beijing Normal University,  
Beijing 100875, P.R. China*

---

## Abstract

In this paper, we propose a block nonlocal total variational (TV) regularization for image restoration. We extend the existing nonlocal TV method into two aspects: firstly, some block nonlocal operators are introduced to extend the point based nonlocal diffusion as a block based nonlocal diffusion process; secondly, the weighting function in the nonlocal methods can be adaptively determined by the cost functional itself. The proposed method is derived from a block based maximum a posteriori (MAP) estimation. By the assumption of the self similarity of small patches, we formulate a regularization term as a log-likelihood functional of a mixture model. To optimize this regularization term efficiently, we employ the idea of the expectation maximum algorithm and give a variational framework to propose a block based nonlocal TV regularization. The weighting function occurred in our model can be regarded as a probability of the similarity for images patches, and it can be updated adaptively according to the newest estimation. Beside, we mathematically prove the existence of minimizer for some of the proposed models. Compared with the nonlocal TV method, numerical results show that our method can greatly improve the quality of the restored images, especially under heavy noise.

---

## 1. Introduction

Image denoising is a fundamental and active research field in image processing. The goal of image denoising is to recover a true image from some noisy observed data. The classical noise model is additive Gaussian noise, that is,

$$v = u + n,$$

where  $v$  is the observed image,  $u$  is the true image and  $n$  stands for noise.

A vast variety of methods for image denoising are available after several decades of developments[3]. Gaussian (Gabor) filters[15] and anisotropic diffusion[20] are the early denoising approaches. Statistical methods (e.g.,[1]) considering the noise as a random variable is a popular denoising technique. They are mainly based on maximum likelihood estimator (MLE) and Bayesian maximum a posteriori (MAP). Variational approaches (e.g.,[21, 22]) is another popular and powerful technique for image denoising. The variational methods obtain a clean image by minimizing a cost functional which typically consists of a fidelity term and a regularization term. Of course, all these methods are not separated from each other, such as many methods can be viewed within the variational framework. But variational method is more highly valued since the total variation (TV) based-ROF model was proposed in [22]. The ROF model is a classical and efficient model to remove Gaussian noise, and many variants [19, 16] based on TV had been designed for different denoising tasks due to its good edges-preserving properties. However, the results obtained with TV

---

\*Jun Liu is currently visiting Department of Mathematics, UCLA, this work was finished during his visit at UCLA. The authors would like to thank Professor Stanley Osher for his valuable comments and suggestions.

*Email addresses:* jliu@bnu.edu.cn (Jun Liu), xjzheng@mail.bnu.edu.cn (Xiaojun Zheng)

could be piecewise constant and the image details such as textures could be removed together with noise. In order to better preserve image textures and repetitive structures, nonlocal denoising methods have gained considerable attention in recent years. The concept of nonlocal averaging was introduced by Yaroslavsky via an nonlocal neighborhood filter [26]. Nonlocal neighborhood filter averages the value at point  $x$  based on points  $y$  with similar values which can be located far from  $x$ . A well-known nonlocal method based on patch distances extending this idea is the nonlocal means [2] proposed by Buades *etc.*. The novelty of the nonlocal mean is to introduce the patch distances, i.e. the value at point  $y$  is used to denoise that at  $x$  if the local value pattern near  $y$  is similar to the local value pattern near  $x$  [23]. A variational interpretation of nonlocal means was given in [14, 9]. Motivated by the nonlocal mean and the graph theory, the nonlocal TV variational framework base on nonlocal operators was proposed in [10]. The weighting function occurred in these nonlocal methods play a very important role in the task of denoising. Although the above nonlocal methods show superior performance in textured or structure repeated images, the weights in the nonlocal method are given and fixed. It may give a bad estimation for the weights and lead to a undesirable restoration under the heavy noise. Furthermore, the estimation in these methods is point by point, it may not robust for heavy noise. Many block based methods [7, 6, 12, 13] have shown that the block based estimation can improve the precision and be more robust to noise. Based on overlapping-patches technique and sparsity, block based nonlocal image denoising methods have been proposed in recent years, e.g. patch-based near-optimal image denoising (PLOW)[5], locally learned dictionaries (K-LLD) [4], clustering-based sparse representation (CSR) [8], a weighted dictionary learning model[17], the expected patch log likelihood (EPLL) [27] and so on.

In this work, we formulate a block based nonlocal maximum a posterior estimation (MAP) framework, in which we construct a nonparametric mixture model to describe the prior probability density function for the small image patches. Instead of optimizing such a complicated regularization term derived from the mixture model, we propose a variant which comes from a well-known expectation maximum (EM) process. It lead to a block diffusion based  $BNLH^1$  model with an adaptive weighting function. The  $BNLH^1$  model can be regarded as an extension of nonlocal means method. We also prove the existence of minimizer for such a  $BNLH^1$  model with minimizing sequence method. It is well-known that the  $L^1$  norm has a better performance on noncontinuity preservation than  $L^2$ , so we define a block nonlocal TV operator, and propose a block nonlocal (BNLTV) TV model . The BNLTV can be efficiently solved by many recently popular splitting algorithms such as augmented Lagrangian [25], splitting Bregman [11] methods and ADMM[24]. Compared with the existing nonlocal TV, the experimental results in this paper show that our method can greatly improve the quality of the reconstruction.

The rest of the paper is organized as follows: in section 2 we review some basic definitions and properties of the existing point based nonlocal TV operator. A block based maximum a posterior estimation (MAP) model and the two block nonlocal variational models  $BNLH^1$  and BNLTV are proposed in section 3; section 4 presents the algorithms and some details about the implementation; experimental results are shown in section 5; we summarize our method and conclude the paper in section 6.

## 2. Nonlocal TV

In this section we will review some basic defintions of point based nonlocal derivative, nonlocal gradient, nonlocal divergence and Laplace operator as following [10].

Let  $\Omega \subset \mathbb{R}^n$ ,  $u : \Omega \rightarrow \mathbb{R}$ ,  $w : \Omega \times \Omega \rightarrow \mathbb{R}$  be a symmetric and nonnegative weighting function.

Nonlocal derivative of  $u$  at  $x$  with respect to  $y$  can be defined as

$$\partial_y u(x) := (u(y) - u(x))\sqrt{w(x, y)}, x, y \in \Omega,$$

then the nonlocal gradient operator  $\nabla_w : F_1(\Omega; \mathbb{R}) \rightarrow F_2(\Omega \times \Omega; \mathbb{R})$ , where  $F_1(\Omega; \mathbb{R}) \doteq \{u : \Omega \rightarrow \mathbb{R}\}$ ,  $F_2(\Omega \times \Omega; \mathbb{R}) \doteq \{v : \Omega \times \Omega \rightarrow \mathbb{R}\}$ , is defined as

$$(\nabla_w u)(x, y) = (u(x) - u(y))\sqrt{w(x, y)},$$

the nonlocal divergence operator  $\text{div}_w : F_2(\Omega \times \Omega; \mathbb{R}) \rightarrow F_1(\Omega; \mathbb{R})$  defined as the adjoint of the nonlocal gradient is given by :

$$(\text{div}_w v)(x) = \int_{\Omega} (v(x, y) - v(y, x))w(x, y)dy,$$

obviously, the Laplace operator  $\Delta_w : F_1(\Omega; \mathbb{R}) \rightarrow F_1(\Omega; \mathbb{R})$  will be

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

Noting that  $w(x, y) = w(y, x)$ , one can easily verify the adjoint equality

$$\langle \nabla_w u, v \rangle = \langle u, -\operatorname{div}_w v \rangle,$$

and obtain the following results:

$$\begin{aligned} \int_{\Omega \times \Omega} \operatorname{div}_w v dx dy &= 0. \\ \langle \Delta_w u, v \rangle &= \langle u, \Delta_w v \rangle, \\ \langle \Delta_w u, u \rangle &= - \langle \nabla_w u, \nabla_w v \rangle \leq 0. \end{aligned}$$

In the definition of nonlocal operator, the weighting function is always assumed as a known one or selected empirically. However, it is easy to see that the weighting function plays a key role in image denoising. If the empirically chosen weighting function is not good, the restoration would be very bad. Besides, the diffusion of  $\Delta_w$  is point based one and this leads to that the pixel reconstruction would be point by point, thus it may be not robust enough if the noise level is high. In order to solve these problems, we will extend the above definitions to some block based new ones.

### 3. The Proposed Methods

#### 3.1. Statistical Methods

In this section, we shall propose a block based maximum a posteriori estimation (MAP) for nonlocal method, in which the weighting function can be determined by the cost functional itself and also can be updated in the iterations.

##### 3.1.1. Block Based Maximum a Posteriori (MAP) Estimation

Let  $\Omega$  be a 2 dimensional discrete set.  $v : \Omega \rightarrow \mathbb{R}$  is the observed noisy image, and  $u : \Omega \rightarrow \mathbb{R}$  stands for the latent clear image. Denote  $B$  as a small symmetrical neighborhood centered at 0, i.e.  $B = \{x : \|x\| < r\}$ . Moreover, we use  $u_B(x)$  to represent a small image block of  $u$ , namely,  $u_B(x) = \{u(x+z) : z \in B\}$ . All the observed image blocks  $v_B(x), x \in \Omega$  can be regarded as some realizations of a random vector  $\xi$ . The MAP estimation problem for  $u$  can be written by

$$u^* = \arg \max_u p(u_B | v_B) = \arg \max_u \frac{p(v_B | u_B)p(u_B)}{p(v_B)}.$$

Since  $p(v_B)$  is known for any given noisy image  $v$ , thus the MAP problem is equivalent to

$$u^* = \arg \max_u \{ \ln p(v_B | u_B) + \ln p(u_B) \}. \quad (1)$$

From the noisy model, one can easily find

$$p(v_B | u_B) \propto \exp\left(-\frac{(u-v)^2}{2\sigma^2}\right),$$

where  $\sigma$  is the standard deviation of the Gaussian noise.

$p(u_B)$  is a priority term for  $u$ . As to natural images, there are usually many similar and repeated structures or textures in the sense of small blocks. We will construct a nonparametric mixture model to describe such a priority. Once  $u$  is estimated, the histogram of all the block  $u_B(y), y \in \Omega$  can be calculated with the formulation

$$\sum_{y \in \Omega} \delta(\mathbf{z} - u_B(y)).$$

Here  $\delta$  is a vector-valued function which count the number of blocks whose structures are the same as  $\mathbf{z}$ , i.e.

$$\delta(\mathbf{x}) = \begin{cases} 1, & \mathbf{x} = \mathbf{0}, \\ 0, & \text{else.} \end{cases}$$

For calculation convenience, usually  $\delta$  can be approximated by a Gaussian function

$$\delta_h(\mathbf{x}) = \frac{1}{(\pi h)^{\frac{|\mathbf{B}|}{2}}} \exp\left(-\frac{\|\mathbf{x}\|^2}{h}\right),$$

where  $h > 0$  is a parameter which controls the the precision of this approximation.

Assume all the clear image blocks  $u_B(x), x \in \Omega$  are some realizations of a random vector with the probability density function

$$p(\mathbf{z}) = \frac{1}{|\Omega|} \sum_{y \in \Omega} \delta_h(\mathbf{z} - u_B(y)),$$

then according to the independent identify distribution assumption, one can obtain

$$p(u_B) = \frac{1}{|\Omega|^{|\Omega|}} \prod_{x \in \Omega} \sum_{y \in \Omega} \delta_h(u_B(x) - u_B(y)).$$

Therefore, ignoring any constant terms, the problem (1) becomes

$$u^* = \arg \min_u \left\{ E(u) = \frac{\lambda}{2} \sum_{x \in \Omega} (u(x) - v(x))^2 - \sum_{x \in \Omega} \ln \sum_{y \in \Omega} \delta_h(u_B(x) - u_B(y)) \right\}. \quad (2)$$

The first term in the above cost functional is the fidelity term, which is used to measure the difference between the noisy image  $v$  and clear image  $u$  in terms of noisy model. The second term can be regraded as a regularization term, which describes the similarity among the small image blocks  $u_B(x)$ .  $\lambda > 0$  is a regularization parameter which balances these two terms.

To optimize  $E(u)$  directly is not easy since the second regularization term, which is derived from a mixture model, is associated to a complex nonlinear equation if we employ the first order optimization conditions. However, this problem can be greatly simplified by the well-known expectation maximum (EM) algorithm. For convenience, we denote

$$E_1(u) = - \sum_{x \in \Omega} \ln \sum_{y \in \Omega} \delta_h(u_B(x) - u_B(y)).$$

The key idea of EM algorithm to solve such a problem as following: in order to minimize  $E_1(u)$ , we turn to construct another function  $H(u; u^n)$  which can be easily optimized and has a proposition

**Proposition 3.1.** *If  $H(u^{n+1}; u^n) \leq H(u^n; u^n)$ , then  $E_1(u^{n+1}) \leq E_1(u^n)$ , where  $n$  is an iteration number.*

According to this proposition, the value of  $E_1$  would be descended during the iteration of minimizing  $H$ . Therefore, the original optimization problem of  $E_1$  can be replaced by finding the minimizer of  $H$ , which is easier to be minimized.

In the next sections, we will give the formulation of  $H$  and give a variational interpretation for such a EM process.

### 3.1.2. Expectation Maximum (EM) Process and Weighting Function

Please note to minimize  $E_1$  is a standard parameters estimation problem of mixture model. We assume that all the small image blocks  $u_B(x), x \in \Omega$  can be divided into  $|\Omega|$  classes (Some classes may be empty). Let we introduce a latent integer random variable, whose values are in the range of  $\{1, 2, \dots, |\Omega|\}$  and each  $\eta(x)$  indicates that each image block  $u_B(x)$  centered at  $x$  should belongs to the  $\eta(x)$ -th classes. We use a function  $\omega(x, y)$  to represent the probability of each block  $u_B(x)$  belongs to  $y$ -th classes. According to the standard EM method, the expectation step can be written as

E-step:

$$\omega^n(x, y) = P\{\eta(x) = y | u_B^n(x)\} = \frac{\delta_h(u_B^n(x) - u_B^n(y))}{\sum_{y \in \Omega} \delta_h(u_B^n(x) - u_B^n(y))}. \quad (3)$$

Then the related maximum step is

M-step:

$$\begin{aligned} u^{n+1} &= \arg \max_u \left\{ -\sum_{x \in \Omega} \sum_{y \in \Omega} \ln(\delta_h(u_B(x) - u_B(y))) \omega^n(x, y) \right\} \\ &= \arg \min_u \left\{ \sum_{x \in \Omega} \sum_{y \in \Omega} \ln(\delta_h(u_B(x) - u_B(y))) \omega^n(x, y) := H(u; u^n) \right\}. \end{aligned} \quad (4)$$

Please note  $H$  is easy to optimized since it is a quadratic term.

In our previous work [18], we reinterpreted the above EM process in a variational framework. The E-step and M-step can be regarded as two steps of alternating direction minimization method. For completeness, we list the main results of [18] here:

**Lemma 3.1 (Commutativity of log-sum operations).** *Given a function  $\delta_h(x, y) > 0$ , we have*

$$-\sum_{x \in \Omega} \ln \sum_{y \in \Omega} \delta_h(x, y) = \min_{\omega \in \mathbb{C}} \left\{ -\sum_{x \in \Omega} \sum_{y \in \Omega} \ln \delta_h(x, y) \omega(x, y) + \sum_{x \in \Omega} \sum_{y \in \Omega} \omega(x, y) \ln \omega(x, y) \right\},$$

where  $\mathbb{C} = \left\{ \omega(x, y) : 0 \leq \omega(x, y) \leq 1, \sum_{y \in \Omega} \omega(x, y) = 1, \forall x \in \Omega \right\}$ .

Hence, we define

$$H(u, \omega) = -\sum_{x \in \Omega} \sum_{y \in \Omega} \ln \delta_h(u_B(x) - u_B(y)) \omega(x, y) + \sum_{x \in \Omega} \sum_{y \in \Omega} \omega(x, y) \ln \omega(x, y),$$

and can get

$$\min_u E_1(u) = \min_u \left\{ \min_{\omega \in \mathbb{C}} H(u, \omega) \right\}.$$

One can choose the alternating direction minimization method to solve such a problem and get an iteration scheme

$$\begin{cases} \omega^{n+1} &= \arg \min_{\omega \in \mathbb{C}} H(u^n, \omega), \\ u^{n+1} &= \arg \min_u H(u, \omega^{n+1}). \end{cases} \quad (5)$$

It is easy to check that the two problems in (5) are equal to the E-step (3) and M-step (4), respectively. Moreover, for the scheme (5), we have

**Lemma 3.2 (Energy Descent).** *The sequence  $u^n$  produced by iteration scheme (5) satisfies*

$$E_1(u^{n+1}) \leq E_1(u^n).$$

Based on the former analysis, we employ the functional  $H(u, \omega)$  as the regularization term and propose the following image restoration functional

$$(u^*, \omega^*) = \arg \min_{u, \omega \in \mathbb{C}} \left\{ \frac{\lambda}{2} \sum_{x \in \Omega} (u(x) - v(x))^2 + h \sum_{x \in \Omega} \sum_{y \in \Omega} \omega(x, y) \ln \omega(x, y) + \sum_{x \in \Omega} \sum_{y \in \Omega} \|u_B(x) - u_B(y)\|^2 \omega(x, y) \right\}.$$

By setting  $\lambda = 0$ , the above model can be regarded as an extension of nonlocal means method. To be different from the nonlocal means in which the weighting functions are manually chosen according to some experience, here the weighting functions are determined by the restoration cost functional itself. Besides, the estimation in nonlocal means is point by point but here is blocks based. Generally speaking, the block based estimation can improve the precision and be more robust to noise. It is well known that the nonlocal

means is related to the nonlocal  $H^1$  regularization problem, hence we can extend the block based method to the nonlocal TV version. In the next section, from the view of the variational method, we will propose a block nonlocal TV model with adaptive weighting functions.

### 3.2. Variational Framework

#### 3.2.1. Some Definitions and Notations

Let  $\Omega \subset \mathbb{R}^2$  be a bounded and open set,  $B_r \subset \mathbb{R}^2$  be a small symmetrical region centered at 0 with radius  $r$ , i.e.  $B_r = \{z : \|z\| < r\}$ .  $\omega : \Omega \times \Omega \rightarrow \mathbb{R}^+$  is a nonnegative smooth weighting function,  $u : \Omega \rightarrow \mathbb{R}$  is an image function, and  $g : B_r \rightarrow \mathbb{R}^+, \rho_\varepsilon : B_\varepsilon \rightarrow \mathbb{R}^+$  are two others smooth weighting function with  $g(z) = g(-z), \int_{B_r} g(z) dz = 1$  and  $\int_{B_\varepsilon} \rho_\varepsilon(z) dz = 1$ . Denote  $\tilde{\Omega} = \Omega + B_r + B_\varepsilon = \{x | x = x_1 + x_2 + x_3, x_1 \in \Omega, x_2 \in B_r, x_3 \in B_\varepsilon\}$ . We use the symbol “ $\tilde{\cdot}$ ” to represent an expansion of a function. For example,  $\tilde{u}$  is an expansion of  $u$  with

$$\tilde{u}(x) = \begin{cases} u(x), & x \in \Omega, \\ u_0(x), & x \in \tilde{\Omega} \setminus \Omega. \end{cases}$$

In which  $u_0(x)$  is a given function associated to the so-called boundary condition. To simplify the following formulations, here we set the boundary condition  $u_0(x) = 0$ . Other boundary conditions such as Neumann boundary condition also can be discussed as the similar way. In the next discussion, without special statement, the boundary conditions are all set to 0 if we need the expansions of any functions. The symbol “ $*$ ” is used to represent the convolution operator, namely,  $(\rho_\varepsilon * u)(x) = \int_{B_\varepsilon} \rho_\varepsilon(y) u(x - y) dy$ .

We define a block nonlocal gradient operator

$$\nabla_\omega^g : L^2(\Omega) \rightarrow L^2(\Omega \times \Omega \times B_r),$$

which has the following representation:

$$\nabla_\omega^g u(x) = \sqrt{g(z)} [(\rho_\varepsilon * \tilde{u})(y + z) - (\rho_\varepsilon * \tilde{u})(x + z)] \sqrt{\omega(x, y)}, \quad \forall x \in \Omega, y \in \Omega, z \in B_r.$$

Here we employ two smooth function  $g$  and  $\rho_\varepsilon$  in the definition of the block nonlocal gradient since we need  $\rho_\varepsilon$  to keep the later constructed regularization term to possess some better mathematical properties, which can theoretically guarantee that we can proof the existence of solutions for our proposed model (please see in section 3.2.4). Hence, the related block nonlocal divergence operator  $\text{div}_\omega^g : L^2(\Omega \times \Omega \times B_r) \rightarrow L^2(\Omega)$  can be given by

$$\text{div}_\omega^g p(x, y, z) = \int_\Omega \int_{B_r} \int_{B_\varepsilon} \sqrt{g(z)} \rho_\varepsilon(s) \left[ \tilde{p}(x - z + s, y, z) \sqrt{\tilde{\omega}(x - z + s, y)} - \tilde{p}(y, x - z + s, z) \sqrt{\tilde{\omega}(y, x - z + s)} \right] ds dz dy.$$

It is easy to check

$$\langle \nabla_\omega^g u, p \rangle = - \langle \text{div}_\omega^g p, u \rangle$$

with the 0 boundary conditions for  $p, \omega$  and  $u$ . Here  $\langle, \rangle$  is the standard inner product in  $L^2$ .

The related block nonlocal Laplacian operator  $\Delta_\omega^g : L^2(\Omega) \rightarrow L^2(\Omega)$  is given by

$$\Delta_\omega^g u(x) = \int_\Omega \int_{B_r} \int_{B_\varepsilon} g(z) \rho_\varepsilon(s) [(\rho_\varepsilon * \tilde{u})(y + z) - (\rho_\varepsilon * \tilde{u})(x + s)] [\tilde{\omega}(y, x - z + s) + \tilde{\omega}(x - z + s, y)] ds dz dy$$

which satisfies  $\Delta_\omega^g = \text{div}_\omega^g \nabla_\omega^g$ .

The proposed blocked based nonlocal operators can be considered as the vector-valued extensions of the existing nonlocal operators. Please see the figure 1 for some comparisons.

#### 3.2.2. Block Nonlocal $H^1$ with Adaptive Weighting Function

The block nonlocal  $H^1$  norm of  $u$  can be given as

$$(H^1)_\omega^g(u) = \int_\Omega \int_\Omega \int_{B_r} g(z) [(\rho_\varepsilon * \tilde{u})(y + z) - (\rho_\varepsilon * \tilde{u})(x + z)]^2 \omega(x, y) dz dy dx.$$

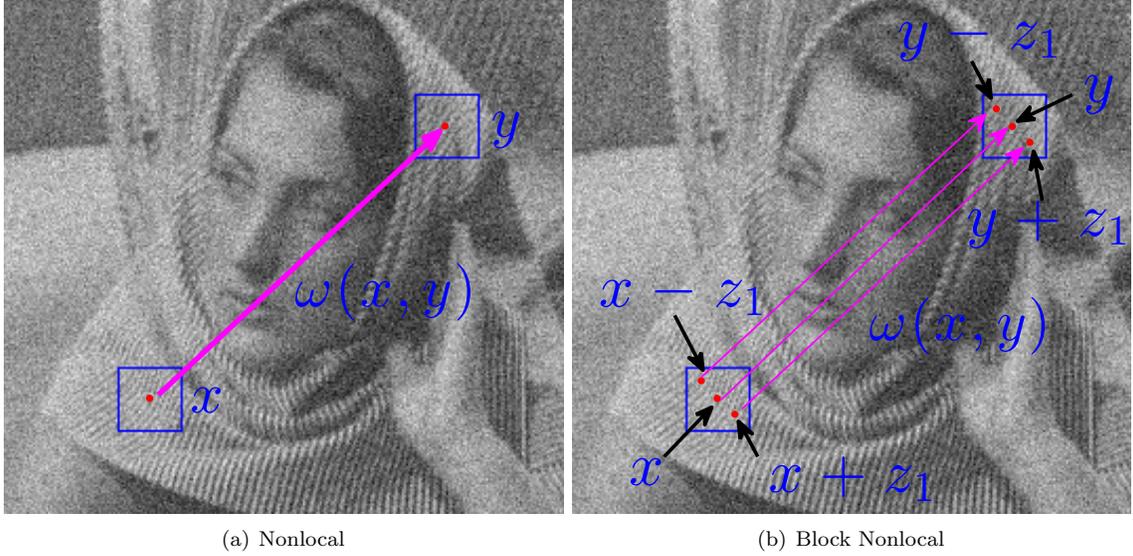


Figure 1: Comparison of point based nonlocal operator and the proposed block based nonlocal operator.

Thus, the previous model can be modified as  $BNLH^1$  model:

$$(u^*, \omega^*) = \arg \min_{u, \omega \in \mathbb{C}} \left\{ \frac{\lambda}{2} \int_{\Omega} (u(x) - v(x))^2 dx + h \int_{\Omega} \int_{\Omega} \omega(x, y) \ln \omega(x, y) dy dx + (H^1)_{\omega}^g(u) \right\},$$

where  $\mathbb{C} = \{\omega(x, y) : 0 \leq \omega(x, y) \leq 1, \int_{\Omega} \omega(x, y) dy = 1, \forall x \in \Omega\}$

### 3.2.3. Block Nonlocal TV

By defining

$$TV_{\omega}^g(u) = \int_{\Omega} \sqrt{\int_{\Omega} \int_{B_r} g(z) [(\rho_{\varepsilon} * \tilde{u})(y+z) - (\rho_{\varepsilon} * \tilde{u})(x+z)]^2 \omega(x, y) dz dy dx}.$$

we get a BNLTV model:

$$(u^*, \omega^*) = \arg \min_{u, \omega \in \mathbb{C}} \left\{ \frac{\lambda}{2} \int_{\Omega} (u(x) - v(x))^2 dx + h \int_{\Omega} \int_{\Omega} \omega(x, y) \ln \omega(x, y) dy dx + TV_{\omega}^g(u) \right\}, \quad (6)$$

where  $\mathbb{C} = \{\omega(x, y) : 0 \leq \omega(x, y) \leq 1, \int_{\Omega} \omega(x, y) dy = 1, \forall x \in \Omega\}$ .

In the proposed method, there are some convolutions  $\rho_{\varepsilon} * \tilde{u}$  in regularization term. As mentioned earlier, the reason for such a design is that these convolutions operators can help us to theoretically proof the existence of minimizers. From the viewpoint of statistics, it requires that there are some self-similarity blocks in the smooth version of image  $u$ . This requirement is reasonable. In practice computation, we can let  $\varepsilon$  be small enough, and then  $\rho_{\varepsilon}$  becomes a delta function.

### 3.2.4. Existence of Minimizer for the Proposed Models

In this section, we will proof the existence of minimizer for  $BNLH^1$ . As to the BNLTV model, we have not yet investigated a theoretical result and let us leave it as a future work.

Let us consider the energy functional for  $BNLH^1$  model

$$J(u, \omega) = \frac{\lambda}{2} \int_{\Omega} (u(x) - v(x))^2 dx + h \int_{\Omega} \int_{\Omega} \omega(x, y) \ln \omega(x, y) dy dx + (H^1)_{\omega}^g(u) \quad (1)$$

We will show the existence of minimizer for (1) in the following space

$$X := \{(u, \omega) : u \in L^2(\Omega), 0 \leq \omega \leq 1, \text{ a.e.}, \text{ and } \int_{\Omega} \omega(x, y) dy = 1, \text{ a.e. } x \in \Omega\}.$$

**Proposition** There exists a solution for BNLH<sup>1</sup> model

$$\min_{(u, \omega) \in X} J(u, \omega) \quad (2).$$

**Proof:** Let

$$\begin{aligned} J_1(u) &= \frac{\lambda}{2} \int_{\Omega} (u(x) - v(x))^2 dx, \\ J_2(\omega) &= h \int_{\Omega} \int_{\Omega} \omega(x, y) \ln \omega(x, y) dy dx, \\ J_3(u, \omega) &= \int_{\Omega} \int_{\Omega} \int_B g(z) (\rho_{\varepsilon} * \tilde{u}(y+z) - \rho_{\varepsilon} * \tilde{u}(x+z))^2 \omega(x, y) dz dy dx. \end{aligned}$$

It is obviously that  $J_1(u) \geq 0$  and  $J_3(u, \omega) \geq 0$ . Furthermore, since  $x \ln x \geq -\frac{1}{e}$  whenever  $x \geq 0$  and  $\Omega$  is a bounded domain,  $J_2(\omega) \geq -\frac{h}{e} |\Omega|^2$ . That is,  $J(u, \omega)$  has a lower bound and  $\inf_{(u, \omega) \in X} J(u, \omega)$  exists. Let  $\{(u_n, \omega_n)\}$  be a minimizing sequence of (2), i.e. a sequence such that

$$J(u_n, \omega_n) \rightarrow \inf_{(u, \omega)} J(u, \omega).$$

It is clear that  $\|\omega_n\|_{L^\infty} \leq 1$  and  $\omega_n \in L^\infty(\Omega \times \Omega)$ , as  $L^\infty$  is the conjugate space of  $L^1$  which is a separable linear normed space, by the Banach-Alaoglu Theorem, there is a \* weakly convergent subsequence (which we relabel by  $n$ ), and a \* weak limit  $\omega \in L^\infty(\Omega \times \Omega)$  such that

$$\omega_n \overset{*}{\rightharpoonup} \omega \quad \text{in } L^\infty(\Omega \times \Omega),$$

That is, for any  $\varphi \in L^1(\Omega \times \Omega)$ ,

$$\int_{\Omega \times \Omega} \omega_n(x, y) \varphi(x, y) dx dy \rightarrow \int_{\Omega \times \Omega} \omega(x, y) \varphi(x, y) dx dy, \quad n \rightarrow +\infty.$$

Since  $\omega \ln \omega$  is continuous and convex with respect to  $\omega$ , it is quasiconvex and  $J_2(\omega)$  is \* weakly lower semicontinuous, i.e.

$$\liminf_{n \rightarrow \infty} \int_{\Omega} \int_{\Omega} \omega_n(x, y) \ln \omega_n(x, y) dy dx \geq \int_{\Omega} \int_{\Omega} \omega(x, y) \ln \omega(x, y) dy dx.$$

We also need to verify the \* weak limit  $\omega$  is in  $X$ , that is,  $0 \leq \omega \leq 1$  a.e., and  $\int_{\Omega} \omega(x, y) dy = 1$ , a.e.  $x \in \Omega$ .

Owing to  $\omega_n \overset{*}{\rightharpoonup} \omega$  in  $L^\infty$ , then, for any  $\varphi \in L^1$ ,

$$\int_{\Omega} \int_{\Omega} \omega_n(x, y) \varphi(x, y) dy dx \rightarrow \int_{\Omega} \int_{\Omega} \omega(x, y) \varphi(x, y) dy dx \quad (3)$$

Set  $A = \{(x, y) \in \Omega \times \Omega : \omega(x, y) > 1\}$ ,  $B_r = \{(x, y) \in \Omega \times \Omega : \omega(x, y) < 0\}$ ,  $\varphi_1(x, y) = \chi_A(x, y)$ ,  $\varphi_2(x, y) = \chi_{B_r}(x, y)$ , then  $\varphi_1, \varphi_2 \in L^1(\Omega \times \Omega)$ . We will show that  $|A| = 0$  and  $|B_r| = 0$ .

Substituting  $\varphi_1$  for  $\varphi$  in (3), we obtain

$$\int_A \omega_n(x, y) dy dx \rightarrow \int_A \omega(x, y) dy dx \quad (4)$$

If  $|A| \neq 0$ , the right hand side of (4) is bigger than  $|A|$ , by sign-preserving theorem of limit, there is a  $N > 0$ ,

for any  $n > N$ ,  $\int_A \omega_n(x, y) dy dx > |A|$ . However, since  $\omega_n \leq 1$ , the left hand side of (4) should be smaller than or equal to  $|A|$  for any  $n$ , which is a contradiction. Thus, we have  $|A| = 0$ . In the same way, we obtain  $|B_r| = 0$ . The results show that  $0 \leq \omega \leq 1$  a.e..

Let  $f(x) = \int_\Omega \omega(x, y) dy$ ,  $\varphi(x, y) = \text{sign}(f(x) - 1)$ , then  $\varphi \in L^1$ , and

$$\int_\Omega \left( \int_\Omega \omega_n(x, y) dy \right) \varphi(x) dx \rightarrow \int_\Omega f(x) \varphi(x) dx$$

i.e.

$$\int_\Omega |f(x) - 1| dx = 0$$

hence,  $\int_\Omega \omega(x, y) dy = 1$  for a.e.  $x \in \Omega$ . Above all, we have  $\omega \in X$ , and

$$\liminf_{n \rightarrow \infty} J_2(\omega_n) \geq J_2(\omega).$$

Since  $J(u, \omega)$  is coercive,  $\{u_n\}$  must be bounded, i.e. there is a constant  $M > 0$  such that

$$\|u_n\|_{L^2(\Omega)} \leq M.$$

Boundedness of the sequence in a reflexive space implies the existence of a weakly convergent subsequence, which we still denote by  $\{u_n\}$ , then, there is a weak limit  $u \in L^2(\Omega)$  such that

$$u_n \rightharpoonup u.$$

Since  $J_1 : u \mapsto \frac{\lambda}{2} \int_\Omega (u(x) - v(x))^2 dx$  is continuous and convex, and  $u_n \rightharpoonup u$  in  $L^2(\Omega)$ , then

$$\liminf_{n \rightarrow \infty} \frac{\lambda}{2} \int_\Omega (u_n(x) - v(x))^2 dx \geq \frac{\lambda}{2} \int_\Omega (u(x) - v(x))^2 dx$$

i.e.

$$\lim_{n \rightarrow \infty} J_1(u_n) = J_1(u).$$

Now, we consider the convergence of  $J_3(u_n, \omega_n)$ .

Here, we assume the support of  $g$  is a proper subset of  $B$  and choose  $\varepsilon$  equals to one half of the distance between the support and  $B$ .

Set  $b_n(x, y) = \int_{B_r} g(z) (\rho_\varepsilon * \tilde{u}_n(x+z) - \rho_\varepsilon * \tilde{u}_n(y+z))^2 dz$ . By the smoothness of convolution operation, integral  $b_n$  depending on parameters  $x, y$  is in  $C^1(\Omega \times \Omega)$ , and

$$\frac{\partial b_n}{\partial x}(x, y) = 2 \int_{B_r} g(z) (\rho_\varepsilon * \tilde{u}_n(x+z) - \rho_\varepsilon * \tilde{u}_n(y+z)) \left( \frac{\partial \rho_\varepsilon}{\partial x} * \tilde{u}_n(x+z) \right) dz,$$

$$\frac{\partial b_n}{\partial y}(x, y) = 2 \int_{B_r} g(z) (\rho_\varepsilon * \tilde{u}_n(x+z) - \rho_\varepsilon * \tilde{u}_n(y+z)) \left( \frac{\partial \rho_\varepsilon}{\partial y} * \tilde{u}_n(x+z) \right) dz.$$

As

$$\begin{aligned} \int_\Omega \int_\Omega |b_n(x, y)| dx dy &\leq 2 \int_\Omega \int_\Omega \int_{B_r} g(z) [(\rho_\varepsilon * \tilde{u}_n(x+z))^2 + (\rho_\varepsilon * \tilde{u}_n(y+z))^2] dz dx dy \\ &\leq 4|\Omega| \|g\|_\infty \int_\Omega \int_{B_r} (\rho_\varepsilon * \tilde{u}_n(x+z))^2 dz dx \leq 4|\Omega| \|B\| \|g\|_\infty M^2, \end{aligned}$$

$$\begin{aligned} \int_\Omega \int_\Omega |\nabla_x b_n(x, y)| dx dy &\leq 2 \int_\Omega \int_\Omega \int_{B_r} g(z) |\rho_\varepsilon * \tilde{u}_n(x+z) - \rho_\varepsilon * \tilde{u}_n(y+z)| \left| \frac{\partial \rho_\varepsilon}{\partial x} * \tilde{u}_n(x+z) \right| dz dx dy \\ &\leq 4|\Omega| \|B\| \|g\|_\infty \|\rho_\varepsilon\|_{1,\infty} M^2, \end{aligned}$$

then  $\{b_n(x, y)\}$  is a bounded sequence in  $W^{1,1}(\Omega \times \Omega)$ . By Rellich-Kondrachov Compactness Theorem,  $W^{1,1}(\Omega \times \Omega)$  is compactly embedded in  $L^1(\Omega \times \Omega)$ , i.e. there exists a subsequence  $\{b_n(x, y)\}$  which converges to  $b(x, y)$  in  $L^1(\Omega \times \Omega)$ .

Apparently,  $J_3(u, \omega)$  is continuous and convex in  $u$ , then it is weakly lower semicontinuous, that is,

$$\liminf_{n \rightarrow +\infty} \int_{\Omega} \int_{\Omega} b_n(x, y) \omega(x, y) dz dx dy \geq \int_{\Omega} \int_{\Omega} \int_{B_r} g(z) (\rho_{\varepsilon} * \tilde{u}(x+z) - \rho_{\varepsilon} * \tilde{u}(y+z))^2 \omega(x, y) dz dx dy.$$

Furthermore,

$$\begin{aligned} & \int_{\Omega} \int_{\Omega} b_n(x, y) (\omega_n(x, y) - \omega(x, y)) dy dx \\ &= \int_{\Omega} \int_{\Omega} (b_n(x, y) \omega_n(x, y) - b(x, y) \omega(x, y)) dy dx + \int_{\Omega} \int_{\Omega} (b(x, y) \omega(x, y) - b_n(x, y) \omega(x, y)) dy dx \\ &= \int_{\Omega} \int_{\Omega} (b_n(x, y) - b(x, y)) \omega_n(x, y) dy dx + \int_{\Omega} \int_{\Omega} b(x, y) (\omega_n(x, y) - \omega(x, y)) dy dx \\ &+ \int_{\Omega} \int_{\Omega} (b(x, y) - b_n(x, y)) \omega(x, y) dy dx \end{aligned}$$

Because of  $\omega_n \xrightarrow{*} \omega$  in  $L^{\infty}(\Omega \times \Omega)$ ,  $b_n \rightarrow b$  in  $L^1(\Omega \times \Omega)$ , we have

$$\lim_{n \rightarrow +\infty} \int_{\Omega} \int_{\Omega} b_n(x, y) (\omega_n(x, y) - \omega(x, y)) dy dx = 0.$$

As

$$J_3(u_n, \omega_n) = \int_{\Omega} \int_{\Omega} b_n(x, y) \omega_n(x, y) dx dy = \int_{\Omega} \int_{\Omega} b_n(x, y) (\omega_n(x, y) - \omega(x, y)) dx dy + \int_{\Omega} \int_{\Omega} b_n(x, y) \omega(x, y) dx dy,$$

we have

$$\liminf_{n \rightarrow +\infty} J_3(u_n, \omega_n) \geq J_3(u, \omega).$$

In a word,

$$J(u, \omega) = J_1(u) + J_2(\omega) + J_3(u, \omega) \leq \liminf_{n \rightarrow \infty} J(u_n, \omega_n),$$

and

$$J(u, \omega) = \inf_{(u, \omega) \in X} J(u, \omega),$$

which completes the proof.

#### 4. Algorithms

To simplify the computation, we can set  $\rho_{\varepsilon}$  to the delta function  $\delta$ , thus  $\rho_{\varepsilon} * \tilde{u} = \delta * \tilde{u} = \tilde{u}$ . In the following discussion, we set  $\rho_{\varepsilon} = \delta$ .

The BNLH<sup>1</sup> model can be directly minimized by an alternating direction minimization algorithm. We summary the steps in algorithm (4.1)

**Algorithm 4.1.** Given an initial value  $u^0 = v$ , for  $n = 1, 2, 3, \dots$ , do  
Step1, calculating the weighting function by

$$\omega^{n+1}(x, y) = \frac{\exp\left(-\frac{\int_{B_r} g(z) (\tilde{u}^n(y+z) - \tilde{u}^n(x+z))^2 dz}{h}\right)}{\int_{\Omega} \exp\left(-\frac{\int_{B_r} g(z) (\tilde{u}^n(y+z) - \tilde{u}^n(x+z))^2 dz}{h}\right) dy}.$$

Step 2, restoring  $u^{n+1}$  by solving the following equation

$$-\Delta_{\omega^{n+1}}^g u + \lambda u = \lambda v.$$

Step 3, if the convergence condition  $\frac{\|u^{n+1}-u^n\|^2}{\|u^n\|^2}$  less than a tolerant error, then stop. Otherwise, go to the step 1.

As to BNLTV, we need to employ some more efficient algorithms, such as augmented Lagrangian method an splitting Bregman method. Here, we write the iteration scheme as augmented Lagrangian method.

Let us introduce a function  $q = (\nabla_{\omega}^g)u$ , the BNLTV model is equal to the constraint minimization problem

$$(u^*, \omega^*, q^*) = \arg \min_{u, \omega \in \mathbb{C}, q = (\nabla_{\omega}^g)u} \left\{ \frac{\lambda}{2} \int_{\Omega} (u(x) - v(x))^2 dx + h \int_{\Omega} \int_{\Omega} \omega(x, y) \ln \omega(x, y) dy dx + \int_{\Omega} \|q(x, \cdot, \cdot)\|_2 dx \right\}.$$

By applying standard augmented Lagrangian method, one can get a Lagrangian function

$$L(u, \omega, q, d) = \frac{\lambda}{2} \int_{\Omega} (u(x) - v(x))^2 dx + h \int_{\Omega} \int_{\Omega} \omega(x, y) \ln \omega(x, y) dy dx + \int_{\Omega} \|q(x, \cdot, \cdot)\|_2 dx + \langle d, q - (\nabla_{\omega}^g)u \rangle + \frac{\eta}{2} \|q - (\nabla_{\omega}^g)u\|_2^2.$$

The saddle of  $L$  can be split as 4 subproblems, each subproblem is easy to solve.

#### Subproblem 1: expectation maximum step

$$\omega^{n+1} = \arg \min_{\omega \in \mathbb{C}} L(u^{n+1}, \omega, q^n, d^n) = \arg \min_{\omega \in \mathbb{C}} \left\{ h \int_{\Omega} \int_{\Omega} \omega(x, y) \ln \omega(x, y) dy dx + \frac{\eta}{2} \|q^n - (\nabla_{\omega}^g)u^{n+1} + \frac{d^n}{\eta}\|_2^2 \right\}.$$

#### Subproblem 2: restoration step

$$u^{n+1} = \arg \min_u L(u, \omega^n, q^n, d^n) = \arg \min_u \left\{ \frac{\lambda}{2} \int_{\Omega} (u(x) - v(x))^2 dx + \frac{\eta}{2} \|q^n - (\nabla_{\omega^n}^g)u + \frac{d^n}{\eta}\|_2^2 \right\}.$$

#### Subproblem 3: gradient updating

$$q^{n+1} = \arg \min_q L(u^{n+1}, \omega^{n+1}, q, d^n) = \arg \min_q \left\{ \int_{\Omega} \|q(x, \cdot, \cdot)\|_2 dx + \frac{\eta}{2} \|q - (\nabla_{\omega^{n+1}}^g)u^{n+1} + \frac{d^n}{\eta}\|_2^2 \right\}.$$

#### Subproblem 4: Lagrangian multiplier updating

$$d^{n+1} = \arg \max_d L(u^{n+1}, \omega^{n+1}, q^{n+1}, d) = \arg \max_d \left\{ \langle d, q^{n+1} - (\nabla_{\omega^{n+1}}^g)u^{n+1} \rangle \right\}.$$

As to the subproblem 1, we can use the first order optimization condition to get an approximation solution. For simplifying the representation, let us denote

$$\mathcal{A}^n(x, y) = \exp \left\{ -\frac{\eta}{2h} \int_{B_r} \left[ g(z)(\tilde{u}^{n+1}(y+z) - \tilde{u}^{n+1}(x+z))^2 + \frac{1}{\sqrt{\omega^n(x, y)}} \sqrt{g(z)}(\tilde{u}^{n+1}(y+z) - \tilde{u}^{n+1}(x+z))(q^n(x, y, z) + \frac{d^n(x, y, z)}{\eta}) \right] dz \right\}$$

Then we obtain

$$\omega^{n+1}(x, y) = \frac{\mathcal{A}^n(x, y)}{\int_{\Omega} \mathcal{A}^n(x, y) dy}. \quad (7)$$

To get the solution of subproblem 2, we need to solve equation

$$\lambda u - \eta \Delta_{\omega}^g u = \lambda v - \eta \operatorname{div}_{\omega}^g (q^n + \frac{d^n}{\eta}). \quad (8)$$

This equation can be solved iteratively by many methods such as Gauss-Seidel iteration.  
 $q^{n+1}$  can be given by a shrinkage process

$$q^{n+1}(x, y, z) = \mathcal{S}((\nabla_{\omega^{n+1}}^g)u^{n+1}(x, y, z) - \frac{d^n(x, y, z)}{\eta}, \eta), \quad (9)$$

where the shrinkage operator

$$\mathcal{S}(f(x, y, z), \eta) = \begin{cases} \frac{f(x, y, z)}{\|f(x, \cdot, \cdot)\|_2} (\|f(x, \cdot, \cdot)\|_2 - \frac{1}{\eta}), & \|f(x, \cdot, \cdot)\|_2 \geq \frac{1}{\eta}, \\ 0, & \|f(x, \cdot, \cdot)\|_2 < \frac{1}{\eta}, \end{cases}$$

and  $\|f(x, \cdot, \cdot)\|_2 = \sqrt{\int_{\Omega} \int_{B_r} f^2(x, y, z) dz dy}$ . The last subproblem can be easily maximized by

$$d^{n+1}(x, y, z) = d^n(x, y, z) + \tau(q^{n+1}(x, y, z) - (\nabla_{\omega^{n+1}}^g)u^{n+1}(x, y, z)), \quad (10)$$

where  $0 < \tau < 2\eta$  is a parameter which controls the convergence of the augmented Lagrangian algorithm. Usually, it can be chosen as  $\tau = \eta$ .

We summarize the steps as algorithm 4.2

**Algorithm 4.2.** Given an initial value  $u^0 = v, \omega^0 = \frac{1}{|\Omega|}, d^0 = q^0 = 0$ , for  $n = 1, 2, 3, \dots$ , do

Step1, weighting function updating.:

Calculating the weighting function by (7).

Step 2, image restoration:

Restoring  $u^{n+1}$  by solving the equation (8).

Step 3, gradient updating:

Computing  $q^{n+1}$  according to (9).

Step 4, Lagrangian multiplier updating:

Renewing  $d^{n+1}$  by formulation (10).

Step 5, convergence checking:

If the convergence condition  $\frac{\|u^{n+1} - u^n\|^2}{\|u^n\|^2}$  less than a tolerant error, then stop. Otherwise, go to the step 1.

## 5. Experimental Results

In this section, we will show some numerical results for the proposed method and make some comparison with some related methods. Without special statements, in the proposed method, some parameters are chosen as following:  $h = 0.25^2, \sigma = 3.0, \eta = 10, B_r$  is set to a  $7 \times 7$  rectangle path.  $\lambda$  is set as different values according to the noise levels, we will list the exact values later. For computational efficiency and saving memory storage, we do not let  $y$  go through the entire  $\Omega$  and just set a  $13 \times 13$  search window for every  $x$ .

According to the definition of the block gradient operator  $\nabla_{\omega}^g$ , both local and nonlocal factors are considered by the variables  $y$  and  $z$ , respectively. Thus, the proposed block based ones are more robust for heavy noise than the point based nonlocal operators [10]. Let us give an intuitive interpretation for such a diffusion mechanism. Let  $\rho_{\varepsilon} = \delta$  and  $\omega$  be symmetrical, then the Gauss-Seide iteration for  $-\Delta_{\omega}^g u$  would be

$$u^{n+1}(x) = \frac{\int_{\Omega} \int_{B_r} u^n(y+z) \omega(x-z, y) g(z) dz dy}{\int_{\Omega} \int_{B_r} g(z) \omega(x-z, y) dz dy}. \quad (11)$$

This equation means that all of the pixels at  $y$  and  $y+z$  are averaged with the weighting function  $\omega(x-z, y)g(z)$ . The pixels located at  $y$  could be far way from  $x$ , these pixels can be regarded as nonlocal ones. Since  $z$  only can be taken in a local neighborhood centered at 0, thus  $y+z$  are local positions of  $y$ . Furthermore, please note that the path centered at  $y$  with large weight would be very similar to the one centered at  $x$ , then  $u(y+z)$  can be regarded as similar local pixels of  $u(x)$ . In this sense, the block nonlocal diffusion are both local and nonlocal. Please see figure 2 for the ideas.

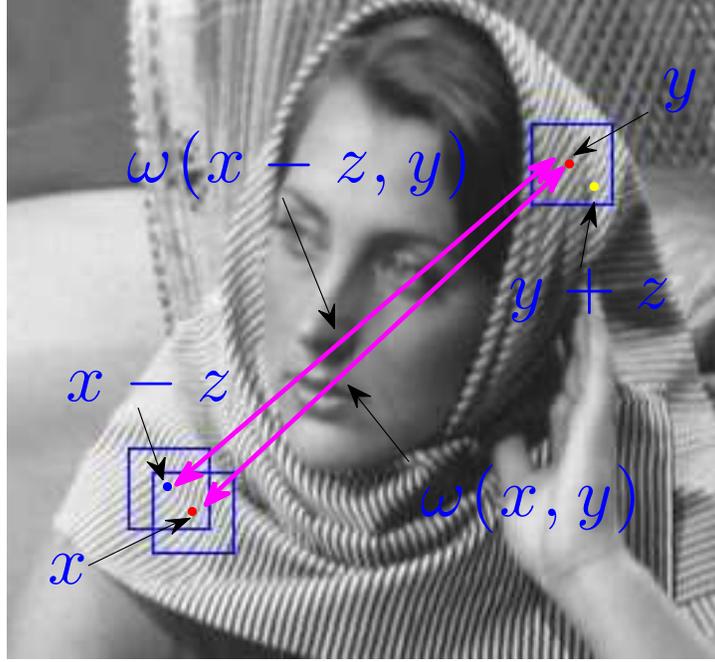


Figure 2: The schematic plot of the Gauss-Seide iteration for  $-\Delta_{\omega}^g u$

The equation (11) also can be reformulated as

$$u^{n+1}(x) = \frac{\int_{\Omega} \int_{B_r} u^n(y) \omega(x-z, y-z) g(z) dz dy}{\int_{\Omega} \int_{B_r} g(z) \omega(x-z, y-z) dz dy}$$

with some appropriate boundary condition. From this viewpoint, the above formulation means that one pixel  $u(x)$  may be contained in many overlapping patches in the block based mode, thus  $u(x)$  would be estimated for many times. Our formulation give a weighting function to average all the repeatedly estimated  $\hat{u}(x)$ . Thus, the block based operators are more robust for noise. It can be found in the following experiments.

In the first experiment, we show the superiority of block diffusion operators. The clean original “barbara” image in the figure 3(a) was destroyed by the additive Gaussian white noise with standard deviation  $\sigma = 80$ , and the noisy image was shown in the figure 3(b). We recover it with different denoising methods. For each algorithm, we try different parameters and select the best reconstructed results (with the highest PSNR) for comparison. Let us first show the differences of local, nonlocal and block based nonlocal operators. To see this clearly, we fix the weighting functions and do not update them in the related nonlocal methods. The result produced by TV [22] was displayed in the figure 3(c), one can find this restored image is almost piecewise constant and loses a lot of texture structures. The PSNR for TV is 22.08 dB. Compared with TV, the NLTV [10] can keep more texture details since the repeat structure information was used in this model. However, some recovered textures demonstrated in the figure 3(d) are still not very clear since the noise is very heavy. Please find the texture details comparison in the figure 4 which contains the related enlargements of white rectangle areas in figure 3. The PSNR of NLTV is 22.45 dB, which is higher than TV’s. In the figure 4(g), we show the result provided by the proposed block based nonlocal method BNLTV. One can find it can produce better texture restorations with PSNR 22.90 dB (please see figure 4 for texture details) under heavy noise because both of the local and nonlocal information are used in this algorithm. In this experiment, the fidelity parameters for TV, NLTV, and BNLTV are 3.6, 1.0 and 2.0, respectively.

In the next, the advantages of the weighting function updating will be illustrated. In figure 6, we restore the same noisy image in figure 4 by using the NLTV and BNLTV (algorithm 4.2) with 10 times iteration. The related result are shown as in the figure 5(a) and 5(b), respectively. Obviously, the reconstructed images have a much better quality than the previous ones, please also see the details comparison in figure 4. It is

Images	Methods	$\sigma$					
		10	30	50	80	100	120
Barbara (512 × 512)	NLTV [10]	33.08	27.41	24.70	22.45	21.52	20.95
	BNLTV	33.53	28.36	25.76	23.62	22.92	22.15
Lena (256 × 256)	NLTV [10]	33.42	27.50	25.16	23.00	21.99	21.21
	BNLTV	33.69	28.39	25.77	24.31	23.52	22.79
House (256 × 256)	NLTV [10]	35.20	30.20	27.30	24.63	23.46	22.65
	BNLTV	35.30	30.79	28.42	26.21	25.29	24.35

Table 1: PSNRs (dB) for NLTV and the proposed BNLTV under Gaussian noise with different levels.

reasonable because the weighting function could be improved by some new estimated images. We display the PSNR values of the BNLN method during the iteration in the figure 6. The final PSNR of NLTV and BNLTV are 23.22 dB and 23.62 dB, respectively. Compared to the previous result without weighting function updating, the PSNRs are improved more than 0.7 dB. Based on some experimental experience, we find that it would be better to choose a big fidelity parameter to prevent the recovered image over smooth, thus we set the fidelity parameter  $\lambda$  to be 4.0 for NLTV and 2.0 for BNLTV.

For more comparison results of different test images under a variety of noise levels, please see in table 1. It can be found that the proposed BNLTV has a better performance than NLTV in all the cases. Generally speaking, the heavier the noise, the larger increasement for the PSNR.

## 6. Conclusion and Discussion

In this paper, we proposed a block based nonlocal method for image restoration. In our model, the weighting function in nonlocal method is not fixed but adaptively determined by the cost functional itself. The weighting function  $\omega(x, y)$  also has a statistical meaning, and it stands for a probability of the similarity between patches centered at  $x$  and  $y$ . By updating the weighting function, the quality of the restored images will be greatly improved, especially under the heavy noise. Besides, we also extend the existing point based nonlocal diffusion to a block based one. The numerical results show that the proposed block nonlocal operators are more robust for noise than the previous ones. It can help us to get a better reconstruction images.

The proposed method can be applied in many image tasks. One of the possible application is the inpainting and compressed sensing, in which the weights are more important and our method would greatly improve the results. Due to the page limitation, let us left them as a future work in the oncoming another paper.

Though the proposed method can produce better result than the existing NLTV, the computational cost is higher than NLTV. Generally speaking, for  $256 \times 256$  images, it would cost about 11 seconds CPU time on MacBook Pro 2011 for one outer iteration (including the weight calculation and restoration).

## 7. Acknowledgement

Liu is partially supported by National Natural Science Foundation of China (No. 11201032) and the Fundamental Research Funds for the Central Universities. Liu is also supported by the China Scholarship Council for a one year visiting at UCLA.

## References

- [1] Biemond, J., Tekalp, A. M., Lagendijk, R. L., 1990. Maximum likelihood image and blur identification: a unifying approach. *Optical Engineering* 29 (5), 422–435.
- [2] Buades, A., Coll, B., Morel, J.-M., 2005. A non-local algorithm for image denoising. In: *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. Vol. 2. IEEE, pp. 60–65.



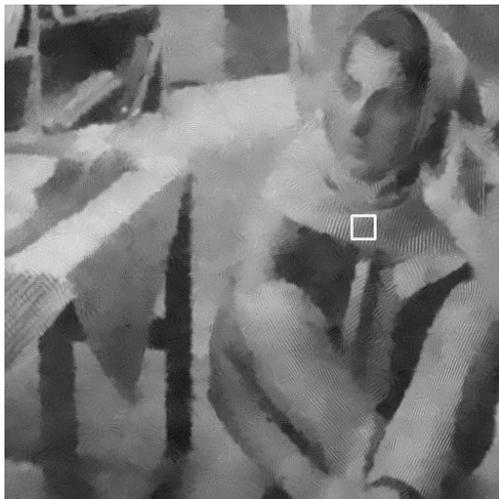
(a) Clean



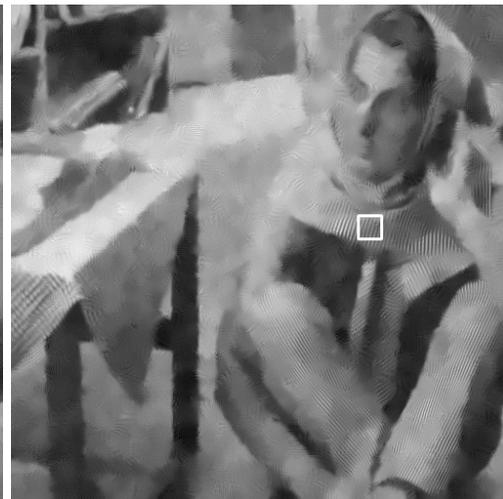
(b) Noisy,  $\sigma = 80$



(c) TV [22], 22.08 dB



(d) NLTV [10], 22.45 dB



(e) BNLT, 22.90 dB

Figure 3: Comparison of point based nonlocal operator and the proposed block based nonlocal operator without weight updating.

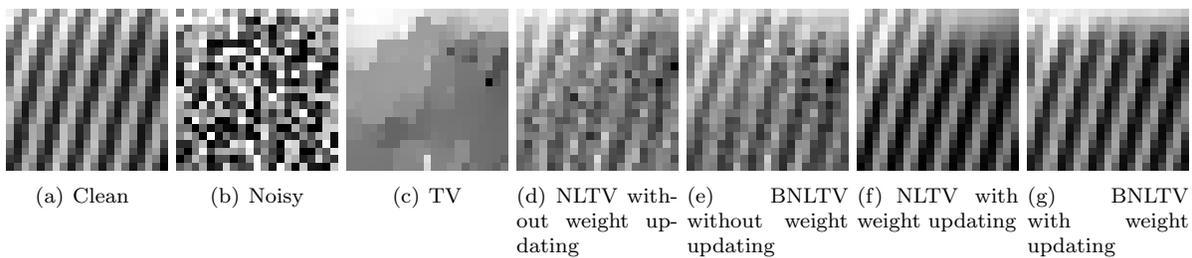


Figure 4: The enlargement of the white rectangle texture area in figure 3 and 5.

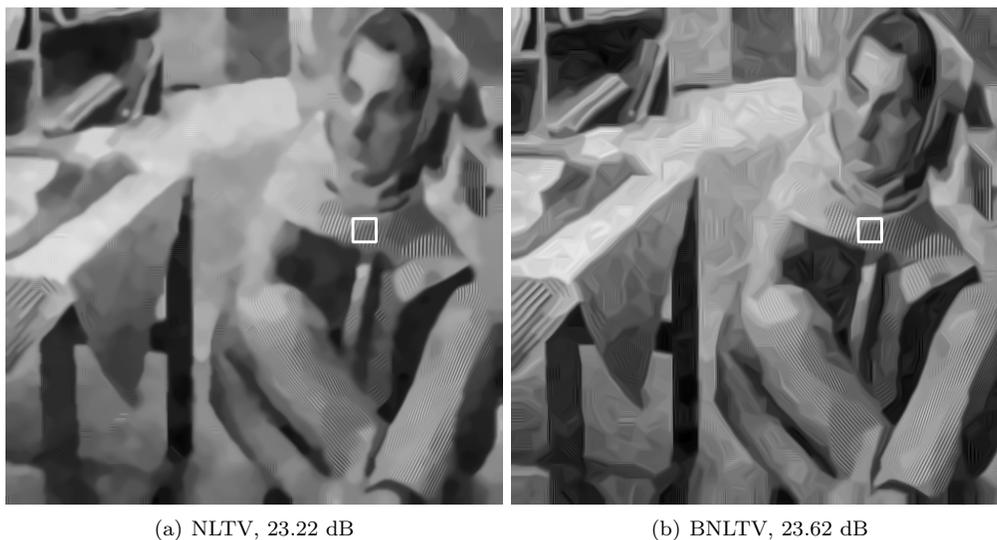


Figure 5: Comparison of point based nonlocal operator and the proposed block based nonlocal operator with weight updating 10 times.

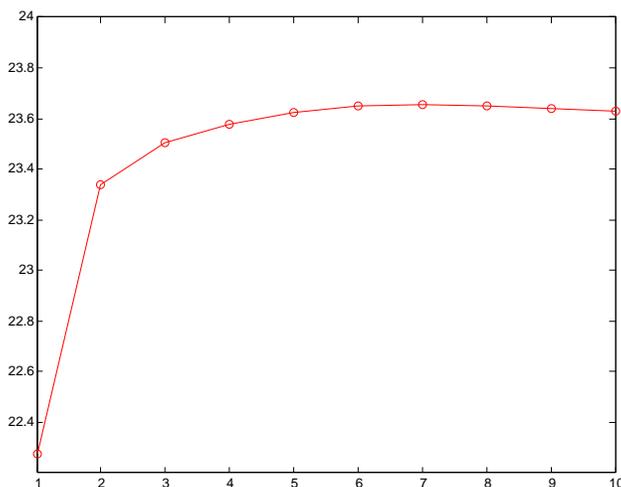


Figure 6: The PSNR values of BNLTV during the iteration. (x-axis: iteration number, y-axis: PSNR values, dB).

- [3] Buades, A., Coll, B., Morel, J.-M., 2005. A review of image denoising algorithms, with a new one. *Multiscale Modeling & Simulation* 4 (2), 490–530.
- [4] Chatterjee, P., Milanfar, P., 2009. Clustering-based denoising with locally learned dictionaries. *Image Processing, IEEE Transactions on* 18 (7), 1438–1451.
- [5] Chatterjee, P., Milanfar, P., 2012. Patch-based near-optimal image denoising. *Image Processing, IEEE Transactions on* 21 (4), 1635–1649.
- [6] Dabov, K., Foi, A., Egiazarian, K., 2007. Video denoising by sparse 3d transform-domain collaborative filtering. In: *Signal Processing Conference, 2007 15th European*. IEEE, pp. 145–149.
- [7] Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K., 2007. Image denoising by sparse 3-d transform-domain collaborative filtering. *Image Processing, IEEE Transactions on* 16 (8), 2080–2095.
- [8] Dong, W., Li, X., Zhang, L., Shi, G., 2011. Sparsity-based image denoising via dictionary learning and structural clustering. In: *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, pp. 457–464.
- [9] Gilboa, G., Osher, S., 2007. Nonlocal linear image regularization and supervised segmentation. *Multiscale Modeling & Simulation* 6 (2), 595–630.
- [10] Gilboa, G., Osher, S., 2008. Nonlocal operators with applications to image processing. *Multiscale Modeling & Simulation* 7 (3), 1005–1028.
- [11] Goldstein, T., Osher, S., 2009. The split bregman method for l1-regularized problems. *SIAM Journal on Imaging Sciences* 2 (2), 323–343.
- [12] Ji, H., Huang, S., Shen, Z., Xu, Y., 2011. Robust video restoration by joint sparse and low rank matrix approximation. *SIAM Journal on Imaging Sciences* 4 (4), 1122–1142.
- [13] Katkovnik, V., Foi, A., Egiazarian, K., Astola, J., 2010. From local kernel to nonlocal multiple-model image denoising. *International journal of computer vision* 86 (1), 1–32.
- [14] Kindermann, S., Osher, S., Jones, P. W., 2005. Deblurring and denoising of images by nonlocal functionals. *Multiscale Modeling & Simulation* 4 (4), 1091–1115.
- [15] Lindenbaum, M., Fischer, M., Bruckstein, A., 1994. On gabor’s contribution to image enhancement. *Pattern Recognition* 27 (1), 1–8.
- [16] Liu, J., Huang, H., Huan, Z., Zhang, H., 2010. Adaptive variational method for restoring color images with high density impulse noise. *International journal of computer vision* 90 (2), 131–149.
- [17] Liu, J., Tai, X.-C., Huang, H., Huan, Z., 2013. A weighted dictionary learning model for denoising images corrupted by mixed noise. *Image Processing, IEEE Transactions on* 22 (3), 1108–1120.
- [18] Liu, J., Zhang, H., 2012. Image segmentation using a local gmm in a variational framework. *J. Math. Imaging Vis.*, DOI:10.1007/s10851-012-0376-5.
- [19] Nikolova, M., 2004. A variational approach to remove outliers and impulse noise. *Journal of Mathematical Imaging and Vision* 20 (1-2), 99–120.
- [20] Perona, P., Malik, J., 1990. Scale-space and edge detection using anisotropic diffusion. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 12 (7), 629–639.
- [21] Rudin, L., Osher, S., et al., 1994. Total variation based image restoration with free local constraints. In: *Image Processing, 1994. Proceedings. ICIP-94., IEEE International Conference*. Vol. 1. IEEE, pp. 31–35.
- [22] Rudin, L. I., Osher, S., Fatemi, E., 1992. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena* 60 (1), 259–268.

- [23] Singer, A., Shkolnisky, Y., Nadler, B., 2009. Diffusion interpretation of nonlocal neighborhood filters for signal denoising. *SIAM Journal on Imaging Sciences* 2 (1), 118–139.
- [24] Wahlberg, B., Boyd, S., Annergren, M., Wang, Y., 2012. An admm algorithm for a class of total variation regularized estimation problems. arXiv preprint arXiv:1203.1828.
- [25] Wu, C., Tai, X.-C., 2010. Augmented lagrangian method, dual methods, and split bregman iteration for rof, vectorial tv, and high order models. *SIAM Journal on Imaging Sciences* 3 (3), 300–339.
- [26] Yaroslavsky, L. P., Yaroslavskij, L., 1985. Digital picture processing. an introduction. Digital picture processing. An introduction.. LP Yaroslavsky (LP Yaroslavskij). Springer Series in Information Sciences, Vol. 9. Springer-Verlag, Berlin-Heidelberg-New York-Tokyo. 12+ 276 pp. Price DM 112.00 (1985). ISBN 3-540-11934-5 (FR Germany), ISBN 0-387-11934-5 (USA). 1.
- [27] Zoran, D., Weiss, Y., 2011. From learning models of natural image patches to whole image restoration. In: *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, pp. 479–486.