

# Nonlocal Adaptive Direction-Guided Structure Tensor Total Variation For Image Recovery

Ezgi Demircan-Tureyen · Mustafa E. Kamasak

Received: date / Accepted: date

**Abstract** A common strategy in variational image recovery is utilizing the nonlocal self-similarity (NSS) property, when designing energy functionals. One such contribution is nonlocal structure tensor total variation (NLSTV), which lies at the core of this study. This paper is concerned with boosting the NLSTV regularization term through the use of directional priors. More specifically, NLSTV is leveraged so that, at each image point, it gains more sensitivity in the direction that is presumed to have the minimum local variation. The actual difficulty here is capturing this directional information from the corrupted image. In this regard, we propose a method that employs anisotropic Gaussian kernels to estimate directional features to be later used by our proposed model. The experiments validate that our entire two-stage framework achieves better results than the NLSTV model and two other competing local models, in terms of visual and quantitative evaluation.

**Keywords** Directional total variation · Image recovery · Nonlocal regularization · Orientation field estimation · Structure tensor.

## 1 Introduction

The general inverse imaging problems seek the recovery of the underlying image  $\mathbf{f} \in \mathbb{R}^{NC}$  (assuming that each  $N$ -pixel channel is vectorized and all these  $C$  channels are stacked together into a single vector) from an observation model of the form  $\mathbf{g} = \mathbf{H}\mathbf{f} + \boldsymbol{\eta}$ , where  $\boldsymbol{\eta}$  is the noise and  $\mathbf{H}$  is a known linear operator. The type of the recovery problem is determined by the almost always ill-conditioned  $\mathbf{H}$ , which mainly corresponds to the impulse response of the imaging device. If  $\mathbf{H} = \mathbf{I}$ , then the problem is denoising. Besides, it may be a matrix that functions as a mask (i.e., inpainting), or a circulant matrix with all shifts of a blurring kernel (i.e., deblurring). It may also be a composite operator, i.e.,  $\mathbf{H} = \mathbf{S}\mathbf{F}$ , where  $\mathbf{F}$  blurs and  $\mathbf{S}$  down-samples (i.e., single-image super-resolution problem), or  $\mathbf{F}$  applies the Fourier transform and  $\mathbf{S}$  works as a mask that retains only a subset of the Fourier coefficients (i.e., reconstruction from sparse Fourier measurements).

The variational approach to such recovery tasks aims at minimizing an energy function of the form:

$$E(\mathbf{f}) = \frac{1}{2} \|\mathbf{g} - \mathbf{H}\mathbf{f}\|_2^2 + \tau R(\mathbf{f}) \quad (1)$$

where the former term is known as data fidelity and the latter one is the regularization term that encodes a prior. The parameter  $\tau$  is the regularization parameter that balances the contribution of each term to the cost. The regularization criteria has a critical role in model-driven recovery, thus its proper selection has been immensely studied. One old (but still functional with its extended versions) regularizer is total variation (TV) by [19], which favors piecewise-constant solutions. For years, TV has been extended in many aspects. Some researchers brought anisotropy to better handle angled boundaries [13, 17, 2, 15],

This work was supported by The Scientific and Technological Research Council of Turkey (TUBITAK) under 115R285.

E. Demircan-Tureyen  
Computer Engineering, Istanbul Kultur University, 34158  
Computer Engineering, Istanbul Technical University, 34467  
Istanbul, Turkey  
Tel.: +90-212-4984719  
E-mail: e.demircan@iku.edu.tr

M. E. Kamasak  
Computer Engineering, Istanbul Technical University, 34467 Istanbul, Turkey  
E-mail: kamasak@itu.edu.tr

some designed higher-order functionals to overcome staircase artifacts by favoring piecewise-smoothness [7, 5, 18], and some others employed nonlocal priors to benefit inherent nonlocal self similarity (NSS) as well [12, 14]. Structure tensor total variation (STV) [15] is one such TV variant, which redesigns TV in semi-local fashion. Rather than penalizing the gradient, it penalizes a summary of the gradients within a local neighborhood. This neighborhood-awareness provides a better representation of the local image variations, and significantly boosts the performance. Recently in [8, 9], we have shown that penalizing the summary of the directional gradients within a patch instead of the gradients increases the performance as long as the local directions are accurately estimated. In this paper, we design a nonlocal counterpart to our prior direction-guided regularizer, under the influence of the nonlocal STV (NLSTV [14]), so that the long-distance dependencies across the image can also be modelled. We also design an algorithm that employs anisotropic Gaussian kernels and nonlocal structure tensors to estimate the directional parameters that are needed by the proposed regularizer.

## 2 Background

In [14], the nonlocal counterpart of the STV was proposed. The extension has leveraged the STV term, which had already robustly encoded the local structural variation in semi-local fashion, by widening its scope to involve the nonlocal variations as well. This idea takes its source from the fact that a local patch often has similar nonlocal patches across the image. This NSS prior was first revealed in the form of nonlocal means (NLM) filtering [6], which approximates the underlying intensity value at a spatial coordinate  $i$  by taking the weighted average of the patches similar to the one around  $i$ th pixel. In [12] and [11], the idea behind NLM was adopted to variational approach to enrich TV based models. They treated images as a graphs, which allowed them to define nonlocal gradients and the interaction of two distant image points became possible. The nonlocal TV (NLTV) term involves the following nonlocal gradient operator:

$$(\nabla_w \mathbf{f})[i] = \sqrt{w(i, j)}(\mathbf{f}[j] - \mathbf{f}[i]) \quad (2)$$

$\forall j \in \{1, 2, \dots, N\}$ , where  $w(i, j)$  is a function that assigns weights between points by considering the pairwise similarities of the intensities and the relative distances. Therefore, the NLTV prior is defined as follows:

$$NLTV(\mathbf{f}) = \sum_{i=1}^N \|\nabla_w \mathbf{f}[i]\|_2 \quad (3)$$

The rationale behind NLTV is to penalize the weighted averages of the gradient magnitudes gathered from the similar patches under the assumption that the gradients of the similar patches are also similar. The NLTV regularization term is designed only for the scalar-valued images, i.e.,  $C = 1$ . If we turn back to NLSTV [14], pursuing the same idea with NLTV, it involves a nonlocal structure tensor operator primarily defined as

$$(S_w \mathbf{f})[i] = \sum_{j \in \mathcal{N}[i]} w(i, j) (J\mathbf{f})[j] (J\mathbf{f})[j]^T \quad (4)$$

where  $\mathcal{N}[i] = \{j : w(i, j) > 0\}$  and  $J$  is the Jacobian operator. It was later reformulated as  $(S_w \mathbf{f})[i] = (J_w \mathbf{f})[i] (J_w \mathbf{f})[i]^T$  to be able to decompose the NLSTV into linear functionals (See [15, 14]). Here, the operator  $J_w : \mathbb{R}^{NC} \mapsto \mathbb{R}^{N \times 2 \times NC}$  is nonlocal Jacobian of the form:

$$(J_w \mathbf{f})[i] = [(D_w \mathbf{f}^1)[i] \dots (D_w \mathbf{f}^C)[i]] \in \mathbb{R}^{2 \times NC} \quad (5)$$

where each superscript referring to a certain channel and  $D_w : \mathbb{R}^N \mapsto \mathbb{R}^{N \times 2 \times N}$  is nonlocal gradient with the definition:

$$(D_w \mathbf{f})[i]^c = [\sqrt{w(i, 1)}(\nabla \mathbf{f}^c)[1] \dots \sqrt{w(i, N)}(\nabla \mathbf{f}^c)[N]] \in \mathbb{R}^{2 \times N} \quad (6)$$

This definition differs from Eq. (2) in the sense that it can also exploit the underlying structure as opposed to the graph-based approach. Having defined the nonlocal Jacobian, the NLSTV is formulated as

$$NLSTV(\mathbf{f}) = \sum_{i=1}^N \|(J_w \mathbf{f})[i]\|_{\mathcal{S}_p} = \|J_w \mathbf{f}\|_{1,p} \quad (7)$$

where  $\mathcal{S}_p$  is Schatten norm of order  $p$ , i.e.,  $\|\mathbf{X}\|_{\mathcal{S}_p} = \|\sigma(\mathbf{X})\|_p$  for a generic matrix  $\mathbf{X}$  and  $\sigma(\cdot)$  returning the singular values [4].

## 3 Method

Our recovery algorithm involves two stages: (1) the stage that estimates the directional parameters required by the subsequent stage, (2) the stage that solves the inverse problem at hand by employing the proposed regularizer. The following subsection starts with the explanation of the proposed regularizer. Since we used the term adaptive direction-guided STV (ADSTV) in [9] to refer the extended version of DSTV [8], which gained the ability of handling multi-directional images, the regularization term that we propose here will be called as nonlocal ADSTV (NLADSTV).

### 3.1 NLADSTV Regularizer

From the NLSTV's point of view, we can also further extend our ADSTV to NLADSTV by pursuing the assumption that the directional gradients of the similar patches are also similar. In this respect, we define a *directional nonlocal gradient* that involves directional parameters estimated beforehand, i.e.,

$$(\tilde{D}_w^{(\alpha, \theta)} \mathbf{f}^c)[i] = [\sqrt{w(i, 1)}(\tilde{H}_{(\alpha, \theta)} \nabla \mathbf{f}^c)[1] \cdots \sqrt{w(i, N)}(\tilde{H}_{(\alpha, \theta)} \nabla \mathbf{f}^c)[N]] \in \mathbb{R}^{2 \times N} \quad (8)$$

where  $(\tilde{H}_{(\alpha, \theta)} \nabla \mathbf{f}^c)[i] = \tilde{A}_{(\alpha^+, \alpha^-[i])} \mathbf{R}_{-\theta[i]} (\nabla \mathbf{f}^c)[i]$  for the rotation and scale matrices, i.e.,

$$\mathbf{R}_\beta = \begin{bmatrix} \cos \beta & -\sin \beta \\ \sin \beta & \cos \beta \end{bmatrix} \quad \tilde{A}_{(\alpha^+, \alpha^-[i])} = \begin{bmatrix} \alpha^+ & 0 \\ 0 & \alpha^-[i] \end{bmatrix} \quad (9)$$

with  $\theta \in [0, \pi)^N$ ,  $\alpha^- \in [1, \alpha^+]^N$ , and  $\alpha^+ > 1$ . The estimation of these parameters is explained in Section III. B. Thus the *directional nonlocal Jacobian* takes the shape of:

$$(\tilde{J}_w^{(\alpha, \theta)} \mathbf{f})[i] = [(\tilde{D}_w^{(\alpha, \theta)} \mathbf{f}^1)[i] \cdots (\tilde{D}_w^{(\alpha, \theta)} \mathbf{f}^C)[i]] \in \mathbb{R}^{2 \times NC} \quad (10)$$

and our NLADSTV is nothing but the mixed  $\ell_1 - S_p$  norm of Eq.(10), i.e.,

$$NLADSTV(\mathbf{f}) = \sum_{i=1}^N \|(\tilde{J}_w^{(\alpha, \theta)} \mathbf{f})[i]\|_{S_p} = \|\tilde{J}_w^{(\alpha, \theta)} \mathbf{f}\|_{1, p} \quad (11)$$

Since our NLADSTV preserves NLSTV's convexity, the numerical optimization of NLADSTV based Eq. (1) is performed by applying the same steps of NLSTV model as described in [14]. It was solved by employing the augmented Lagrangian [3] convex optimization tool. Due to its straightforwardness and to avoid repetition, we skip the numerical optimization part in this paper by referring the reader to [14]. However, since the adjoint of the directional nonlocal Jacobian is needed by the optimization procedure, adopting Proposition 1 in [14], it is expressed as

$$(\tilde{J}_w^{(\alpha, \theta)*} \mathbf{X})[i] = [(\tilde{D}_w^{(\alpha, \theta)*} \mathbf{X}^1)[i] \cdots (\tilde{D}_w^{(\alpha, \theta)*} \mathbf{X}^C)[i]] \in \mathbb{R}^C \quad (12)$$

where  $\mathbf{X} \in \mathbb{R}^{N \times 2 \times NC}$  is an arbitrary matrix field and the adjoint NL gradient acts on  $\mathbf{X}^c \in \mathbb{R}^{N \times 2 \times N}$  as follows:

$$(\tilde{D}_w^{(\alpha, \theta)*} \mathbf{X}^c)[i] = -\text{div} \sum_{j=1}^N \left( \tilde{H}_{(\alpha^-[i], \theta[i])}^T \sqrt{w(j, i)} \mathbf{X}^c[j, :, i] \right) \quad (13)$$

where  $\mathbf{X}^c[j, :, i] \in \mathbb{R}^2$  refers to the  $i$ th column of the  $j$ th matrix of the matrix field  $\mathbf{X}^c$ ,  $\text{div}$  is the discrete divergence operator, and  $\tilde{H}_{(\alpha^-[i], \theta[i])}^T = R_{\theta[i]} \tilde{A}_{(\alpha^+, \alpha^-[i])}$ .

When it comes to the nonlocal weights, similar to the NLM, the following function is used:

$$w(i, j) = e^{-d_\rho(i, j)/\beta^2} \quad (14)$$

where  $d_\rho(\cdot)$  is the distance function of patches and  $\beta$  is filtering parameter. Note that, in practical realization of NLSTV, the target space of the nonlocal gradient was shrunk to  $\mathbb{R}^{N \times 2 \times L}$ , where  $L \ll N$ , by using a sparse version of weighting strategy. This strategy seeks the similar patches inside a search window rather than the entire image. Next, it takes only  $L$  most similar pair of patches into consideration [14]. The distance function that serves this purpose is given below:

$$d_\rho(i, j) = \sum_{l=-s/2}^{s/2} G_\rho[l] |\mathbf{f}[i-l] - \mathbf{f}[i+d(i, j)-l]|^2 \quad (15)$$

where  $G_\rho$  is a symmetric weighting kernel of size  $s \times s$  and  $-r/2 \leq d(i, j) \leq r/2$  returns the relative distance between the pixels  $i$  and  $j$ . The same realization of the weighting function is used by our NLADSTV regularizer.

### 3.2 Parameter Estimation

Our regularization term requires three additional parameters at each spatial image point:  $\theta[i]$ ,  $\alpha^-[i]$ , and a constant  $\alpha^+$ . The parameter  $\theta$  corresponds to the orientation map of the image, while  $\alpha^-$  and  $\alpha^+$  together determines the anisotropic behaviour

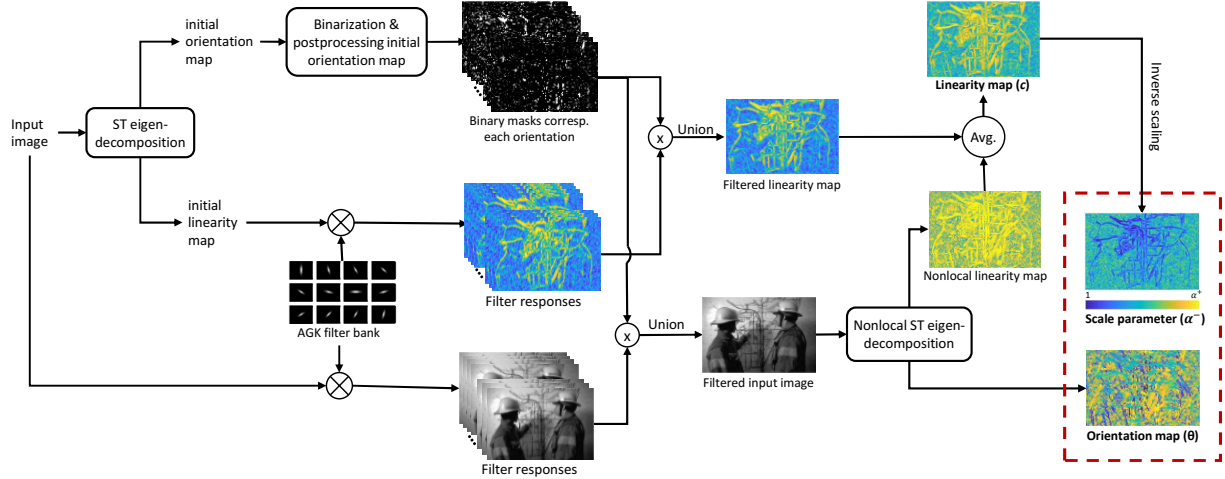


Figure 1: The flowchart of our DPE procedure for the estimation of directional parameters.

of our regularizer. If at a point  $i$ ,  $\alpha^- [i]$  is equal to  $\alpha^+$ , then the regularizer behaves like NLSTV. Otherwise, as  $\alpha^- [i]$  gets closer to one, it gains more anisotropic behaviour, in other words; the sensitivity to the changes towards  $\theta [i]$  increases.

In [9], we suggested an algorithm to estimate these parameters. To put it simply, it was applying successive eigen-decomposition on the structure tensor (ST) of the observed image (or the luminance information of the observed image for vector-valued images), and TV based smoothing on the parameter fields obtained at multiple scales. In this paper, in order to boost the parameter estimation performance, we propose a different procedure that employs anisotropic Gaussian kernels and nonlocal structure tensor. This subsection elaborates our directional parameter estimation procedure (hereinafter referred to as DPE) by dividing it into steps (See Fig. 1 for the illustration). Note that, the term “input image” refers to the luminance information of the observed image, unless otherwise stated. We denote this image as  $\tilde{\mathbf{g}}$ .

### 3.2.1 Structure Tensor Eigendecomposition

We start by obtaining two initial maps through the eigendecomposition of the structure tensor (ST) of the input image. The initial orientation map  $\theta^0$  is a map that is composed of the entries  $\theta^0 [i] \angle \mathbf{v}^- [i]$ , where  $\mathbf{v}^- [i]$  is the eigenvector associated with the smallest eigenvalue  $\lambda^- [i]$  at an image point  $i$ . The initial linearity map  $\mathbf{c}^0$ , on the other side, is made of the entries computed as the difference of eigenvalues normalized by the largest eigenvalue  $\lambda^+ [i]$ , i.e.,  $\mathbf{c}^0 [i] = (\lambda^+ [i] - \lambda^- [i]) / \lambda^+ [i]$ . This non-negative ratio serves as a measure of local linearity.

### 3.2.2 Postprocessing Initial Orientation Map

The initial orientation map is converted into binary masks each of which is filtering out the orientations except a certain (main) orientation and the orientations falling within the interval  $[-\Delta\theta, +\Delta\theta]$  around it. In our work, the main orientations are sampled at 15 degree intervals between  $[0, 180)$ , thus  $\Delta\theta = 7.5$  degree.

Each orientation mask is further processed by using morphological closings in order to fill in the cracks and the holes, so that these filtered-out pixels are also classified by the corresponding orientation. The cracks along the corresponding orientation are closed by using oriented line segments of 5-pixel length as structuring element (SE). The orientation of the line segment matches with the main orientation of the mask. The holes, on the other hand, are closed by using 3-by-3 square SE.

### 3.2.3 Filtering with Anisotropic Gaussian Kernels

Anisotropic Gaussian kernels (AGK) and their derivatives have usually been applied to the edge [23, 20], line [21], and corner [22] detection tasks. These kernels are nothing but the elongated versions of the isotropic counterpart and defined as

$$g(\mathbf{g}_L; \sigma, \psi, \beta) = \frac{1}{2\pi\sigma^2} \exp \left( -\frac{1}{2\sigma^2} \mathbf{g}_L^T \mathbf{R}_\beta \begin{bmatrix} \psi^2 & 0 \\ 0 & \psi^{-2} \end{bmatrix} \mathbf{R}_\beta^T \mathbf{g}_L \right) \quad (16)$$

where  $\mathbf{R}_\beta$  is the rotation matrix given in Eq. (9),  $\sigma$  represents the scale,  $\psi > 1$  denotes the factor of anisotropy ( $\psi = 1$  for isotropic kernel), and  $\beta$  gives the direction of the kernel.

We convolve both the input image  $\mathbf{g}_L$  and the initial linearity map  $\mathbf{c}^0$  with a filter bank of discrete AGKs covering the main orientations. As mentioned above, the main orientations are sampled at 15 degree intervals between  $[0, 180)$ . We also fixed the scale parameter as  $\sigma = 0.75$  and the anisotropy factor as  $\psi = 4$ . Among all orientations, we select the ones coinciding with the angles in the postprocessed version of the initial orientation map by performing element-wise multiplication of each filter response with the corresponding mask and by taking the union of those multiplications. Let us call the resulting filtered versions of  $\tilde{\mathbf{g}}$  and  $\mathbf{c}^0$  as filtered input image  $\tilde{\mathbf{g}}_{agk}$  and filtered linearity map  $\mathbf{c}_{agk}$ , respectively.

### 3.2.4 Nonlocal Structure Tensor Eigendecomposition

This step is very similar to the first step, except this time the structure tensor is nonlocal and computed on the filtered input image  $\tilde{\mathbf{g}}_{agk}$ , rather than  $\tilde{\mathbf{g}}$ . The rationale behind preferring not to apply the nonlocal ST eigendecomposition in the first step is due to its less robustness to the noise or artifacts, when compared to its local counterpart. The orientation map  $\theta$  with the entries  $\theta[i] \angle \mathbf{v}_{nl}[i]$ , where the subscript  $nl$  indicates that the eigenvector is derived from the nonlocal ST, is the final orientation map now, which will be fed into the restoration process as it is. On the other side, the entries of the nonlocal linearity map is obtained as  $\mathbf{c}_{nl}[i] = (\lambda_{nl}^+[i] - \lambda_{nl}^-[i]) / \lambda_{nl}^+[i]$  and averaged with the filtered linearity map to reach the final linearity map, i.e.,  $\mathbf{c} = (\mathbf{c}_{agk} + \mathbf{c}_{nl}) / 2$ .

### 3.2.5 The Final Directional Parameters

The orientation map  $\theta$  to be used by the NLADSTV regularizer is the direct output of nonlocal ST eigendecomposition as stated in the previous step. The parameter  $\alpha^-$ , on the other hand, is computed by inversely scaling the entries of  $\mathbf{c} \in \mathbb{R}^N$  onto the range  $[1, \alpha^+]$ , i.e.,

$$\alpha^-[i] = \frac{\alpha^+ - 1}{\max(\mathbf{c}) - \min(\mathbf{c})} (\max(\mathbf{c}) - \mathbf{c}[i]) + 1 \quad (17)$$

In Fig. 1 dashed red box shows the final directional parameters.

## 3.3 Overall Algorithm

Our recovery framework starts by estimating the directional parameters as described in Subsection B. Next, the NLADSTV regularized inverse problem (See Eq. (18)) is solved by following the optimization scheme in [14]. The overall algorithm is provided in Algorithm 1. This algorithm actually repeats the steps in the Algorithm 1 in [14], except the presence of the DPE procedure (called at third line) and the nonlocal Jacobian operator switched to our directional nonlocal Jacobian.

As mentioned earlier, we don't go into the details of the numerical optimization in this paper, however the Algorithm 1 needs some explanations. It solves the following NLADSTV regularized inverse problem:

$$\hat{\mathbf{f}} = \underset{\mathbf{f}}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{g} - \mathbf{H}\mathbf{f}\|_2^2 + \tau \|\tilde{\mathcal{J}}_w^{(\alpha, \theta)} \mathbf{f}\|_{1,p} + \iota_{\mathcal{C}}(\mathbf{f}) \quad (18)$$

where  $\iota_{\mathcal{C}}$  is the indicator function of a convex set  $\mathcal{C}$  that takes the value 0 if  $\mathbf{f} \in \mathcal{C}$  and  $\infty$  otherwise. The set  $\mathcal{C}$  corresponds to an additional constraint, such as nonnegativity, and if there is no constraint, then  $\mathcal{C} = \mathbb{R}^{NC}$ . The authors of [14] reformulates this problem in constrained form by using auxiliary variables and seeks a solution through augmented Lagrangian methods [3]. They employ alternating-direction method of multipliers (ADMM) [10].

In the line 5, the prox operator at the RHS corresponds to the proximal map of the mixed  $\ell_1 - S_p$  norm weighted by  $\tau/\mu$ . For a function  $h$  evaluated at  $\mathbf{z}$ , it is defined as:

$$\operatorname{prox}_h(\mathbf{z}) = \underset{\mathbf{f}}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{f} - \mathbf{z}\|^2 + h(\mathbf{f}) \quad (19)$$

While deriving a minimization approach in [14] for the NLSTV regularized inverse problem, Eq. (19) is reduced to the proximal map of  $S_p$  norm by inputting the entries  $\Omega[i] \in \mathbb{R}^{2 \times LC}$  of  $\Omega = \tilde{\mathcal{J}}_w^{(\alpha, \theta)} \mathbf{f}^{(t)} + \mathbf{s}_1^t$  separately. Next, through the SVD decomposition of  $\Omega[i]$ , the problem is reduced to the proximal map of  $\ell_p$  norm with an input consisting of the singular values of  $\Omega[i]$ . For the efficient evaluation of  $\operatorname{prox}_p$ , they utilize iterative proximal algorithm proposed in [16]. (See Section IV in [14] for the details).

In the line 6,  $P_{\mathcal{C}}$  is nothing but the projection of  $\mathbf{f}^{(t)} + \mathbf{s}_2^{(t)}$  onto the convex set  $\mathcal{C}$ . It is defined as  $P_{\mathcal{C}} = \min(\max(0, \mathbf{f}), \gamma)$  for the set  $\mathcal{C} = \{\mathbf{f} \in \mathbb{R}^{NC} : 0 \leq \mathbf{f}[i] \leq \gamma, \forall i = 1, \dots, NC\}$  that we consider [14].

The rest of the Algorithm 1 keeps realizing the steps of ADMM iterations and updating the Lagrange multipliers (The reader is referred to [14] for details).

**Algorithm 1** Algorithm for NLADSTV-based recovery (adopted from [14])

---

```

1: INPUT:  $\mathbf{g}, \mathbf{H}, \tau > 0, \mu > 0, p \geq 1, \alpha^+ > 1$ 
2: INIT:  $\mathbf{f}^{(0)} = \mathbf{g}, \mathbf{s}_1^{(0)} = \mathbf{0}, \mathbf{s}_2^{(0)} = \mathbf{0}, t = 0,$ 
3:  $(\theta, \alpha^-) \leftarrow \text{DPE}(\mathbf{g}, \alpha^+)$ 
4: while stopping criterion is not satisfied do
5:    $\mathbf{z}_1^{(t+1)} \leftarrow \text{prox}_{\frac{\tau}{\mu} \|\cdot\|_{1,p}}(\tilde{\mathcal{J}}_w^{(\alpha,\theta)} \mathbf{f}^{(t)} + \mathbf{s}_1^t)$ 
6:    $\mathbf{z}_2^{(t+1)} \leftarrow P_{\mathcal{C}}(\mathbf{f}^{(t)} + \mathbf{s}_2^{(t)})$ 
7:    $\mathbf{B} \leftarrow (\frac{1}{\mu} \mathbf{H}^T \mathbf{H} + \tilde{\mathcal{J}}_w^{(\alpha,\theta)} \tilde{\mathcal{J}}_w^{(\alpha,\theta)} + \mathbf{I})$ 
8:    $\omega_1^{(t+1)} \leftarrow \mathbf{z}_1^{(t+1)} - \mathbf{s}_1^{(t)}$ 
9:    $\omega_2^{(t+1)} \leftarrow \mathbf{z}_2^{(t+1)} - \mathbf{s}_2^{(t)}$ 
10:   $\mathbf{f}^{(t+1)} \leftarrow \mathbf{B}^{-1}(\frac{1}{\mu} \mathbf{H}^T \mathbf{g} + \tilde{\mathcal{J}}_w^{(\alpha,\theta)} \omega_1^{(t+1)} + \omega_2^{(t+1)})$ 
11:   $\mathbf{s}_1^{(t+1)} \leftarrow \mathbf{s}_1^{(t)} + \tilde{\mathcal{J}}_w^{(\alpha,\theta)} \mathbf{f}^{(t+1)} - \mathbf{z}_1^{(t+1)}$ 
12:   $\mathbf{s}_2^{(t+1)} \leftarrow \mathbf{s}_2^{(t)} + \mathbf{f}^{(t+1)} - \mathbf{z}_2^{(t+1)}$ 
13:   $t \leftarrow t + 1$ 
14: end while

```

---



Figure 2: Thumbnails of the grayscale and color images used in the experiments. From left to right and top to bottom: *Chateau, Dog, Mural, Workers, Buildings, Deer, Arch, Man, Monarch, and Structure*.

## 4 Experimental Results

We consider the problems of image denoising and image deblurring by confronting four related methods: STV, ADSTV, NLSTV, and NLADSTV. The experiments are conducted on the vector-valued images shown in Fig. 2. The images are taken from the popular Berkeley Segmentation Dataset (BSD) [1]. For the quantitative evaluation, we use the peak signal-to-noise ratio (PSNR). The source codes of STV<sup>1</sup> and NLSTV<sup>2</sup> that we use were made publicly available by the authors on GitHub. Our NLADSTV is implemented on top of NLSTV.

Since the algorithms under consideration are all aiming to minimize Eq. (1), they need regularization parameter  $\tau$  to be tuned. We fine-tuned  $\tau$  such that it leads to the best PSNR in all experiments. For both STV and ADSTV, the Gaussian kernel of support  $3 \times 3$  pixel with the standard deviation  $\sigma = 0.5$  is used. For NLSTV and NLADSTV, on the other hand, the sizes of the neighborhood (used for the nonlocal weights) and the search window are set to  $7 \times 7$  and  $11 \times 11$ , respectively. Also, the number of the similar patches  $L = 9$ , as suggested in [14]. For the sake of fairness, we fixed the free parameter  $\alpha^+$  of ADSTV and NLADSTV as 4, but it is possible to increase the restoration quality of these methods by fine-tuning  $\alpha^+$ . When it comes to the setting of the DPE procedure, the Gaussian kernel used for the computation of the local structure tensor has  $9 \times 9$  pixel support and the standard deviation of 1.5, while the size of the neighborhood is  $5 \times 5$ , the size of the search window is  $11 \times 11$ , and  $L = 13$ . In addition to them, note that the order of the Schatten norm  $p = 1$  for all regularizers under consideration.

### 4.1 Image Denoising

We consider additive i.i.d. Gaussian noise with four noise levels  $\sigma_\eta = \{0.05, 0.1, 0.15, 0.2\}$ . Table 1 reports the PSNR scores obtained by using four competing methods. By inspecting the results, one observes that the local regularizers perform worse than their nonlocal counterparts, except for the *Deer* image, which may not be exhibiting NSS property. STV is the least performing regularizer for all four noise levels. As was shown in [8], ADSTV systematically outperforms STV with the improvement around 0.4 dBs. Another observation is, as the noise level increases, ADSTV also produces better results than

<sup>1</sup> [https://github.com/cig-skoltech/Structure\\_Tensor\\_Total\\_Variation](https://github.com/cig-skoltech/Structure_Tensor_Total_Variation)

<sup>2</sup> <https://github.com/cig-skoltech/NLSTV>

Table 1: PSNR Comparisons on Image Denoising

$\sigma_\eta$ :	0.05				0.1				0.15				0.2			
Method:	STV	ADSTV	NLSTV	NLADSTV	STV	ADSTV	NLSTV	NLADSTV	STV	ADSTV	NLSTV	NLADSTV	STV	ADSTV	NLSTV	NLADSTV
Chateau	32.38	32.73	32.92	<b>33.15</b>	28.79	29.33	29.41	<b>29.71</b>	26.93	27.60	27.56	<b>27.93</b>	25.74	26.37	26.29	<b>26.71</b>
Dog	32.19	32.56	32.25	<b>32.61</b>	29.00	29.38	29.09	<b>29.44</b>	27.44	27.82	27.59	<b>27.87</b>	26.46	26.78	26.63	<b>26.85</b>
Mural	30.16	30.41	30.20	<b>30.46</b>	26.59	26.95	26.74	<b>27.14</b>	24.78	25.19	25.03	<b>25.47</b>	23.61	24.00	23.88	<b>24.35</b>
Workers	31.90	32.38	32.40	<b>32.87</b>	28.04	28.69	28.66	<b>29.35</b>	25.94	26.65	26.60	<b>27.35</b>	24.55	25.20	25.13	<b>25.92</b>
Buildings	31.24	31.82	31.35	<b>32.13</b>	27.44	28.15	27.72	<b>28.55</b>	25.47	26.25	25.90	<b>26.74</b>	24.20	24.93	24.64	<b>25.50</b>
Deer	31.49	<b>31.76</b>	31.31	31.69	28.30	<b>28.65</b>	28.22	<b>28.54</b>	26.69	<b>27.04</b>	26.67	26.90	25.64	<b>25.97</b>	25.66	25.80
Arch	30.86	31.03	30.92	<b>31.07</b>	27.63	27.90	27.76	<b>28.14</b>	26.08	26.38	26.27	<b>26.61</b>	25.15	25.34	25.36	<b>25.63</b>
Man	33.42	33.64	33.65	<b>33.97</b>	29.54	29.85	29.88	<b>30.31</b>	27.38	27.71	27.71	<b>28.24</b>	25.90	26.19	26.16	<b>26.75</b>
Monarch	35.16	35.52	35.42	<b>35.73</b>	31.68	32.09	32.03	<b>32.37</b>	29.73	30.17	30.21	<b>30.51</b>	28.39	28.78	28.88	<b>29.23</b>
Structure	30.00	30.31	30.32	<b>30.65</b>	26.27	26.68	26.58	<b>27.05</b>	24.51	24.92	24.82	<b>25.29</b>	23.45	23.82	23.73	<b>24.18</b>
<b>Avg.</b>	<b>31.88</b>	<b>32.22</b>	<b>32.07</b>	<b>32.43</b>	<b>28.33</b>	<b>28.77</b>	<b>28.61</b>	<b>29.06</b>	<b>26.50</b>	<b>26.97</b>	<b>26.84</b>	<b>27.29</b>	<b>25.31</b>	<b>25.74</b>	<b>25.64</b>	<b>26.09</b>

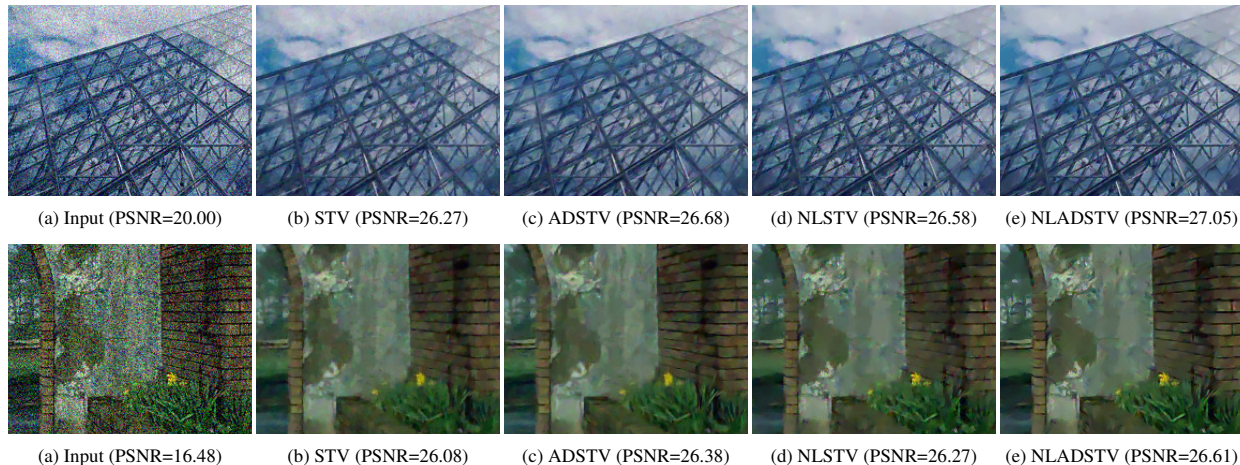


Figure 3: The detail patches cropped from restored versions of the noisy *Structure* and *Arch* images. *Structure* image is degraded the noise level of  $\sigma_\eta = 0.1$ . *Arch* image is degraded by the noise level of  $\sigma_\eta = 0.15$ .

NLSTV. Our NLADSTV, on the other hand, is the best performing regularizer with 0.4 dBs improvement over NLSTV and 0.3 dBs over its local counterpart ADSTV, on average.

Fig. 3 is also provided for the visual judgement. We demonstrate two representative detail patches. As we observe from the patch taken from the *Structure* image, the straight lines seem smoother in the NLADSTV reconstruction. The oil-painting-like artifacts present in the STV and NLSTV are far less visible in ADSTV and NLADSTV. When compared to the ADSTV, the NLADSTV is better at removing noise while preserving details, thanks to its nonlocality. In the *Arch* image, the details of the plant and the bricks around the arch are more distinguishably reconstructed by our NLADSTV than the other methods.

## 4.2 Image Deblurring

In deblurring problem, instead of feeding the direct observation (blurred image) to the DPE procedure, we input the deblurred version of the image by a Wiener filter, since otherwise the blur effect causes the poor localization of edges and the loss of some small gaps between parallel lines. By means of Wiener filter, our DPE procedure can better estimate the directional parameters.

Table 2 reports the deblurring performances of the methods under consideration for two types of blurring kernels. We consider a Gaussian kernel of support  $9 \times 9$  pixel with  $\sigma_b = 6$  and a motion kernel of support 19 pixel. The blurred images are also degraded with Gaussian noise corresponding to blurred SNR (BSNR = variance of  $\mathbf{Hf}$  divided by  $\sigma_\eta^2$ ) of 20 and 30 dBs. While the deblurring results for motion blur are on par with the denoising results, the improvement that we obtain with NLADSTV is less notable (around 0.15 dBs over both ADSTV and NLSTV on average) when it comes to the Gaussian blur.

We demonstrate two exemplary results in Fig. 4. As one can observe from the detail patch of the *Buildings* image, the artifacts between the closely parallel lines are far less visible in the NLADSTV regularized solution. In the *Workers* image again, the the cloudy look around the lines is more apparent in the other methods than in our NLADSTV.

Table 2: PSNR Comparisons on Image Deblurring

BSNR:	Gaussian PSF								Motion PSF							
	20 dB				30 dB				20 dB				30 dB			
	STV	ADSTV	NLSTV	NLADSTV	STV	ADSTV	NLSTV	NLADSTV	STV	ADSTV	NLSTV	NLADSTV	STV	ADSTV	NLSTV	NLADSTV
Chateau	25.62	25.92	25.83	<b>26.04</b>	27.85	28.10	27.77	<b>28.27</b>	27.73	28.14	28.15	<b>28.43</b>	31.81	32.28	32.44	<b>32.78</b>
Dog	27.18	<b>27.25</b>	27.19	<b>27.25</b>	28.70	28.84	28.76	<b>28.90</b>	29.54	29.76	29.59	<b>29.85</b>	33.81	34.14	33.88	<b>34.24</b>
Mural	23.91	24.13	24.08	<b>24.20</b>	26.12	26.25	26.34	<b>26.42</b>	26.30	26.71	26.56	<b>26.93</b>	31.17	31.51	31.34	<b>31.66</b>
Workers	23.80	24.09	24.04	<b>24.37</b>	26.26	26.52	26.62	<b>26.89</b>	26.38	26.99	26.96	<b>27.74</b>	31.42	31.99	32.16	<b>32.78</b>
Buildings	24.02	24.21	24.37	<b>24.58</b>	25.98	26.25	26.24	<b>26.51</b>	27.22	27.71	27.50	<b>28.08</b>	31.81	32.42	31.89	<b>32.65</b>
Deer	27.39	<b>27.54</b>	27.30	27.43	29.28	<b>29.44</b>	29.28	29.39	29.25	<b>29.54</b>	29.20	29.51	33.90	<b>34.20</b>	33.81	34.19
Arch	25.93	26.00	26.00	<b>26.11</b>	27.58	27.58	27.63	<b>27.68</b>	27.64	27.84	27.73	<b>28.06</b>	31.88	32.07	31.96	<b>32.27</b>
Man	25.38	<b>25.45</b>	25.40	<b>25.45</b>	28.15	28.28	28.35	<b>28.44</b>	27.34	27.63	27.74	<b>28.09</b>	32.40	32.72	32.86	<b>33.23</b>
Monarch	30.15	30.15	<b>30.39</b>	30.31	32.54	32.63	<b>32.97</b>	<b>32.97</b>	33.20	33.64	33.70	<b>34.00</b>	38.40	38.76	38.88	<b>39.20</b>
Structure	22.99	23.10	23.02	<b>23.14</b>	24.22	24.35	24.30	<b>24.46</b>	25.32	25.64	25.44	<b>25.90</b>	29.79	30.13	30.06	<b>30.43</b>
Avg.	25.64	25.79	25.76	<b>25.89</b>	27.67	27.82	27.83	<b>27.99</b>	27.99	28.36	28.26	<b>28.66</b>	32.64	33.02	32.93	<b>33.34</b>

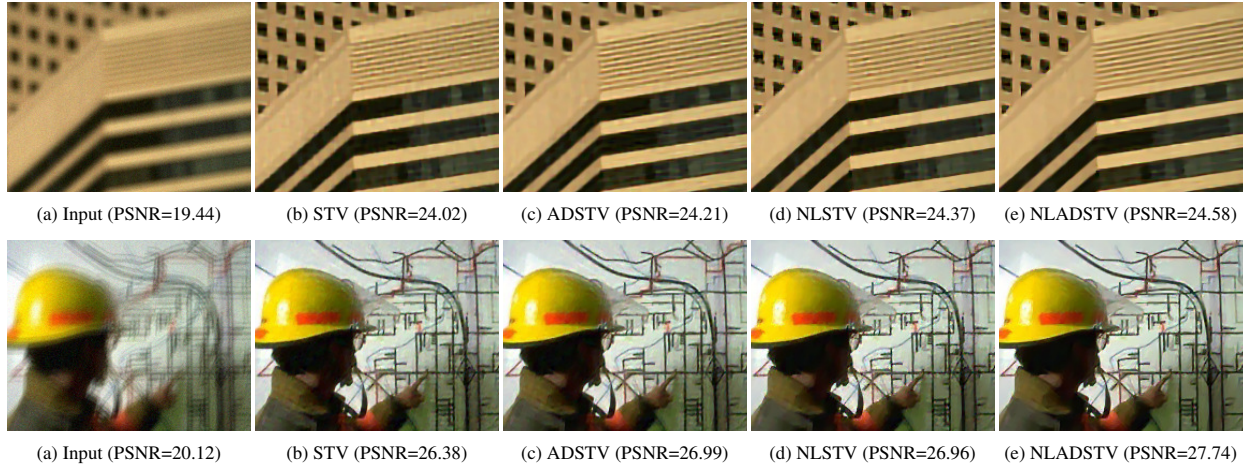


Figure 4: The detail patches cropped from restored versions of the blurred and noisy *Buildings* and *Workers* images. *Buildings* image is degraded by Gaussian blur and noise level of BSNR = 20 dBs. *Workers* image is degraded by motion blur and BSNR = 20 dBs.

## 5 Discussion

According to the experiments conducted, our approach seems to produce promising results. However, the effectiveness of the NLADSTV based regularization highly depends on the performance of the directional parameter estimation (DPE) procedure. In the case of Gaussian blur for instance, since our DPE procedure was able to estimate less reliable parameters, the PSNR gain was not as satisfying as it was in the other experiments. Fortunately, because the structure tensor is a positive semi-definite matrix, the linearity factor  $\mathbf{c}[i]$  also behaves like a coherence factor, which measures the reliability of the estimated orientations. Thus for the unreliable orientation estimations,  $\mathbf{c}[i]$  becomes small and the NLADSTV starts working like the NLSTV. Therefore, the restoration quality never falls below the one that we obtain by using NLSTV.

The NLADSTV comes up with additional computational demands. The DPE procedure takes around 10 sec. on average, on a computer equipped with Intel Core Processor i7-7500U (2.70-GHz) with 16 GB of memory. Besides, the convergence of the NLADSTV based problem requires more iterations than the NLSTV, which makes the entire framework run 2 times slower than the NLSTV based image recovery.

## 6 Conclusion

This paper proposes a competitive framework for variational image recovery. We establish a two-stage algorithm to advance the nonlocal structure tensor total variation (NLSTV) regularizer, which already exploits local structural regularity and non-local self-similarity properties. We further enrich these priors by encoding some spatially varying directional characteristics of the underlying image. In order to understand these characteristics, we design an algorithm that estimates the directional parameters required by our regularizer. The experiments prove that the proposed method achieves significant performance improvements over its local counterpart and the NLSTV.

## References

- [1] Arbelaez P, Maire M, Fowlkes C, Malik J (2010) Contour detection and hierarchical image segmentation. *IEEE transactions on pattern analysis and machine intelligence* 33(5):898–916
- [2] Bayram I, Kamasak ME (2012) Directional total variation. *IEEE Signal Processing Letters* 19(12):781–784
- [3] Bertsekas DP (2014) *Constrained optimization and Lagrange multiplier methods*. Academic press
- [4] Bhatia R (2013) *Matrix analysis*, vol 169. Springer Science & Business Media
- [5] Bredies K, Kunisch K, Pock T (2010) Total generalized variation. *SIAM Journal on Imaging Sciences* 3(3):492–526
- [6] Buades A, Coll B, Morel JM (2011) Non-local means denoising. *Image Processing On Line* 1:208–212
- [7] Chan T, Marquina A, Mulet P (2000) High-order total variation-based image restoration. *SIAM Journal on Scientific Computing* 22(2):503–516
- [8] Demircan-Tureyen E, Kamasak ME (2019) On the direction guidance in structure tensor total variation based denoising. In: *Iberian Conference on Pattern Recognition and Image Analysis*, Springer, pp 89–100
- [9] Demircan-Tureyen E, Kamasak ME (2020) Adaptive direction-guided structure tensor total variation. *arXiv preprint arXiv:200105717*
- [10] Eckstein J, Bertsekas DP (1992) On the douglas-rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming* 55(1-3):293–318
- [11] Elmoataz A, Lezoray O, Boughleux S (2008) Nonlocal discrete regularization on weighted graphs: a framework for image and manifold processing. *IEEE transactions on Image Processing* 17(7):1047–1060
- [12] Gilboa G, Osher S (2008) Nonlocal operators with applications to image processing. *Multiscale Modeling & Simulation* 7(3):1005–1028
- [13] Grasmair M, Lenzen F (2010) Anisotropic total variation filtering. *Applied Mathematics & Optimization* 62(3):323–339
- [14] Lefkimmiatis S, Osher S (2015) Nonlocal structure tensor functionals for image regularization. *IEEE Transactions on Computational Imaging* 1(1):16–29
- [15] Lefkimmiatis S, Roussos A, Maragos P, Unser M (2015) Structure tensor total variation. *SIAM Journal on Imaging Sciences* 8(2):1090–1122
- [16] Liu J, Ye J (2010) Efficient  $l_1/l_q$  norm regularization. *arXiv preprint arXiv:10094766*
- [17] Lou Y, Zeng T, Osher S, Xin J (2015) A weighted difference of anisotropic and isotropic total variation model for image processing. *SIAM Journal on Imaging Sciences* 8(3):1798–1823
- [18] Papafitsoros K, Schönlieb CB (2014) A combined first and second order variational approach for image reconstruction. *Journal of mathematical imaging and vision* 48(2):308–338
- [19] Rudin LI, Osher S, Fatemi E (1992) Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena* 60(1-4):259–268
- [20] Wang G, Lopez-Molina C, De Baets B (2019) Multiscale edge detection using first-order derivative of anisotropic gaussian kernels. *Journal of Mathematical Imaging and Vision* 61(8):1096–1111
- [21] Wang G, Lopez-Molina C, de Ulzurrun GVD, De Baets B (2019) Noise-robust line detection using normalized and adaptive second-order anisotropic gaussian kernels. *Signal Processing* 160:252–262
- [22] Zhang W, Sun C (2020) Corner detection using multi-directional structure tensor with multiple scales. *International Journal of Computer Vision* 128(2):438–459
- [23] Zhang W, Zhao Y, Breckon TP, Chen L (2017) Noise robust image edge detection based upon the automatic anisotropic gaussian kernels. *Pattern Recognition* 63:193–205