



On the Direction Guidance in Structure Tensor Total Variation Based Denoising

Ezgi Demircan-Tureyen^{1,2}  and Mustafa E. Kamasak² 

¹ Faculty of Engineering, Department of Computer Engineering,
Istanbul Kultur University, 34156 Istanbul, Turkey
e.demircan@iku.edu.tr

² Faculty of Computer and Informatics, Department of Computer Engineering,
Istanbul Technical University, 34390 Istanbul, Turkey
{edemircan,kamasak}@itu.edu.tr

Abstract. This paper introduces a new analysis-based regularizer, which incorporates the neighborhood-awareness of the structure tensor total variation (STV) and the tunability of the directional total variation (DTV), in favor of a pre-selected direction with a pre-selected dose of penalization. In order to show the utility of the proposed regularizer, we consider the problem of denoising uni-directional images. Since the regularizer is convex, we develop a simple optimization algorithm by realizing its proximal map. The quantitative and the visual experiments demonstrate the superiority of our regularizer over DTV (only for scalar-valued images) and STV.

Keywords: Image denoising · Total variation · Structure tensor · Directional total variation · Convex optimization · Inverse problems

1 Introduction

From capturing to encoding and transmission, different stages of an image acquisition pipeline produce different types of noise (additive vs. multiplicative, structured vs. random, signal dependent vs. independent, etc.). The noise existing on an image not only creates visual disruption, but also obscures significant details. Either to meet the quality expectations or to extract valuable information from the image, denoising has become an indispensable part of the processing. And yet, various denoising techniques have been proposed so far. These techniques most particularly aimed at recovering the latent signal $f \in \mathbb{R}^{NC}$ from the observation $g \in \mathbb{R}^{NC}$, by considering the forward corruption model of the form: $g = f + \eta$, where $\eta \in \mathbb{R}^{NC}$ is the additive noise that is assumed to be i.i.d. and normally distributed ($\eta \sim \mathcal{N}(\underline{0}, \sigma_\eta^2 I)$). Note that, the images are assumed to be stacked into vectors, where N and C are respectively denoting the number of pixels in a single channel, and the number of channels. For the scalar-valued images $C = 1$.

This work is supported by Istanbul Kultur University under ULEP-2018-2019/49.

The variational methods inverse the given forward model by minimizing an energy functional of the form:

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

where $R(f)$ stands for the regularizer, whose role is encoding a prior on the unknown image. The prior may reflect some real characteristics, or may just presume them to reduce the size of search space. $\tau \geq 0$ is a variable used to tune the contribution of $R(f)$ to $E(f)$, and for the small values of it, the fidelity to the observed data dominates the prior. The choice of $R(f)$ is crucial. Over the years, intensive research has been pursued on designing good regularizers, which can adequately characterize the latent image in an efficiently solvable way. The concept of compressive sensing (CS) proves that the recovery of a deficient signal is possible, as long as it exhibits sparsity in a certain domain. This sparsity can be encoded by a regularizer. The total variation (TV) is such a regularizer, where the image's gradient is assumed to be sparse (yielding a piecewise constant – PWC image) as a special case of CS. It is a convex functional, thus ensures a feasible solution. TV has given rise to a diverse range of TV inspired regularizers aiming to cope with TV's limitations (especially with the staircase artifacts caused due to the PWC assumption), while preserving its convexity. One of the most popular methods is the total generalized variation (TGV) [4], which favors piecewise smoothness rather than PWC, by encoding not only first-, but also higher-order information. Another such TV variant is structure tensor total variation (STV) [8], which forms the basis for this paper together with another variant: directional total variation (DTV) [1] designed for uni-directional images. STV penalizes the non-linear combinations of the first-order derivatives within a neighborhood (i.e. eigenvalues of the structure tensor), while DTV penalizes weighted and rotated gradients of each pixel. STV applies semi-local regularization, while DTV does local. Moreover, STV can intrinsically decide and adaptively apply the dose of penalization, while DTV requires priorly determined parameters for this purpose.

STV and DTV are both analysis-based convex regularizers. Here the term *analysis-based* refers to the regularizers of the form (in discrete setting):

$$R(f) = \sum_{i=1}^N \phi((\Gamma f)[i]) \quad (2)$$

which directly run on the image, rather than applying the regularization in a transform domain (i.e. *synthesis-based* regularizers such as wavelet, curvelet and shearlet.) In Eq. (2), $\Gamma : \mathbb{R}^N \rightarrow \mathbb{R}^{N \times D}$ is regularization operator with $(\Gamma f)[i]$ denoting the D -dimensional i -th element of Γf mapping, and $\phi : \mathbb{R}^M \rightarrow \mathbb{R}$ is a potential function. For TV regularization, $\Gamma = \nabla$ that denotes the gradient operator, and $\phi = \|\cdot\|_p$, where $p = 2$ for the native isotropic TV [10]. There are also anisotropic alternatives, designed to catch convex shapes with angled boundaries. For example in [7], a convex TV variant that uses $p = 1$ was suggested, which was later referred to as anisotropic TV. In [6], the authors preferred using $p \in (0, 1)$, and in the recent work [9], a weighted mixed-norm design like

$\phi = w_1 \|\cdot\|_1 + w_2 \|\cdot\|_2$ was proposed. The common drawback of the last two is their non-convexity. Also, there are applications of having $p = 0$, as in [12], which is intrinsically NP-hard. This work aims at designing such a regularizer by combining the ideas behind STV and DTV regularizers, which will be introduced in the following section.

2 Background

By recalling Eq. (2), STV plugs $\Gamma = \mathcal{S}_k$ and $\phi = \|\sqrt{\lambda(\cdot)}\|_p$. $(\mathcal{S}_k f)[i]$ stands for the *structure tensor*, which is a symmetric positive semi-definite (PSD) matrix normally defined as $(\mathcal{S}_k f)[i] = k_\sigma * ((Jf)[i](Jf)[i]^T)$, where k_σ is a Gaussian kernel of standard deviation σ ; and λ returns the eigenvalues of the given matrix. Here J denotes the Jacobian operator defined by extending the gradient for vector-valued images: $(Jf)[i] = [(\nabla f^1)[i] \ (\nabla f^2)[i] \ \cdots \ (\nabla f^C)[i]]^T$, where superscripted f 's are denoting the channels. Thus for scalar-valued images, $J(\cdot) = \nabla(\cdot)$. When compared to the local differential operators, semi-local structure tensor provides a better way to characterize the directional variations, since it also takes the neighboring pixels into account. This way lies in its rooted eigenvalues and their associated unit eigenvectors, which summarize the gradients within a patch supported by k_σ and centered at i -th pixel.

In [8], the authors also defined the structure tensor in terms of another operator that they proposed and called as *patch-based Jacobian* $J_k : \mathbb{R}^{N \times C} \mapsto \mathbb{R}^{N \times LC \times 2}$, which embodies the convolution kernel of size L . It is defined as:

$$(J_k f)[i] = [(\tilde{\nabla} f^1)[i], (\tilde{\nabla} f^2)[i] \ \cdots \ (\tilde{\nabla} f^C)[i]]^T \quad (3)$$

where $(\tilde{\nabla} f^c)[i] = [(\mathcal{T}_1 \nabla f^c)[i], (\mathcal{T}_2 \nabla f^c)[i], \dots, (\mathcal{T}_L \nabla f^c)[i]]$. Each l -th entity corresponds to shifting and weighting on the gradient as $(\mathcal{T}_l \nabla f^c)[i] = \omega[p_l] (\nabla f)[x_i - p_l]$, where x_i denotes the actual 2D coordinates of the i -th pixel, $p_l \in \mathcal{P}$ is the shift amount so that each $x_i - p_l$ is within the \mathcal{P} -neighborhood of x_i , and the weight is determined by the smoothing kernel k_σ as $\omega[p_l] = \sqrt{k_\sigma[p_l]}$. This new operator allows us to decompose the nonlinear $(\mathcal{S}_k f)[i]$ into linear $(J_k f)[i]$'s as follows:

$$(\mathcal{S}_k f)[i] = (J_k f)[i](J_k f)[i]^T \quad (4)$$

so that the new formulation reveals an easier way of employing optimization tools. The rooted eigenvalues of $(\mathcal{S}_k f)[i]$ coincide with the singular values of $(J_k f)[i]$, and this paves the way of using *Schatten p -norms* to redefine STV by this time plugging $\Gamma = J_k$ and $\phi = \|\cdot\|_{S_p}$ in Eq. (2). S_p matrix norm is nothing but the ℓ_p -norm of the vector that contains the singular values of the subjected matrix, which is $(J_k f)[i]$ in our case. Therefore, STV regularizer in its first and second form is given below:

$$STV(f) = \sum_{i=1}^N \|\sqrt{\lambda((\mathcal{S}_k f)[i])}\|_p = \sum_{i=1}^N \|(J_k f)[i]\|_{S_p} \quad (5)$$

where $\lambda((\mathcal{S}_k f)[i]) = [\lambda^+((\mathcal{S}_k f)[i]), \lambda^-((\mathcal{S}_k f)[i])]^T$.

The DTV, on the other hand, simply substitutes $\Gamma = \nabla$ and $\phi = \|A_\alpha R_{-\theta}(\cdot)\|_2$, where $R_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ and $A_\alpha = \begin{bmatrix} \alpha & 0 \\ 0 & 1 \end{bmatrix}$. Here θ corresponds to the dominant direction (that should be pre-determined), and α is used to tune the dose of penalization throughout that direction. DTV stipulates that the image to be recovered exhibits a global directional dominance. It is only designed for the scalar-valued images.

The structure tensor has the ability of capturing the first-order information within a local neighborhood. This neighborhood-awareness makes it more robust to the noise (or the other types of deterioration), when compared to the gradient. Since STV is designed by using the rooted eigenvalues of the structure tensor, it better codifies the image variation at a point than the regularizers that aims to penalize the gradient magnitude (such as TV and DTV). However, STV may not always distinguish the edges under excessive noise, thus may smooth out them. Even though DTV's preference of penalizing the gradient magnitude makes it vulnerable to the noise, it is good at struggling with it along the edges, since its gradients are rotated to a favorable direction and scaled according to a favorable dose of smoothness.

3 Method

Let us define a new operator Π (as being a composition of R_θ and A_α), which can act on the gradients within the patch, when applied to the patch-based Jacobian, as follows:

$$(\Pi J_k f)[i] = \begin{bmatrix} (\mathcal{T}_1 A_\alpha R_{-\theta} \nabla f^1)[i] & (\mathcal{T}_2 A_\alpha R_{-\theta} \nabla f^1)[i] & \cdots & (\mathcal{T}_L A_\alpha R_{-\theta} \nabla f^1)[i] \\ (\mathcal{T}_1 A_\alpha R_{-\theta} \nabla f^2)[i] & (\mathcal{T}_2 A_\alpha R_{-\theta} \nabla f^2)[i] & \cdots & (\mathcal{T}_L A_\alpha R_{-\theta} \nabla f^2)[i] \\ \vdots & \vdots & \vdots & \vdots \\ (\mathcal{T}_1 A_\alpha R_{-\theta} \nabla f^c)[i] & (\mathcal{T}_2 A_\alpha R_{-\theta} \nabla f^c)[i] & \cdots & (\mathcal{T}_L A_\alpha R_{-\theta} \nabla f^c)[i] \end{bmatrix} \quad (6)$$

Let it be called as *directional patch-based Jacobian*. Therefore, we propose a new regularizer of the form: DSTV functional of the form:

$$DSTV(f) = \sum_{i=1}^N \|(\Pi J_k f)[i]\|_{S_p} \quad (7)$$

where DSTV abbreviates the direction guided STV, here and hereafter. By means of Eq. (7), one has the chance of penalizing the eigenvalues of the structure tensor, in favor of the predetermined direction. By means of Eq. (7), the eigenvalues of the structure tensor this time summarize the rotated and scaled gradients with the incorporation of prior knowledge. Also, the presence of the $A_\alpha R_\theta$'s doesn't destroy the convexity of the regularizer.

Equation (7) can equivalently be written as a support function of the set $B_F = \{\mathcal{Y} \in \mathbb{R}^{L^C \times 2} : \|\mathcal{Y}\|_{S_2} \leq 1\}$, which is the unit ball of S_2 (Frobenius) matrix norm and takes the shape of:

$$DSTV(f) = \sum_{i=1}^N \sup_{\mathcal{Y} \in B_F} \langle \Pi J_k f \rangle [i], \mathcal{Y} \quad (8)$$

where $\langle \cdot \rangle$ denotes the inner product. The right-hand side of Eq. (8) can be rewritten in the compact form of $\sup_{\Psi[i] \in B_F} \langle \Pi J_k f, \Psi \rangle$, by introducing another variable

$\Psi \in \mathbb{R}^{N \times LC \times 2}$, where $\Psi[i]$ is the i -th submatrix. This expression allows us to use the available algorithms for orthogonal projection onto the convex B_F .

Therefore, a DSTV-regularized denoising problem will require to solve the following minimization problem:

$$\min_{f \in \mathcal{C}} \frac{1}{2} \|g - f\|_2^2 + \tau \sup_{\Psi[i] \in B_F} \langle \Pi J_k f, \Psi \rangle \quad (9)$$

where \mathcal{C} is nothing but a set that corresponds to an additional constraint on f (e.g. box constraint), which is equal to \mathbb{R}^N for unconstrained case.

Due to DSTV's nonsmoothness, solving Eq. (9) is nontrivial. But since the convexity of the STV (proven in [8]) is preserved in Eq. (8), the proximal map of DSTV corresponds to the minimizer, thus can be employed as:

$$\hat{f} = \text{prox}_{\tau DSTV(f)}(g) := \underset{f \in \mathcal{C}}{\text{argmin}} \frac{1}{2} \|g - f\|_2^2 + \tau \sup_{\Psi[i] \in B_F} \langle f, J_k^* \Pi^* \Psi \rangle \quad (10)$$

where J_k^* and Π^* arising after we leave f alone, denote the adjoints. J_k^* was defined in [8] as:

$$(J_k^* X)[k] = \sum_{l=1}^L -\text{div}(\mathcal{T}_l^* X[i, s]) \quad (11)$$

where $s = (c - 1)L + l$ and $k = (c - 1)N + n$ with $1 \leq n \leq N$ and $1 \leq c \leq C$. \mathcal{T}_l^* corresponds to the adjoint of \mathcal{T}_l , which scans the $X[i, s]$ in column-wise manner, where $X[i, s] \in \mathbb{R}^2$ is the s -th row of the i -th submatrix of an arbitrary $X \in \mathbb{R}^{N \times LC \times 2}$. In Eq. (10), $X = \Pi^* \Psi$ for the operator Π^* acting the same way with Π , except that the operator pair $A_\alpha R_{-\theta}$ applied to the two-dimensional vector elements ($\Psi[i, s]$ in this case) is replaced by $R_\theta A_\alpha$. Also in Eq. (11), div is discrete divergence, whose definition depends on the discretization scheme used for the gradient. In [8], it is defined using backward differences, since the gradient is discretized using forward differences.

From now on, we will be following the fast gradient projection (FGP) method [2], which combines the dual approach introduced in [5] to solve TV-based denoising, and the convergence accelerator FISTA [3]. Equation (10) can first be expressed as a minimax problem for the objective $\mathcal{L}(f, \Psi) = \frac{1}{2} \|g - f\|_2^2 + \tau \langle f, J_k^* \Pi^* \Psi \rangle$, which is convex w.r.t. f and concave w.r.t. Ψ ; thus, has a common saddle point that doesn't change when the minimum and the maximum are swapped as shown below:

$$\min_{f \in \mathcal{C}} \max_{\Psi[i] \in B_F} \mathcal{L}(f, \Psi) = \mathcal{L}(\hat{f}, \hat{\Psi}) = \max_{\Psi[i] \in B_F} \min_{f \in \mathcal{C}} \mathcal{L}(f, \Psi) \quad (12)$$

Algorithm 1. Algorithm for DSTV-based denoising

```

1:  $\epsilon \in (0, \min(1, L(d)))$ ,  $\tau \in [\epsilon, 1]$ 
2:  $u^{(j=1)} := [0, 0]$ ,  $t^{(j=1)} := 1$ 
3: while Eq. (10) is not  $\epsilon$ -converged do
4:    $u^{(j)} \leftarrow P_{B_F} \left( u^{(j)} + (8\sqrt{2}(\alpha^2 + 1)\tau)^{-1} (\Pi J_k P_C (g - \tau J_k^* \Pi^* u^{(j)})) \right)$ 
5:    $(u^{(j+1)}, t^{(j+1)}) \leftarrow \text{FISTA} (t^{(j)}, u^{(j)}, u^{(j-1)})$ 
6: end while
7:  $\hat{f} \leftarrow P_C (g - \gamma \tau J_k^* \Pi^* u^{(j)})$ 

```

Maximization of the dual problem $d(\Psi) = \min_{f \in C} \mathcal{L}(f, \Psi)$ at the right-hand side is same with the minimization of the primal problem at the left-hand side. When the maximizer $\hat{\Psi}$ of $d(\Psi)$ is found, it can be used to find the minimizer \hat{f} of the primal problem. In order to find $\hat{\Psi}$, we first derive the minimizer \hat{f} , in terms of Ψ , as:

$$\hat{f} = \underset{f \in C}{\operatorname{argmin}} \mathcal{L}(f, \Psi) \quad (13)$$

where $\mathcal{L}(f, \Psi) = \|f - (g - \tau J_k^*(\Pi^*\Psi))\|_2^2 - M$ is found by expanding the equation, collecting the constants at the end, and subsumed under the term M . The solution of Eq. (13) is $\hat{f} = P_C(g - \tau J_k^* \Pi^* \Psi)$, where $P_C(\cdot)$ is the orthogonal projection onto the convex set C . Then we proceed by plugging \hat{f} in $\mathcal{L}(f, \Psi)$ to get the dual problem $d(\Psi) = \mathcal{L}(\hat{f}, \Psi) = \frac{1}{2} \|w - P_C(w)\|_2^2 + \frac{1}{2} (\|z\|_2^2 - \|w\|_2^2)$, where $w = g - \tau J_k^* \Pi^* \Psi$. As opposed to the primal problem, $d(\Psi)$ is smooth with a well-defined gradient:

$$\nabla d(\Psi) = \tau \Pi J_k P_C (g - \gamma \tau J_k^* \Pi^* \Psi) \quad (14)$$

obtained based on the derivation in Lemma 4.1 of [2]. From now on, finding the maximizer $\hat{\Psi}$ of $d(\Psi)$ is trivial by using the idea of the projected gradients [5]. In case of minimization, it iteratively performs decoupled sequences of gradient descent and gradient projection on a set. In our case, each $\Psi[i]$ will be projected onto the set B_F , where the projection is defined as:

$$P_{B_F}(\Psi[i]) = \frac{\Psi[i]}{\max(1, \|\Psi[i]\|_F)} \quad (15)$$

in [8], to which the readers are referred for the derivation. When it comes to gradient ascent, an appropriate step size that ensures the convergence is needed. Since Eq. (14) is a Lipschitz continuous gradient, a constant step size $1/L(d)$ can be used, where $L(d)$ is the Lipschitz constant having the upper bound $L(d) \leq 8\sqrt{2}(\alpha^2 + 1)\gamma^2\tau^2$. For the derivation, see Appendix A.

As a consequence, the overall algorithm is shown in Algorithm 1. FISTA procedure is referring to the fast iterative shrinkage-thresholding algorithm proposed in [3], and attached to Appendix B.

4 Experimental Results

In this section, we perform experiments to compare the performance of our regularizer, with its predecessor STV and its influencer DTV. Since STV’s superiority over the baseline TV and the TGV has already been shown in [8], they are not included to the competing algorithms. In Fig. 1, the uni-directional grayscale and color images used in the experiments are all shown. Among those images; *reed*, *sea shell*, *cotton bud*, and *feather* were taken from Amsterdam Library of Textures (ALOT) image dataset while the others were public domain images. We use peak signal to noise ratio (PSNR) and structural similarity index (SSIM) [11] to assess the performances of the algorithms. DSTV has been implemented on top of STV, whose source code is publicly available on the author’s (Lefkimiatis, S.) website. All methods were written in MATLAB by only making use of the CPU. The runtimes were computed on a computer equipped with Intel Core Processor i7-7500U (2.70-GHz) with 16 GB of memory.



Fig. 1. Above row respectively shows the thumbnails of the grayscale *cotton bud*, *feather*, *straw*; and color *reed*, *grass*, *paper*, *sea shell*, *palm*, *cat fur* images of size 256×256 used in the experiments. (Color figure online)

For the convolution kernel k_σ of the structure tensor, we choose it to be a 5×5 pixel Gaussian window with standard deviation $\sigma = 0.5$ pixels, for both STV and DSTV. This decision is taken by prioritizing the reconstruction quality; however, by selecting a 3×3 window, the computational complexity, thus the reported runtimes can be reduced (nearly up to the DTV’s runtimes) at the cost of a little loss from the quality. On the other hand, in all experiments, α for DTV and DSTV, and the regularization parameter τ for all competing regularizers are optimized, such that they yield the best possible PSNR. Furthermore, to be used by DTV and DSTV, the θ value is estimated from the observed image by computing the DTV measures for a range of angles, while keeping $\alpha > 1$ constant. The angle that yields the smallest DTV is picked as the dominant direction.

The grayscale results of the DSTV against to the DTV and the STV are reported in Table 1. The PSNR and SSIM performances are listed, along with their respective runtimes, by considering four different noise levels applied on the first three images shown in Fig. 1. Although it is obvious that the DSTV yields better recovery than the others, for the cotton bud and the feather images, DTV almost catches up DSTV, especially in terms of the PSNR measure. This is due to the fact that those images are coherently uni-directional (except the bottom left and top right corners of the feather image), with smooth, non-textural, and continuous stripes. Denoising this kind of images is a trivial task for DTV, since its enforced direction of smoothness coincides with the correct pattern. Even so,

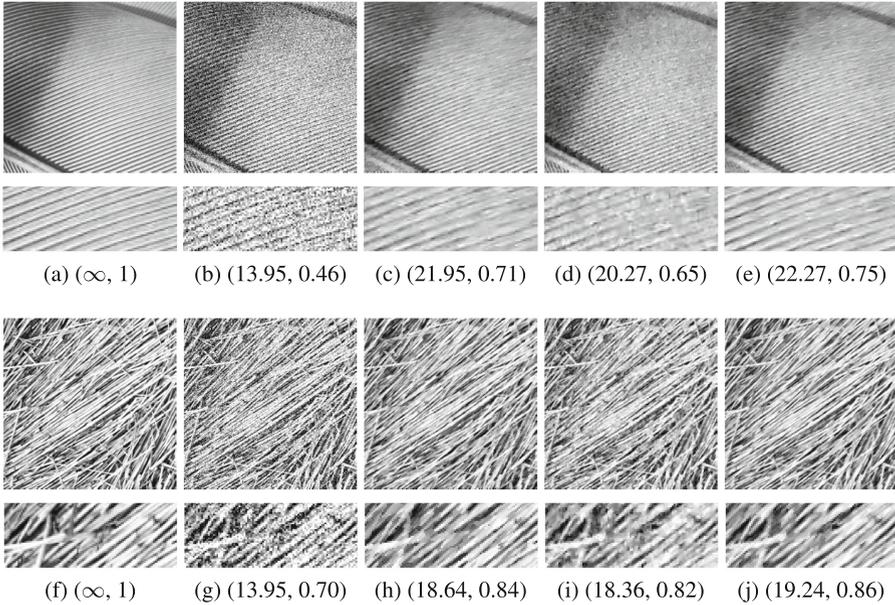


Fig. 2. The restored versions of the noisy ($\sigma_\eta = 0.2$) *feather* and *straw* images (col-2), by using DTV (col-3), STV (col-4) and DSTV (col-5) regularizers. The quantity pairs shown at the bottom of each image are corresponding to the (PSNR, SSIM) values. Also the detail patches cropped from each image are provided.

in higher frequency feather image, the SSIM gap between DTV and DSTV is concretely higher than the one obtained in cotton bud image. When it comes to the straw image, deviations from the dominant direction work against both DTV and DSTV. The results on straw image can better reflect the DSTV's superiority arising from the semi-localness of the structure tensor, over DTV. This semi-localness make DSTV more robust to the errors in pre-determined θ . For the visual assessment, Fig. 2 shows the restored versions of the feather and straw images by three subjected regularizers. As one can observe, STV causes to the oil-painting like artifacts on the edges, whereas it is the only regularizer that can manage to restore the texture at the bottom left corner of the feather image, where the direction does not match with the dominant one. On the other side, the DTV over-smooths by damaging rough textures on the details, which are well preserved by the DSTV (see the detail images in Fig. 2).

In Table 2 and in Fig. 3, the results obtained from the color images are reported. Since DTV is only designed for the scalar-valued images, it is excluded from these experiments. According to the results presented in the table, DSTV systematically outperforms STV for our uni-directional image denoising problem. Its superiority becomes more apparent in the high-frequency images with less deviations from the pre-determined dominant direction (e.g. palm). In sea shell image on the other hand, the PSNR and the SSIM values obtained by both

Table 1. Assessing the PSNR/SSIM performances of DTV [1], STV [8], and DSTV regularizers (along with their elapsed runtimes) on scalar-valued (grayscale) images shown in Fig. 2.

	σ_η	0.1	0.15	0.2	0.25
Cotton bud	DTV	29.80/ 0.90 4.55 s	27.91/0.86 4.02 s	26.58/0.83 4.10 s	25.57/0.80 2.16 s
	STV	27.85/0.86 25.08 s	25.78/0.80 29.01 s	24.35/0.75 22.70 s	23.28/0.70 28.98 s
	DSTV	29.87/0.90 25.61 s	28.04/0.87 25.46 s	26.89/0.84 16.60 s	25.98/0.81 29.24 s
Feather	DTV	25.17/0.86 3.06 s	23.23/0.78 4.26 s	21.95/0.71 3.92 s	21.04/0.66 4.42 s
	STV	23.83/0.82 29.77 s	21.69/0.72 23.91 s	20.27/0.65 26.06 s	19.30/0.56 23.46 s
	DSTV	25.60/0.87 17.82 s	27.13/0.81 23.66 s	22.27/0.75 28.50 s	21.43/0.70 31.1 s
Straw	DTV	22.59/0.93 1.62 s	20.19/0.89 2.31 s	18.64/0.84 2.68 s	17.51/0.79 4.07 s
	STV	22.47/0.93 27.16 s	19.97/0.88 28.45 s	18.36/0.82 30.52 s	17.20/0.77 34.48 s
	DSTV	23.17/0.94 32.28 s	20.81/0.90 33.61 s	19.24/0.86 28.52 s	18.11/0.82 35.85 s

Table 2. Assessing the PSNR/SSIM performances of STV [8] and DSTV denoisers (along with their elapsed runtimes) on vector-valued (color) images shown in Fig. 3.

Method	STV			DSTV		
	σ_η	0.1	0.15	0.2	0.1	0.15
Reed	28.31/0.95	26.73/0.93	25.62/0.92	29.17/0.96	27.75/0.94	26.77/0.93
	28.14 s	27.44 s	28.88 s	30.39 s	30.12 s	31.98 s
Grass	28.13/0.97	26.09/0.96	24.72/ 0.95	29.42/0.98	27.26/0.97	25.63/0.95
	28.52 s	31.17 s	46.52 s	29.91 s	31.12 s	30.17 s
Paper	27.38/0.95	25.03/0.91	23.47/0.88	29.73/0.97	27.57/0.95	26.04/0.93
	32.85 s	27.60 s	46.30 s	31.73 s	48.21 s	47.41 s
Sea shell	31.68/ 0.90	30.00/ 0.87	29.45/ 0.85	31.72/0.90	30.45/0.87	29.54/0.85
	33.85 s	31.46 s	28.60 s	33.50 s	37.19 s	32.29 s
Palm	24.34/0.91	22.02/0.85	20.52/0.79	25.87/0.94	23.67/0.90	22.19/0.86
	30.81 s	31.36 s	50.55 s	27.92 s	29.80 s	23.65 s
Cat fur	24.81/0.76	23.08/0.65	22.11/0.57	24.87/0.77	23.28/0.68	22.31/0.60
	30.23 s	30.85 s	30.44 s	29.83 s	27.49 s	32.75 s

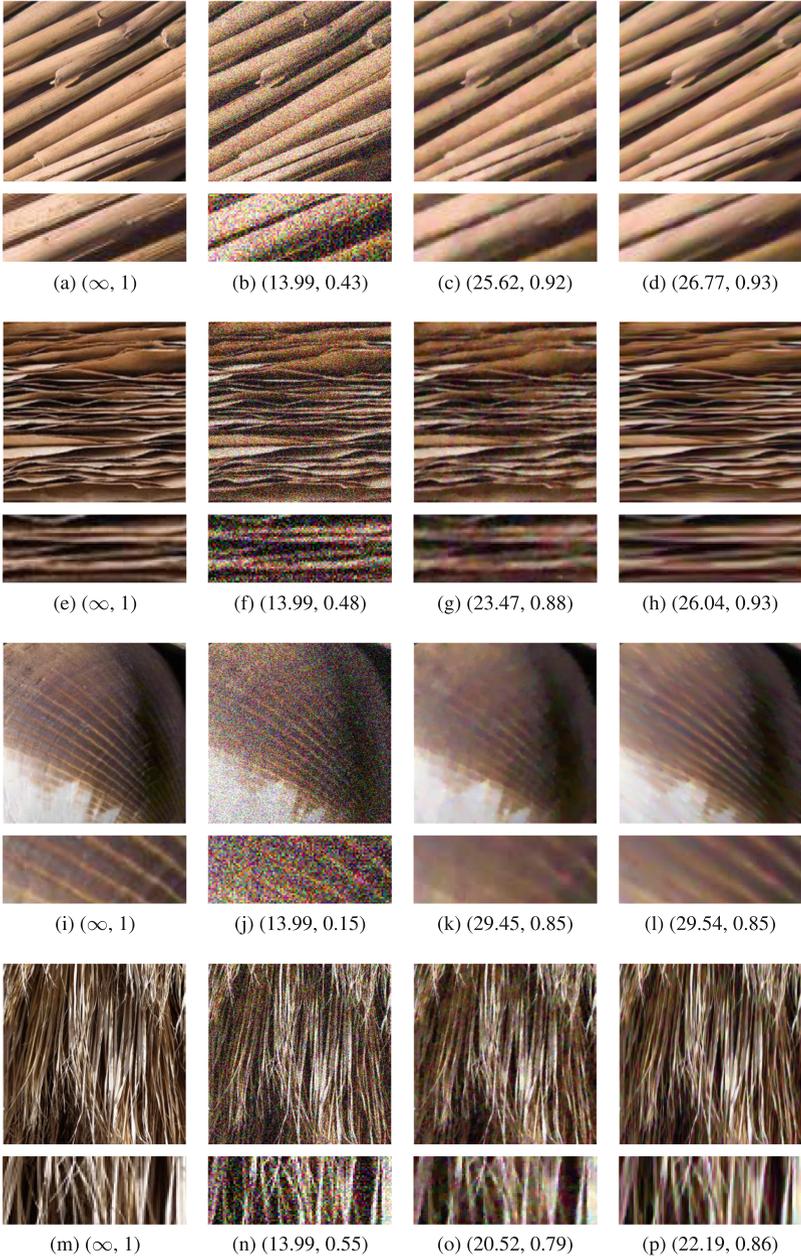


Fig. 3. The restored versions of the noisy ($\sigma_\eta = 0.2$) *reed*, *paper*, *sea shell*, and *palm* images (col-2), by using STV (col-3) and DSTV (col-4) regularizers. The quantity pairs shown at the bottom of each image are corresponding to the (PSNR, SSIM) values. Also the detail patches cropped from each image are provided. (Color figure online)

regularizers are almost same. This result is due to the smooth transitions of the stripes on the shell from one direction to another. However, for the same image, the value of DSTV can visually be seen by comparing the detail images given in Fig. 3(k) and (l). In contrast to the quantitative results, these images reflect the DSTV’s ability of texture preservation. The same applies for the cat fur image, which substantially may not be categorized as a uni-directional image. Also to evaluate the DSTV’s robustness to the small deviations from the dominant direction, one can observe the first row of Fig. 3. Overall, DSTV’s idea of contributing STV minimization process in favor of the dominant direction succeeds denoising in more appealing way for nearly uni-directional images; while STV fails to restore high-frequency details under excessive amount of noise, since the fair contributions of the rapidly changing gradients within a neighborhood may mislead the process.

5 Conclusion and Future Work

In this paper, we discoursed the denoising problem from variational perspective, and proposed a new analysis-based regularizer to be used on uni-directional images. We utilized the DTV’s idea of incorporating the prior knowledge on the directional dominance of the latent image to the inversion process, while designing our regularizer DSTV as a variant of STV. Throughout the paper, only the uni-directional images were our concern. Even though this scenario is biased to the DTV and DSTV, it provides valuable insight on how the guidance of a directional prior can improve one of the state-of-the-art regularizers. As shown by the experimental results, DSTV systematically outperformed its two predecessors. This encourages us about a possible extension to DSTV that can work with spatially varying θ ’s and α ’s, extracted from the observed image. A relevant research question is if it is possible to develop such a preprocessor algorithm that can extract robust derivative-based directional descriptors, and can feed them into DSTV-based denoising machinery.

Appendix A

In order to find an upper bound for Lipschitz constant, we follow the derivation in [2] and adapt it to our formulation.

$$\begin{aligned}
 \|\nabla d(v) - \nabla d(u)\| &= \tau \|(\Pi J_k P_C(g - \tau J_k^* \Pi^* u) - \Pi J_k P_C(g - \tau J_k^* \Pi^* v))\| \\
 &\leq \tau \|\Pi J_k\| \|P_C(g - \tau J_k^* \Pi^* u) - P_C(g - \tau J_k^* \Pi^* v)\| \\
 &\leq \tau \|\Pi J_k\| \|\tau J_k^* \Pi^*(u - v)\| \\
 &\leq \tau^2 \|\Pi\|^2 \|J_k\|^2 \|u - v\| \\
 &\leq \tau^2 \|\Pi\|^2 \|\nabla\|^2 \|T\|^2 \|u - v\|
 \end{aligned} \tag{16}$$

where $T = \sum_{l=1}^L (\mathcal{T}_l^* \mathcal{T}_l)$. Knowing from [2] that $\|\nabla\|^2 \leq 8$, from [8] that $\|T\|^2 \leq \sqrt{2}$, and further showing that $\|\Pi^*\|^2 \leq (\alpha^2 + 1)$, we come up with the Lipschitz constant $L(d) \leq 8\sqrt{2}(\alpha^2 + 1)\tau^2$.

Appendix B

Algorithm 2. FISTA [3]

- 1: **function:** $(f^{(next)}, t^{(next)}) = \text{FISTA}(t, f^{(cur)}, f^{(prev)})$
 - 2: $t^{(next)} \leftarrow (1 + \sqrt{1 + 4t^2})/2$
 - 3: $f^{(next)} \leftarrow f^{(cur)} + (\frac{t-1}{t^{(next)}})(f^{(cur)} - f^{(prev)})$
 - 4: **end**
-

References

1. Bayram, I., Kamasak, M.E.: Directional total variation. *IEEE Signal Process. Lett.* **19**(12), 781–784 (2012)
2. Beck, A., Teboulle, M.: Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. *IEEE Trans. Image Process.* **18**(11), 2419–2434 (2009)
3. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.* **2**(1), 183–202 (2009)
4. Bredies, K., Kunisch, K., Pock, T.: Total generalized variation. *SIAM J. Imaging Sci.* **3**(3), 492–526 (2010)
5. Chambolle, A.: An algorithm for total variation minimization and applications. *J. Math. Imaging Vis.* **20**(1–2), 89–97 (2004)
6. Chartrand, R.: Exact reconstruction of sparse signals via nonconvex minimization. *IEEE Signal Process. Lett.* **14**(10), 707–710 (2007)
7. Esedoğlu, S., Osher, S.J.: Decomposition of images by the anisotropic Rudin-Osher-Fatemi model. *Commun. Pure Appl. Math.* **57**(12), 1609–1626 (2004)
8. Lefkimmatis, S., Roussos, A., Maragos, P., Unser, M.: Structure tensor total variation. *SIAM J. Imaging Sci.* **8**(2), 1090–1122 (2015)
9. Lou, Y., Zeng, T., Osher, S., Xin, J.: A weighted difference of anisotropic and isotropic total variation model for image processing. *SIAM J. Imaging Sci.* **8**(3), 1798–1823 (2015)
10. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. *Phys. D* **60**(1–4), 259–268 (1992)
11. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* **13**(4), 600–612 (2004)
12. Xu, L., Lu, C., Xu, Y., Jia, J.: Image smoothing via L_0 gradient minimization. *ACM Trans. Graph. (TOG)* **30**, 174 (2011)