

Adaptive Direction-Guided Structure Tensor Total Variation

Ezgi Demircan-Tureyen^{a,b,*}, Mustafa E. Kamasak^b

^aDept. of Computer Engineering, Istanbul Kultur University, Istanbul 34156, Turkey

^bDept. of Computer Engineering, Istanbul Technical University, Istanbul 34390, Turkey

Abstract

Direction-guided structure tensor total variation (DSTV) is a recently proposed regularization term that aims at increasing the sensitivity of the structure tensor total variation (STV) to the changes towards a predetermined direction. Despite of the plausible results obtained on the uni-directional images, the DSTV model is not applicable to the multi-directional images of real-world. In this study, we build a two-stage framework that brings adaptivity to DSTV. We design an alternative to STV, which encodes the first-order information within a local neighborhood under the guidance of spatially varying directional descriptors (i.e., orientation and the dose of anisotropy). In order to estimate those descriptors, we propose an efficient preprocessor that captures the local geometry based on the structure tensor. Through the extensive experiments, we demonstrate how beneficial the involvement of the directional information in STV is, by comparing the proposed method with the state-of-the-art analysis-based denoising models, both in terms of restoration quality and computational efficiency.

Keywords: variational models, image restoration, directional total variation, structure tensor, orientation field estimation, inverse problems

1. Introduction

Restoration of an image from its noisy linear transformations is an active field of research in imaging science. This field involves a variety of linear inverse problems; such as denoising, deblurring, inpainting, and super-resolution. Among them, the denoising problem is extensively studied; not only due to having a wide spectrum of application fields, but also because it serves as a building block for the other, more complex problems. The fundamental drawback of denoising (and the others) is ill-posedness. The knowledge deduced from the observed image is often insufficient and requires to take some prior knowledge on the latent image into account. When such a problem is handled from a variational perspective, its reinterpretation in the form of an optimization problem involves a regularization term (i.e., regularizer), which is responsible for encoding that prior knowledge.

A lot of effort has been dedicated to design good regularizers, which can characterize the unknown image well and in an efficiently solvable way. Rudin-Osher-Fatemi's total variation (TV) [1] has been considered to be a *good* regularizer, due to its well-characterization of the images (as piecewise constant signals up to the edge regions), and its convex definition. Inevitably, TV's piecewise constancy (PWC) assumption does not always hold. The real-world images may be piecewise smooth rather than constant, and when this is the case, TV yields artificial edges, known as staircase artifacts [2]. As being the initiator of the notion of edge-preserving regularization, TV has

given rise to a diverse range of TV inspired regularizers. Many of these regularizers extend TV to stretch its PWC assumption, by bring anisotropy on the TV-term to better catch angled boundaries (e.g., [3, 4, 5]), and/or by incorporating the higher order information to promote piecewise smoothness (e.g., total generalized variation – TGV [6], and the other notable works [2, 7, 8]).

There are also TV extensions that steer anisotropy by a desired direction. Directional total variation (DTV) [9] is such a regularizer that aims at penalizing the local changes along a single dominant direction more than the others. For this purpose, it employs rotated and scaled gradients. Even though it performs well on the uni-directional images, DTV model fails to restore arbitrary, real-world images. It is applicable only when the image to be restored exhibits a single directional dominance. In [10], the authors suggested an edge-adaptive DTV (EADTV) to make DTV handle several dominant directions. They proposed a two-stage approach: (1) extracting the tangential directions of the edges, (2) feeding them into the EADTV model. However, the noise heavily affects the edges and mostly the wrong directions are penalized. A very recent work [11] proposed an extended DTV model that follows a similar two-stage approach. Distinctively, to extract directions, they employed a more sophisticated orientation field estimation (OFE) method used for fingerprint identification in [12]. They also made use of the mixed $\ell_p - \ell_2$ norm, where $p \in (0, 1)$, while $p = 1$ in DTV.

The direction-guided regularizers mentioned above requires a preprocessor that provides the directions to them. So, the erroneous outputs of the preprocessor mislead the succeeding inversion process. In [13], anisotropic TV filtering (ATV) was

*Corresponding author

Email addresses: e.demircan@iku.edu.tr (Ezgi Demircan-Tureyen), kamasak@itu.edu.tr (Mustafa E. Kamasak)

proposed as a combination of TV and anisotropic diffusion. It embodies the *diffusion tensor* designed according to the *structure tensor*. It intrinsically depends on the underlying image, rather than the observation. The downsides of ATV are its non-convexity and heuristic design. The objective function mostly has multiple local or global minimizers, and it may result in a suboptimal solution. The recent structure tensor total variation (STV) [14], which forms the basis of this study in collaboration with DTV [9], proposed a convex objective by following the same idea behind ATV. It suggests to penalize the rooted eigenvalues of the structure tensor. In this way, it can capture the variations within a local neighborhood in contrast to TV and its localized extensions. In [14], its superiority has been shown over vectorial TV (VTV) [15], which extends TV to handle vector-valued images, and TGV [6].

Above mentioned regularizers are all local, except ATV and STV, which are classified as semi-local. There are also variational models that benefit from the non-local (NL) self similarity by adopting the idea of [16] (e.g., NLTV [17], NLSTV [18]). They usually perform better than their local or semi-local counterparts. Moreover, the regularizers mentioned here are all categorized as analysis-based priors, where the regularization directly runs on the image. Contrary to this, there are also synthesis-based priors that perform the regularization in a sparsifying transform domain (e.g., wavelet [19, 20], shearlets [21]). Also there is a recent paper [22] that proposes a semi-local regularizer along with its non-local version. Their prior depends on a new paradigm: It employs sparsity of the corners rather than the edges. They showed that their method (Noble’s Corner Detector based Regularization – NCDR) is superior to STV (and NLSTV for the non-local versions) in terms of restoration quality, apart from the fact that it is non-convex, works heuristically and slow.

1.1. Contributions

Our intention is proposing a variational denoising framework which produces physically satisfying and statistically meaningful results that are superior to the state-of-the-art analysis-based variational methods. In this respect, our contributions are listed below:

1. We design an adaptive direction-guided penalty term by extending STV.
2. We propose a preprocessing method that estimates the parameters required by our regularizer.
3. We develop an efficient algorithm based on the proximal map of the proposed regularizer in order to solve the introduced convex optimization problem.
4. We assessed the performance of our framework by comparing the results with those obtained by TV, EADTV [9, 10], STV [14], and NCDR [22].

Note that, we only focus on extending the semi-local STV to semi-local adaptive DSTV; thus, we won’t consider designing a non-local counterpart of our regularizer, which can straightforwardly be accomplished by exploiting NLSTV [18]. It would surely boost the restoration quality.

2. Problem Formulation

We deal with the recovery of an underlying function $f : \Omega \mapsto \mathbb{R}^C$ from an observation $g : \Omega \mapsto \mathbb{R}^C$, where the forward corruption is modeled as: $g(x) = f(x) + \eta(x)$, where $\eta : \Omega \mapsto \mathbb{R}^C$ referring to the additive noise. Here, C denotes the number of channels, which is more than one for the vector-valued images, and $x \in \Omega$, where the domain $\Omega \subseteq \mathbb{R}^2$ is a 2-dimensional image space. The noise at each pixel is modeled as an *i.i.d.* random variable that follows Gaussian distribution with $\eta(x) \sim \mathcal{N}(0, \sigma_\eta^2)$.

The variational models combine a data fidelity term with a regularization term $R(f)$ in the form of an optimization problem. For the Gaussian denoising, the fidelity term is nothing but the least squares fitting term to stick by the observed data, thus, one has the objective function of the form:

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

to be minimized with respect to f . Here, τ is the regularization parameter that is used to tune the contribution of $R(f)$ to the cost. Note that, there are studies in the literature that makes τ adaptive to give rise to spatially dependent regularization (e.g., [23], [24], [25]).

2.1. Notation

Throughout the paper, both of the scalar-valued grayscale images ($C = 1$) and the vector-valued RGB images ($C = 3$) will be our concern. We will only focus on the discrete domain, where f is a discrete-space function, thus we switch the notation from $f(x)$ to $f[i]$, for $i \in \{1, 2, \dots, N\}$ referring to the index of a certain pixel, where N is the number of pixels in a single channel. Note that, we treat the images as being lexicographically stacked into vectors ($f \in \mathbb{R}^{NC}$).

$\|X\|_{p,q}$ stands for either the mixed $\ell_p - \ell_q$ norm defined as $(\sum_{i=1}^N \|X[i]\|_q^p)^{1/p}$, or the mixed $\ell_p - \mathcal{S}_q$ norm when X is a matrix. Here \mathcal{S}_q refers to the Schatten norm of order q ¹. The symbols \odot and \circ are used to denote entrywise product and entrywise power, respectively; while $\langle \cdot \rangle$ refers to the inner product. Furthermore, the vectors/matrices are denoted by using lowercase/uppercase bold letters thereafter.

3. Background

The analysis-based regularizers can generically be defined as follows:

$$R(\mathbf{f}) = \|\phi((\Gamma \mathbf{f})[i])\|_p \quad (2)$$

TV seminorm is nothing but the ℓ_1 norm of the gradient magnitude, i.e., according to Eq. (2), $\Gamma = \nabla$, $\phi = \|\cdot\|_2$, and $p = 1$, where ∇ denotes the gradient operator. As mentioned in Sect. 1, many alternatives to this native isotropic TV have been proposed, e.g., $\phi = \|\cdot\|_1$ in [3], $\phi = \|\cdot\|_q$ and $q \in (0, 1)$ in [4], $\phi = w_1 \|\cdot\|_1 + w_2 \|\cdot\|_2$ in [5], where $\Gamma = \nabla$ and $p = 1$ were kept for all.

¹Schatten norm of a matrix \mathbf{X} is ℓ_q norm of the vector $\sigma(\mathbf{X})$, where σ returns the singular values, i.e., $\|\mathbf{X}\|_{\mathcal{S}_q} = \|\sigma(\mathbf{X})\|_q$ [26]

3.1. Directional Total Variation (DTV)

DTV [9] is a local analysis-based regularizer that simply changes the TV's rotation invariant ∇ in Eq. (2) with $\Gamma = \Lambda_\alpha \mathbf{R}_{-\theta} \nabla$, where $\mathbf{R}_\theta = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$ and $\Lambda_\alpha = \begin{bmatrix} \alpha & 0 \\ 0 & 1 \end{bmatrix}$ with $\theta \in [0, \pi)$ corresponding to the dominant direction $(\cos \theta, \sin \theta)$, and α that is used to tune the dose of the penalization. In this way, it penalizes the magnitudes of the scaled and rotated gradients, towards a predetermined direction. As mentioned earlier, DTV can only be applied to the uni-directional images. EADTV [10] is its adaptive extension that suggests to use spatially varying angles $\theta[i]$. Even it is trivial to estimate a scalar θ from an observed uni-directional image, when θ represents an *orientation field*, it becomes a challenging problem. EADTV simply computes each $\theta[i]$ such that,

$$(\cos \theta[i], \sin \theta[i]) \perp (\nabla \mathbf{g}_\sigma)[i] \quad (3)$$

where the subscript σ indicates that the image \mathbf{g} is pre-smoothed by a Gaussian filter of standard deviation σ . However, the gradient is not a convenient descriptor under extensive amount of noise. The estimation of the gradient is mostly misguided, and since DTV is sensitive along the angles $\theta[i]$, the structure is destroyed by the incorrect estimates. Moreover, rippled values of the angles cause unstable output.

By recalling Eq. (2), the recent paper [11] plugs $\Gamma = \Lambda_\alpha(\mathbf{g})\mathbf{R}_{-\theta(\mathbf{g})}\nabla$, $\phi = \|\cdot\|_2$, and $p \in (0, 1)$ where Λ_α and θ are both functions of \mathbf{g} . Except the preference of the outer norm, this definition differs from the EADTV in the sense that it better recognizes the local edge orientations due to the more complex OFE method that it employs, and uses an adaptive stretching matrix weighted according to the first-order local edge information.

3.2. Structure Tensor Total Variation (STV)

The *structure tensor* of an image \mathbf{f} at a point i is a 2×2 symmetric positive semi-definite (PSD) matrix of the form:

$$(S_K \mathbf{f})[i] = \mathbf{K}_{\hat{\sigma}} * ((J\mathbf{f})[i](J\mathbf{f})[i]^T) \quad (4)$$

where $\mathbf{K}_{\hat{\sigma}}$ is a Gaussian kernel of standard deviation $\hat{\sigma}$ centered at the i -th pixel, and J denotes the Jacobian operator that extends the gradient for vector-valued images, i.e.,

$$(J\mathbf{f})[i] = [(\nabla \mathbf{f}^1)[i] \ (\nabla \mathbf{f}^2)[i] \ \cdots \ (\nabla \mathbf{f}^C)[i]] \quad (5)$$

The superscripted \mathbf{f} 's are denoting the channels, and for scalar-valued images, $J(\cdot) = \nabla(\cdot)$. This semi-local descriptor can summarize all the gradients within the patch supported by $\mathbf{K}_{\hat{\sigma}}$, thus provides a better way of characterizing the variations, when compared to the local differential operators. The eigenvectors of it correspond to the directions of maximum and minimum vectorial variations, and the rooted eigenvalues serve as a measure of variations towards the respective directions. At this point, the STV prior [14] presents an innovatory way of describing the total amount of variations. It replaces the TV's ∇ in Eq. (2) with $\Gamma = \sqrt{\lambda(\cdot)}$, where $(\lambda \mathbf{f})[i]$ is a 2D vector composed of the maximum $\lambda^+(\cdot)$ and the minimum $\lambda^-(\cdot)$ eigenvalues of the structure

tensor at the point i , i.e., $(\lambda \mathbf{f})[i] = [\lambda^+((S_K \mathbf{f})[i]), \lambda^-((S_K \mathbf{f})[i])]^T$. STV also considers a more general case by using $\phi = \|\cdot\|_q$ ($q > 1$). Since this design of STV is nonlinear and Eq. (4) involves a convolution kernel, the authors of [14] reformulated $(S_K \mathbf{f})[i]$ in terms of another operator that they named 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 , and enables to express the STV functional such that it can be decomposed into linear functionals. $(J_K \mathbf{f})[i]$ is defined as:

$$(J_K \mathbf{f})[i] = [(\tilde{\nabla} \mathbf{f}^1)[i], (\tilde{\nabla} \mathbf{f}^2)[i] \ \cdots \ (\tilde{\nabla} \mathbf{f}^C)[i]]^T \quad (6)$$

where $(\tilde{\nabla} \mathbf{f}^c)[i] = [(\mathcal{T}_1 \nabla \mathbf{f}^c)[i], (\mathcal{T}_2 \nabla \mathbf{f}^c)[i], \dots, (\mathcal{T}_L \nabla \mathbf{f}^c)[i]]$. Each l -th entity applies shifting and weighting on the gradient as $(\mathcal{T}_l \nabla \mathbf{f}^c)[i] = \sqrt{K_\sigma[p_l]}(\nabla \mathbf{f}^c)[x_i - p_l]$, where x_i denotes the actual 2D coordinates of the i -th pixel, and for $L = (2L_K + 1)^2$, $p_l \in \{-L_K, \dots, L_K\}^2$ is the shift amount. That way, the structure tensor in terms of the patch-based Jacobian is:

$$(S_K \mathbf{f})[i] = (J_K \mathbf{f})[i](J_K \mathbf{f})[i]^T \quad (7)$$

Now, the rooted eigenvalues of $(S_K \mathbf{f})[i]$ coincide with the singular values of $(J_K \mathbf{f})[i]$, and by employing S_q matrix norm, STV is redefined as:

$$STV(\mathbf{f}) = \sum_{i=1}^N \|(J_K \mathbf{f})[i]\|_{S_q} \quad (8)$$

This redefinition allows one to find an efficient numerical solution to STV regularized problems, by using convex optimization tools.

4. Proposed Method

Let us first point out the relation between the DTV and STV measures. By expanding the DTV measure, one can draw the following conclusion:

$$DTV(\mathbf{f}) \triangleq \|\Lambda_\alpha \mathbf{R}_{-\theta} \nabla \mathbf{f}\|_{1,2} = \left\| \begin{bmatrix} \alpha D_\theta \mathbf{f} \\ D_{\theta^\perp} \mathbf{f} \end{bmatrix} \right\|_{1,2} \quad (9)$$

where D_θ denotes the *directional derivative* of \mathbf{f} in the direction characterized by θ , and $\theta^\perp = \theta \pm \pi/2$. Eq. (9) shows that DTV actually works by increasing the dose of penalization (with the weight α) in the direction that it presumes the variation is minimum. The penalty dose in the direction of maximum variation, which is orthogonal to that of the minimum, is only determined by the regularization parameter τ . This idea of penalizing the maximum and minimum directional variations coincides with the STV. However, instead of making assumptions, STV codifies the maximum and the minimum directional variations with the maximum and the minimum eigenvectors and the square roots of the corresponding eigenvalues ($\sqrt{\lambda^+}$ and $\sqrt{\lambda^-}$) of the structure tensor. In other words, it deduces the directional variation from a summary of all derivatives within a local neighborhood.

$$STV(\mathbf{f}) \triangleq \left\| \begin{bmatrix} \sqrt{\lambda^+(S_K \mathbf{f})} \\ \sqrt{\lambda^-(S_K \mathbf{f})} \end{bmatrix} \right\|_{1,q} \quad (10)$$

From this point of view, the question that motivated us to design a direction-guided STV was: Can STV more accurately deduce the directional variation if the summarized derivatives are directional in a pre-determined direction?

4.1. Direction-Guided STV (DSTV)

While it has been shown that the STV produces the state-of-the-art results among the other analysis-based local regularizers; under excessive amount of noise, it may not distinguish the edges and may smooth out them. At this point, the DSTV [27] comes into play. It aims at incorporating the prior knowledge on the underlying image's local geometry into STV. Under the inspiration of DTV, DSTV employs directional derivatives by applying the operator $\mathbf{\Pi}_{(\alpha,\theta)}$ that can act on the gradient at each image point as $(\mathbf{\Pi}_{(\alpha,\theta)}\nabla f^c)[i] = \mathbf{\Lambda}_\alpha \mathbf{R}_{-\theta}(\nabla f^c)[i]$, while gathering the neighboring gradients to constitute the patch-based Jacobian as:

$$(J_K^{(\alpha,\theta)} f)[i] = \begin{bmatrix} (\mathcal{T}_1 \mathbf{\Pi}_{(\alpha,\theta)} \nabla f^1)[i] & (\mathcal{T}_2 \mathbf{\Pi}_{(\alpha,\theta)} \nabla f^1)[i] & \cdots & (\mathcal{T}_L \mathbf{\Pi}_{(\alpha,\theta)} \nabla f^1)[i] \\ (\mathcal{T}_1 \mathbf{\Pi}_{(\alpha,\theta)} \nabla f^2)[i] & (\mathcal{T}_2 \mathbf{\Pi}_{(\alpha,\theta)} \nabla f^2)[i] & \cdots & (\mathcal{T}_L \mathbf{\Pi}_{(\alpha,\theta)} \nabla f^2)[i] \\ \vdots & \vdots & \vdots & \vdots \\ (\mathcal{T}_1 \mathbf{\Pi}_{(\alpha,\theta)} \nabla f^c)[i] & (\mathcal{T}_2 \mathbf{\Pi}_{(\alpha,\theta)} \nabla f^c)[i] & \cdots & (\mathcal{T}_L \mathbf{\Pi}_{(\alpha,\theta)} \nabla f^c)[i] \end{bmatrix} \quad (11)$$

$J_K^{(\alpha,\theta)}$ is named as *directional patch-based Jacobian* [27], and involved by the DSTV functional, i.e.,

$$DSTV(\mathbf{f}) = \sum_{i=1}^N \|(J_K^{(\alpha,\theta)} \mathbf{f})[i]\|_{S_q} \quad (12)$$

By means of Eq. (12), one has the chance of leading the STV machinery in favor of a predetermined direction. It has been experimentally shown on uni-directional images in [27] that this incorporation produces prominently better results. Also, since the convexity is preserved in DSTV, the same convex optimization tools used to solve STV can be applied to DSTV based problems.

4.2. Adaptive DSTV (ADSTV)

In natural images, the dominant directions vary in different regions, which makes DSTV inapplicable. In order to handle arbitrary natural images, we employ the same way that EADTV employed when bringing adaptivity to DTV. We consider θ as a spatially varying field of orientation, i.e., $\theta : \Omega \mapsto [0, \pi)^C$, and this changes the rotation matrix to $\mathbf{R}_{-\theta[i]}$.

Furthermore, in natural images, the anisotropic structures are not homogeneously distributed. Also among the anisotropic structures, their degrees of anisotropy spatially change. DSTV's stretching matrix $\mathbf{\Lambda}_\alpha$ uses a constant factor α for the entire image. Just as done to the rotation parameter, one may employ a field of spatially changing stretching factors, so that the dose of penalization at a point can be tuned. However, since the regularization parameter τ is a constant scalar, with varying stretching factor the total amount of regularization changes from region to region. The fittest value of τ for a directional region that requires larger values of $\alpha[i]$ may produce poor results

in the regions with isotropic structures, where the corresponding values of $\alpha[i]$ are small. To compensate it, we suggest to use a scaling matrix $\tilde{\mathbf{\Lambda}}_{(\alpha^+, \alpha^-[i])}$ with two parameters, one of which is coming from a spatially varying field of $\alpha^- : \Omega \mapsto [1, \alpha^+)^C$, i.e.,

$$\tilde{\mathbf{\Lambda}}_{(\alpha^+, \alpha^-[i])} = \begin{bmatrix} \alpha^+ & 0 \\ 0 & \alpha^-[i] \end{bmatrix} \quad (13)$$

While at a point i within a highly directional region, $\alpha^-[i] = 1$; in a flat region, it switches to isotropic behavior: $\alpha^-[i] = \alpha^+$. When it comes to the values between these two end points, as they get closer to one, the sensitivity to the changes towards $\theta[i]$ increases.

As a consequence, the directional patch-based Jacobian operator given in Eq. (11) is extended to its adaptive version (which will be denoted as $\tilde{J}_K^{(\alpha,\theta)}$) that employs $(\tilde{\mathbf{\Pi}}_{(\alpha,\theta)}\nabla f^c)[i] = \tilde{\mathbf{\Lambda}}_{(\alpha^+, \alpha^-[i])} \mathbf{R}_{-\theta[i]}(\nabla f^c)[i]$ to perform directional derivation. Therefore, our ADSTV regularizer is defined as follows:

$$ADSTV(\mathbf{f}) \triangleq \sum_{i=1}^N \|(\tilde{J}_K^{(\alpha,\theta)} \mathbf{f})[i]\|_{S_q} = \|\tilde{J}_K^{(\alpha,\theta)} \mathbf{f}\|_{1,q} \quad (14)$$

4.3. Parameter Estimation

Estimation of the parameters α^- and θ is not a trivial task. In a noisy image, line-like structures are often interrupted by the noise, which results in deviations from the correct direction. When the conventional smoothing filters are applied to suppress noise, the same line-like structures get thicker, and the edges are dislocated. The reconnection of the interrupted lines is concerned in many studies, such as the ones that employ the morphological operators for the curvilinear structures in biomedical images [28, 29], or for the ridges in fingerprint images [30]. Their principle goal is detection rather than restoration. In the proposed method, we focus finding a layout for the orientations rather than a precise field of them. By means of that, the preprocessor does not put much computational overhead.

In our preprocessor, we employ eigenvectors and eigenvalues of the structure tensor to extract the knowledge about the orientations and the dose of anisotropy, in semi-local fashion, in contrast to EADTV. We also employ isotropic TV regularization to circumvent the problem of interrupted lines and to ignore the insignificant deviations. Below, we explain our parameter estimation procedure, which we will refer to as directional parameter estimator (DPE), step by step in detail. Note that our preprocessor employs Sobel operator as the discretization scheme used for the gradient. Also note that, our parameter estimator uses the luminance information of the input image (\mathbf{g}_L), therefore for the vector-valued images, each channel of the θ will be a copy of each other, i.e., $\theta^1 = \theta^2 = \cdots = \theta^C$. The same applies to α^- .

As a measure of anisotropy, we equip a field of *coherence* measures $\mathbf{c} : \Omega \mapsto [0, 1]$ computed at each image point as follows:

$$\mathbf{c}_\sigma[j] = \frac{\lambda_\sigma^+[j] - \lambda_\sigma^-[j]}{\lambda_\sigma^+[j]} \quad (15)$$

where $\lambda_\sigma^\pm[j] = \lambda^\pm((S_K \mathbf{g}_{L\sigma})[j])$ and $j = \{1, 2, \dots, N\}$. Note that in this section, the variables subscripted by σ denote the

pre-smoothed version of \mathbf{g}_L not only with the Gaussian kernel of standard deviation σ , but also with the support $\sigma^2 \times \sigma^2$. Moreover, the kernel $K_{\bar{\sigma}}$ embodied by the structure tensor that we employ here operates $\bar{\sigma}$ larger than $\hat{\sigma}$, again with the support $\bar{\sigma}^2 \times \bar{\sigma}^2$. Eq. (15) is nothing but a measure of the contrast between the directions of the highest and lowest fluctuations. For a pure local orientation ($\lambda_{\sigma}^+ > 0, \lambda_{\sigma}^- = 0$) the coherence is one, while for an isotropic structure ($\lambda_{\sigma}^+ = \lambda_{\sigma}^- > 0$) it is zero. After computing \mathbf{c}_{σ} , by treating it as a corrupted image, we apply TV-based regularization. This way, the fluctuating intensities are suppressed, thus the uncertain responses are refused. The ROF model [1] that serves our purpose is given below:

$$\hat{\kappa}_{\sigma} = \operatorname{argmin}_{\kappa_{\sigma} \in [0,1]^N} \|\mathbf{c}_{\sigma} - \kappa_{\sigma}\|_2^2 + TV(\kappa_{\sigma}) \quad (16)$$

To be able to detect all desired structures, the response of TV regularized coherence $\hat{\kappa}_{\sigma}$ is analyzed at different scales (i.e., $\sigma_k^2 = 2k - 1$, where $k = \{1, 2, \dots, K\}$). Since as the scale parameter σ_k gets larger, the false positive edge responses come into being; our method follows the iterative procedure shown below by giving more credits to the responses obtained at the fine-scales:

$$\kappa^{(k)}[j] = \begin{cases} \kappa^{(k-1)}[j], & \text{if } \hat{\kappa}_{\sigma_k}[j] \leq \kappa^{(k-1)}[j] \\ \frac{1}{2}(\kappa^{(k-1)}[j] + \hat{\kappa}_{\sigma_k}[j]), & \text{otherwise} \end{cases} \quad (17)$$

We suggest using $K = 2$ for the noise levels less than 0.2, while $K = 3$ for the higher ones. If the image at hand is thought to be either highly directional or highly nondirectional, our algorithm further improves the coherence values. It proceeds by enhancing the values higher than the mean in a directional image, while diminishing the values less than the mean in a non-directional one. This procedure is also executed at each scale, i.e.,

$$\phi^{(k)}[j] = \begin{cases} \mathcal{V}(\kappa^{(k)}[j], 1), & \text{if } \gamma_1^{(k)} > 1 \\ \mathcal{V}(\sqrt{\kappa^{(k)}[j]}, -1), & \text{else if } \gamma_1^{(k)} < -1 \\ \kappa^{(k)}[j], & \text{otherwise} \end{cases} \quad (18)$$

Here the function $\mathcal{V}(X, s) = \begin{cases} X^2, & \text{if } sX < s\mu^{(k)} \\ X, & \text{otherwise} \end{cases}$, where $\mu^{(k)}$

is the mean of the TV regularized coherence field, i.e., $\mu^{(k)} = E[\hat{\kappa}^{(k)}]$, and the variable $\gamma_1^{(k)}$ denotes the skewness of the TV regularized coherence histogram at the scale σ_k , i.e., $\gamma_1^{(k)} = \text{Skew}[\kappa^{(k)}]$. If the skewness is greater than 1, the distribution of the coherence values is considered to be highly skewed to the right, which implies that the nondirectional regions predominate the directional ones. If it is less than -1, the distribution is highly left skewed and it can be inferred that the image is highly directional. Note that in our implementation, we used *skewness*(\cdot) function of Matlab.

Finally, our algorithm obtains the entries of α^- , by inverse scaling the entries of $\phi \in [0, 1]^N$ onto the range $[1, \alpha^+]$, so that the regions with higher response to the above mentioned coherence estimation procedure will be regularized by using smaller

$\alpha^-[j]$.

$$\alpha^-[j] = \frac{\alpha^+ - 1}{\max(\phi) - \min(\phi)} (\max(\phi) - \phi[j]) + 1 \quad (19)$$

When it comes to the estimation of θ , we first consider the eigendirections of the structure tensor, computed at a scale with the highest response to Eq. (17), i.e.,

$$\vartheta[j] = \angle \mathbf{v}_{\hat{\sigma}_k}^-[j], \quad \hat{\sigma}_k = \operatorname{argmax}_{\sigma_1 < \dots < \sigma_K} \hat{\kappa}_{\sigma_k}[j] \quad (20)$$

where $\mathbf{v}_{\hat{\sigma}_k}^-[j]$ is the eigenvector corresponding to the smallest eigenvalue $\lambda_{\hat{\sigma}_k}^+[j]$. Next, as was done to the coherence, the estimated angles are TV regularized as follows:

$$\hat{\theta} = \operatorname{argmin}_{\theta \in [0, \pi]^N} \frac{1}{2} \|\vartheta - \theta\|_2^2 + \tau TV(\theta) \quad (21)$$

where we set $\tau = 0.02$ for all our experiments. Setting τ to a larger value may cause deviations from the correct directions due to the contrast loss.

In Figure 1, we present colormaps taken from the different stages of our DPE procedure applied to the exemplary *Zebra* image from Berkeley Segmentation Dataset (BSD) [31], contaminated with the noise level $\sigma_{\eta} = 0.15$. Figure 1 (a) shows the coherence field extracted from the luminance image by using $\bar{\sigma} = \sqrt{11}$. Thanks to the neighborhood-awareness of the structure tensor, the directions of the small-scale structures are reliably distinguished without the need of pre-smoothing. Figure 1 (b) demonstrates the TV regularized coherence field after the application of Eq. (16). One can see that the image points within highly directional 11×11 neighborhood respond with very high coherence values (e.g., the bodies of the zebras), while the response attenuates as the point gets closer to a non-directional region (e.g. the contours of the bodies). This gradual decrease prevents the affects of the sharp transitions from the anisotropic behaviour to the isotropic one. Due to the uncertainties and the larger-scale directional structures that couldn't be extracted, the background is all suppressed. In Figure 1 (c), we show the colormap of the final α^- , obtained after applying one more sequence of Eq. (15), Eq. (16), Eq. (17), and Eq. (18) at the scale with σ_2 , and finally Eq. (19). Additional to the result obtained at the scale σ_1 , one can see the inclusion of some directional structures at the background encoded in orange colors. Eventually, Figure 1 (d) shows the final distribution of the orientations, i.e., $\theta = \hat{\theta}$. Furthermore, in Figure 2, we roughly illustrate the idea behind our two stage framework. The directional descriptors are visualized as an ellipse, whose major and minor axes respectively corresponds to the α^+ and $\alpha^-[i]$ at a point i , and its orientation coincides with the orientation to be penalized.

4.4. Numerical Optimization

According to the objective function given in Eq. (1), the problem to be solved is:

$$\hat{\mathbf{f}} = \operatorname{argmin}_{\mathbf{f} \in \mathcal{C}} \frac{1}{2} \|\mathbf{g} - \mathbf{f}\|_2^2 + \tau \|\tilde{J}_K^{(\alpha, \theta)} \mathbf{f}\|_{1,q} \quad (22)$$

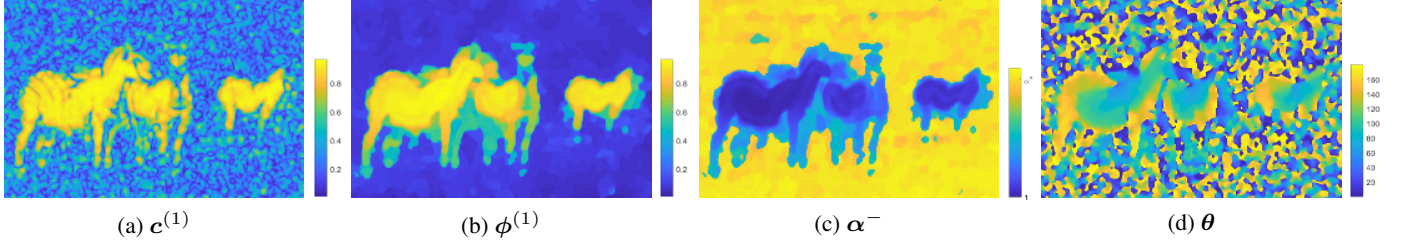


Figure 1: Exemplary outputs of the DPE procedure. The colormaps respectively show (a) the coherence field of unsmoothed \mathbf{g}_L , (b) its TV regularized version, (c) the final weights, and (d) the final directions for the noise level $\sigma_\eta = 0.15$.

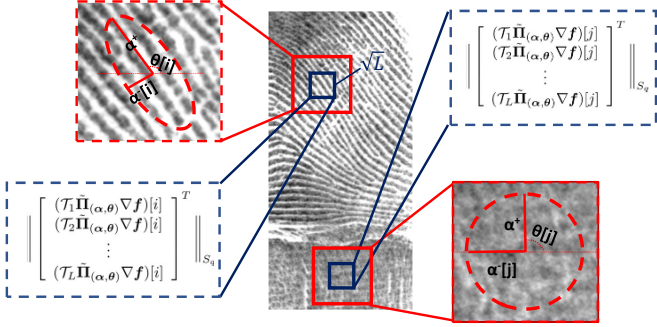


Figure 2: A rough illustration of the DPE procedure vs. ADSTV denoiser.

where \mathcal{C} is a set corresponding to an additional constraint ($\mathcal{C} = \mathbb{R}^{NC}$ in unconstrained case). The solution to this problem corresponds to the proximal map of $ADSTV(\mathbf{f})$, i.e., $\text{prox}_{\tau ADSTV}(\mathbf{g})$. In [27], $\text{prox}_{\tau DSTV}(\mathbf{g})$ was solved through the derivation of dual problem in detail. Having ADSTV instead of DSTV does not change the primal problem, hence the way that we derive dual will remain same. However, DSTV was only considering Frobenius norm ($q = 2$). We define ADSTV in more general form, considering $q \geq 1$, as a direct extension of STV. In this respect, to come up with a concise derivation, using the fact that the dual of the norm $\|\cdot\|_{S_q}$ is $\|\cdot\|_{S_p}$, where $\frac{1}{q} + \frac{1}{p} = 1$, one can redefine ADSTV in terms of the support function of the form:

$$ADSTV(\mathbf{f}) = \sup_{\Psi[i] \in B_{S_p}} \langle \tilde{J}_K^{(\alpha, \theta)} \mathbf{f}, \Psi \rangle \quad (23)$$

by introducing a variable $\Psi \in \mathbb{R}^{N \times LC \times 2}$, and the set $B_{S_p} = \{\mathbf{X} \in \mathbb{R}^{LC \times 2} : \|\mathbf{X}\|_{S_p} \leq 1\}$. Here $\Psi[i]$ refers to the i -th submatrix of Ψ . This paves the way for rewriting the problem in Eq. (22) as a minimax problem:

$$\min_{\mathbf{f} \in \mathcal{C}} \max_{\Psi[i] \in B_{S_p}} \frac{1}{2} \|\mathbf{g} - \mathbf{f}\|_2^2 + \tau \langle \mathbf{f}, \tilde{J}_K^{(\alpha, \theta)*} \Psi \rangle \quad (24)$$

where $\tilde{J}_K^{(\alpha, \theta)*}$ arising after we leave \mathbf{f} alone in the second term denotes the adjoint. The adjoint of the patch-based Jacobian $J_K^* : \mathbb{R}^{N \times LC \times 2} \mapsto \mathbb{R}^{NC}$ was defined in [14] with its derivation:

$$(J_K^* \mathbf{X})[t] = \sum_{l=1}^L -\text{div}(\mathcal{T}_l^* \mathbf{X}[i, s]) \quad (25)$$

where $s = (c-1)L + l$ and $t = (c-1)N + n$ 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 $\mathbf{X}[i, s]$ in column-wise manner, where $\mathbf{X}[i, s] \in \mathbb{R}^2$ is the s -th row of the i -th submatrix of an arbitrary $\mathbf{X} \in \mathbb{R}^{N \times LC \times 2}$. Also, the operator div is discrete divergence, applied by using backward differences, since the gradient is discretized using forward differences, as in [14]. From this point of view, in order to define our $\tilde{J}_K^{(\alpha, \theta)*} : \mathbb{R}^{N \times LC \times 2} \mapsto \mathbb{R}^{NC}$, one should only change the divergence operator in Eq. (25) with the directional divergence operator $\widetilde{\text{div}}_{(\alpha[i], \theta[i])}$, i.e.,

$$(\tilde{J}_K^{(\alpha, \theta)*} \mathbf{X})[t] = \sum_{l=1}^L -\widetilde{\text{div}}_{(\alpha[i], \theta[i])}(\mathcal{T}_l^* \mathbf{X}[i, s]) \quad (26)$$

whose definition is given below:

$$\begin{aligned} \widetilde{\text{div}}_{(\alpha[i], \theta[i])}(\cdot) &\triangleq \text{div}(\tilde{\Pi}_{(\alpha[i], \theta[i])}^T(\cdot)) \\ &= \text{div}(\mathbf{R}_{\theta[i]} \hat{\Lambda}_{(\alpha^+, \alpha^+)}(\cdot)) \end{aligned} \quad (27)$$

Let us call the objective function in Eq. (24) as $\mathcal{L}(\mathbf{f}, \Psi)$. It is convex w.r.t. \mathbf{f} and concave w.r.t. Ψ . By following the fast gradient projection (FGP) method [32], which combines the dual approach introduced in [33] to solve TV-based denoising, and the fast iterative shrinkage-thresholding (FISTA) [34] to accelerate convergence, one can swap min and max, i.e.,

$$\min_{\mathbf{f} \in \mathcal{C}} \max_{\Psi[i] \in B_{S_p}} \mathcal{L}(\mathbf{f}, \Psi) = \mathcal{L}(\hat{\mathbf{f}}, \hat{\Psi}) = \max_{\Psi[i] \in B_{S_p}} \min_{\mathbf{f} \in \mathcal{C}} \mathcal{L}(\mathbf{f}, \Psi) \quad (28)$$

since the common saddle point is not affected. Maximization of the dual problem $d(\Psi) = \min_{\mathbf{f} \in \mathcal{C}} \mathcal{L}(\mathbf{f}, \Psi)$ at the right-hand side is same with the minimization of the primal problem $P(\mathbf{f}) = \max_{\Psi[i] \in B_{S_p}} \mathcal{L}(\mathbf{f}, \Psi)$ at the left-hand side. Therefore, finding the maximizer $\hat{\Psi}$ of $d(\Psi)$ serves to find the minimizer $\hat{\mathbf{f}}$ of $P(\Psi)$. One can rewrite $\hat{\mathbf{f}}$ in terms of Ψ , as:

$$\hat{\mathbf{f}} = \underset{\mathbf{f} \in \mathcal{C}}{\text{argmin}} \|\mathbf{f} - (\mathbf{g} - \tau \tilde{J}_K^{(\alpha, \theta)*} \Psi)\|_2^2 - M \quad (29)$$

by expanding $\mathcal{L}(\mathbf{f}, \Psi)$ and collecting the constants under the term M . The solution to Eq. (29) is:

$$\hat{\mathbf{f}} = P_C(\mathbf{g} - \tau \tilde{J}_K^{(\alpha, \theta)*} \Psi) \quad (30)$$

where P_C is the orthogonal projection onto the set \mathcal{C} . Then we

Algorithm 1 Algorithm for ADSTV-based denoising

INPUT: \mathbf{g} , $\alpha^+ > 1$, $\tau > 0$, $p > 1$, P_C
INIT: $\Psi^{(0)} = \mathbf{0} \in \mathbb{R}^{N \times LC \times 2}$, $t^{(1)} = 1$, $i = 1$
 $(\theta, \alpha^-) \leftarrow \text{DPE}(\mathbf{g}, \alpha^+)$ [Section 4.3]
 $\mathbf{L} \leftarrow 8\sqrt{2}\tau((\alpha^+)^2 + (\alpha^-)^2)$
while stopping criterion is not satisfied **do**
 $\mathbf{z} \leftarrow P_C(\mathbf{g} - \tau\tilde{J}_K^{(\alpha, \theta)^*} \Psi^{(i-1)})$
 $\Psi^{(i)} \leftarrow P_{B_{S_p}}(\Psi^{(i-1)} + \mathbf{L}^{\circ-1} \odot (\tilde{J}_K^{(\alpha, \theta)} \mathbf{z}))$
 $(\Psi^{(i+1)}, t^{(i+1)}) \leftarrow \text{FISTA}(t^{(i)}, \Psi^{(i)}, \Psi^{(i-1)})$
 $i \leftarrow i + 1$
end while
return $P_C(\mathbf{g}^{(k)} - \tau\tilde{J}_K^{(\alpha, \theta)^*} \Psi^{(j)})$

function FISTA ($t, \mathbf{f}^{(\text{cur})}, \mathbf{f}^{(\text{prev})}$) [34]
 $t^{(\text{next})} \leftarrow (1 + \sqrt{1 + 4t^2})/2$
 $\mathbf{f}^{(\text{next})} \leftarrow \mathbf{f}^{(\text{cur})} + (\frac{t-1}{t^{(\text{next})}})(\mathbf{f}^{(\text{cur})} - \mathbf{f}^{(\text{prev})})$
 return $(\mathbf{f}^{(\text{next})}, t^{(\text{next})})$
end

proceed by plugging $\hat{\mathbf{f}}$ in $\mathcal{L}(\mathbf{f}, \Psi)$ to get $d(\Psi) = \mathcal{L}(\hat{\mathbf{f}}, \Psi)$, i.e.,

$$d(\Psi) = \frac{1}{2}\|\mathbf{w} - P_C(\mathbf{w})\|_2^2 + \frac{1}{2}(\|\mathbf{g}\|_2^2 - \|\mathbf{w}\|_2^2) \quad (31)$$

where $\mathbf{w} = \mathbf{g} - \tau\tilde{J}_K^{(\alpha, \theta)^*} \Psi$. Dual problem is smooth with a well-defined gradient computed:

$$\nabla d(\Psi) = \tau\tilde{J}_K^{(\alpha, \theta)} P_C(\mathbf{g} - \tau\tilde{J}_K^{(\alpha, \theta)^*} \Psi) \quad (32)$$

based on the derivation in Lemma 4.1 of [32]. Therefore, in order to find the maximizer $\hat{\Psi}$ of $d(\Psi)$, one can employ the projected gradients [33]. It performs decoupled sequences of gradient descent/ascent and projection onto a set. For the gradient ascent in our case, an appropriate step size that ensures the convergence needs to be selected. For the gradient in Eq. (32), a $LC \times 2$ matrix filled by a constant step size $1/\mathbf{L}[i]$ can be used for each image point, where $\mathbf{L} \in \mathbb{R}^{N \times LC \times 2}$ is a multidimensional array of Lipschitz constants each of which having the upper bound $\mathbf{L}[i] \leq 8\sqrt{2}\tau^2((\alpha^+)^2 + (\alpha^-)^2)$. One can see Appendix A for the derivation.

When it comes to the projection, as an efficient way, the authors of [14] reduced the projection of each $\Psi[i]$ onto B_{S_p} to the projection of the singular values onto the ℓ_p unit ball $B_p = \{\boldsymbol{\sigma} \in \mathbb{R}_+^2 : \|\boldsymbol{\sigma}\|_p \leq 1\}$. Therefore, for the SVD $\Psi[i] = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T$ and $\boldsymbol{\Sigma} = \text{diag}(\boldsymbol{\sigma})$, where $\boldsymbol{\sigma} = [\sigma_1, \sigma_2]$, the projection was defined as:

$$P_{B_{S_p}}(\Psi[i]) = \Psi[i]\mathbf{V}\boldsymbol{\Sigma}^+\boldsymbol{\Sigma}_p\mathbf{V}^T \quad (33)$$

where $\boldsymbol{\Sigma}^+$ is pseudoinverse of $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}_p = \text{diag}(\boldsymbol{\sigma}_p)$ for $\boldsymbol{\sigma}_p$ are the projected singular values of $\boldsymbol{\Sigma}$. The readers are referred to [14] for the details about the derivation and the realization. As a consequence, the overall algorithm is shown in Algorithm 1.

5. Experiments

In this section, we assess qualitative and quantitative performances of the proposed variational denoising framework. We compare it with the other related analysis-based regularizers: TV (as a baseline), EADTV (as a representative of previous attempts to make DTV [9] adaptive), STV [14] (as a predecessor of ADSTV), and NCDR [22] (as a recent analysis-based regularization scheme depending on a new paradigm). We don't include TGV [6] in the competing algorithms since, STV's superiority over it has already been demonstrated in [14]. Owing to the fact that TV, EADTV, and NCDR are only applicable to the scalar-valued images, they are only involved in the grayscale environment. Thus, the experiments on the vector-valued images merely compare STV and DSTV regularizers. The source codes of STV² and NCDR³ that we use were made publicly available by the authors on GitHub. Our ADSTV is implemented on top of STV, while TV and EADTV are written from scratch. 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.

To assess the quantitative performances of the methods, we use peak signal-to-noise ratio (PSNR) in dB and structural similarity index (SSIM). The experiments are performed on the images shown in Figure 3. The first row shows the thumbnails of seven classical grayscale test images of sizes 256×256 (*Monarch*, *Shells*, and *Feather*) and 512×512 (the others). In the other two rows, we have RGB color images. *Plant*, *Spiral*, *Rope* and *Corns* of sizes 256×256 are more textural with different directional characteristics. *Rope* and *Corns* are taken from the popular Berkeley Segmentation Dataset (BSD) [31], while the others are public domain images. The rest of the color images are natural images taken from BSD. They have either no or limited amount of directional parts. Note that, the intensities of all test images have been normalized to the range [0, 1].

The algorithms under consideration are all aiming to minimize Eq. (1). Therefore, the critical regularization parameter τ is in common, and should be chosen precisely. For a fair comparison, we fine-tuned τ such that it leads to the best PSNR in all experiments. For STV, ADSTV, and NCDR priors; one should also choose the convolution kernel. We used a Gaussian kernel of support 3×3 for all three algorithms. The standard deviations (σ) were set to 0.5 for STV and DSTV, while $\sigma = 1$ was used for NCDR, as suggested in the algorithms' respective papers: [14] and [22]. Both STV and ADSTV use the nuclear norm of the patch-based Jacobian, since it was chosen as the best performing norm. Similar to the convolution kernel, all the other parameters of NCDR were kept same as they were used in [22]. The EADTV's α is fine-tuned by changing it from 2 (since it reduces to TV when $\alpha = 1$) to 30 with 1-unit intervals. This is also the case for the ADSTV's α^+ , unless otherwise stated. DSTV requires an additional convolution kernel

²https://github.com/cig-skoltech/Structure_Tensor_Total_Variation

³<https://github.com/HUST-Tan/NCDR>

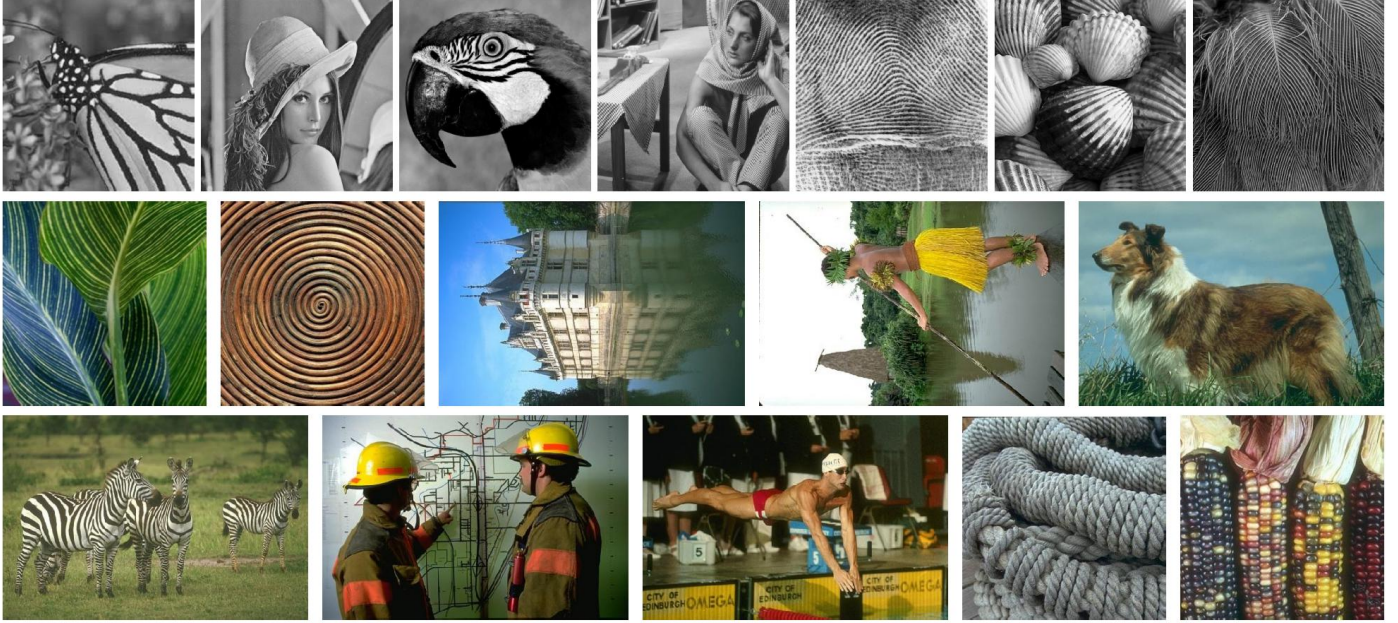


Figure 3: Thumbnails of the grayscale and color images used in the experiments. From left to right and top to bottom: *Monarch*, *Lena*, *Parrot*, *Barbara*, *Fingerprint*, *Shells*, *Feather*, *Plant*, *Spiral*, *Chateau*, *Indigenous*, *Dog*, *Zebra*, *Workers*, *Swimmer*, *Rope*, and *Corns*.

too, to be used by the preprocessor DPE. We fixed it as a Gaussian kernel with variance and support $\bar{\sigma}^2 = 7$ for 256×256 images, while $\bar{\sigma}^2 = 11$ for 963×481 natural images from BSD, and $\bar{\sigma}^2 = 15$ for 512×512 images. For the optimization of all TV-based algorithms, the stopping criterion was either having a relative normed difference between two successive estimates that reaches a threshold of 10^{-5} , or fulfilling a number of iterations, equals to 100 in our case. When it comes to NCDR, since it uses gradient descent to solve its non-convex optimization problem, it is quite hard to ensure that whether a reported result is due to a local or a global minimum. The number of iterations for NCDR was experimentally set to 500.

We consider additive i.i.d. Gaussian noise with five noise levels of $\sigma_\eta = \{0.05, 0.1, 0.15, 0.2, 0.25\}$. In Table 1, we report PSNR and SSIM measures obtained by using TV, EADTV, STV, NCDR, and ADSTV priors applied to the grayscale images. From the results, we observe that the EADTV does not offer significant improvements over TV in the images with less directional patches such as *Lena* and *Parrot*. But, it even performs worse than TV for *Barbara* and *Feather* images, in which highly directional parts present, as the noise level increases. This is due to the fact that, those directional parts involve very high frequency components. Thus, even small deviations in the direction estimation extremely affects the result. In almost all of the experiments, EADTV preferred to use $\alpha = 2$ (except *Fingerprint* for which larger α values are selected), and this is also due to the need of compensating the sheer amount of mistakes made by its direction estimator. STV produced significantly better results when compared to TV and EADTV for all cases and images, except *Fingerprint*, where for all the noise levels $\sigma_\eta > 0.1$, EADTV outperformed STV. When it comes to NCDR, it yielded superior PSNR results to all competing algorithms including ADSTV, except the experiments on *Parrot*

and *Shell*. However, as one may see in the last column, these results are obtained at the cost of computational load. The average runtimes of NCDR at all noise levels are pretty high. On the other hand, when the quality is measured by SSIM, our ADSTV systematically outperformed NCDR (and the others). Figure 4 is provided for the visual judgement. It can be inferred that NCDR smooths the structural regions better than ADSTV at the risk of loss of details (See (d)-(e), (i)-(j), and the scarf of *Barbara* in (s)-(t)). NCDR also causes artifacts in the junctions more apparent than those of ADSTV (e.g., (d)-(e)). Also the ADSTV is better at smoothing flat region as one can see on *Lena*'s face in (n)-(o). These may clarify why the SSIM values obtained by ADSTV are higher.

Table 2 on the other side, shows PSNR/SSIM values obtained by the application of STV and ADSTV to vector-valued images. According to the results, ADSTV systematically outperformed STV in terms of PSNR and SSIM measures, even in the images that do not exhibit directional dominance. However, in terms of the runtimes that are reported in the last row, ADSTV seems to bring almost twice as much load to the computation. Again in Figure 5, we demonstrate exemplary detail patches cropped from the original images (first column), noisy versions (second column), and the restored versions by STV (third column) and ADSTV (last column). The results obtained by STV have oil-painting-like artifacts, whereas this effect is far less visible in ADSTV's results. With the incorporation of the directional information, the edges became more apparent, (see *Rope* and *Chateau*), more smoothing towards the desired direction (see *Spiral*), and the gaps between closely parallel lines could better be distinguished (e.g. the feather in *Dog* image, the details of the leaves in *Corns* image, and skirt in *Indigenous* image).

One of the ADSTV's disadvantages is introducing an extra free parameter α^+ that needs to be tuned. In Figure 6, we

Table 1: PSNR/SSIM Comparison of TV, EADTV, STV, NCDR, and ADSTV on grayscale images. Last column shows the average runtimes.

		Monarch	Lena	Parrot	Barbara	Fingerp.	Shell	Feather	Avg.	t (sec.)
$\sigma_\eta = 0.05$	TV	31.54/0.91	32.88/0.87	33.17/0.88	29.48/0.84	29.14/0.93	28.69/0.89	26.79/0.93	30.24/0.89	1.51
	EADTV	31.91/0.92	33.09/0.87	33.16/0.88	29.59/0.83	29.73/0.94	28.81/0.89	26.86/0.93	30.45/0.89	1.82
	STV	32.32/ 0.93	33.59/0.88	33.56/0.90	30.19/0.88	30.11/0.94	29.01/0.89	27.16/0.94	30.85/0.91	17.41
	NCDR	32.66/0.93	34.26/0.89	33.40/0.90	31.44/0.89	30.08/0.94	29.24/0.89	28.47/0.94	31.36/0.91	276.98
	ADSTV	32.61/ 0.93	34.03/ 0.89	33.77/0.91	31.09/ 0.89	30.09/0.94	29.36/0.90	27.89/0.95	31.25/0.92	38.24
$\sigma_\eta = 0.1$	TV	27.88/0.85	29.86/0.81	29.74/0.82	25.57/0.71	25.33/0.86	24.94/0.77	21.95/0.81	26.47/0.80	2.08
	EADTV	28.31/0.84	30.07/0.81	29.81/0.81	25.51/0.70	25.83/0.87	24.45/0.76	21.65/0.81	26.52/0.80	2.77
	STV	28.65/0.87	30.54/0.83	30.22/0.84	26.45/0.77	26.21/0.87	25.34/0.79	22.40/0.83	27.12/0.83	16.92
	NCDR	29.29/0.88	31.36/0.84	30.31/ 0.85	27.56/0.80	26.54/0.88	25.74/0.79	24.32/0.84	27.88/0.84	256.77
	ADSTV	29.07/ 0.88	31.15/ 0.84	30.47/0.85	27.34/ 0.81	26.48/ 0.88	25.83/0.81	23.60/0.86	27.70/0.85	34.59
$\sigma_\eta = 0.15$	TV	25.84/0.80	28.23/0.77	27.91/0.77	23.88/0.64	23.30/0.79	23.06/0.68	19.58/0.67	24.54/0.73	2.28
	EADTV	26.24/0.80	28.38/0.77	28.02/0.77	23.78/0.63	24.22/0.82	23.22/0.69	19.03/0.67	24.70/0.74	3.07
	STV	26.58/0.82	28.85/0.79	28.43/0.79	24.47/0.69	24.20/0.81	23.48/0.71	20.04/0.70	25.15/0.76	16.77
	NCDR	27.34/0.84	29.61/0.80	28.69/0.80	25.47/0.72	24.64/0.82	24.00/0.70	22.00/0.72	25.96/0.77	270.70
	ADSTV	27.02/ 0.84	29.45/ 0.81	28.73/0.81	25.40/ 0.74	24.46/0.81	24.04/0.73	21.30/0.76	25.76/0.79	31.63
$\sigma_\eta = 0.2$	TV	24.43/0.75	27.14/0.74	26.68/0.73	23.04/0.61	21.94/0.72	21.87/0.61	18.21/0.54	23.33/0.67	2.78
	EADTV	24.78/0.76	27.22/0.74	26.77/0.73	22.96/0.60	22.80/0.77	22.00/0.62	18.12/0.54	23.50/0.68	2.58
	STV	25.13/0.78	27.69/0.76	27.22/0.76	23.45/0.65	22.68/0.76	22.27/0.64	18.65/0.58	23.87/0.70	16.21
	NCDR	25.92/0.80	28.35/0.78	27.56/0.77	24.21/0.66	23.28/0.77	22.81/0.63	20.35/0.60	24.64/0.72	269.34
	ADSTV	25.61/ 0.80	28.29/ 0.78	27.59/0.78	23.98/ 0.67	23.12/ 0.77	22.79/0.66	19.66/0.65	24.43/0.73	36.77
$\sigma_\eta = 0.25$	TV	23.35/0.72	26.33/0.71	25.75/0.71	22.52/0.58	20.93/0.67	21.02/0.55	17.38/0.42	22.47/0.62	2.76
	EADTV	23.64/0.72	26.35/0.72	25.82/0.71	22.48/0.58	21.68/0.72	21.13/0.57	17.23/0.41	22.62/0.63	2.82
	STV	24.01/0.75	26.82/0.74	26.30/0.74	22.90/0.61	21.58/0.70	21.40/0.59	17.75/0.47	22.97/0.66	16.60
	NCDR	24.80/0.77	27.37/0.75	26.65/0.75	23.33/0.62	22.16/0.71	21.88/0.57	19.11/0.48	23.61/0.66	274.48
	ADSTV	24.53/ 0.77	27.37/0.76	26.78/0.76	23.22/ 0.63	22.11/ 0.71	21.90/0.61	18.54/0.54	23.49/0.68	36.70

Table 2: PSNR/SSIM Comparison of STV and ADSTV on color image denoising. Last row shows the average runtimes.

	$\sigma_\eta = 0.05$		$\sigma_\eta = 0.1$		$\sigma_\eta = 0.15$		$\sigma_\eta = 0.2$		$\sigma_\eta = 0.25$	
	STV	ADSTV	STV	ADSTV	STV	ADSTV	STV	ADSTV	STV	ADSTV
Plant	29.23/0.96	30.32/0.97	25.27/0.90	26.63/0.93	23.28/0.85	24.60/0.89	22.05/0.80	23.24/0.85	21.21/0.76	22.27/0.81
Spiral	27.82/0.97	28.39/0.98	24.02/0.93	24.76/0.94	22.04/0.89	23.06/0.91	20.73/0.86	21.95/0.89	19.74/0.83	21.08/0.87
Chateau	32.43/0.95	32.74/0.96	28.83/0.91	29.39/0.92	26.93/0.87	27.63/0.89	25.75/0.85	26.46/0.87	24.88/0.83	25.51/0.85
Indigenous	32.70/0.95	33.04/0.96	29.33/0.90	29.74/0.92	27.71/0.88	28.07/0.89	26.70/0.85	26.99/0.87	25.99/0.84	26.25/0.85
Dog	32.19/0.95	32.53/0.96	29.00/0.90	29.40/0.91	27.44/0.87	27.80/0.88	26.46/0.84	26.76/0.85	25.78/0.82	26.05/0.83
Zebra	30.70/ 0.97	30.87/0.97	26.86/0.93	27.46/0.94	24.84/0.90	25.58/0.91	23.46/0.86	24.26/0.89	22.47/0.84	23.34/0.87
Workers	31.92/0.95	32.41/0.96	28.06/0.90	28.77/0.92	25.96/0.85	26.77/0.89	24.57/0.81	25.39/0.85	23.56/0.77	24.37/0.82
Swimmer	33.63/0.97	33.92/0.98	30.17/ 0.95	30.57/0.95	28.29/0.92	28.73/0.93	27.03/0.90	27.43/0.92	26.09/0.88	26.50/0.90
Rope	28.37/ 0.91	28.53/0.91	24.68/0.80	24.96/0.81	22.87/0.70	23.18/0.72	21.73/0.61	22.06/0.63	20.93/0.55	21.23/0.58
Corns	29.74/ 0.95	29.90/0.95	25.94/0.89	26.16/0.90	23.92/0.83	24.06/0.84	22.59/0.78	22.74/0.79	21.63/ 0.74	21.77/0.74
Avg.	30.87/0.95	31.27/0.96	27.22/0.90	27.78/0.91	25.33/0.86	25.95/0.88	24.11/0.82	24.73/0.84	23.23/0.79	23.84/0.81
t (sec.)	20.62	40.18	21.32	33.81	18.97	33.39	21.12	36.27	22.50	35.54

present two graphs to show how sensitive the gain over STV by using our method with respect to α^+ is. Figure 6 (a) serves the purpose by testing varying α^+ values on five different grayscale and color images at the noise level $\sigma_\eta = 0.15$, while Figure 6 (b) examines them on *Barbara* image at five different noise levels. As one can infer from the results, even for the incorrect selections of α^+ , the gain has almost always been above zero, which shows the robustness of our approach. Yet a relevant research question here might be if it is possible to calculate α^+ automatically, as a function of an estimated noise level by using variance estimators and coherence.

6. Discussion

In the previous section, we compared the proposed denoising framework with TV [1], EADTV [9, 10], STV [14], and NCDR [22] based denoising models. STV was holding the state-of-the-art records among the other TV-related local analysis-based priors. NCDR came up with a new paradigm regarding the sparsity of the corner points, in 2018. Although the NCDR puts plausible results out, its non-convexity and heuristic nature make it disadvantageous. Moreover, it's only applicable to the scalar-valued images. The qualitative results and SSIM measurements show the effectiveness of our ADSTV over NCDR in terms of

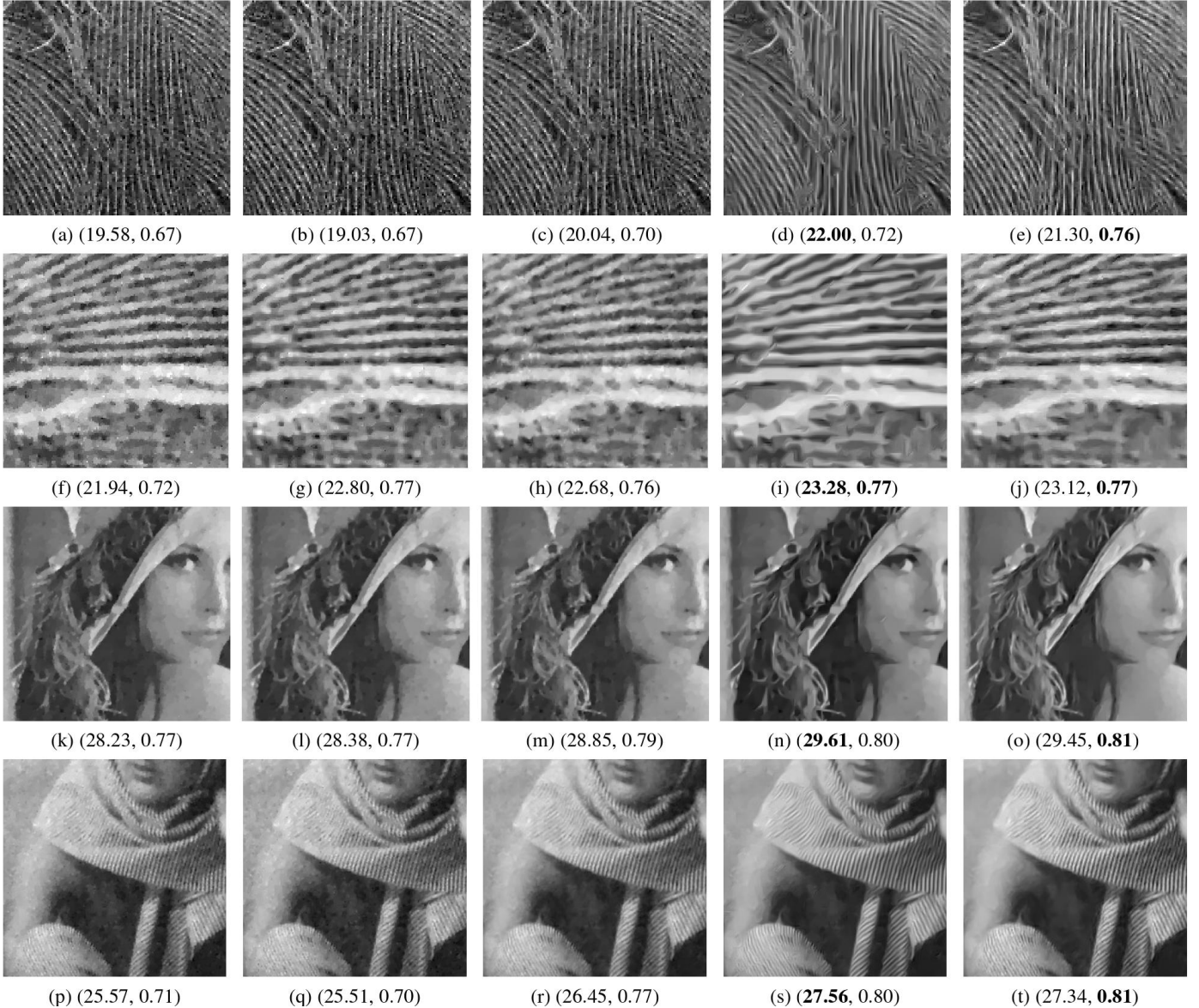


Figure 4: The detail patches showing grayscale image denoising results. The noisy *Barbara* ($\sigma_\eta = 0.1$), *Feather*, *Lena* ($\sigma_\eta = 0.15$), and *Fingerprint* ($\sigma_\eta = 0.2$) images are restored by using TV (col-1), EADTV (col-2), STV (col-3), NCDR (col-4), and ADSTV (col-5) regularizers. The quantity pairs shown at the bottom of each image are corresponding to the (PSNR, SSIM) values.

restoration quality, while NCDR results in better PSNR. But when we consider the computational performances, NCDR machinery is around 7 times slower than ADSTV. We believe that our framework provides a good balance between the restoration quality and the computational performance.

On the other hand, as mentioned in Sect. 1 and Sect. 3.1, a closely related paper [11] was published as preprint (to be published in 2020) at the moment that we were studying on this paper. That work couldn't be involved in our experiments but, due to the common experiment performed on the same *Lena* image, we think it is worth talking over it here. The authors reported that the restored versions of *Lena* contaminated by Gaussian noise with $\sigma = \{0.05, 0.1, 0.15\}$ has the signal to noise ratio (SNR) around 17.5, 15 and, 13, respectively; whereas the corresponding results of ADSTV are 19.77, 16.84, and 15.07.

We think this difference is due to the incorporation of structure tensor.

Finally, in order not to finish discussion without referring to today's learning based approaches on restoration, we want to touch upon briefly how close our results to those obtained by the popular feed-forward denoising convolutional neural networks (DnCNN) [35]. On the same *Parrot*, *Lena*, and *Barbara* images of the range $[0, 255]$, the authors of DnCNN experimented three noise levels $\{15, 25, 50\}$ and reported the PSNR scores. When we perform the same experiments on the *Parrot* image, we observed that the results of ADSTV were almost 1 dB better than DnCNN. However on *Lena* and *Barbara* images it was not the case, and DnCNN performed around 1 dB and 1.5 dB better than ADSTV, respectively, as might be expected. Similar to the classification tasks, data-driven meth-

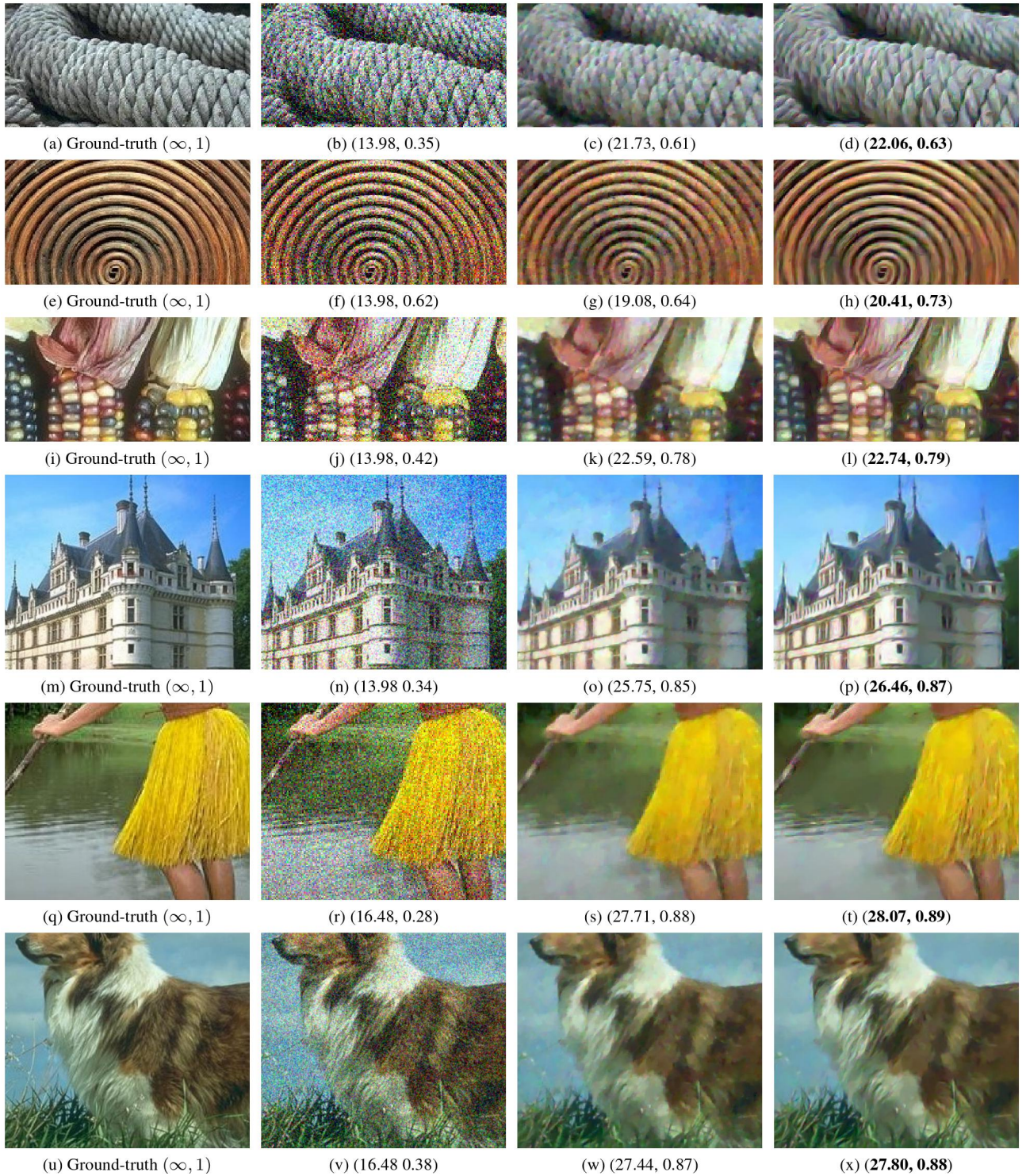


Figure 5: The detail patches showing color image denoising results. The noisy *Indigenous*, *Dog* ($\sigma_\eta = 0.15$), and *Rope*, *Spiral*, *Corns*, *Chateau*, *Indigenous* ($\sigma_\eta = 0.2$) images (col-2) are restored by using STV (col-3) and ADSTV (col-4) regularizers. The quantity pairs shown at the bottom of each image are corresponding to the (PSNR, SSIM) values.

ods seems to predominate the model-driven ones in restoration tasks. They can themselves capture the underlying geometry within the data. Moreover, they can run very fast once trained. However, beside those advantages, they also have limitations such as the lack of the suitable ground truth images that would

be used for training purpose, difficulty of incorporating prior domain knowledge, and the strong dependency to the imaging parameters (e.g. noise level) of the training data. To tackle with these limitations, new efforts try to involve the variational models in the learning-based methods [22, 36, 37]. Also as

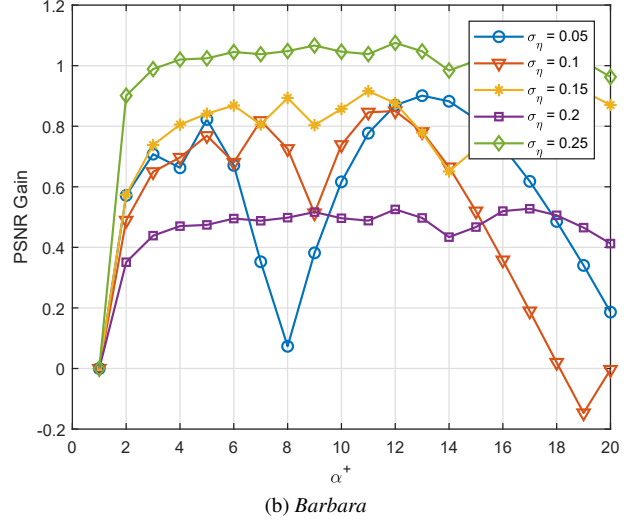
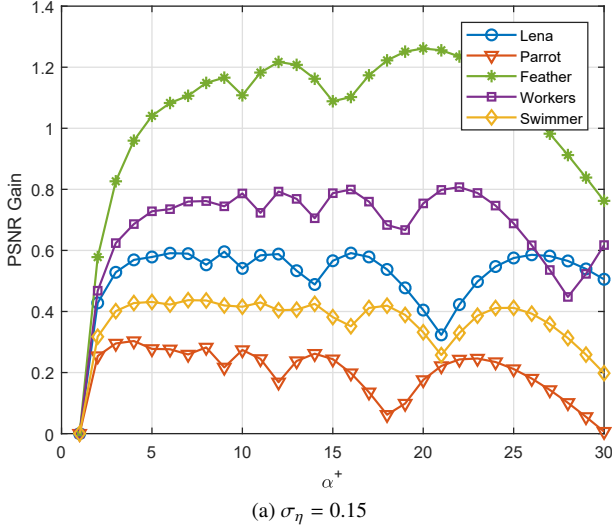


Figure 6: PSNR gain over STV with respect to α^+ .

remarked in [38], the learning-based approaches comprise the variational approaches as special cases. These facts and the non-negligible drawbacks motivate the researchers to produce in both fields with the hope that their achievements might further be combined.

7. Conclusion

In this study, we've described a two-stage variational framework for denoising scalar- and vector-valued images. First stage is the preprocessor that estimates some directional parameters that will guide to the second stage, and the second stage is the one that performs denoising. Our preprocessor attempts to circumvent the chicken-and-egg problem of capturing local geometry from a noisy image to be used while denoise it, by employing the eigenvalues and the orthogonal eigenvectors of the structure tensor, and making assumptions on the piecewise-constancy of the underlying directional parameters. When it comes to the denoising stage, it encodes the problem as a convex optimization problem that involves the proposed adaptive direction-guided structure tensor total variation (ADSTV) as regularizer. Our ADSTV is designed by extending the structure tensor total variation (STV), so that it works biased in a certain direction varying at each point in the image domain. This approach results in improved restoration quality that has been verified by extensive experiments. We compared our method with the related denoising models involving various analysis-based priors.

We believe that the ADSTV regularizer can be used in a large spectrum of inverse imaging problems, even though the effectiveness of our preprocessor may change from problem to problem. Furthermore, there might be appropriate application fields where the source image that is used to estimate the directional parameters differs from the target image to be restored (e.g., video restoration where the source and target frames are different, pansharpening where the chromatic image can be used as

a source for the parameters). We plan to pursue this issue as future work.

Appendix A

In order to find an upper bound for Lipschitz constants, we follow the derivation in [32] and adapt it to our formulation.

$$\begin{aligned}
\|\nabla d(\mathbf{u}) - \nabla d(\mathbf{v})\| &= \tau \|\tilde{J}_K^{(\alpha, \theta)} P_C(\mathbf{g} - \tau \tilde{J}_K^{(\alpha, \theta)*} \mathbf{u}) - \tilde{J}_K^{(\alpha, \theta)} P_C(\mathbf{g} - \tau \tilde{J}_K^{(\alpha, \theta)*} \mathbf{v})\| \\
&\leq \tau \|\tilde{J}_K^{(\alpha, \theta)}\| \|P_C(\mathbf{g} - \tau \tilde{J}_K^{(\alpha, \theta)*} \mathbf{u}) - P_C(\mathbf{g} - \tau \tilde{J}_K^{(\alpha, \theta)*} \mathbf{v})\| \\
&\leq \tau \|\tilde{J}_K^{(\alpha, \theta)}\| \|\tau \tilde{J}_K^{(\alpha, \theta)*}(\mathbf{u} - \mathbf{v})\| \\
&\leq \tau^2 \|\tilde{\Gamma}_{(\alpha, \theta)}\|^2 \|J_K\|^2 \|\mathbf{u} - \mathbf{v}\| \\
&\leq \tau^2 \|\tilde{\Gamma}_{(\alpha, \theta)}\|^2 \|\nabla\|^2 \|T\|^2 \|\mathbf{u} - \mathbf{v}\|
\end{aligned} \tag{34}$$

where $T = \sum_{l=1}^L (\mathcal{T}_l^* \mathcal{T}_l)$. Knowing from [32] that $\|\nabla\|^2 \leq 8$, from [14] that $\|T\|^2 \leq \sqrt{2}$, and further showing that $\|\tilde{\Gamma}_{(\alpha, \theta)}\|^2 \leq (\alpha^+)^2 + (\alpha^-)^2$, we come up with the spatially varying Lipschitz constants $\mathbf{L}[i] \leq 8 \sqrt{2} \tau^2 ((\alpha^+)^2 + (\alpha^-)^2)$.

Acknowledgements

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions. This work was supported by The Scientific and Technological Research Council of Turkey (TUBITAK) under 115R285.

References

- [1] L. I. Rudin, S. Osher, E. Fatemi, Nonlinear total variation based noise removal algorithms, *Physica D: nonlinear phenomena* 60 (1992) 259–268.
- [2] T. Chan, A. Marquina, P. Mulet, High-order total variation-based image restoration, *SIAM Journal on Scientific Computing* 22 (2000) 503–516.
- [3] S. Esedoglu, S. J. Osher, Decomposition of images by the anisotropic rudin-osher-fatemi model, *Communications on pure and applied mathematics* 57 (2004) 1609–1626.

- [4] R. Chartrand, Exact reconstruction of sparse signals via nonconvex minimization, *IEEE Signal Processing Letters* 14 (2007) 707–710.
- [5] Y. Lou, T. Zeng, S. Osher, J. Xin, A weighted difference of anisotropic and isotropic total variation model for image processing, *SIAM Journal on Imaging Sciences* 8 (2015) 1798–1823.
- [6] K. Bredies, K. Kunisch, T. Pock, Total generalized variation, *SIAM Journal on Imaging Sciences* 3 (2010) 492–526.
- [7] S. Lefkimmatis, J. P. Ward, M. Unser, Hessian schatten-norm regularization for linear inverse problems, *IEEE transactions on image processing* 22 (2013) 1873–1888.
- [8] K. Papafitsoros, C.-B. Schönlieb, A combined first and second order variational approach for image reconstruction, *Journal of mathematical imaging and vision* 48 (2014) 308–338.
- [9] I. Bayram, M. E. Kamasak, Directional total variation, *IEEE Signal Processing Letters* 19 (2012) 781–784.
- [10] H. Zhang, Y. Wang, Edge adaptive directional total variation, *The Journal of Engineering* 2013 (2013) 61–62.
- [11] Z.-F. Pang, H.-L. Zhang, S. Luo, T. Zeng, Image denoising based on the adaptive weighted tvp regularization, *Signal Processing* 167 (2020) 107325.
- [12] L. Hong, Y. Wan, A. Jain, Fingerprint image enhancement: algorithm and performance evaluation, *IEEE transactions on pattern analysis and machine intelligence* 20 (1998) 777–789.
- [13] M. Grasmair, F. Lenzen, Anisotropic total variation filtering, *Applied Mathematics & Optimization* 62 (2010) 323–339.
- [14] S. Lefkimmatis, A. Roussos, P. Maragos, M. Unser, Structure tensor total variation, *SIAM Journal on Imaging Sciences* 8 (2015) 1090–1122.
- [15] P. Blongren, T. F. Chan, Color tv: total variation methods for restoration of vector-valued images, *IEEE transactions on image processing* 7 (1998) 304–309.
- [16] A. Buades, B. Coll, J.-M. Morel, Non-local means denoising, *Image Processing On Line* 1 (2011) 208–212.
- [17] G. Gilboa, S. Osher, Nonlocal operators with applications to image processing, *Multiscale Modeling & Simulation* 7 (2008) 1005–1028.
- [18] S. Lefkimmatis, S. Osher, Nonlocal structure tensor functionals for image regularization, *IEEE Transactions on Computational Imaging* 1 (2015) 16–29.
- [19] A. Chambolle, R. A. De Vore, N.-Y. Lee, B. J. Lucier, Nonlinear wavelet image processing: variational problems, compression, and noise removal through wavelet shrinkage, *IEEE Transactions on Image Processing* 7 (1998) 319–335.
- [20] M. A. Figueiredo, J. M. Bioucas-Dias, R. D. Nowak, Majorization–minimization algorithms for wavelet-based image restoration, *IEEE Transactions on Image processing* 16 (2007) 2980–2991.
- [21] G. Kutyniok, M. Shahram, X. Zhuang, Shearlab: A rational design of a digital parabolic scaling algorithm, *SIAM Journal on Imaging Sciences* 5 (2012) 1291–1332.
- [22] H. Liu, S. Tan, Image regularizations based on the sparsity of corner points, *IEEE Transactions on Image Processing* 28 (2018) 72–87.
- [23] G. Gilboa, N. Sochen, Y. Y. Zeevi, Variational denoising of partly textured images by spatially varying constraints, *IEEE Transactions on Image Processing* 15 (2006) 2281–2289.
- [24] M. Grasmair, Locally adaptive total variation regularization, in: *International Conference on Scale Space and Variational Methods in Computer Vision*, Springer, 2009, pp. 331–342.
- [25] Y. Dong, M. Hintermüller, Multi-scale total variation with automated regularization parameter selection for color image restoration, in: *International Conference on Scale Space and Variational Methods in Computer Vision*, Springer, 2009, pp. 271–281.
- [26] R. Bhatia, *Matrix analysis*, volume 169, Springer Science & Business Media, 2013.
- [27] E. Demircan-Tureyen, M. E. Kamasak, On the direction guidance in structure tensor total variation based denoising, in: *Iberian Conference on Pattern Recognition and Image Analysis*, Springer, 2019, pp. 89–100.
- [28] O. Merveille, B. Naegel, H. Talbot, L. Najman, N. Passat, 2d filtering of curvilinear structures by ranking the orientation responses of path operators (rorpo) (2017).
- [29] O. Merveille, H. Talbot, L. Najman, N. Passat, Curvilinear structure analysis by ranking the orientation responses of path operators, *IEEE transactions on pattern analysis and machine intelligence* 40 (2017) 304–317.
- [30] M. A. Oliveira, N. J. Leite, A multiscale directional operator and morphological tools for reconnecting broken ridges in fingerprint images, *Pattern Recognition* 41 (2008) 367–377.
- [31] P. Arbelaez, M. Maire, C. Fowlkes, J. Malik, Contour detection and hierarchical image segmentation, *IEEE Trans. Pattern Anal. Mach. Intell.* 33 (2011) 898–916. URL: <http://dx.doi.org/10.1109/TPAMI.2010.161>. doi:10.1109/TPAMI.2010.161.
- [32] A. Beck, M. Teboulle, Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems, *IEEE Transactions on Image Processing* 18 (2009) 2419–2434.
- [33] A. Chambolle, An algorithm for total variation minimization and applications, *Journal of Mathematical imaging and vision* 20 (2004) 89–97.
- [34] A. Beck, M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, *SIAM journal on imaging sciences* 2 (2009) 183–202.
- [35] K. Zhang, W. Zuo, Y. Chen, D. Meng, L. Zhang, Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising, *IEEE Transactions on Image Processing* 26 (2017) 3142–3155.
- [36] S. Lefkimmatis, Universal denoising networks: A novel cnn architecture for image denoising, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 3204–3213.
- [37] D. Ulyanov, A. Vedaldi, V. Lempitsky, Deep image prior, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 9446–9454.
- [38] M. T. McCann, K. H. Jin, M. Unser, Convolutional neural networks for inverse problems in imaging: A review, *IEEE Signal Processing Magazine* 34 (2017) 85–95.