VECTORIAL TOTAL VARIATION BASED ON ARRANGED STRUCTURE TENSOR FOR MULTICHANNEL IMAGE RESTORATION

Shunsuke Ono[†], *Keiichiro Shirai*[††], *and Masahiro Okuda*[†††]

[†]Tokyo Institute of Technology, [††]Shinshu University, [†††]The University of Kitakyushu

ABSTRACT

We propose a new regularization function, named as *Arranged Structure tensor Total Variation* (ASTV), for multichannel image restoration. Since the standard structure tensor is a matrix whose eigenvalues well encodes local neighborhood information of an image, there has been proposed vectorial total variation based on the structure tensor for image regularization. However, the correlation among the channels cannot be measured by the structure tensor because the discrete differences of all the channels are just summed up in the entries of the structure tensor. On the other hand, ASTV is based on a newly-defined *arranged structure tensor* that becomes an approximately low-rank matrix when multichannel images have strong correlation among their channels. This suggests that penalizing the nuclear norm of the arranged structure tensor is a reasonable regularization for multichannel images, leading to the definition of ASTV. Experimental results illustrate the advantage of ASTV over a state-of-the-art vectorial total variation based on the structure tensor.

***Index Terms*—** Multichannel image restoration, regularization, structure tensor

1. INTRODUCTION

The restoration of multichannel images, such as color image denoising/deblurring, demosaicking, multispectral/hyperspectral imaging, and compressed sensing, is an important task in many signal processing applications. Such restoration problems are usually ill-posed or ill-conditioned inverse problems, so that one requires some *regularization* based on underlying properties of multichannel images. A successful class of regularization techniques for multichannel images would be vectorial total variation (VTV) [1, 2, 3, 4, 5] and its higher-order/semilocal/nonlocal generalizations [6, 7, 8, 9, 10].

Among them, we focus on the Structure tensor Total Variation (STV) [8][1] because of the following reasons. First, as mentioned in [8], STV exploits *local neighborhood information*, so that it can avoid several drawbacks of VTV such as producing the staircasing effect. Second, since STV is not a nonlocal regularization, STV is free from chicken-and-egg self-similarity evaluation.

As the name indicates, STV is defined as a function of the eigenvalues of the so-called *structure tensor* [11, 12], a matrix whose eigenvalues summarize the prevailing direction of the gradient of an image. The structure tensor has been used in many applications, such as anisotropic diffusion [13], optical flow [14], and corner detection [15]. Specifically, the structure tensor at a pixel location of a multichannel image is a 2×2 matrix constructed from vertical and horizontal differences in the local neighborhood (e.g., 3×3 window) of the pixel location. Thereby, its eigenvalues have a rich information on local spatial variations.

However, for a multichannel image, the *correlation among the channels* cannot be fully evaluated by the structure tensor because the discrete differences of all the channels are just summed up in the entries of the structure tensor (see Sec. 2.1 for details). Since multichannel images usually have strong correlation among their channels, this should be properly incorporated into regularization.

We should remark that several existing VTVs [4, 5] explicitly take the correlation into account. However, the one proposed in [4] is *anisotropic*, i.e., the vertical and horizontal gradients are decoupled, resulting in the generation of blocky artifacts around contours. The one proposed in [5] overcame this drawback but it can be applied only to color images since it uses a color transform. In addition to the above things, structure-tensor-based approaches, which leverage information on local spatial variations, are not considered in [4, 5].

Based on the above discussion, we propose a new vectorial total variation with a newly-defined *arranged structure tensor* for multichannel image restoration, which is termed as *Arranged Structure tensor TV* (ASTV). The arranged structure tensor is a $2M \times 2M$ matrix with $M \in \mathbb{N}$ being the number of channels, so that it has $2M$ eigenvalues. As will be explained in Sec. 2.1, when a multichannel image of interest has strong correlation among its channels, the arranged structure tensor becomes an approximately (but not exactly) low-rank matrix. This observation suggests that penalizing the nuclear norm, the tightest convex relaxation of the rank function [16], of the arranged structure tensor is a reasonable regularization for multichannel images, leading to the definition of ASTV. The advantage of ASTV over STV is demonstrated by experiments on denoising and compressed sensing reconstruction.

2. PROPOSED METHOD

2.1. Arranged structure tensor

Let $\mathbf{u} \in \mathbb{R}^{MN}$ be an image with M channels $\mathbf{u}_1, \ldots, \mathbf{u}_M \in \mathbb{R}^N$ (N is the number of pixels), e.g., $M = 3$ in the case of color images. Note that we treat an image/channel as a vector by stacking its columns on top of one another. Also let $\mathbf{D}_v$ and $\mathbf{D}_h$ be vertical and horizontal discrete difference operators that map one channel in $\mathbb{R}^N$ to its (vectorized) vertical/horizontal gradient map in $\mathbb{R}^N$, respectively. We denote pixel locations by $n \in \{1, \ldots, N\}$, the set of pixel locations in local neighborhood (usually a square window) at the pixel location n by $\mathcal{I}_n$ (NOTE: $n \in \mathcal{I}_n$), and a sub-vector of a given vector $\mathbf{x} \in \mathbb{R}^N$ consisting of its weighted entries at the pixel locations in $\mathcal{I}_n$ by $\mathbf{x}^{\mathbf{w}}_{\mathcal{I}_n} \in \mathbb{R}^{|\mathcal{I}|}$ with the weight vector $\mathbf{w} \in \mathbb{R}^{|\mathcal{I}|}_+$ ($\mathbb{R}_+$ stands for the set of all positive real numbers). Here we assume that the same shape of local neighborhood and the same weight vector are applied to every pixel location, so that the cardinalities (the number of pixels in local neighborhood) of $\mathcal{I}_1, \ldots, \mathcal{I}_N$ are equiva-

The work was supported by JSPS Grants-in-Aid: 15H06197; 15K06076.

[1]Both grayscale and multichannel images are considered in [8], but we are only interested in multichannel cases.

lent, denoted by $|\mathcal{I}|$ (NOTE: To handle local neighborhood around image boundaries, we use periodic boundary extension).

Then, the *arranged structure tensor* of $\mathbf{u}$ at the pixel location n is defined by

$$\mathbf{S}_{\mathbf{u},\mathbf{w}}^{(n)} := \mathbf{L}_{\mathbf{u},\mathbf{w}}^{(n)\top}\mathbf{L}_{\mathbf{u},\mathbf{w}}^{(n)} \in \mathbb{R}^{2M\times 2M}, \tag{1}$$

$$\mathbf{L}_{\mathbf{u},\mathbf{w}}^{(n)} := ([\mathbf{D}_v\mathbf{u}_1]_{\mathcal{I}_n}^{\mathbf{w}}\ [\mathbf{D}_h\mathbf{u}_1]_{\mathcal{I}_n}^{\mathbf{w}} \cdots [\mathbf{D}_v\mathbf{u}_M]_{\mathcal{I}_n}^{\mathbf{w}}\ [\mathbf{D}_h\mathbf{u}_M]_{\mathcal{I}_n}^{\mathbf{w}}). \tag{2}$$

We remark that the arranged structure tensor defined in (1) is different from the standard structure tensor of $\mathbf{u}$ at the pixel location n: $\widetilde{\mathbf{S}}_{\mathbf{u},\mathbf{w}}^{(n)} := \widetilde{\mathbf{L}}_{\mathbf{u},\mathbf{w}}^{(n)\top}\widetilde{\mathbf{L}}_{\mathbf{u},\mathbf{w}}^{(n)} \in \mathbb{R}^{2\times 2}$, where

$$\widetilde{\mathbf{L}}_{\mathbf{u},\mathbf{w}}^{(n)} := \begin{pmatrix} [\mathbf{D}_v\mathbf{u}_1]_{\mathcal{I}_n}^{\mathbf{w}\top} & \cdots & [\mathbf{D}_v\mathbf{u}_M]_{\mathcal{I}_n}^{\mathbf{w}\top} \\ [\mathbf{D}_h\mathbf{u}_1]_{\mathcal{I}_n}^{\mathbf{w}\top} & \cdots & [\mathbf{D}_h\mathbf{u}_M]_{\mathcal{I}_n}^{\mathbf{w}\top} \end{pmatrix}^{\top} \in \mathbb{R}^{M|\mathcal{I}|\times 2}. \tag{3}$$

The difference between the standard and arranged structure tensors are depicted in Fig. 1 (left). Note that there are totally N arranged (standard) structure tensors for one image, i.e., $\mathbf{L}_{\mathbf{u},\mathbf{w}}^{(1)}, \ldots, \mathbf{L}_{\mathbf{u},\mathbf{w}}^{(N)}$.

Since the square root of each eigenvalue of the arranged (or standard) structure tensor equals to each singular value of $\mathbf{L}_{\mathbf{u},\mathbf{w}}^{(n)}$ (or $\widetilde{\mathbf{L}}_{\mathbf{u},\mathbf{w}}^{(n)}$), we can discuss the difference of their properties through $\mathbf{L}_{\mathbf{u},\mathbf{w}}^{(n)}$ and $\widetilde{\mathbf{L}}_{\mathbf{u},\mathbf{w}}^{(n)}$. First, it is clear that every singular value of both $\mathbf{L}_{\mathbf{u},\mathbf{w}}^{(n)}$ and $\widetilde{\mathbf{L}}_{\mathbf{u},\mathbf{w}}^{(n)}$ becomes small if the local neighborhood is smooth, so that essentially, suppressing some norm of $\mathbf{L}_{\mathbf{u},\mathbf{w}}^{(n)}$ or $\widetilde{\mathbf{L}}_{\mathbf{u},\mathbf{w}}^{(n)}$ results in a smoothing effect on $\mathbf{u}$. Indeed, the Frobenius norms of $\mathbf{L}_{\mathbf{u},\mathbf{w}}^{(n)}$ and $\widetilde{\mathbf{L}}_{\mathbf{u},\mathbf{w}}^{(n)}$ take the same value because they consist of the same entries (but their arrangements are different).

Things change when we focus on their singular values from the view of the correlation of channels. We see in (3) that the (vertical/horizontal) discrete differences of all the channels are stacked into one column in $\widetilde{\mathbf{L}}_{\mathbf{u},\mathbf{w}}^{(n)}$, implying that information on the correlation is almost lost in the singular values of $\widetilde{\mathbf{L}}_{\mathbf{u},\mathbf{w}}^{(n)}$. On the other hand, the information is still *alive* in the singular values of $\mathbf{L}_{\mathbf{u},\mathbf{w}}^{(n)}$ because the discrete differences of each channel are arranged horizontally in (2). More specifically, if the channels of $\mathbf{u}$ have strong correlation then the columns of $\mathbf{L}_{\mathbf{u},\mathbf{w}}^{(n)}$ become approximately linearly dependent, so that the singular values except the first one are expected to be very small. This observation naturally leads to the definition of our regularization function in the next subsection.

2.2. Vectorial total variation based on arranged structure tensor

To promote the spatial smoothness of multichannel images with considering the correlation among channels, we propose a regularization function based on the arranged structure tensor as follows:

$$\mathrm{ASTV}_{\mathbf{w}}(\mathbf{u}) := \textstyle\sum_{n=1}^{N} \|\mathbf{L}_{\mathbf{u},\mathbf{w}}^{(n)}\|_*, \tag{4}$$

where $\|\cdot\|_*$ is the nuclear norm, i.e., the sum of all the singular values of $(\cdot)$. Following the prior work [8], we name this function as *Arranged Structure tensor Total Variation* (ASTV). We remark that for single channel images, ASTV and STV (the one proposed in [8]) are equivalent since $\mathbf{L}_{\mathbf{u},\mathbf{w}}^{(n)} = \widetilde{\mathbf{L}}_{\mathbf{u},\mathbf{w}}^{(n)}$.

The set of discussions in the previous subsection suggests the two things: (i) suppressing all the singular values of $\mathbf{L}_{\mathbf{u},\mathbf{w}}^{(n)}$ makes restored images smooth, and (ii) promoting the approximate low-rankness of $\mathbf{L}_{\mathbf{u},\mathbf{w}}^{(n)}$ is suitable for images with strong correlation among channels. Hence, we adopt the nuclear norm for evaluating $\mathbf{L}_{\mathbf{u},\mathbf{w}}^{(n)}$ because the nuclear norm is the sum of the singular values and the tightest convex relaxation of the rank function [16].

2.3. Multichannel image restoration by ASTV

2.3.1. Problem formulation

Consider to restore an original multichannel image $\bar{\mathbf{u}} \in \mathbb{R}^{MN}$ from observation data, which is cast as inverse problems of the form: $\mathbf{v} = \mathcal{D}(\mathbf{\Phi}\bar{\mathbf{u}})$, where $\mathbf{\Phi} \in \mathbb{R}^{R\times MN}$ ($R \leq MN$) is a matrix representing some degradation (e.g., blur and/or random sampling), $\mathcal{D}: \mathbb{R}^R \to \mathbb{R}^R$ is a noise contamination process, and $\mathbf{v} \in \mathbb{R}^R$ is an observation.

Based on the above model, we formulate multichannel image restoration by ASTV as the following convex optimization problem:

$$\min_{\mathbf{u}} \mathrm{ASTV}(\mathbf{u}) + \mathcal{F}_{\mathbf{v}}(\mathbf{\Phi}\mathbf{u}) \ \text{ s.t. } \mathbf{u} \in C, \tag{5}$$

where $\mathcal{F}_{\mathbf{v}} \in \Gamma_0(\mathbb{R}^R)$[2] is a data-fidelity function, and $C \subset \mathbb{R}^{MN}$ is a closed convex constraint on $\mathbf{u}$. We assume that the *proximity operator*[3] [17] of $\mathcal{F}_{\mathbf{v}}$ can be computed efficiently. We also assume that the computation of the projection[4] onto C is efficient.

We give several examples of $\mathcal{F}_{\mathbf{v}}$ in Remark 1. Meanwhile, a typical example of C is a box constraint, a known dynamic range of $\bar{\mathbf{u}}$ (e.g., $C := [0, 255]^{MN}$ for eight-bit images).

Remark 1 (Examples of $\mathcal{F}_{\mathbf{v}}$). The ℓ_2 norm data-fidelity, given by $\mathcal{F}_{\mathbf{v}}(\mathbf{x}) := \frac{\mu}{2}\|\mathbf{x}-\mathbf{v}\|^2$, would be the most popular choice for Gaussian noise cases. The ℓ_1 norm is a useful data-fidelity measure for impulse noise cases, which is given by $\mathcal{F}_{\mathbf{v}}(\mathbf{x}) := \mu\|\mathbf{x}-\mathbf{v}\|_1$. They can also be used as *data-fidelity constraints*, i.e., $\mathcal{F}_{\mathbf{v}}(\mathbf{x}) := \iota_B(\mathbf{x})$, where $B := \{\mathbf{x} \in \mathbb{R}^R |\ \|\mathbf{x}-\mathbf{v}\|_{1 \text{ or } 2} \leq \varepsilon\}$, and ι_B is the *indicator function*[5] of B defined by $\iota_B(\mathbf{x}) := 0$, if $\mathbf{x} \in B$; ∞, otherwise. It is worth noting that such a constraint-type data-fidelity facilitates the parameter setting because ε has a clearer meaning than μ, as addressed in [18, 19, 20]. For Poisson noise cases, the *generalized Kullback-Leibler divergence* is known as a suitable data-fidelity function (the definition can be found in [21]). The proximity operators of these examples can be computed efficiently.

2.3.2. Optimization

Since Prob. (5) is a highly nonsmooth optimization problem, we have to use some iterative algorithms for solving it. In this paper, we adopt a primal-dual splitting method [22], which does not require matrix inversion. It solves convex optimization problems of the form:

$$\min_{\mathbf{x}\in\mathcal{X}} g(\mathbf{x}) + h(\mathbf{A}\mathbf{x}), \tag{6}$$

where $g \in \Gamma_0(\mathcal{X})$ and $h \in \Gamma_0(\mathcal{Y})$ ($\mathcal{X}$ and $\mathcal{Y}$ are some Euclidean spaces) and $\mathbf{A}: \mathcal{X} \to \mathcal{Y}$ is a linear operator.

The algorithm is given by

$$\left\lfloor \begin{array}{l} \mathbf{x}^{(k+1)} = \mathrm{prox}_{\gamma_1 g}(\mathbf{x}^{(k)} - \gamma_1\mathbf{A}^\top\mathbf{y}^{(k)}), \\ \mathbf{y}^{(k+1)} = \mathrm{prox}_{\gamma_2 h^*}(\mathbf{y}^{(k)} + \gamma_2\mathbf{A}(2\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)})), \end{array} \right.$$

where h^* the convex conjugate function[6] of h, and $\gamma_1, \gamma_2 > 0$ satisfy

[2] The set of all proper lower semicontinuous convex functions on $\mathbb{R}^N$ is denoted by $\Gamma_0(\mathbb{R}^N)$.

[3] The proximity operator of index $\gamma > 0$ of $f \in \Gamma_0(\mathbb{R}^N)$ is defined by $\mathrm{prox}_{\gamma f}: \mathbb{R}^N \to \mathbb{R}^N : \mathbf{x} \mapsto \underset{\mathbf{y}}{\mathrm{argmin}}\, f(\mathbf{y}) + \frac{1}{2\gamma}\|\mathbf{y}-\mathbf{x}\|^2$.

[4] Given a nonempty closed convex set $C \subset \mathbb{R}^N$, the *projection onto* C is defined by $P_C: \mathbb{R}^N \to \mathbb{R}^N : \mathbf{x} \mapsto \mathrm{argmin}\, \|\mathbf{x}-\mathbf{y}\|$ s.t. $\mathbf{y} \in C$.

[5] The proximity operator of the indicator function of a nonempty closed convex set C equals to the projection onto C, i.e., $\mathrm{prox}_{\gamma\iota_C} = P_C$.

[6] The proximity operator of f^* can be computed via that of f, i.e., $\mathrm{prox}_{\gamma f^*}(\mathbf{x}) = \mathbf{x} - \gamma\,\mathrm{prox}_{\gamma^{-1}f}(\gamma^{-1}\mathbf{x})$ (see, e.g., [23, Theorem 14.3(ii)].

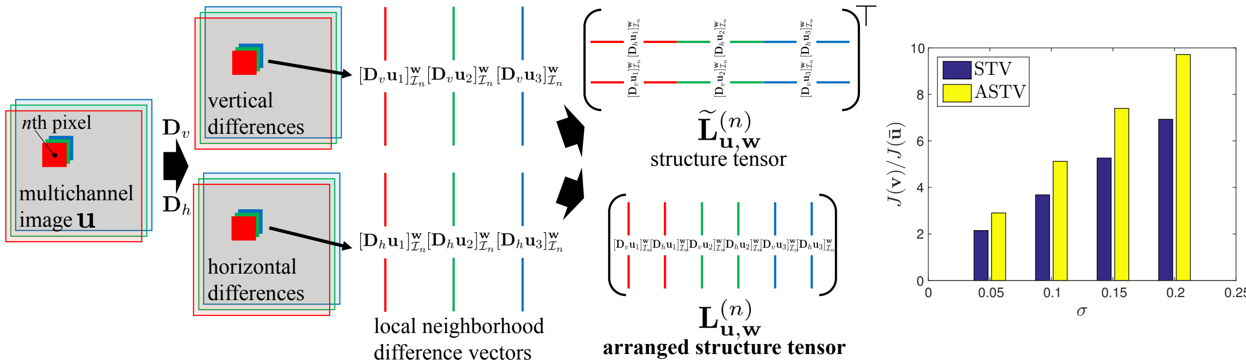

Fig. 1. Construction of the standard and arranged structure tensors (left) and comparison of STV [8] and ASTV (proposed) in terms of the ratio of the function values on noisy and clean images (right).

$\gamma_1\gamma_2\|A\|_{op}^2 \leq 1$ ($\|\cdot\|_{op}$ stands for the operator norm of $\cdot$). Under some mild conditions on g, h, and $\mathbf{A}$, the sequence $(\mathbf{x}^{(k)})_{k\in\mathbb{N}}$ converges to a solution of Prob. (6). To apply the primal-dual splitting method to Prob. (5), we reformulate it into Prob. (6).

First, since the definition of ASTV in (4) is not amenable to optimization due to its structure involving several linear operations to $\mathbf{u}$, we give an alternative expression of ASTV. Specifically, we separate these operations and define them as matrices as follows:

$$\mathrm{ASTV}_{\mathbf{w}}(\mathbf{u}) := \|\mathbf{WPDu}\|_{*,N}. \tag{7}$$

Here, $\mathbf{D} : \mathbb{R}^{MN} \to \mathbb{R}^{2MN}$ is a discrete difference operator that maps all the channels of $\mathbf{u}$ to their vertical and horizontal difference images, $\mathbf{P} : \mathbb{R}^{2MN} \to \mathbb{R}^{2|\mathcal{I}|MN}$ is an *expansion* operator that makes $|\mathcal{I}|$ copies of $\mathbf{Du}$, $\mathbf{W} : \mathbb{R}^{2|\mathcal{I}|MN} \to \mathbb{R}^{2|\mathcal{I}|MN}$ is a weighting operator that applies the weights in $\mathbf{w}$ to local neighborhood at every location, and $\|\cdot\|_{*,N} : \mathbb{R}^{2|\mathcal{I}|MN} \to \mathbb{R}$ is the sum of the nuclear norm of the arranged structure tensor at every location. One sees that in $\mathbf{WPDu}$, local neighborhood at any location does not overlap each other, which means that the arranged structure tensor at every location can be constructed from $\mathbf{WPDu}$ without reusing the entries, i.e., $\mathbf{WPDu}$ and $(\mathbf{L}^{(1)}_{\mathbf{u},\mathbf{w}}), \ldots, (\mathbf{L}^{(1)}_{\mathbf{u},\mathbf{w}})$ are bijective. This makes the proximity operator of $\|\cdot\|_{*,N}$ readily available by computing the proximity operator of the nuclear norm of the arranged structure tensor at each location. Specifically, for $\mathbf{L}^{(n)}_{\mathbf{u},\mathbf{w}}$ with its singular value decomposition $\mathbf{U}^{(n)}\,\mathrm{diag}(\sigma_1^{(n)}, \ldots, \sigma_{2M}^{(n)})\mathbf{V}^{(n)\top}$, the proximity operator of the nuclear norm of the arranged structure tensor at the location n is given by

$$\begin{aligned} &\mathrm{prox}_{\gamma\|\cdot\|_*}(\mathbf{L}^{(n)}_{\mathbf{u},\mathbf{w}}) = \mathbf{U}^{(n)}\boldsymbol{\Sigma}^{(n)}_{\gamma}\mathbf{V}^{(n)\top}, \\ &\boldsymbol{\Sigma}^{(n)}_{\gamma} := \mathrm{diag}(\max\{\sigma_1^{(n)} - \gamma, 0\}, \ldots, \max\{\sigma_{2M}^{(n)} - \gamma, 0\}). \end{aligned} \tag{8}$$

Second, by introducing the indicator function of C, Prob (5) can be rewritten as

$$\min_{\mathbf{u}} \|\mathbf{WPDu}\|_{*,N} + \mathcal{F}_{\mathbf{v}}(\boldsymbol{\Phi}\mathbf{u}) + \iota_C(\mathbf{u}). \tag{9}$$

Finally, by letting

$$\begin{aligned} &g : \mathbb{R}^{MN} \to \mathbb{R}\cup\{\infty\} : \mathbf{u} \mapsto \iota_C(\mathbf{u}), \\ &h : \mathbb{R}^{2|\mathcal{I}|MN+R} \to \mathbb{R}\cup\{\infty\} : (\mathbf{y}_1, \mathbf{y}_2) \mapsto \|\mathbf{y}_1\|_{*,N} + \mathcal{F}_{\mathbf{v}}(\mathbf{y}_2), \\ &\mathbf{A} : \mathbb{R}^{MN} \to \mathbb{R}^{2|\mathcal{I}|MN+R} : \mathbf{u} \mapsto (\mathbf{WPDu}, \boldsymbol{\Phi}\mathbf{u}), \end{aligned}$$

Prob. (9) is reduced to Prob. (6). The resulting algorithm based on the primal-dual splitting method is summarized in Alg. 1, where Step 2 and 6 are computable from the assumptions on Prob. (5), and see (8) for Step 5. We also note that clearly, $\mathbf{D}$, $\mathbf{P}$, $\mathbf{W}$ and thier transposes can be computed efficiently.

Algorithm 1: Primal-dual splitting method for Prob. (4)

input : $\mathbf{u}^{(0)}, \mathbf{y}_1^{(0)}, \mathbf{y}_2^{(0)}$
1 **while** *A stopping criterion is not satisfied* **do**
2 $\quad \mathbf{u}^{(k+1)} = P_C(\mathbf{u}^{(k)} - \gamma_1(\mathbf{D}^\top\mathbf{P}^\top\mathbf{W}^\top\mathbf{y}_1^{(k)} + \boldsymbol{\Phi}^\top\mathbf{y}_2^{(k)}))$;
3 $\quad \mathbf{y}_1^{(k)} \leftarrow \mathbf{y}_1^{(n)} + \gamma_2\mathbf{WPD}(2\mathbf{u}^{(k+1)} - \mathbf{u}^{(k)})$;
4 $\quad \mathbf{y}_2^{(k)} \leftarrow \mathbf{y}_2^{(n)} + \gamma_2\boldsymbol{\Phi}(2\mathbf{u}^{(k+1)} - \mathbf{u}^{(k)})$;
5 $\quad \mathbf{y}_1^{(k+1)} = \mathbf{y}_1^{(k)} - \gamma_2\,\mathrm{prox}_{\frac{1}{\gamma_2}\|\cdot\|_{*,N}}(\frac{1}{\gamma_2}\mathbf{y}_1^{(k)})$;
6 $\quad \mathbf{y}_2^{(k+1)} = \mathbf{y}_2^{(k)} - \gamma_2\,\mathrm{prox}_{\frac{1}{\gamma_2}F_{\mathbf{v}}}(\frac{1}{\gamma_2}\mathbf{y}_2^{(k)})$;
7 $\quad n \leftarrow n + 1$;

3. EXPERIMENTS

ASTV can serve as a building block in various multichannel image restoration scenarios. In the experiments, we apply ASTV to two specific problems: denoising and compressed sensing (CS) reconstruction, and compare it with STV [8].

All the experiments were performed using MATLAB (R2014a, 64bit), on a Windows 8.1 (64bit) laptop computer with an Intel Core i7 2.1 GHz processor and 8 GB of RAM. For test images, we took color images (i.e., $M = 3$) from the *Berkeley Segmentation Database*[7] [24], and their dynamic range was normalized, i.e., every pixel value is in $[0, 1]$. We use PSNR (Peak Signal-to-Noise Ratio)[8] for objective evaluation of restored images. The shape of local neighborhood in ASTV and STV was set to a 3×3 square window, and we consider two cases for the entries of the weight vector $\mathbf{w}$: (i) uniform (all the weights set to $1/9$) and (ii) a 3×3 Gaussian kernel with standard deviation $\sigma_{\mathbf{w}} = \sqrt{0.5}$, which is the same setting suggested in [8].

3.1. Denoising

First, we conducted Gaussian noise removal experiments, where clean test images were contaminated by an additive white Gaussian

[7] For each image, the center region of size 256×256 is cropped.
[8] PSNR is defined by $10\log_{10}(MN/\|\mathbf{u} - \bar{\mathbf{u}}\|^2)$.

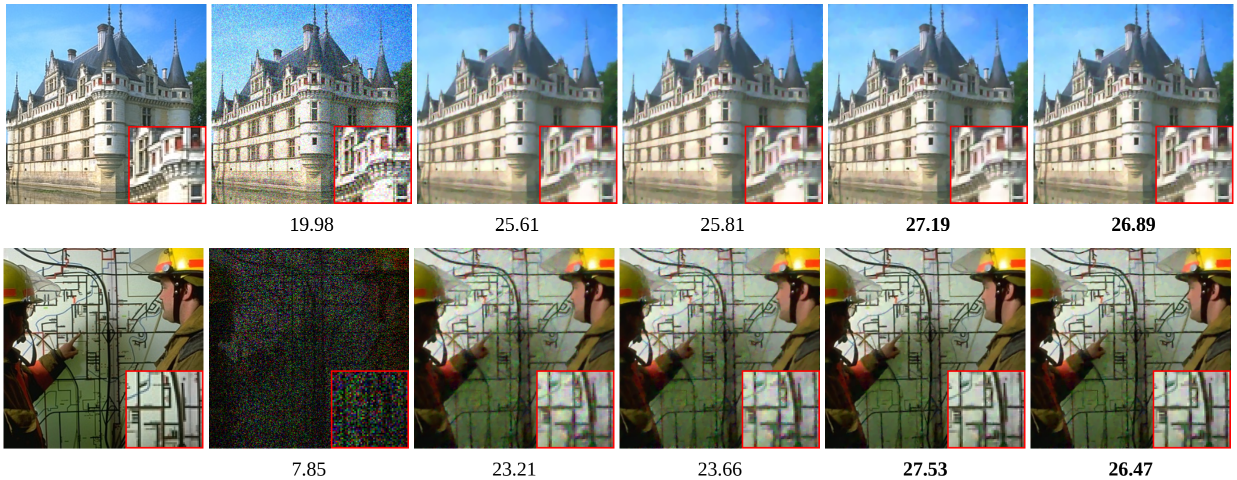

Fig. 2. Resulting images on denoising (top) and CS reconstruction (bottom) experiments: From left to right, original, observation, STV (uniform $\mathbf{w}$), STV (Gaussian $\mathbf{w}$), **ASTV** (uniform $\mathbf{w}$), and **ASTV** (Gaussian $\mathbf{w}$).

noise $\mathbf{n}$ with standard deviation $\sigma = 0.1$, i.e., $\mathbf{v} = \bar{\mathbf{u}} + \mathbf{n}$. Following the discussion in Remark 1, the ℓ_2-norm data-fidelity constraint was adopted, where for a fair comparison, we set the radius ε as the oracle value for each image, i.e., $\varepsilon = \|\bar{\mathbf{u}} - \mathbf{v}\|$. Specifically, we solve the following problem:

$$\min_{\mathbf{u}\in\mathbb{R}^{MN}} J(\mathbf{u}) + \iota_{B_{\mathbf{v},\varepsilon}}(\mathbf{u}) \quad \text{s.t. } \mathbf{u} \in [0,1]^{MN}, \tag{10}$$

where J denotes STV or ASTV, and $B_{\mathbf{v},\varepsilon} := \{\mathbf{x} \mid \|\mathbf{x} - \mathbf{v}\|_2 \leq \varepsilon\}$. Clearly, this problem is a special case of Prob. (5).

Results on *Castle* image are shown in Fig. 2 (top) with their PSNR [dB]. One can see that the images restored by ASTV are better than those by STV in terms of PSNR, and that ASTV well reduces color smearing in restored images. Aside from the visualized images, we measured the gain by ASTV from STV in terms of PSNR averaged over 10 test images, and the result was 1.41 [dB], which also illustrates the effectiveness of ASTV over STV for denoising. Interestingly, the uniform weight $\mathbf{w}$ is preferable for ASTV, whereas STV favors the Gaussian weight $\mathbf{w}$ as addressed in [8]. This suggests that the inter-channel correlation in local neighborhood should be evaluated without spatial weighting.

To demonstrate the suitability of ASTV as a regularization function for multichannel images, we evaluated the function values of STV and ASTV both on clean and noisy images. Specifically, since the scale of STV and ASTV are different, we computed the ratio of STV or ASTV on noisy images and that on clean images, i.e., $J(\mathbf{v})/J(\bar{\mathbf{u}})$ (J denotes STV or ASTV), for measuring how much the function value is increased by noise. Figure 1 (right) indicates the average of $J(\mathbf{v})/J(\bar{\mathbf{u}})$ for STV or ASTV based on 300 images ($\sigma = 0.05, 0.1, 0.15, 0.2$). One observes the function value of ASTV is rapidly increased by noise compared with STV, which implies that ASTV well distinguishes clean and noisy images.

The computational difference between STV and ASTV in optimization only lies in the associated proximity operator. The CPU time of the computation of the proximity operator in the case of STV is 2.32 sec, and that in the case of ASTV is 4.64 sec ($N = 65536$ and $M = 3$), i.e., ASTV is more expensive than STV. This is because the size of the arranged structure tensor is M times larger than that of the standard structure tensor, which is a limitation of ASTV compared with STV. Note that all the program codes were implemented by MATLAB without parallelization.

3.2. Compressed sensing reconstruction

We also conducted experiments on compressed sensing (CS) reconstruction [25, 26] that arises in imaging problems, such as coded aperture imaging and computational photography [27, 28]. Here, we try to recover an original image $\bar{\mathbf{u}}$ from its incomplete measurements $\mathbf{v} = \boldsymbol{\Phi}\bar{\mathbf{u}} + \mathbf{n}$, where $\boldsymbol{\Phi} \in \mathbb{R}^{R\times MN}$ ($R = 0.2MN$) is a random Noiselet measurement matrix [29], and $\mathbf{n} \in \mathbb{R}^M$ is an additive white Gaussian noise with standard deviation $\sigma = 0.1$. Since CS reconstruction is a highly ill-posed problem, we need some regularization, leading to the following optimization problem

$$\min_{\mathbf{u}\in\mathbb{R}^{MN}} J(\mathbf{u}) + \iota_{B_{\mathbf{v},\varepsilon}}(\boldsymbol{\Phi}\mathbf{u}) \quad \text{s.t. } \mathbf{u} \in [0,1]^{MN}, \tag{11}$$

which is also a special case of Prob. (5). The radius ε was set to the oracle value, i.e., $\varepsilon = \|\boldsymbol{\Phi}\bar{\mathbf{u}} - \mathbf{v}\|$.

Figure 2 (bottom) is a showcase of results on *Map* image, where the use of ASTV results in higher PSNR than the use of STV. One can see that the resulting images obtained by ASTV have less false-color-like artifacts than those obtained by STV. As in the case of denoising, adopting the uniform weight $\mathbf{w}$ is suitable for ASTV. Finally, we note that the gain of ASTV from STV in terms of PSNR averaged over 10 test images was 3.26 [dB], i.e., ASTV is also a better regularization than STV for CS reconstruction.

4. CONCLUDING REMARKS

We have proposed a new vectorial total variation with the arranged structure tensor for multichannel image restoration. The arranged structure tensor has a notable property: it becomes an approximately low-rank matrix when a multichannel image of interest has strong correlation among its channels. Thanks to this property, our proposed VTV, named the arranged structure tensor total variation (ASTV), properly incorporates both local spatial variations and inter-channel correlation, resulting in a reasonable regularization for multichannel images. Combining ASTV with cartoon-texture decomposition [30, 31] is an interesting future work.

5. REFERENCES

[1] P. Blomgren and T. F. Chan, "Color TV: Total variation methods for restoration of vector valued images," *IEEE Trans. Image Process.*, vol. 7, no. 3, pp. 304–309, 1998.

[2] X. Bresson and T. F. Chan, "Fast dual minimization of the vectorial total variation norm and applications to color image processing," *Inverse Probl. Imag.*, vol. 2, no. 4, pp. 455–484, 2008.

[3] B. Goldluecke, E. Strekalovskiy, and D. Cremers, "The natural vectorial total variation which arises from geometric measure theory," *SIAM J. Imag. Sci.*, vol. 5, no. 2, pp. 537–563, 2012.

[4] T. Miyata, "Total variation defined by weighted L infinity norm for utilizing inter channel dependency," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, 2012.

[5] S. Ono and I. Yamada, "Decorrelated vectorial total variation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2014.

[6] K. Bredies, "Recovering piecewise smooth multichannel images by minimization of convex functionals with total generalized variation penalty," in *Efficient Algorithms for Global Optimization Methods in Computer Vision*, pp. 44–77. Springer, 2014.

[7] L. Condat, "Semi-local total variation for regularization of inverse problems," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, 2014, pp. 1806–1810.

[8] S. Lefkimmiatis and A. Roussos.and P. Maragos.and M. Unser, "Structure tensor total variation," *SIAM J. Imag. Sci.*, vol. 8, no. 2, pp. 1090–1122, 2015.

[9] G. Chierchia, N. Pustelnik, B. Pesquet-Popescu, and J.-C. Pesquet, "A nonlocal structure tensor-based approach for multicomponent image recovery problems," *IEEE Trans. Image Process.*, vol. 23, no. 12, pp. 5531–5544, 2014.

[10] S. Lefkimmiatis and S. Osher, "Nonlocal structure tensor functionals for image regularization," *IEEE Trans. Comput. Imag.*, vol. 1, no. 1, pp. 16–29, 2015.

[11] S. Di Zenzo, "A note on the gradient of a multi-image," *Comput. vis. graph. image process.*, vol. 33, no. 1, pp. 116–125, 1986.

[12] J. Bigun and G. H. Granlund, "Optimal orientation detection of linear symmetry," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 1987, pp. 433–438.

[13] J. Weickert, *Anisotropic diffusion in image processing*, vol. 1, Teubner Stuttgart, 1998.

[14] A. Bruhn, J. Weickert, and C. Schnörr, "Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods," *Int. J. Comput. Vis.*, vol. 61, no. 3, pp. 211–231, 2005.

[15] C. Harris and M. Stephens, "A combined corner and edge detector.," in *Alvey vision conference*, 1988.

[16] M. Fazel, *Matrix Rank Minimization with Applications*, Ph.D. thesis, Stanford University, 2002.

[17] J. J. Moreau, "Fonctions convexes duales et points proximaux dans un espace hilbertien," *C. R. Acad. Sci. Paris Ser. A Math.*, vol. 255, pp. 2897–2899, 1962.

[18] M. Afonso, J. Bioucas-Dias, and M. Figueiredo, "An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems," *IEEE Trans. Image Process.*, vol. 20, no. 3, pp. 681–695, 2011.

[19] G. Chierchia, N. Pustelnik, J.-C. Pesquet, and B. Pesquet-Popescu, "Epigraphical projection and proximal tools for solving constrained convex optimization problems," *Signal, Image and Video Process.*, pp. 1–13, 2014.

[20] S. Ono and I. Yamada, "Signal recovery with certain involved convex data-fidelity constraints," *IEEE Trans. Signal Process.*, vol. 99, no. 99, pp. –, 2015, (Early Access).

[21] P. L. Combettes and J.-C. Pesquet, "A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery," *IEEE J. Sel. Topics in Signal Process.*, vol. 1, pp. 564–574, 2007.

[22] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *J. Math. Imaging and Vision*, vol. 40, no. 1, pp. 120–145, 2010.

[23] H. H. Bauschke and P. L. Combettes, *Convex analysis and monotone operator theory in Hilbert spaces*, Springer, New York, 2011.

[24] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2001.

[25] R. G. Baraniuk, "Compressive sensing," *IEEE Signal Process. Magazine*, vol. 24, no. 4, 2007.

[26] E. Candès and M. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Magazine*, vol. 25, no. 2, pp. 21–30, 2008.

[27] J. Romberg, "Imaging via compressive sampling," *IEEE Signal Process. Magazine*, vol. 25, no. 2, pp. 14–20, 2008.

[28] M.F. Duarte, M.A. Davenport, D. Takhar, J.N. Laska, Ting Sun, K.F. Kelly, and R.G. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE Signal Process. Magazine*, vol. 25, no. 2, pp. 83–91, 2008.

[29] R. Coifman, F. Geshwind, and Y. Meyer, "Noiselets," *Applied and Computational Harmonic Analysis*, vol. 10, pp. 27–44, 2001.

[30] J.-F. Aujol, G. Gilboa, T. Chan, and S. Osher, "Structure-texture image decomposition - modeling, algorithms, and parameter selection," *Int. J. Comput. Vis.*, vol. 67, no. 1, pp. 111–136, 2006.

[31] S. Ono, T. Miyata, and I. Yamada, "Cartoon-texture image decomposition using blockwise low-rank texture characterization," *IEEE Trans. Image Process.*, vol. 23, no. 3, pp. 1128–1142, 2014.