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 encode 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] 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 $2 \times 2$ matrix constructed from vertical and horizontal differences in the local neighborhood (e.g., $3 \times 3$ window).

The work was supported by JSPS Grants-in-Aid: 15H06197; 15K06076. Both grayscale and multichannel images are considered in [8], but we are only interested in multichannel cases.

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 $u \in \mathbb{R}^{MN}$ be an image with $M$ channels $u_1, \ldots, u_M \in \mathbb{R}^N$ ($N$ 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 $D_v$ and $D_h$ be vertical and horizontal discrete difference operators that map one channel 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 $x \in \mathbb{R}^N$ consisting of its weighted entries at the pixel locations in $\mathcal{I}_n$ by $x_{\mathcal{I}_n} \in \mathbb{R}^{|\mathcal{I}_n|}$ with the weight vector $w \in \mathbb{R}^{|\mathcal{I}_n|}$ (where $|\cdot|$ 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-
lent, denoted by $|Z|$ (NOTE: To handle local neighborhood around image boundaries, we use periodic boundary extension).

Then, the arranged structure tensor of $u$ at the pixel location $n$ is defined by
\[
S_{n,w}^{(u)} := \sum_{n=1}^{N} \|L_{n,w}^{(u)}\|_1,
\]
where $L_{n,w}^{(u)} := \Phi(u_n) \in \mathbb{R}^{M \times N}$.

We remark that the arranged structure tensor defined in (1) is different from the standard structure tensor for one image, i.e.,
\[
S_{n,w}^{(u)} := \sum_{n=1}^{N} \|L_{n,w}^{(u)}\|_1 \in \mathbb{R}^{2 \times 2},
\]
where
\[
L_{n,w}^{(u)} := \begin{bmatrix} D_u u_1 w_{Z_n} \cdots D_u u_M w_{Z_n} \\ D_u u_1 w_{Z_n} \cdots D_u u_M w_{Z_n} \end{bmatrix} \in \mathbb{R}^{M \times 2}.
\]

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

Since the square root of each eigenvalue of the arranged (or standard) structure tensor equals to each singular value of $L_{n,w}^{(u)}$ or $L_{n,w}^{(u)}$, we can discuss the difference of their properties through $L_{n,w}^{(u)}$ and $L_{n,w}^{(u)}$. First, it is clear that every singular value of both $L_{n,w}^{(u)}$ and $L_{n,w}^{(u)}$ becomes small if the local neighborhood is smooth, so that essentially, suppressing some norm of $L_{n,w}^{(u)}$ or $L_{n,w}^{(u)}$ results in a smoothing effect on $u$. Indeed, the Frobenius norms of $L_{n,w}^{(u)}$ and $L_{n,w}^{(u)}$ 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 $L_{n,w}^{(u)}$, implying that information on the correlation is almost lost in the singular values of $L_{n,w}^{(u)}$. On the other hand, the information is still alive in the singular values of $L_{n,w}^{(u)}$ because the discrete differences of each channel are arranged horizontally in (2). More specifically, if the channels of $u$ have strong correlation then the columns of $L_{n,w}^{(u)}$ 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:
\[
\text{ASTV}_w(u) := \sum_{n=1}^{N} \|L_{n,w}^{(u)}\|_1,
\]
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 $L_{n,w}^{(u)} = L_{n,w}^{(u)}$.

The set of discussions in the previous subsection suggests the two things: (i) suppressing all the singular values of $L_{n,w}^{(u)}$ makes restored images smooth, and (ii) promoting the approximate low-rankness of $L_{n,w}^{(u)}$ is suitable for images with strong correlation among channels. Hence, we adopt the nuclear norm for evaluating $L_{n,w}^{(u)}$ 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 $u \in \mathbb{R}^{MN}$ from observation data, which is cast as inverse problems of the form: $v = D(\Phi(u))$, where $\Phi \in \mathbb{R}^{RM \times MN}$ (a matrix representing some degradation e.g., blur and/or random sampling), $D : \mathbb{R}^{R} \rightarrow \mathbb{R}^{R}$ is a noise contamination process, and $v \in \mathbb{R}^{R}$ is an observation.

Based on the above model, we formulate multichannel image restoration by ASTV as the following convex optimization problem:
\[
\min \text{ASTV}(u) + F_v(\Phi u) \quad \text{s.t.} \quad u \in C,
\]
where $F_v \in \mathbb{R}_0^0(\mathbb{R}^{MN})$ is a data-fidelity function, and $C \subset \mathbb{R}^{MN}$ is a closed convex constraint on $u$. We assume that the proximity operator $^7$ of $F_v$ can be computed efficiently. We also assume that the computation of the projection $^7$ on $C$ is efficient.

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

Remark 1 (Examples of $F_v$). The $\ell_0$ norm data-fidelity, given by $F_v(x) := \|x - v\|_2^2$, would be the most popular choice for Gaussian noise cases. The $\ell_1$ norm is a useful data-fidelity measure for impulse noise cases, which is given by $F_v(x) := \mu \|x - v\|_1$. They can also be used as data-fidelity constraints, i.e., $F_v(x) := \|x\|_2^2$.

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:
\[
\min g(x) + h(Ax),
\]
where $g \in \mathbb{R}_0(\mathbb{X})$ and $h \in \mathbb{R}_0(\mathbb{Y})$ (\mathbb{X} and \mathbb{Y} are some Euclidean spaces) and $A : \mathbb{X} \rightarrow \mathbb{Y}$ is a linear operator.

The algorithm is given by
\[
\begin{align*}
x(k+1) &= \text{prox}_{\gamma_1 h^*}(x(k) - \gamma_1 A^T y(k)), \\
y(k+1) &= \text{prox}_{\gamma_2 h^*}(y(k) + \gamma_2 A(2x(k+1) - x(k)));
\end{align*}
\]
where $h^*$ the convex conjugate function\footnote{The set of all proper lower semicontinuous convex functions on $\mathbb{R}^N$ is denoted by $\mathbb{R}_0^0(\mathbb{R}^N)$.} of $h$, and $\gamma_1, \gamma_2 > 0$ satisfy\footnote{The proximity operator of $\gamma_1 h^*$ is expressed as $\text{prox}_{\gamma h^*}(x) = x - \gamma_1 A^T y$ (see, e.g., [23, Theorem 14.3(ii)]).} $\gamma > 0$.

Given a nonempty closed convex set $C \subset \mathbb{R}$, the projection onto $C$ is defined by $P_C : \mathbb{R} \rightarrow \mathbb{R}$ : $x \rightarrow \arg \min \{f(y) : y \in C\}$.

The proximity operator of the indicator function of a nonempty closed convex set $C$ equals to the projection onto $C$, i.e., $\text{prox}_{\gamma_1 h}(x) = P_C$.

The proximity operator of $f^*$ can be computed via that of $f$, i.e., $\text{prox}_{\gamma_1 h^*}(x) = x - \gamma_1 A^T y$ (see, e.g., [23, Theorem 14.3(ii)]).
\[ \gamma_1 \gamma_2 \| A \|_{op}^2 \leq 1 \] (\(\| \cdot \|_{op}\) stands for the operator norm of \(\cdot\)). Under some mild conditions on \(g\), \(h\), and \(A\), the sequence \(x(k)\) 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 \(u\), we give an alternative expression of ASTV. Specifically, we separate these operations and define them as matrices as follows:

\[
\text{ASTV}_w(u) := \| \text{WDPU} \|_{*,N}. \tag{7}
\]

Here, \(D : \mathbb{R}^{MN} \rightarrow \mathbb{R}^{2MN}\) is a discrete difference operator that maps all the channels of \(u\) to their vertical and horizontal difference images, \(P : \mathbb{R}^{2MN} \rightarrow \mathbb{R}^{2|Z|M}\) is an expansion operator that makes \(|Z|\) copies of \(Du\), \(W : \mathbb{R}^{2|Z|M} \rightarrow \mathbb{R}^{2|Z|M}\) is a weighting operator that applies the weights in \(w\) to local neighborhood at every location, and \(\| \cdot \|_{*,N} : \mathbb{R}^{2|Z|M} \rightarrow \mathbb{R}\) is the sum of the nuclear norm of the arranged structure tensor at every location. One sees that in \(\text{WDPU}\), local neighborhood at any location does not overlap each other, which means that the arranged structure tensor at every location can be constructed from \(\text{WDPU}\) without reusing the entries, i.e., \(\text{WDPU}\) and \((L_{u,w}^{(1)}, \ldots, L_{u,w}^{(|Z|)})\) 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 \(L_{u,w}^{(n)}\) with its singular value decomposition \(U^{(n)} \Sigma^{(n)} \psi^{(n)}\), the proximity operator of the nuclear norm of the arranged structure tensor at the location \(n\) is given by

\[
\text{prox}_{\gamma \| \cdot \|_{*,N}}(L_{u,w}^{(n)}) = U^{(n)} \Sigma^{(n)} \psi^{(n)\top}. \tag{8}
\]

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

\[
\min_u \| \text{WDPU} \|_{*,N} + \mathcal{F}_\nu(\Phi u) + \iota C(u). \tag{9}
\]

Finally, by letting

\[
\begin{align*}
g & : \mathbb{R}^{MN} \rightarrow \mathbb{R} \cup \{0\} : u \mapsto \iota_C(u), \\
h & : \mathbb{R}^{2|Z|M} \rightarrow \mathbb{R} \cup \{0\} : (y_1, y_2) \mapsto \| y_1 \|_{*,N} + \mathcal{F}_\nu(y_2), \\
A & : \mathbb{R}^{MN} \rightarrow \mathbb{R}^{2|Z|M} : u \mapsto (\text{WDPU}, \Phi u),
\end{align*}
\]

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, \(D, P, W\), and their transposes can be computed efficiently.

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, 64-bit), on a Windows 8.1 (64-bit) 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\(^3\) [24], and their dynamic range was normalized, i.e., every pixel value is in \([0, 1]\). We use PSNR (Peak Signal-to-Noise Ratio)\(^4\) for objective evaluation of restored images. The shape of local neighborhood in ASTV and STV was set to a \(3 \times 3\) square window, and we consider two cases for the entries of the weight vector \(w\): (i) uniform (all the weights set to \(1/9\)) and (ii) a \(3 \times 3\) Gaussian kernel with standard deviation \(\sigma_w = \sqrt{0.5}\), 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

\[\sqrt{0.5} \text{ for each image, the center region of size } 256 \times 256 \text{ is cropped.}\]

\[\text{PSNR is defined by } 10 \log_{10}(MN/\|u - \bar{u}\|^2).\]
noise \( n \) with standard deviation \( \sigma = 0.1 \), i.e., \( v = \tilde{u} + n \). Following the discussion in Remark 1, the \( \ell_2 \)-norm data-fidelity constraint was adopted, where for a fair comparison, we set the radius \( \varepsilon \) as the oracle value for each image, i.e., \( \varepsilon = \| \tilde{u} - v \| \). Specifically, we solve the following problem:

\[
\min_{u \in \mathbb{R}^{MN}} J(u) + \lambda B_{\varepsilon,x}(u) \quad \text{s.t.} \quad u \in [0, 1]^{MN},
\]

where \( J \) denotes STV or ASTV, and \( B_{\varepsilon,x} := \{ x : \| x - v \|_2 \leq \varepsilon \} \).

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 3.26 (dB), i.e., ASTV is also a better regularization than STV for CS reconstruction.

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 \( \tilde{u} \) from its incomplete measurements \( v = \Phi \tilde{u} + n \), where \( \Phi \in \mathbb{R}^{R \times MN} \) (\( R = 0.2MN \)) is a random Noiselet measurement matrix [29], and \( n \in \mathbb{R}^{M} \) 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

\[
\min_{u \in \mathbb{R}^{MN}} J(u) + \lambda B_{\varepsilon,x}(\Phi u) \quad \text{s.t.} \quad u \in [0, 1]^{MN},
\]

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

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 \( w \) is preferable for ASTV, whereas STV favors the Gaussian weight \( w \) 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 the function value is increased by noise. Figure 1 (right) indicates the average of \( J(v) / J(\tilde{u}) \) 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.

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 leading to the following optimization problem

Averages of PSNR (dB) for different images:

- Castle: 20.19, 26.66
- Map: 19.98, 26.11
- Color: 25.81, 27.19
- Image: 27.53, 26.47

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

- Clearly, this problem is a special case of Prob. (5) where \( \varepsilon \) denotes STV or ASTV, and \( \varepsilon \) denotes STV or ASTV.

- Specifically, we measured the gain by ASTV from STV in terms of PSNR.

- One can see that the images restored by ASTV are better than those by STV in terms of PSNR.

- 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.
5. REFERENCES


