MULTISCALE STRUCTURE TENSOR TOTAL VARIATION FOR IMAGE RECOVERY

Makoto Watanabe[1], *Ryo Matsuoka*[2], *Seisuke Kyochi*[1*], *Shunsuke Ono*[3], *and Masahiro Okuda*[1]

1: The University of Kitakyushu, Fukuoka, Japan
2: Kagawa University, Kagawa, Japan
3: Tokyo Institute of Technology, Tokyo, Japan
Email: s-kyochi@kitakyu-u.ac.jp

ABSTRACT

This paper proposes multiscale structure-tensor total variation (MSTV) for image recovery. Gradient vectors in local patches usually have similar directions, and thus each local gradient matrix (the set of gradient vectors) tends to be low rank. STV introduces this property by calculating the sum of nuclear norms from all the local gradient matrices over the input image. By STV regularization, fine textures are recovered efficiently. However, since STV only considers differences of vertically and horizontally adjacent pixels, if neighboring samples are not reliable due to severe degradation, a latent image cannot be recovered efficiently. In this work, we assume that, for any two target pixels in a local patch, two vectors consisting of multiple differences not only between each target and adjacent pixels but also each target and further distant pixels exhibit a similar direction. According to this assumption, our MSTV firstly applies wavelet-based multiscale decomposition to vertical/horizontal gradient vectors and then evaluates the sum of nuclear norms of all the local wavelet coefficients. Experimental results show that the MSTV improves both numerical reconstruction error and subjective visual quality, compared with the conventional STV.

Index Terms— Convex optimization, image recovery, structure tensor total variation, nuclear norm, wavelet transform

1. INTRODUCTION

Image recovery (e.g., denoising, deblurring, missing pixel recovery, super-resolution, and so on) is a crucial task for accurate imaging in situations where it is difficult to acquire desired images. Convex optimization plays an essential role in image recovery and has been extensively studied [1, 2, 3, 4]. To obtain high quality restored images in each task, we should design a suitable convex prior which mathematically characterizes desired properties of ideal images, such as smoothness, patterns, and sparsity in some transformed domain. For example, total variation (TV) [5, 6, 7, 8, 9] and its extensions of higher-order, semi-local, and non-local versions [10, 11, 12, 13], local linearity of color components [14, 15], and sparse representation by (local/non-local) frame/dictionary [16, 17] have been proposed.

One of the efficient convex priors for image recovery is structure-tensor total variation (STV) [11] and focused in this paper. Since STV utilizes only semilocal similarity of local gradient vectors, it does not suffer from 1) the staircasing effect problem of TV, and 2) chicken-and-egg self-similarity evaluation as in nonlocal approaches. STV is defined by structure tensor [18, 19] applied to many applications in the field of computer vision [20]. After the original STV was proposed, several extensions for multichannel images (e.g., color and hyper-spectral ones) [21, 22, 23] were studied. In natural images, it is often the case that gradient vectors in each local patch tend to have similar directions (see the dashed box "1: STV" in Fig. 1). According to this property, STV calculates the nuclear norm of the *patch-based Jacobian matrix* [11] (the set of local gradient vectors) for each local patch, then sums them up over the image. By integrating STV into the cost function and minimizing it, the low-rankness of local gradients can be promoted, and thus fine textures can be recovered efficiently.

In this work, we extend the conventional STV to multiscale STV (MSTV)[1]. It is often the case that, in local periodic-pattern texture regions, for given two target pixels, two vectors consisting of multiple differences between not only each target and (horizontally and vertically) adjacent pixels but also each target and further distant pixels have a similar direction (see the dashed box "2: MSTV" in Fig. 1). Our MSTV consists of *patch-based multiscale Jacobian matrices* that includes multiscale difference components related to the pixel in the local patch, and thus more robust measure of image variation can be designed (particularly for periodic-texture regions).

To construct patch-based multiscale Jacobian matrices, we further decompose horizontal and vertical differences into multiple scales by using an overcomplete Parseval tight frame (Haar wavelet transform), then take the sum of the nuclear norms of all the patch-based multiscale Jacobian matrices. Furthermore, we introduce a subband-wise weighting operation into MSTV, which is termed as weighted MSTV (WMSTV).

The rest of this paper is organized as follows. Sec. 2 reviews STV and primal-dual splitting (PDS) [24, 25, 26], which is a solver of a class of convex optimization used in this paper. Then, MSTV and WMSTV are explained in Sec. 3. The proposed method is evaluated in the experiments of image denoising and compressed image sensing in Sec. 4. Finally, this paper is concluded in Sec. 5.

1.1. Notations

Let $\mathbb{N}$, $\mathbb{R}$, and $\mathbb{R}_+$ be the sets of positive integers, real numbers, nonnegative real numbers, respectively. Boldfaced large and small letters are matrices and vectors, respectively. A set of N_r [row] and N_c [column] ($N_r,\ N_c \in \mathbb{N}$) real matrices is described as $\mathbb{R}^{N_r \times N_c}$. The matrix $\mathbf{I} \in \mathbb{R}^{N \times N}$ is reserved for the identity matrix. The transpose of a matrix $\mathbf{A} \in \mathbb{R}^{N_c \times N_r}$ is $\mathbf{A}^\top \in \mathbb{R}^{N_r \times N_c}$. A block diagonal matrix of $\{\mathbf{A}_i\}_{i=0}^{N-1}$ is denoted as $\mathrm{diag}(\mathbf{A}_0, \ldots, \mathbf{A}_{N-1}) \in \mathbb{R}^{\sum M_i \times \sum N_i}$ $(\mathbf{A}_i \in \mathbb{R}^{M_i \times N_i})$ and, if all the matrices are scalar, a block diagonal matrix becomes a diagonal matrix. The element-wise multiplication is $\odot$.

*This work was supported in part by JSPS Grants-in-Aid (16H04362, 17K12710, 17K14683, 18K18073) and JST-PRESTO.

[1]For simple discussion, we only consider grayscale images in this paper.

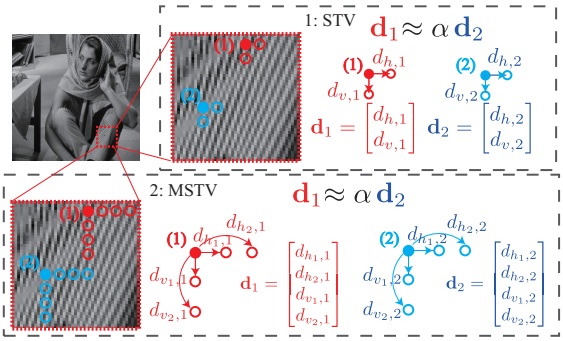


Fig. 1: Similarity of gradient/multiscale difference vectors of STV and MSTV.

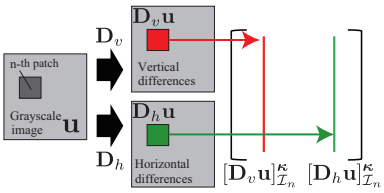


Fig. 2: Construction of patch-based Jacobian matrix $\mathbf{J}^{(n)}_{\mathbf{u},\boldsymbol{\kappa}}$.

2. PRELIMINARIES

2.1. Structure tensor total variation

Since the structure tensor captures first-order information around a local region, it carries more flexible and robust measures of image variation than TV as shown in the following. Let $\mathbf{u} = \left[\mathbf{u}_1^\top, \ldots, \mathbf{u}_M^\top\right]^\top \in \mathbb{R}^{MN}$ be a vectorized image consisting of M channels $\mathbf{u}_1, ..., \mathbf{u}_M \in \mathbb{R}^N$ (N is the number of pixels), e.g., $M = 1$ and $M = 3$ in the case of grayscale and RGB images. We denote pixel indices assigned in the (vertical) raster scan order by $n \in \mathcal{N} := \{1, ..., N\}$, and the set of the pixel indices in the local patch at the pixel location $n \in \mathcal{N}$ by $\mathcal{I}_n$. In addition, for given $\mathbf{u} \in \mathbb{R}^N$, $[\mathbf{u}]^{\boldsymbol{\kappa}}_{\mathcal{I}_n} := \mathbf{P}_{\mathcal{I}_n}(\boldsymbol{\kappa} \odot \mathbf{u})$ is the sub-vector collecting element-wise weighted pixels in the n-th local patch $\mathcal{I}_n$ ($\boldsymbol{\kappa} \in \mathbb{R}_+^{|\mathcal{I}|}$ is a weight vector and $\mathbf{P}_{\mathcal{I}_n} \in \mathbb{R}^{|\mathcal{I}_n| \times N}$ is the extracting matrix). According to these notations, the structure tensor of the n-th local patch of $\mathbf{u} \in \mathbb{R}^{MN}$ is defined as [11]:

$$\mathbf{S}^{(n)}_{\mathbf{u},\boldsymbol{\kappa}} := \mathbf{J}^{(n)\top}_{\mathbf{u},\boldsymbol{\kappa}} \mathbf{J}^{(n)}_{\mathbf{u},\boldsymbol{\kappa}} \in \mathbb{R}^{2\times 2}, \tag{1}$$

$$\mathbf{J}^{(n)}_{\mathbf{u},\boldsymbol{\kappa}} := \begin{bmatrix} [\mathbf{D}_v\mathbf{u}_1]^{\boldsymbol{\kappa}\top}_{\mathcal{I}_n} & \cdots & [\mathbf{D}_v\mathbf{u}_M]^{\boldsymbol{\kappa}\top}_{\mathcal{I}_n} \\ [\mathbf{D}_h\mathbf{u}_1]^{\boldsymbol{\kappa}\top}_{\mathcal{I}_n} & \cdots & [\mathbf{D}_h\mathbf{u}_M]^{\boldsymbol{\kappa}\top}_{\mathcal{I}_n} \end{bmatrix}^\top, \tag{2}$$

where $\mathbf{D}_v$, $\mathbf{D}_h \in \mathbb{R}^{N\times N}$ are vertical and horizontal difference matrices. Finally, the structure-tensor total variation (STV) of $\mathbf{u}$ is defined as:

$$\mathrm{STV}(\mathbf{u}) := \sum_{n=1}^{N} \|\mathbf{J}^{(n)}_{\mathbf{u},\boldsymbol{\kappa}}\|_*, \tag{3}$$

where $\|\cdot\|_*$ is the nuclear norm, i.e., the sum of all the singular values of $(\cdot)$.

2.2. Primal-dual splitting method

Consider the following convex optimization problem to find

$$\mathbf{x}^\star \in \operatorname*{argmin}_{\mathbf{x}\in\mathbb{R}^N} g(\mathbf{x}) + h(\mathbf{L}\mathbf{x}), \tag{4}$$

where $g \in \Gamma_0(\mathbb{R}^N)$, $h \in \Gamma_0(\mathbb{R}^M)$[2], and $\mathbf{L} \in \mathbb{R}^{M\times N}$, respectively. Then the primal-dual splitting algorithm (PDS) [24, 25, 26] for solving (4) is given as follows:

$$\begin{cases} \mathbf{x}^{(n+1)} = \mathrm{prox}_{\gamma_1 g}[\mathbf{x}^{(n)} - \gamma_1 \mathbf{L}^\top \mathbf{z}^{(n)}] \\ \mathbf{z}^{(n+1)} = \mathrm{prox}_{\gamma_2 h^*}[\mathbf{z}^{(n)} + \gamma_2 \mathbf{L}(2\mathbf{x}^{(n+1)} - \mathbf{x}^{(n)})] \end{cases}, \tag{5}$$

where prox denotes the *proximal operator*[3] [27] and h^* is the conjugate function[4] of h [27].

3. MULTISCALE STRUCTURE-TENSOR TOTAL VARIATION

This section proposes MSTV as an extension of the STV. As indicated in Sec. 2.1, STV only considers differences between vertically and horizontally neighboring pixels with respect to a target pixel. Hence, if the adjacent pixels are not accurate due to severe degradation, a latent image might not be sufficiently recovered. Therefore, we propose more robust STV against degradation by assuming that multiple scale differences obtained for each pixel in a local pattern texture region form a low rank matrix. The detail formulation is explained in the following subsection.

3.1. MSTV based on shift-invariant Haar wavelet transform

In order to extend STV to MSTV, we introduce separable two-dimensional shift-invariant Haar wavelet transform (2DHT)[5] because of its computational efficiency and Parseval tight frame property. Specifically, we apply 2DHT to vertical differences and horizontal differences, and create a multiscale difference vector (Fig. 3). Here, recall that j-th level 1DHT is defined as:

$$\begin{cases} \ell_{j+1,n} = \frac{1}{2}(\ell_{j,n} + \ell_{j,n+2^{j-1}}) \\ d_{j+1,n} = \frac{1}{2}(\ell_{j,n} - \ell_{j,n+2^{j-1}}) \end{cases}, \tag{6}$$

where $\ell_{j,n}$ be the j-th lowpass subband coefficients. Thus, applying HT to the horizontal and gradient vectors brings us multiscale difference information of images.

Let $\mathbf{W} = \left[\mathbf{W}_{J,\mathrm{LL}}^\top\ \mathbf{W}_J^\top\ \ldots\ \mathbf{W}_1^\top\right]^\top \in \mathbb{R}^{(3J+1)N\times N}$ be the J-level 2DHT, where $\mathbf{W}_{J,\mathrm{LL}}$ and $\mathbf{W}_j = \left[\mathbf{W}_{j,\mathrm{LH}}^\top\ \mathbf{W}_{j,\mathrm{LH}}^\top\ \mathbf{W}_{j,\mathrm{HH}}^\top\right]^\top$ are the transform matrices computing the J-th lowpass subband coefficients and the j-th highpass subband ones, respectively (note that $\mathbf{W}^\top\mathbf{W} = \mathbf{I}$). For a given input grayscale image $\mathbf{u} \in \mathbb{R}^N$ (N is the

[2]The set of proper lower semicontinuous convex functions [27] on $\mathbb{R}^N$

[3]The proximal operator is defined for a function $f \in \Gamma_0(\mathbb{R}^N)$ and an index $\gamma \in (0, \infty)$ by $\mathrm{prox}_{\gamma f}(\mathbf{x}) := \operatorname*{argmin}_{\mathbf{y}\in\mathbb{R}^N} f(\mathbf{y}) + \frac{1}{2\gamma}\|\mathbf{x} - \mathbf{y}\|_2^2$.

[4]For $^\forall f \in \Gamma_0(\mathbb{R}^p)$, the conjugate function f^* of f is defined as: $f^*(\boldsymbol{\xi}) = \sup_{\mathbf{x}\in\mathbb{R}^N} \langle \mathbf{x}, \boldsymbol{\xi}\rangle - f(\mathbf{x})$, and the proximity operator of the conjugate function is defined as: $\mathrm{prox}_{\gamma f^*}(\mathbf{x}) = \mathbf{x} - \gamma \mathrm{prox}_{\frac{1}{\gamma} f}\left(\frac{1}{\gamma}\mathbf{x}\right)$.

[5]Hereafter, we simply denote two-dimensional shift-invariant Haar wavelet transform as 2DHT.

number of pixels), wavelet-based structure-tensor at the pixel location $n \in \mathcal{N}$ is defined as

$$\begin{aligned}
\widetilde{\mathbf{S}}^{(n)}_{\mathbf{u},\boldsymbol{\kappa}} &:= \widetilde{\mathbf{J}}^{(n)\top}_{\mathbf{u}} \widetilde{\mathbf{J}}^{(n)}_{\mathbf{u}} \in \mathbb{R}^{2(3J+1)\times 2(3J+1)}, \\
\widetilde{\mathbf{J}}^{(n)}_{\mathbf{u},\boldsymbol{\kappa}} &:= \left[\widetilde{\mathbf{J}}^{(n)}_{\mathbf{u},\boldsymbol{\kappa},v}\ \widetilde{\mathbf{J}}^{(n)}_{\mathbf{u},\boldsymbol{\kappa},h}\right] \in \mathbb{R}^{|\mathcal{I}_n|\times 2(3J+1)} \\
\widetilde{\mathbf{J}}^{(n)}_{\mathbf{u},\boldsymbol{\kappa}} &:= [[\mathbf{W}_{J,\mathrm{LL}}\mathbf{D}_v\mathbf{u}]^{\boldsymbol{\kappa}}_{\mathcal{I}_n}\ [\mathbf{W}_{J,\mathrm{LH}}\mathbf{D}_v\mathbf{u}]^{\boldsymbol{\kappa}}_{\mathcal{I}_n}\ [\mathbf{W}_{J,\mathrm{HL}}\mathbf{D}_v\mathbf{u}]^{\boldsymbol{\kappa}}_{\mathcal{I}_n} \\
&\quad [\mathbf{W}_{J,\mathrm{HH}}\mathbf{D}_v\mathbf{u}]^{\boldsymbol{\kappa}}_{\mathcal{I}_n}\ \ldots\ [\mathbf{W}_{1,\mathrm{LH}}\mathbf{D}_v\mathbf{u}]^{\boldsymbol{\kappa}}_{\mathcal{I}_n} \\
&\quad [\mathbf{W}_{1,\mathrm{HL}}\mathbf{D}_v\mathbf{u}]^{\boldsymbol{\kappa}}_{\mathcal{I}_n}\ [\mathbf{W}_{1,\mathrm{HH}}\mathbf{D}_v\mathbf{u}]^{\boldsymbol{\kappa}}_{\mathcal{I}_n}] \in \mathbb{R}^{|\mathcal{I}_n|\times(3J+1)}.
\end{aligned} \tag{7}$$

The other term $\widetilde{\mathbf{J}}^{(n)}_{\mathbf{u},\boldsymbol{\kappa},h}$ is defined in the same fashion of $\widetilde{\mathbf{J}}^{(n)}_{\mathbf{u},\boldsymbol{\kappa},v}$. $\widetilde{\mathbf{J}}^{(n)}_{\mathbf{u},\boldsymbol{\kappa}}$ is termed as *patch-based multiscale Jacobian matrix*.

Finally, multiscale structure-tensor total variation (MSTV) of $\mathbf{u} \in \mathbb{R}^N$ is defined as:

$$\mathrm{MSTV}(\mathbf{u}) := \sum_{n=1}^{N} \|\widetilde{\mathbf{J}}^{(n)}_{\mathbf{u},\boldsymbol{\kappa}}\|_*, \tag{8}$$

In the following, we reformulate the definition of MSTV in (8) to apply it to PDS (Sec. 2.2) because the original definition includes several linear operators implicitly, which disturbs us to compute the proximal operator. We rewrite (8) as:

$$\mathrm{MSTV}(\mathbf{u}) := \|\boldsymbol{\Gamma}_{\boldsymbol{\kappa}}\mathbf{P}\widetilde{\mathbf{W}}\mathbf{D}\mathbf{u}\|_{*,N}, \tag{9}$$

where $\widetilde{\mathbf{W}} = \mathrm{diag}(\mathbf{W},\mathbf{W}) \in \mathbb{R}^{2(3J+1)N\times 2N}$ (J is the decomposition level), $\boldsymbol{\Gamma}_{\boldsymbol{\kappa}}$, $\mathbf{P} \in \mathbb{R}^{2(3J+1)N\times 2(3J+1)N}$ are the weighting matrix for calculating $[\cdot]^{\boldsymbol{\kappa}}_{\mathcal{I}_n}$ and the expansion matrix that copies duplicated elements among patches, respectively. Since the expansion matrix $\mathbf{P}$ makes the local patches non-overlapping, the proximity operator of $\|\cdot\|_{*,N}$ can be decoupled with the proximal operator of the nuclear norm $\mathrm{prox}_{\gamma\|\cdot\|_{*,\boldsymbol{\kappa}}}$ for each patch-based multiscale Jacobian $\widetilde{\mathbf{J}}^{(n)}_{\mathbf{u},\boldsymbol{\kappa}}$, which can be computed by singular value thresholding:

$$\begin{aligned}
&\mathrm{prox}_{\gamma\|\cdot\|_*}(\widetilde{\mathbf{J}}^{(n)}_{\mathbf{u},\boldsymbol{\kappa}}) = \mathbf{U}\Sigma_\gamma\mathbf{V}^\top, \\
&\Sigma_\gamma = \mathrm{diag}(\{\sigma_1-\gamma\}_+,\ldots,\{\sigma_r-\gamma\}_+),
\end{aligned} \tag{10}$$

where $\mathbf{U}$ and $\mathbf{V}$ are orthogonal matrices obtained via singular value decomposition, $\sigma_1,\ldots,\sigma_r$ are the singular values of $\widetilde{\mathbf{J}}^{(n)}_{\mathbf{u},\boldsymbol{\kappa}}$, and $\{a\}_+ := \max\{a,0\}$, respectively.

3.2. Weighted multiscale structure tensor total variation

As for the configuration of MSTV introduced in the previous section, we should remark several points in the following.

Remark 1: The higher level 2DHT coefficients of the vertical and horizontal gradient images ($\mathbf{D}_v\mathbf{u}$, $\mathbf{D}_h\mathbf{u}$) indicate the difference information between a target and distant pixels, while the lower one between a target and close pixels.

Remark 2: Since the vertical and horizontal gradient images ($\mathbf{D}_v\mathbf{u}$, $\mathbf{D}_h\mathbf{u}$) are sparse and their means are almost zero, the absolute values of each 2DHT coefficient becomes lower as the decomposition level increases due to the averaging operation (6).

Remark 3: Singular value thresholding (10) during optimization is considered as approximation of the column vectors $\widetilde{\mathbf{J}}^{(n)}_{\mathbf{u},\boldsymbol{\kappa}}$ towards a certain low dimensional subspace spanned by a subset of right singular vectors $\{\mathbf{v}_i\}_{i=1}^K \subset \mathbf{V}$ (note that $\mathbf{V}$ is designed to optimally represent $\widetilde{\mathbf{J}}^{(n)}_{\mathbf{u},\boldsymbol{\kappa}}$ in any lower dimensionality in ℓ_2-norm sense). In particular, the first principal vector $\mathbf{v}_1 \in \{\mathbf{v}_i\}_{i=1}^K$ tends to lean towards vectors with a large norm.

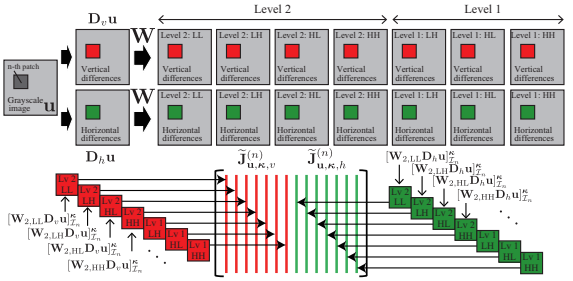


Fig. 3: Construction of patch-based multiscale Jacobian matrix $\widetilde{\mathbf{J}}^{(n)}_{\mathbf{u},\boldsymbol{\kappa}}$ (decomposition level: $J = 2$).

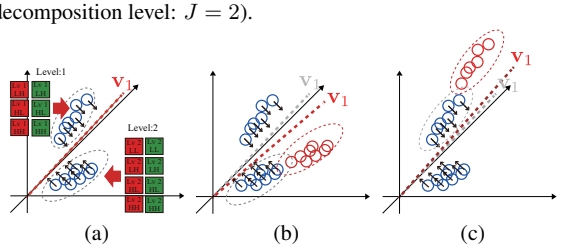


Fig. 4: Effect of amplifying subband coefficients (decomposition level: 2) on the direction of the first principal vector. Each blue and red sample represents the vector containing local subband coefficients, e.g., $[\mathbf{W}_{J,\mathrm{LL}}\mathbf{D}_v\mathbf{u}]^{\boldsymbol{\kappa}}_{\mathcal{I}_n}$ and its amplified version, respectively. (a) no amplification, (b) amplifying the first level subband coefficients, (c) amplifying the second level subband coefficients.

These remarks inform us that we should amplify the 2DHT subband coefficients and change the direction of the right singular vectors $\{\mathbf{v}_i\}$ to preserve important subband coefficients according to the prior information of a latent image or the degree of degradation (see Fig. 4). For example, since images consisting of periodic-pattern textures satisfy nonlocal similarity property, higher level 2DHT coefficients (Remark 1) carry important information. we should amplify higher level 2DHT subband coefficients to preserve them as possible. On the other hand, it is more important to consider adjacent pixels (severe degradation or few pattern textures), it good to slightly amplify lower level 2DHT coefficients.

In order to tune the weight for the 2DHT coefficients, we further introduce weighed MSTV (WMSTV) that involves a subband-wise weighting (diagonal) matrix $\mathbf{Q} \in \mathbb{R}^{2(3J+1))N\times(3J+1))N}$ as:

$$\mathrm{WMSTV}(\mathbf{u}) := \|\boldsymbol{\Gamma}_{\boldsymbol{\kappa}}\mathbf{P}\mathbf{Q}\widetilde{\mathbf{W}}\mathbf{D}\mathbf{u}\|_{*,N}, \tag{11}$$

3.3. Image recovery by MSTV/WMSTV

This section introduces MSTV/WMSTV into the cost function as a regularizer. We assume that the observation $\mathbf{v}$ is obtained through some degradation process (e.g., blur) $\boldsymbol{\Phi} \in \mathbb{R}^{L\times N}$ and (additive/multiplicative) noise contamination $\mathcal{D} : \mathbb{R}^L \to \mathbb{R}^L$ as $\mathbf{v} = \mathcal{D}(\boldsymbol{\Phi}\mathbf{u}^\star)$, where $\mathbf{u}^\star \in \mathbb{R}^N$ is a latent image. Then, we attempt to find $\mathbf{u}^\star$ by solving the following equation:

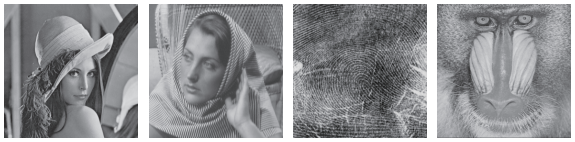
$$\mathbf{u}^\star = \underset{\mathbf{u}\in\mathbb{R}^N}{\mathrm{argmin}}\ F_{\mathbf{v}}(\Phi\mathbf{u}) + \|\boldsymbol{\Gamma}_{\boldsymbol{\kappa}}\mathbf{P}\mathbf{Q}\widetilde{\mathbf{W}}\mathbf{D}\mathbf{u}\|_{*,N} + \iota_{[0,1]^N}(\mathbf{u}) \tag{12}$$

where $\boldsymbol{\Phi} \in \mathbb{R}^{L\times N}$ denotes the degradation process, $F_{\mathbf{v}}$ is some data fidelity function, $\iota_A(\mathbf{x})$ is the indicator function[6] of a set A. $[0,1]^N$

[6]The indicator function of set A is defined as $\iota_A(\mathbf{x}) = 0$, $(\mathbf{x} \in A)$, $\iota_A(\mathbf{x}) = \infty$, $(\mathbf{x} \notin A)$.

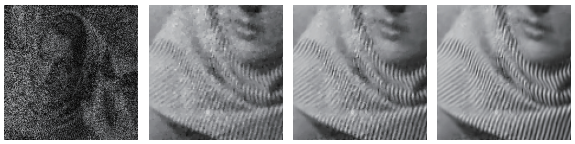
Algorithm 1 Solver for (12)

set $n = 0$ and choose $\mathbf{u}^{(0)}, \mathbf{z}_1^{(0)}, \mathbf{z}_2^{(0)}, \gamma_1, \gamma_2$.
while stop criterion is not satisfied **do**
  $\widetilde{\mathbf{u}}^{(n)} = \mathbf{u}^{(n)} - \gamma_1(\mathbf{\Phi}^\top \mathbf{z}_1^{(n)} + \mathbf{D}^\top \widetilde{\mathbf{W}}^\top \mathbf{Q}^\top \mathbf{P}^\top \mathbf{\Gamma}_{\boldsymbol{\kappa}}^\top \mathbf{z}_2^{(n)})$
  $\mathbf{u}^{(n+1)} = \mathrm{prox}_{\gamma_1 \iota_{[0,1]^N}}(\widetilde{\mathbf{u}}^{(n)})$
  $\mathbf{t}_1^{(n)} = \mathbf{z}_1^{(n)} + \gamma_2 \mathbf{\Phi}(2\mathbf{u}^{(n+1)} - \mathbf{u}^{(n)})$
  $\mathbf{t}_2^{(n)} = \mathbf{z}_2^{(n)} + \gamma_2 \mathbf{\Gamma}_{\boldsymbol{\kappa}} \mathbf{P} \mathbf{Q} \widetilde{\mathbf{W}} \mathbf{D}(2\mathbf{u}^{(n+1)} - \mathbf{u}^{(n)})$.
  $\mathbf{z}_1^{(n+1)} = \mathbf{t}_1^{(n)} - \gamma_2 \mathrm{prox}_{\frac{1}{\gamma_2} F_{\mathbf{v}}}\left(\frac{1}{\gamma_2}\mathbf{t}_1^{(n)}\right)$.
  $\mathbf{z}_2^{(n+1)} = \mathbf{t}_2^{(n)} - \gamma_2 \mathrm{prox}_{\frac{1}{\gamma_2}\|\cdot\|_{*,N}}\left(\frac{1}{\gamma_2}\mathbf{t}_2^{(n)}\right)$.
  $n = n + 1$.
end while
Output $\mathbf{u}^{(n)}$.



(a) *Lena* (b) *Barbara* (c) *Fingerprint* (d) *Mandrill*

Fig. 5: Example of test images.



(a) Observation (b) TV (c) STV (d) WMSTV (i)

Fig. 6: Reconstructed images of compressed image sensing.

is the set of N-dimensional vectors whose entries are in $[0, 1]$. In order to solve (12) by PDS, the functions g and h, and the matrix $\mathbf{L}$ in (4) are set as:

$$g : \mathbb{R}^N \to \{0, \infty\}, \quad \mathbf{u} \mapsto \iota_{[0,1]^N}(\mathbf{u}),$$
$$h : \mathbb{R}^{2|\mathcal{I}|N+L} \to [0, \infty], (\mathbf{z}_1, \mathbf{z}_2) \mapsto F_{\mathbf{v}}(\mathbf{z}_1) + \|\mathbf{z}_2\|_{*,N},$$
$$\mathbf{z}_1 = \mathbf{\Phi}\mathbf{u},\ \mathbf{z}_2 = \mathbf{\Gamma}_{\boldsymbol{\kappa}} \mathbf{P} \mathbf{Q} \widetilde{\mathbf{W}} \mathbf{D} \mathbf{u},\ \mathbf{L} = \begin{bmatrix} \mathbf{\Phi} \\ \mathbf{\Gamma}_{\boldsymbol{\kappa}} \mathbf{P} \mathbf{Q} \widetilde{\mathbf{W}} \mathbf{D} \end{bmatrix}. \quad (13)$$

According to this setting, a solver for (12) can be described as in Algorithm 1, where the proximal operator of the indicator function $\iota_{[0,1]^N}$ is the metric projection onto $[0, 1]^N$, i.e., clip operation into $[0, 1]^N$.

4. EXPERIMENTAL RESULTS

We evaluated the performance of the proposed MSTV and WMSTV in image denoising and compressed sensing reconstructions. TV, TGV, and STV were used as conventional methods. As test images, we use *Lena*, *Barbara*, *Fingerprint*, and *Mandrill*, and the 300 images of the *Berkeley Segmentation Database* (BSDS300) [28]. The size of the images was set to 256×256. Fig. 5 shows the examples of the original images.

In STV, MSTV, and WMSTV, the patch size was set to 3×3 and the weight vector $\boldsymbol{\kappa}$ was uniform (all weights are set to $1/9$). The number of decomposition level for MSTV and WMSTV is set to 3. We tested two sets of weight values for WMSTV: (i) $(q_1, q_2, q_3) = (1.2, 1, 1)$ and (ii) $(q_1, q_2, q_3) = (1, 1, 1.4)$, where q_i is the weight for all the j-th level 2DHT coefficients.

We used the cost function shown in (12), where the data-fidelity function was set as $F_{\mathbf{v}} = \iota_{\mathcal{B}(\mathbf{v},\epsilon)}$ ($\mathcal{B}(\mathbf{v}, \epsilon) := \{\mathbf{x} \in \mathbb{R}^M | \|\mathbf{x} - \mathbf{v}\|_2 \leq \epsilon\}$) is the indicator function defined by the ℓ_2-norm ball. The radius was set as $\epsilon = \|\mathbf{u} - \mathbf{v}\|_2$, where $\mathbf{u}$ is an original image.

Table 1: Numerical results (PSNR [dB])

Image denoising						
		Lena	*Barbara*	*Fingerprint*	*Mandrill*	Ave.
PSNR	TV	27.55	24.78	22.27	25.13	26.34
	TGV	28.17	25.07	22.20	24.96	26.50
	STV	27.78	24.89	22.96	25.49	26.67
	MSTV	**27.86**	25.76	23.43	25.71	26.69
	WMSTV (i)	**27.86**	25.76	**23.49**	**25.74**	**26.75**
	WMSTV (ii)	27.81	**25.79**	23.43	25.68	26.69
Compressed image sensing						
		Lena	*Barbara*	*Fingerprint*	*Mandrill*	Ave.
PSNR	TV	29.52	24.03	21.87	25.93	28.06
	TGV	30.72	24.44	21.44	25.95	28.85
	STV	31.34	25.84	22.13	26.54	28.46
	MSTV	31.60	28.06	22.72	26.89	29.20
	WMSTV (i)	**31.65**	**28.15**	**22.83**	**26.93**	**29.33**
	WMSTV (ii)	31.51	28.07	22.75	26.87	29.05

4.1. Image denoising

In this experiment, we add additive white Gaussian noise with the standard derivation $\sigma = 0.1$ to the original images $\mathbf{v} = \mathbf{u} + \mathbf{n}$ ($\mathbf{\Phi}$ is set as $\mathbf{I}$). The experimental results are shown in Table 1 ("Ave." means the average of all the resulting PSNRs from the BSDS300). The table shows that WMSTV achieved the best performance in numerical reconstruction quality. As mentioned in Sec. 3.2, amplifying 2DHT coefficients at a higher level (WMSTV (ii)) is suitable for *Barbara* richly containing periodic-pattern textures, while WMSTV (i) is better for *Lena*, *Fingerprint*, and *Mandrill* that consist of smooth regions or weak edge regions.

4.2. Compressed image sensing

In compressed image sensing, each incomplete observation $\mathbf{v} = \widetilde{\mathbf{\Phi}}\mathbf{u} + \mathbf{n}$ ($\widetilde{\mathbf{\Phi}} := \mathbf{S}\mathbf{\Phi}$) is obtained by the Noiselet transform [29] $\mathbf{\Phi}$ followed by random downsampling $\mathbf{S} \in \mathbb{R}^{L \times N}$ ($L = 0.3 \times N$) in the presence of additive white Gaussian noise $\mathbf{n}$ with standard derivation $\sigma = 0.1$.

As shown in Table. 1, WMSTV achieved the best reconstruction performance. Since compressed image sensing is a highly ill-posed problem, the difference information between a target and distant pixels are not reliable. Therefore, WMSTV (i) provides better performance for most images. Fig. 6 shows the reconstructed images of *Barbara*. Obviously, the directional lines are accurately recovered by WMSTV (i).

5. CONCLUDING REMARKS

In this paper, we extended from STV to MSTV for more robust image recovery. According to the assumption of multiscale difference low-rankness, we designed the patch-based multiscale Jacobian matrix by applying 2DHT to gradient vectors. Besides, we introduced WMSTV to tune the effect of multiscale difference low-rankness according to observed images. Experimental results showed that WMSTV provided the best performance in image denoising and compressed image sensing. An extension of this work to multi-channel images is our future work.

6. REFERENCES

[1] P. L Combettes and J.-C. Pesquet, "A proximal decomposition method for solving convex variational inverse problems," *Inverse Problems*, vol. 24, no. 6, pp. 065014, Nov. 2008.

[2] T. Goldstein and S. Osher, "The split Bregman method for L1-regularized problems," *SIAM J. Imag. Sci.*, vol. 2, no. 2, pp. 323–343, Apr. 2009.

[3] N. Pustelnik, C. Chaux, and J. Pesquet, "Parallel proximal algorithm for image restoration using hybrid regularization," *IEEE Trans. Image Process.*, vol. 20, no. 9, pp. 2450–2462, Sep. 2011.

[4] M. V. Afonso, J. M. B.-Dias, and M. A. T. 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, Mar. 2011.

[5] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D*, vol. 60, no. 1-4, pp. 259–268, Nov. 1992.

[6] X. Bresson and Tony 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, Nov. 2008.

[7] R. H. Chan, Y. Dong, and M. Hintermuller, "An efficient two-phase L^1-TV method for restoring blurred images with impulse noise," *IEEE Trans. Image Process.*, vol. 19, no. 7, pp. 1731–1739, Jul. 2010.

[8] I. Bayram and M. E. Kamasak, "Directional total variation," *IEEE Signal Process. Letters*, vol. 19, no. 12, pp. 781–784, Sep. 2012.

[9] S. Ono and I. Yamada, "Decorrelated vectorial total variation," in *Proc. 2014 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2014, pp. 4090–4097.

[10] K. Bredies, K. Kunisch, and T. Pock, "Total generalized variation," *SIAM J. Imag. Sci.*, vol. 3, no. 3, pp. 492–526, Sep. 2010.

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

[12] 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, Dec. 2014.

[13] S. Lefkimmiatis and S. Osher, "Nonlocal structure tensor functionals for image regularization," *IEEE Trans. Image Process.*, vol. 1, no. 1, pp. 16–29, Mar. 2015.

[14] S. Ono and I. Yamada, "Color-line regularization for color artifact removal," *IEEE Trans. Comput. Imag.*, vol. 2, no. 3, pp. 204–217, Sep. 2016.

[15] K. Yamanaka, S. Kyochil, S. Ono, and K. Shirai, "Color affine subspace pursuit for color artifact removal," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 1358–1362.

[16] A. Danielyan, V. Katkovnik, and K. Egiazarian, "BM3D frames and variational image deblurring," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1715–1728, Apr. 2012.

[17] J. Sulam, B. Ophir, M. Zibulevsky, and M. Elad, "Trainlets: Dictionary learning in high dimensions," *IEEE Trans. Signal Process.*, vol. 64, no. 12, pp. 3180–3193, Jun. 2016.

[18] S. D. Zenzo, "A note on the gradient of a multi-image," *Computer Vision, Graphics, and Image Processing*, vol. 33, no. 1, pp. 116 – 125, 1986.

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

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

[21] S. Ono, K. Shirai, and M. Okuda, "Vectorial total variation based on arranged structure tensor for multichannel image restoration," in *Proc. IEEE Int. Conf. Acoust, Speech, Signal Process. (ICASSP)*, Mar. 2016, pp. 4528–4532.

[22] R. Kurihara, S. Ono, K. Shirai, and M. Okuda, "Hyperspectral image restoration based on spatio-spectral structure tensor regularization," in *Proc. European Signal Process. Conf. (EUSIPCO)*, Aug. 2017, pp. 488–492.

[23] Zhuo-Xu Cui, Qibin Fan, Yichuan Dong, and Tong Liu, "A nonconvex nonsmooth regularization method with structure tensor total variation," *Journal of Visual Communication and Image Representation*, vol. 43, pp. 30 – 40, 2017.

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

[25] L. Condat, "A primal–dual splitting method for convex optimization involving lipschitzian, proximable and linear composite terms," *Journal of Optimization Theory and Applications*, vol. 158, no. 2, pp. 460–479, Aug. 2013.

[26] B. C. Vu, "A splitting algorithm for dual monotone inclusions involving cocoercive operators," *Adv. Comput. Math.*, vol. 38, no. 3, pp. 667–681, Nov. 2011.

[27] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, New York, NY, USA: Springer-Verlag, 2011.

[28] 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)*, Jul. 2001, vol. 2, pp. 416–423.

[29] R. Coifman, F. Geshwind, and Y. Meyer, "Noiselets," *Applied and Computational Harmonic Analysis*, vol. 10, no. 1, pp. 27 – 44, 2001.