Journal of Southeast University (English Edition) Vol. 35, No. 3, pp. 332 - 340 Sept. 2019 ISSN 1003—7985

Image denoising method with tree-structured group sparse modeling of wavelet coefficients

Zhang Tao Wei Haiguang Mo Xutao

(School of Mathematics and Physics, Anhui University of Technology, Ma'anshan 243000, China)

Abstract: In order to enhance the image contrast and quality, inspired by the interesting observation that an increase in noise intensity tends to narrow the dynamic range of the local standard deviation (LSD) of an image, a tree-structured group sparse optimization model in the wavelet domain is proposed for image denoising. The compressed dynamic range of LSD caused by noise leads to a contrast reduction in the image, as well as the degradation of image quality. To equalize the LSD distribution, sparsity on the LSD matrix is enforced by employing a mixed norm as a regularizer in the image denoising model. This mixed norm introduces a coupling between wavelet coefficients and provides a tree-structured group scheme. The alternating direction method of multipliers (ADMM) and the fast iterative shrinkage-thresholding algorithm (FISTA) are applied to solve the group sparse model based on different cases. Several experiments are conducted to verify the effectiveness of the proposed model. The experimental results indicate that the proposed group sparse model can efficiently equalize the LSD distribution and therefore can improve the image contrast and quality.

Key words: local standard deviation; group sparse; image denoising; mixed norm; texture;

DOI: 10.3969/j.issn.1003-7985.2019.03.009

Image restoration is a major subtopic in image science and has been extensively studied over the past decades. The purpose of image restoration is to reconstruct a high-quality image from its degraded observation. It is typically treated as an ill-posed linear inverse problem that can be generally formulated as

$$\boldsymbol{b} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n} \tag{1}$$

where $\boldsymbol{b}$ is the degraded observation; $\boldsymbol{x}$ is the desired true image; and $\boldsymbol{n}$ is Gaussian white noise with a zero mean. Note that $\boldsymbol{H}$ is a matrix representing a linear degradation operator.

Received 2019-04-23, **Revised** 2019-08-15.

Biography: Zhang Tao (1982—), male, doctor, associate professor, zt9877@163.com.

Foundation items: The National Natural Science Foundation of China (No. 61701004, 11504003), the Natural Science Foundation of Anhui Province (No. 1708085QA15).

Citation: Zhang Tao, Wei Haiguang, Mo Xutao. Image denoising method with tree-structured group sparse modeling of wavelet coefficients [J]. Journal of Southeast University (English Edition), 2019, 35(3): 332 - 340. DOI: 10.3969/j.issn.1003-7985.2019.03.009.

To address the ill-posed nature of the image restoration problem, image prior knowledge is usually employed to regularize the solution of the following minimization problem:

$$\min_{x} \frac{1}{2} \| \boldsymbol{H}\boldsymbol{x} - \boldsymbol{b} \|_2^2 + \lambda \Phi(\boldsymbol{x}) \tag{2}$$

where the first term is the data fidelity term and the second term $\Phi(\boldsymbol{x})$ is known as the regularization term, which regularizes the solution by enforcing certain prior constraints. In addition, $\lambda > 0$ is the regularization parameter, which controls the balance between the fidelity and regularization terms.

How to select a good regularization term is an active area of research. Traditional regularization methods include Tikhonov regularization[1] and total variation (TV) regularization[2]. The total variation measure has been shown to be suitable for preserving sharp edges. Although TV regularization has been proven to be very useful in many applications, it is flawed in that it removes textures and creates staircase artifacts. Texture preserving methods such as sparse representations, non-local methods[3-5] and the deep convolutional neural network (CNN)-based method[6-7] have been introduced as image restoration approaches. Sparse regularization is a recent and successful method to solve image restoration problems[8-9]. Most sparse-based methods represent signals from a given dictionary and then process the coefficients of expansion individually. However, in addition to sparsity, signals may exhibit a "group sparsity" structure. Group sparsity is usually modeled by introducing a coupling between coefficients in the same structure set[10-12]. In the framework of variational formulation, this type of coupling may be introduced by a suitable regularization term[13]. In this study, we introduce a mixed norm regularization term, which enforces the sparsity on the LSD matrix of the restored image.

LSD is an important measure for the structural information of the image[14-15]. Through numerical experiments, we observed that the dynamic range of LSD decreases with an increase in noise intensity. Enforcing the sparsity of LSD matrix can efficiently equalize the LSD distribution. In this study, we examine the wavelet characterization of the LSD matrix of an image, and then propose a group sparse optimization model in the wavelet

domain.

We describe the wavelet characterization of the LSD matrix, and then introduce the tree-structured group sparse model in the wavelet domain. Secondly, we give the solution of the proposed model based on different cases through the ADMM and FISTA techniques. Finally, some numerical experiments are presented to verify the potential of the proposed model.

1 Wavelet Characterization of LSD Matrix

In this study, an image is interpreted as a function defined on the unit square $I=[0,1]^2$. Let Q be the collection of dyadic cubes contained in I.

$$Q=\left\{Q_{j,k}=Q_{j,k_1,k_2}\overset{\Delta}{=\!=}\left[\frac{k_1}{2^j},\frac{k_1+1}{2^j}\right]\times\left[\frac{k_2}{2^j},\frac{k_2+1}{2^j}\right]\right\} \tag{3}$$

where $j=0,1,\ldots;k=(k_1,k_2)\in\Gamma_j=\{0,1,\ldots,2^j-1\}^2$.

Let φ and ψ be a univariate wavelet constructed out of V_j, which is an r-regular multiresolution approximation of $L^2(R^2)$ with $r\geqslant 1$ (for the exact conditions, see Ref. [16]). Two dimensional wavelets can be constructed by a tensor product

$$\psi_Q^{\varepsilon}(x_1,x_2):=\psi_{j,k}^{\varepsilon}(x_1,x_2)=2^j\psi^{\varepsilon_1}(2^jx_1-k_1)\cdot\psi^{\varepsilon_2}(2^jx_2-k_2)$$

where $j\in Z$, $k=(k_1,k_2)\in Z^2$, $\varepsilon\in E=(\varepsilon_1,\varepsilon_2)^2/\{0,0\}$ with $\varepsilon_j=0$ or 1, $\psi^0=\varphi$, and $\psi^1=\psi$.

One can easily construct periodic wavelets on $L^2(I)$ that can be used to decompose periodic function f on $L^2(I)$.

For a compactly supported wavelet $\psi_{j,k}^{\varepsilon}$, we define its periodic version, which is still denoted here by

$$\psi_Q^{\varepsilon}(x_1,x_2):=\psi_{j,k}^{\varepsilon}=\sum_{l\in Z^2}\psi_{j,k}^{\varepsilon}(x-l)$$

For the wavelet decomposition of functions in $L^2(\mathbf{R}^2)$, we do not need to translate all of these periodic wavelets. On level j, we need only the translates $k\in\Gamma_j$. Because the translates of the scaling function form a partition of unity, we can obtain $\tilde{\varphi}=\chi_I$, which is the characteristic function of I. The wavelet expansion for a function $f\in L^2(\mathbf{R}^2)$ is

$$f=\sum_{K\in\Gamma_j}\langle f,\tilde{\varphi}_{j,K}\rangle\tilde{\varphi}_{j,K}+\sum_{j=J}^{\infty}\sum_{k\in\Gamma_j}\sum_{\varepsilon\in E}\langle f,\tilde{\psi}_{j,k}^{\varepsilon}\rangle\tilde{\psi}_{j,k}^{\varepsilon} \tag{4}$$

and the L^2 norm of f can be characterized by wavelet coefficients:

$$\|f\|_{L^2}^2=\sum_{K\in\Gamma_j}|\langle f,\tilde{\varphi}_{j,K}\rangle|^2+\sum_{j=J}^{\infty}\sum_{k\in\Gamma_j}\sum_{\varepsilon\in E}|\langle f,\tilde{\psi}_{j,k}^{\varepsilon}\rangle|^2 \tag{5}$$

In this subsection, we derive the wavelet characterization of the LSD matrix of f on dyadic region $Q_{J,K}$. The LSD matrix is defined as

$$\boldsymbol{P}_{J,K}(f)=\left(\frac{1}{|Q_{J,K}|}\int_{Q_{J,K}}|f(x)-m_{Q_{J,K}}f|^2\mathrm{d}x\right)^{1/2} \tag{6}$$

where $J=0,1,\cdots$, and $K=(K_1,K_2)\in\Gamma_J=\{0,1,\cdots,2^J-1\}^2$ are the indices for characterizing the dyadic cube $Q_{J,K}$ defined in Eq. (3); $m_{Q_{J,K}}f=\frac{1}{|Q_{J,K}|}\int_{Q_{J,K}}f(x)\mathrm{d}x$ is the mean of f on the dyadic region $Q_{J,K}$; $|Q_{J,K}|$ denotes the volume of $Q_{J,K}$ and $|Q_{J,K}|=2^{2J}$.

The LSD matrix characterizes the local oscillating property of the image. Different components of an image may have different LSD. In a homogenous region, the LSD is rather small, while in an edge and texture region, the LSD becomes large. Generally, the LSD of noise is much smaller than that of texture. Thus, LSD is often used to separate homogeneous and textured regions in natural image denoising.

In Ref. [17], we derive the discrete wavelet characterization of the LSD matrix, which can be described by the following theorem.

Theorem 1 The LSD matrix of f on dyadic region $Q_{J,K}$ can be characterized by wavelet coefficients as follows:

$$\boldsymbol{P}_{J,K}(f)=\left(2^{2J}\sum_{\varepsilon\in E}\sum_{Q_{j,k}\in Q_{J,K}}|f_{j,k}^{\varepsilon}|^2\right)^{1/2}=\left(2^{2J}\sum_{\varepsilon\in E}\sum_{j\geqslant J}\sum_{k\in\Lambda^{j-J}(K)}|f_{j,k}^{\varepsilon}|^2\right)^{1/2} \tag{7}$$

Proof Let $C(Q_{J,K})$ denote the mean value of $f(x)$ on $Q_{J,K}$, then

$$C(Q_{J,K})=\frac{1}{Q_{J,K}}\int_{Q_{J,K}}f(x)\mathrm{d}x=2^{2J}\int_{\Omega}f(x)\tilde{\varphi}(2^Jx-K)\mathrm{d}x=2^J\langle f,\tilde{\varphi}_{J,K}\rangle \tag{8}$$

$$\langle C(Q_{J,K}),\tilde{\varphi}_{J,K}\rangle=\int_{\Omega}C(Q_{J,K})\cdot 2^J\tilde{\varphi}(2^Jx-k)\mathrm{d}x=C(Q_{J,K})\cdot 2^J\cdot|Q_{J,K}|=2^J\langle f,\tilde{\varphi}_{J,K}\rangle\cdot 2^J\cdot 2^{-2J}=\langle f,\tilde{\varphi}_{J,K}\rangle \tag{9}$$

We write $f(x)$ as $f(x)=f_1(x)+f_2(x)+C(Q_{J,K})$, where

$$f_1(x)=\begin{cases}f(x)-C(Q_{J,K}) & x\in Q_{J,K}\\ 0 & \text{otherwise}\end{cases} \tag{10}$$

By Eq. (9), we can obtain

$$\langle f_1,\tilde{\varphi}_{J,K}\rangle=\langle f-C(Q_{J,K}),\tilde{\varphi}_{J,K}\rangle=0 \tag{11}$$

Since $f_2(x)=0$ if $x\in Q_{J,K}$, and the support of $\tilde{\psi}_{J,K}^{\varepsilon}$ is contained in $Q_{J,K}$, we have $\langle f_2,\tilde{\psi}_{J,K}^{\varepsilon}\rangle=0$. Thus,

$$\langle f_1,\tilde{\psi}_{J,K}^{\varepsilon}\rangle=\langle f-f_2-C(Q_{J,K}),\tilde{\psi}_{J,K}^{\varepsilon}\rangle=\langle f_1,\tilde{\psi}_{J,K}^{\varepsilon}\rangle \tag{12}$$

Then,

$$\frac{1}{|Q_{J,K}|}\int_{Q_{J,K}} |f(x)-C(Q_{J,K})|^2 \mathrm{d}x = 2^{2J}\|f_1(x)\|_{L^2}^2 = 2^{2J}\Big(\sum_{k\in\Gamma}|\langle f_1,\tilde{\varphi}_{J,K}\rangle|^2 + \sum_{\varepsilon\in E}\sum_{j=J}^{N}\sum_{k\in\Gamma_j}|\langle f_1,\tilde{\psi}^{\varepsilon}_{j,k}\rangle|^2\Big) = 2^{2J}\sum_{\varepsilon\in E}\sum_{Q_{j,k}\subseteq Q_{J,K}}|\langle f,\tilde{\psi}^{\varepsilon}_{j,k}\rangle|^2 = 2^{2J}\sum_{\varepsilon\in E}\sum_{j\geqslant J}\sum_{k\in\Lambda^{(j-J)}(k)}|\langle f,\tilde{\psi}^{\varepsilon}_{j,k}\rangle|^2 \tag{13}$$

Then, we can obtain

$$P_{J,K}(f) = \Big(2^{2J}\sum_{\varepsilon\in E}\sum_{j\geqslant J}\sum_{k\in\Lambda^{(j-J)}(k)}|\langle f,\tilde{\psi}^{\varepsilon}_{j,k}\rangle|^2\Big)^{1/2}$$

2 Group Sparse Model in The Wavelet Domain

Let $f\in\mathbf{R}^{m\times n}$ be an image. We assume that f has the multi-scale wavelet expansion as indicated in Eq. (4). Then, the support of $\{\tilde{\psi}^{\varepsilon}_{j,k}\}_{j,k,\varepsilon}$ partitions the image into different scales of dyadic regions. At scale J, $I=\bigcup_{K\in\Gamma} Q_{J,K}$, where $Q_{J,K}=\sup\tilde{\psi}^{\varepsilon}_{J,K}$. As indicated in Eq. (4), f is decomposed into low-frequency component and high-frequency component, which are denoted by f_{L} and f_{H}, respectively.

For fixed scale J, let $G_K=\{(j,k,\varepsilon): Q_{j,k}\subseteq Q_{J,K}, j\geqslant J, k\subseteq\Gamma_j, \varepsilon\in E\}$ denote the index set. Then, $\alpha_{G_k}=\{\langle f,\tilde{\psi}^{\varepsilon}_{j,k}\rangle:(j,k,\varepsilon)\in G_K\}$ is a group collecting all the descendant wavelet coefficients of $\alpha^{\varepsilon}_{J,K}=\langle f,\tilde{\psi}^{\varepsilon}_{J,K}\rangle$. $\{\boldsymbol{\alpha}_{G_k}\}_{K\in\Gamma_J}$ forms a non-overlapping partition of all high-frequency coefficients. Note that the wavelet coefficients in each group $\{\boldsymbol{\alpha}_{G_k}\}_{K\in\Gamma_J}$ have a parent-child relationship with the tree structure (see Fig. 1). With this grouping strategy, we can rewrite the LSD matrix $\boldsymbol{P}_{J,K}$ as follows:

$$\boldsymbol{P}_{J,K}(f)=2^J\|\boldsymbol{\alpha}_{G_k}\|_2 \qquad K\in\Gamma_J \tag{14}$$

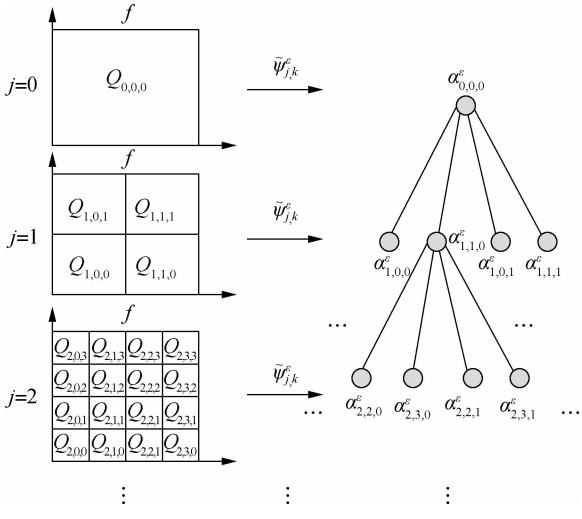

Fig. 1 Tree-structured wavelet coefficients

In this study, we introduce the prior that the LSD matrix $\boldsymbol{P}_{J,K}$ is sparse. This can be obtained by minimizing the l_1 norm of the LSD matrix. In other words, we use the following weighted $l_{2,1}$ norm as a regularizer:

$$\|\boldsymbol{\alpha}\|_{l_{\omega,2,1}}=\sum_{K\in\Gamma_J}2^J\boldsymbol{\omega}_K\|\boldsymbol{\alpha}_{G_k}\|_2 \tag{15}$$

where $\boldsymbol{\omega}_K$ are weights associated with each group. Properly choosing weights may result in improved recovery performance. In this study, we choose $\boldsymbol{\omega}_K$ as

$$\boldsymbol{\omega}_K=1-\frac{\boldsymbol{P}_{J,K}}{\max_K \boldsymbol{P}_{J,K}} \tag{16}$$

The mixed norm in Eq. (15) explicitly introduces the coupling between wavelet coefficients instead of the usual independence assumption of l_{p} norm sparse optimization problems. Thus, we consider the following regression problem to recover the group sparse solution:

$$\min_{\alpha}\sum_{K\in\Gamma_J}\boldsymbol{\omega}_K\|\boldsymbol{\alpha}_{G_k}\|_2+\frac{1}{2\lambda}\|\boldsymbol{H}\boldsymbol{W}^{\mathrm{T}}\boldsymbol{\alpha}-\boldsymbol{b}_{\mathrm{H}}\|_2^2 \tag{17}$$

where $\boldsymbol{b}_{\mathrm{H}}$ is the high-frequency component of observed image b, and $\boldsymbol{W}$ is the wavelet transform. $\boldsymbol{\omega}_K$ controls the group sparsity level adaptively. While the image regions are dominated by texture and edges, $\boldsymbol{P}_{J,K}$ is relatively large and $\boldsymbol{\omega}_K$ is relatively small. Then, the wavelet coefficients group $\boldsymbol{\alpha}_{G_k}$ is less penalized such that the edges and texture are well preserved during denoising. Conversely, in homogeneous image regions, $\boldsymbol{P}_{J,K}$ is relatively small and $\boldsymbol{\omega}_K$ is relatively large. Then, $\boldsymbol{\alpha}_{G_k}$ has more intense shrinkage to enhance the denoising process.

Because most of the noise is concentrated in the high-frequency component, we must only recover the high-frequency component of the image. If $\boldsymbol{\alpha}^*$ is the solution of model (17), then the final restored image is $\boldsymbol{b}^*=\boldsymbol{W}^{\mathrm{T}}\boldsymbol{\alpha}^*+\boldsymbol{b}_{\mathrm{L}}$, where $\boldsymbol{b}_{\mathrm{L}}$ is the low-frequency component of the observed image.

Our intent is not to seek a state-of-the-art denoising method, but rather to investigate the sparsity of LSD as signal priors and compare it to the TV-based, frame-based and non-local mean methods.

3 Solution of the Proposed Model

We apply the ADMM to solve model (17) when $\boldsymbol{H}$ is identity and $\boldsymbol{W}^{\mathrm{T}}$ is a linear operator (for the ADMM algorithm, see Ref. [18]). To accomplish this, we introduce an auxiliary variable and transform model (17) into an equivalent form:

$$\min_{z,\boldsymbol{\alpha}}\sum_{K\in\Gamma_J}\boldsymbol{\omega}_K\|\boldsymbol{Z}_{G_k}\|_2+\frac{\beta_2}{2}\|\boldsymbol{W}^{\mathrm{T}}\boldsymbol{\alpha}-\boldsymbol{b}_{\mathrm{H}}\|_2^2 \tag{18}$$

$$\text{s.t.} \qquad \boldsymbol{Z}=\boldsymbol{\alpha} \tag{19}$$

The augmented Lagrangian problem takes the form:

$$\min_{\boldsymbol{Z},\boldsymbol{\alpha}} \sum_{K\in\Gamma_J} \boldsymbol{\omega}_K \| \boldsymbol{Z}_{G_K} \|_2 - \boldsymbol{\mu}^{\mathrm{T}}(\boldsymbol{Z}-\boldsymbol{\alpha}) + \frac{\beta_1}{2}\| \boldsymbol{Z}-\boldsymbol{\alpha} \|_2^2 + \frac{\beta_2}{2}\| \boldsymbol{W}^{\mathrm{T}}\boldsymbol{\alpha} - \boldsymbol{b}_{\mathrm{H}} \|_2^2 \quad (20)$$

where $\mu > 0$ is a multiplier and β_1 is a penalty parameter. We then apply the ADMM to minimize the Lagrangian problem alternately with respect to $\boldsymbol{Z}$ and $\boldsymbol{\alpha}$.

The α subproblem is given by

$$\min_{\alpha} \boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\alpha} + \frac{\beta_1}{2}\| \boldsymbol{Z}-\boldsymbol{\alpha} \|_2^2 + \frac{\beta_2}{2}\| \boldsymbol{W}^{\mathrm{T}}\boldsymbol{\alpha} - \boldsymbol{b}_{\mathrm{H}} \|_2^2 \quad (21)$$

Note that (21) is a convex quadratic problem. Therefore, its solution is

$$\boldsymbol{\alpha} = \frac{1}{\beta_1+\beta_2}(\beta_1\boldsymbol{Z} - \boldsymbol{\mu} + \beta_2\boldsymbol{W}\boldsymbol{b}_{\mathrm{H}}) \quad (22)$$

For the Z subproblem, it is reduced to minimize (20) with respect to Z:

$$\min_{\boldsymbol{Z}} \left\{ \| \boldsymbol{Z} \|_{l_{w,2,1}} - \boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{Z} + \frac{\beta_1}{2}\| \boldsymbol{Z}-\boldsymbol{\alpha} \|_2^2 \right\} \quad (23)$$

By simple manipulation, (23) is equivalent to the following problem:

$$\min_{\boldsymbol{Z}} \sum_{K=1}^{2^{2J}} \left\{ \boldsymbol{\omega}_K \| \boldsymbol{Z}_{G_K} \|_2 + \frac{\beta_1}{2} \| \boldsymbol{Z}_{G_K} - \boldsymbol{\alpha}_{G_K} - \frac{1}{\beta_1}(\boldsymbol{\mu})_{G_K} \|_2^2 \right\} \quad (24)$$

This can be solved by group-wise soft shrinkage:

$$\boldsymbol{Z}_{G_K} = \max\left\{ \| \boldsymbol{\gamma}_K \|_2 - \frac{\boldsymbol{\omega}_K}{\beta_1}, 0 \right\} \frac{\gamma_K}{\| \gamma_K \|_2} \qquad K = 1,2,\cdots,2^{2J} \quad (25)$$

where $\boldsymbol{\gamma}_K = \boldsymbol{\alpha}_{G_K} + \frac{1}{\beta_1}(\boldsymbol{\mu})_{G_K}$.

The multiplier $\boldsymbol{\mu}$ is updated by

$$\boldsymbol{\mu} \leftarrow \boldsymbol{\mu} + \tau(\boldsymbol{Z}-\boldsymbol{\alpha}) \quad (26)$$

where τ is the step length. In summary, we obtain the algorithm for solving model (17) by ADMM as shown in the following:

Algorithm 1 ADMM for model (17) when $\boldsymbol{H}$ is identity

Input: $\boldsymbol{b}_{\mathrm{H}}, \boldsymbol{Z}^0, \boldsymbol{\mu}^0, \beta_1, \beta_2, \tau$.

Iteration: For $k = 1,2,\cdots$, until a stopping criterion is reached,

$$\boldsymbol{\alpha}^k = \frac{1}{\beta_1+\beta_2}(\beta_1\boldsymbol{Z}^{k-1} - \boldsymbol{\mu}^{k-1} + \beta_2\boldsymbol{W}\boldsymbol{b}_{\mathrm{H}})$$

$$\boldsymbol{Z}^k = \mathrm{GShrink}\left(\boldsymbol{\alpha}^k + \frac{\boldsymbol{\mu}^{k-1}}{\beta_1}, \frac{\boldsymbol{\omega}}{\beta_2} \right)$$

$$\boldsymbol{\mu}^k = \boldsymbol{\mu}^{k-1} + \tau(\boldsymbol{Z}^k - \boldsymbol{\alpha}^k)$$

Output: $\boldsymbol{\alpha}^k$.

Note that regularizer $\| \cdot \|_{l_{w,2,1}}$ is non-smooth but convex, whereas the fidelity term is smooth and convex. Thus, in the following we apply the iterative shrinkage-thresholding algorithm (ISTA) to solve the general case of model (17). The general step of ISTA is described as

$$\boldsymbol{\alpha}_{k+1} = \mathrm{Prox}_{t_k g}(\boldsymbol{\alpha}_k - t_k\boldsymbol{W}\boldsymbol{H}^{\mathrm{T}}(\boldsymbol{H}\boldsymbol{W}^{\mathrm{T}}\boldsymbol{\alpha}_k - \boldsymbol{b}_{\mathrm{H}})) \quad (27)$$

where $g(\cdot) = \lambda \| \cdot \|_{l_{w,2,1}}$ is the mixed norm functional proposed in (15), and Prox is the proximal operator. The proximal algorithm is an efficient tool for non-smooth minimization problems. In proximal algorithms, the base operation is to evaluate the proximal operator of a function, which involves solving a small convex optimization problem. The proximal operator is defined by

$$\mathrm{Prox}_f(x) = \arg\min_u \left\{ f(x) + \frac{1}{2}\| x-u \|_2^2 \right\}$$

Therefore, the $\mathrm{Prox}_{t_k g}$ operation in (27) can be expressed as

$$\mathrm{Prox}_{t_k g}(x) = \arg\min_u \left\{ g(x) + \frac{1}{2t_k}\| x-u \|_2^2 \right\} \quad (28)$$

The Prox operation can be easily solved by group-wise shrinkage, that is

$$\mathrm{Prox}_{t_k g}(x) = \mathrm{GShink}(x, \lambda t_k \omega) \quad (29)$$

Although ISTA has the advantage of simplicity, it has also been recognized to be a slow method. The fast iterative shrinkage thresholding algorithm (FISTA) is the accelerated version of ISTA[19]. It not only preserves the computational simplicity, but also exhibits a significantly faster convergence rate than standard gradient-based methods. Algorithm 2 summarizes the FISTA for solving model (17), which is described as follows:

Algorithm 2 FISTA for model (17)

Input: $\boldsymbol{b}_{\mathrm{H}}, \boldsymbol{\alpha}^0 = \boldsymbol{0}, t_1 = 1, \boldsymbol{y}^1 = \boldsymbol{\alpha}^0, L, \lambda$.

Iteration: For $k = 1,2,\cdots$, until a stopping criterion is reached,

$$\boldsymbol{\alpha}^k = \mathrm{Prox}_{t_k g}\left(\boldsymbol{y}^k - \frac{1}{L}\boldsymbol{W}\boldsymbol{H}^{\mathrm{T}}(\boldsymbol{H}\boldsymbol{W}^{\mathrm{T}}\boldsymbol{y}^k - \boldsymbol{b}_{\mathrm{H}}) \right)$$

$$t_{k+1} = \frac{1+\sqrt{1+4t_k^2}}{2}$$

$$\boldsymbol{y}^{k+1} = \boldsymbol{\alpha}^k + \frac{t_k-1}{t_{k+1}}(\boldsymbol{\alpha}^k - \boldsymbol{\alpha}^{k-1})$$

Output: $\boldsymbol{\alpha}^k$.

4 Experiments

In this section, we present various experiments to show the performance of the proposed method for image denoising. In all the experiments, we chose the periodic "sym4" wavelet and scale $J = 6$. First, we describe the

group sparse approximation performance of the proposed model based on a noiseless case, as shown in Figs. 2 and 3. Fig. 2(a) shows the original "Baboon" image and Fig. 2(b) shows the plot of the vectorization of its LSD matrix. The vectorization is obtained by stacking the columns of the LSD matrix on top of one another. From the plot of vectorized LSD matrix, we can easily observe its sparsity.

Figs. 3 (a) to (d) show the group sparse approximation with parameters $\beta_2 = 0.001$, 0.002, 0.004 and 0.007, respectively. Figs. 3(e) to (h) show the corresponding vectorized LSD matrix. There are a total of 1 024 groups of wavelet coefficients, and the percentages of non-zero groups of the approximation image with different parameters are listed in Tab. 1.

Fig. 4 shows the evolution of the LSD matrix along with the noise level. Figs. 4(a) to (c) show the noisy Baboon images degraded by three levels of Gaussian noise with a mean of zero and standard deviations of 10, 20, and 30, respectively. Figs. 4(d) to (f) show the corresponding vectorized LSD matrix. We can see that the dynamic range of the LSD matrix becomes increasingly narrow with each increase in the standard deviation. The dynamic range of the LSD matrix is defined as

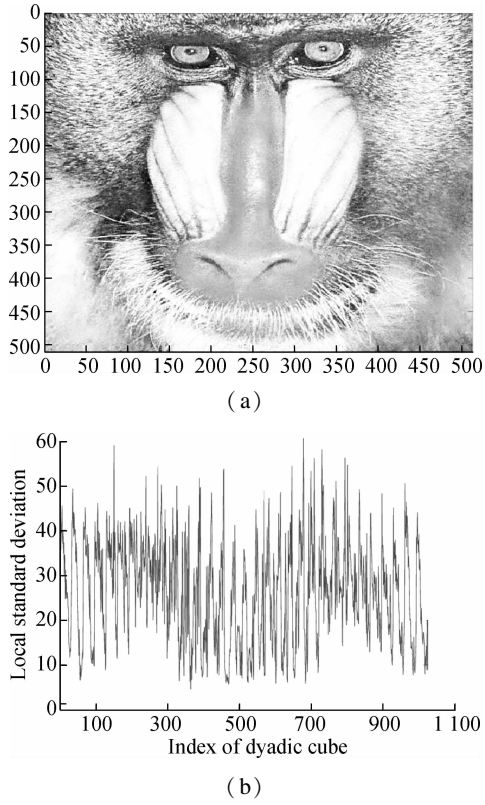

Fig. 2 Test image Baboon and the corresponding LSD matrix. (a) Original 512 × 512 Baboon image; (b) Vectorized LSD matrix of (a)

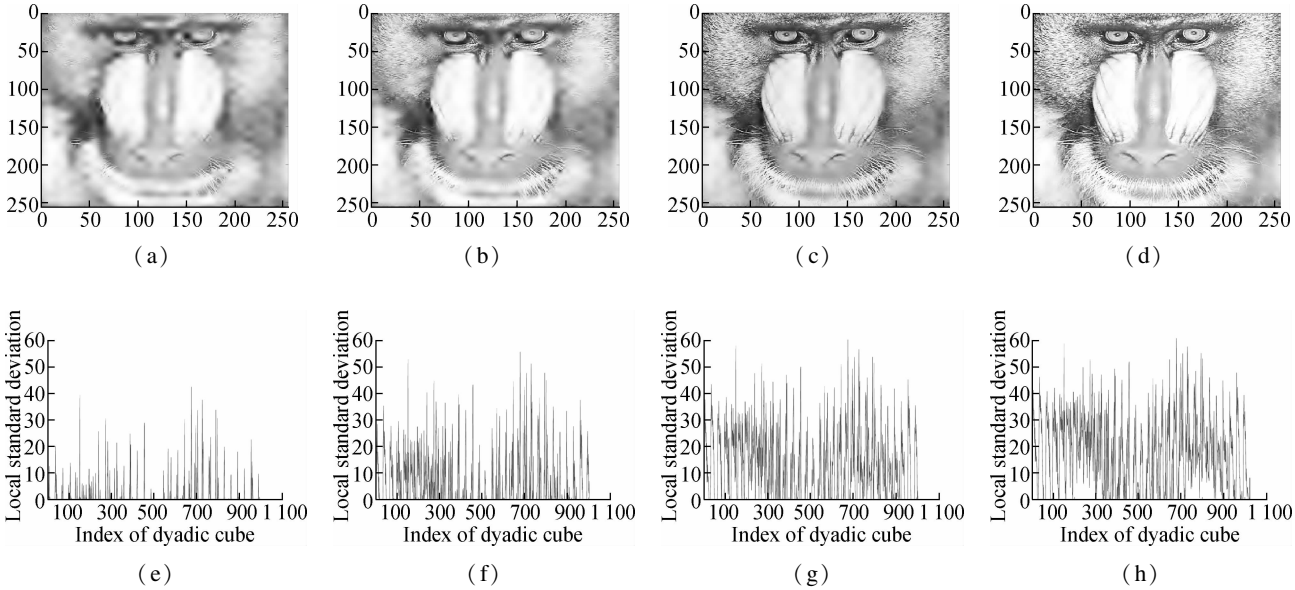

Fig. 3 Group sparse approximation result and the corresponding vectorized LSD matrix on noiseless image Baboon ($\beta_1 = 0.2$). (a) $\beta_2 = 0.001$; (b) $\beta_2 = 0.002$; (c) $\beta_2 = 0.004$; (d) $\beta_2 = 0.007$; (e) Vectorized LSD matrix of (a); (f) Vectorized LSD matrix of (b); (g) Vectorized LSD matrix of (c); (h) Vectorized LSD matrix of (d)

Tab. 1 Group sparse approximation performance with different parameters

Parameter β_2	Number of non-zero groups	Percentage/%
0.001	884	86.33
0.002	587	57.32
0.004	348	33.98
0.007	184	17.97

$$R = [\min_K P_{J,K}, \max_K P_{J,K}]$$

We apply the ADMM algorithm to equalize the LSD matrix of the noisy Baboon image ($\sigma = 15$). In Fig. 5 and Tab. 2, we present the experimental results to show how parameters β_1 and β_2 affect the group sparsity and the range of the LSD matrix. We can see that parameter

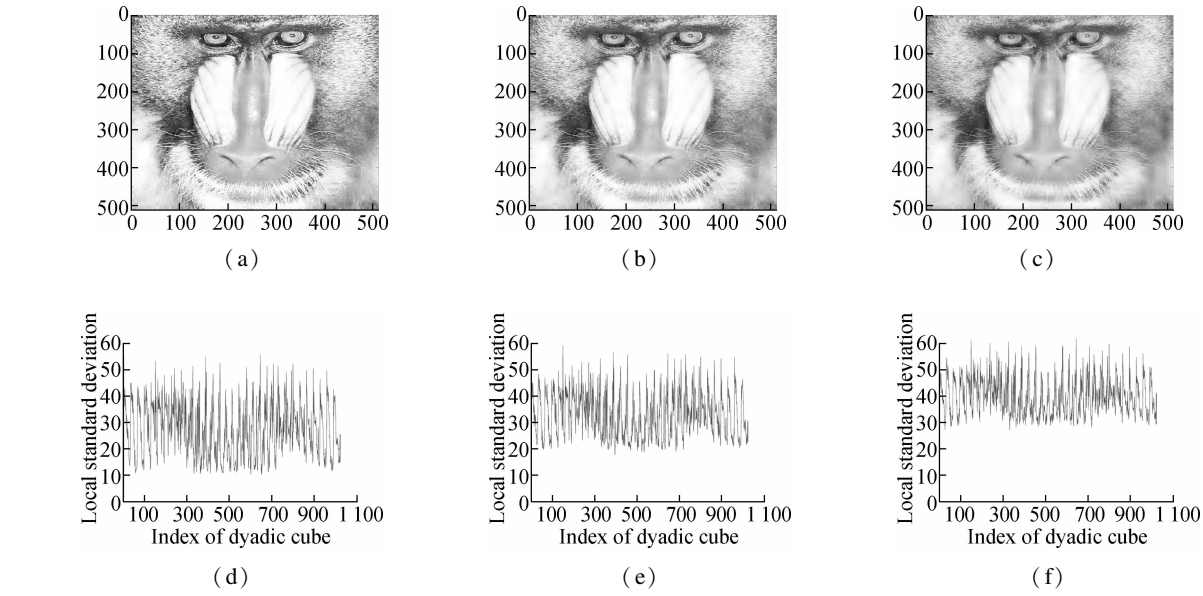

Fig. 4 Evolution of the LSD matrix along with the noise level. (a) Noisy image with noise std of 10; (b) Noisy image with noise std of 20; (c) Noisy images with noise std of 30; (d) Vectorized LSD matrix of (a); (e) Vectorized LSD matrix of (b); (f) Vectorized LSD matrix of (c)

β_2 controls the group sparsity of the solution and parameter β_1 controls the range of the LSD matrix. The group sparsity can be measured by the group sparsity rate ρ, which is the percentage of the non-zero groups.

In Figs. 6 to 8 and Tab. 3, the denoising performance of Algorithm 1 is compared with that of the wavelet shrinkage model, TV model, balanced frame-based model[8], and non-local mean model[5]. We can see that the proposed model outperforms the wavelet shrinkage model and frame-based model, and achieves a competitive performance as compared with the TV model and non-local mean model.

Tab. 2 Impact of parameters β_1 and β_2 on ρ and R

The setting of parameters β_1 and β_2		ρ	R
$\beta_1 = 0.03$	$\beta_2 = 0.002$	56.45	[0, 57.05]
	$\beta_2 = 0.003$	22.56	[0, 56.47]
	$\beta_2 = 0.005$	0	[2.53, 56.59]
	$\beta_2 = 0.007$	0	[5.58, 56.69]
$\beta_2 = 0.005$	$\beta_1 = 0.1$	0	[2.84, 56.31]
	$\beta_1 = 0.5$	0	[2.65, 48.83]
	$\beta_1 = 1.0$	0	[1.92, 35.62]
	$\beta_1 = 1.5$	0	[1.09, 27.93]

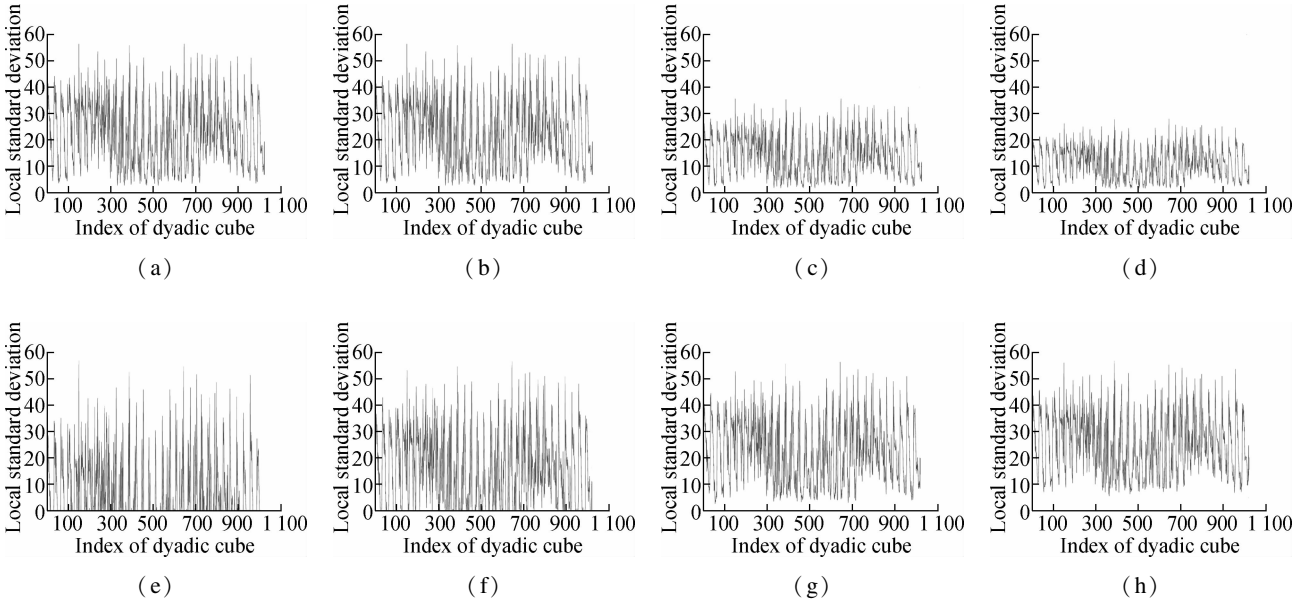

Fig. 5 Performance of parameters β_1 and β_2 for controlling ρ and R. (a) $\beta_1 = 0.1, \beta_2 = 0.005$; (b) $\beta_1 = 0.5, \beta_2 = 0.005$; (c) $\beta_1 = 1$, $\beta_2 = 0.005$; (d) $\beta_1 = 1.5, \beta_2 = 0.005$; (e) $\beta_1 = 0.03, \beta_2 = 0.002$; (f) $\beta_1 = 0.03, \beta_2 = 0.003$; (g) $\beta_1 = 0.03, \beta_2 = 0.005$; (h) $\beta_1 = 0.03, \beta_2 = 0.007$

Tab. 3 Comparisons of PSNR performance with different models

Model	Hill($\sigma=10$)	House($\sigma=10$)	Parrot($\sigma=15$)	Leopard($\sigma=15$)	Baboon($\sigma=15$)	Barbara($\sigma=20$)	Man($\sigma=20$)
Soft shrinkage	26.62	26.45	23.64	17.93	21.18	23.12	24.98
Frame	28.09	30.40	25.87	23.74	22.05	24.31	26.74
TV	29.93	31.73	28.61	25.78	26.03	26.32	29.14
Nonlocal mean	30.12	32.41	29.19	25.40	26.77	30.36	29.16
Proposed model	30.75	32.34	29.05	26.36	26.82	26.86	28.90

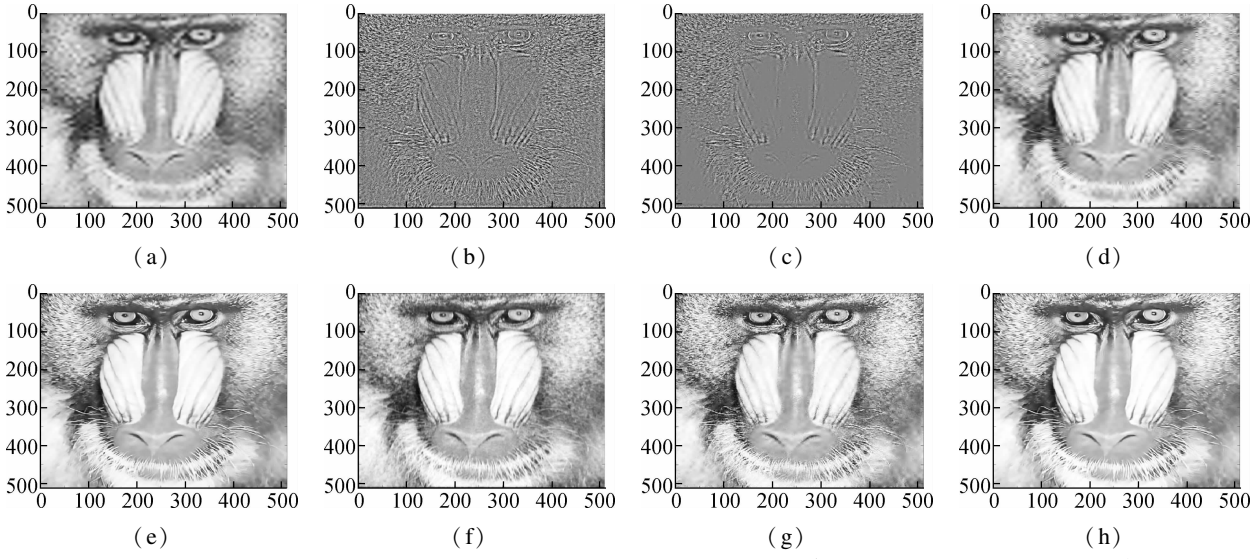

Fig. 6 Comparison of denoising results by different models on the Baboon image (noise standard deviation $\sigma=15$). (a) Low frequency component $\boldsymbol{b}_{\mathrm{L}}$ by wavelet decomposition; (b) High frequency component $\boldsymbol{b}_{\mathrm{H}}$ by wavelet decomposition; (c) Restored high frequency by Algorithm 1; (d) Restored image by soft shrinkage model; (e) Restored image by balanced frame-based model; (f) Restored image by TV model; (g) Restored image by non-local mean model; (h) Restored image by Algorithm 1

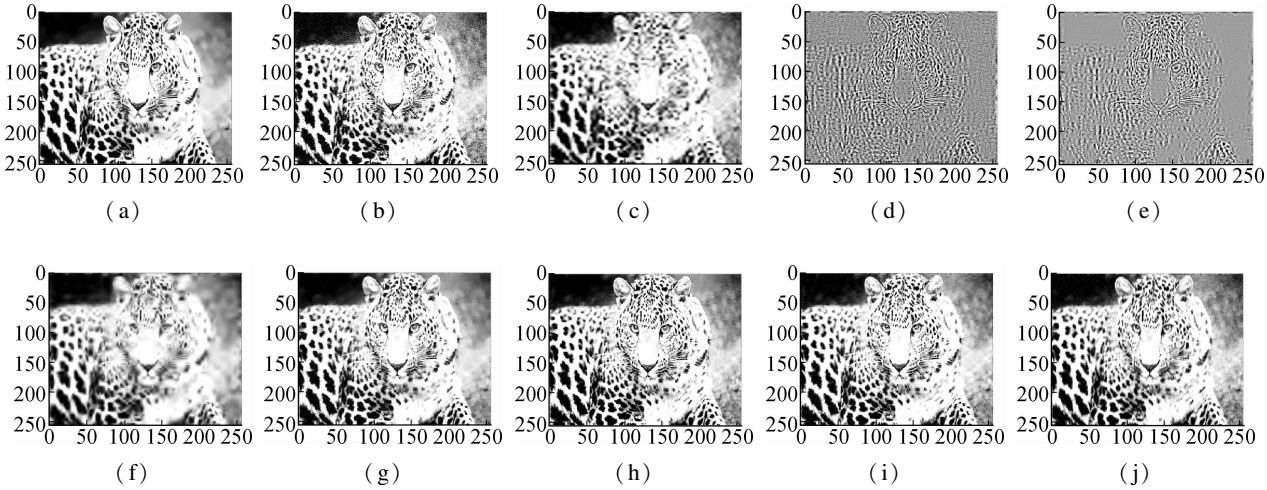

Fig. 7 Comparison of denoising results by different models on the Leopard image (noise standard deviation $\sigma=15$). (a) Original image; (b) Noisy version; (c) Low frequency $\boldsymbol{b}_{\mathrm{L}}$ by wavelet decomposition; (d) High frequency $\boldsymbol{b}_{\mathrm{H}}$ by wavelet decomposition; (e) Restored high frequency by Algorithm 1; (f) Restored image by soft shrinkage model; (g) Restored image by balanced frame-based model; (h) Restored image by TV model; (i) Restored image by non-local mean model; (j) Restored image by Algorithm 1

Fig. 9 shows the results of Algorithm 2 on the woman image. The test image is contaminated by salt and pepper noise with noise intensity $d=10\%$. After wavelet decomposition, most of the noises are concentrated in the high frequency component, which can be seen in Fig. 9(c). Fig. 9(d) shows the restored high-frequency component by the proposed model. One can see that most of the noises are removed and the textures are well preserved. Figs. 9(e) to (g) compare the restored image by our method with that by the balanced frame-based method and non-local mean method. Particularly in the tagged region, we can see that the proposed method preserves more textures and achieves a better subjective visual effect.

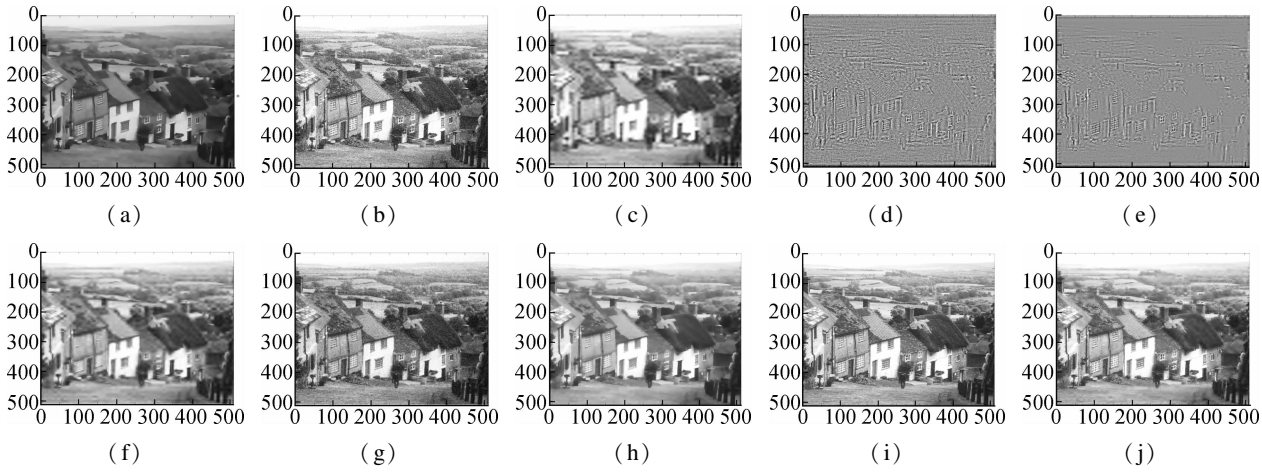

Fig. 8 Comparison of denoising results by different models on the Hill image (noise standard deviation $\sigma = 10$). (a) Original image; (b) Noisy version; (c) Low frequency component $\boldsymbol{b}_{\mathrm{L}}$ by wavelet decomposition; (d) High frequency component $\boldsymbol{b}_{\mathrm{H}}$ by wavelet decomposition; (e) Restored high frequency component by Algorithm 1; (f) Restored image by soft shrinkage model; (g) Restored image by balanced frame based model; (h) Restored image by TV model; (i) Restored image by non-local mean model; (j) Restored image by Algorithm 1

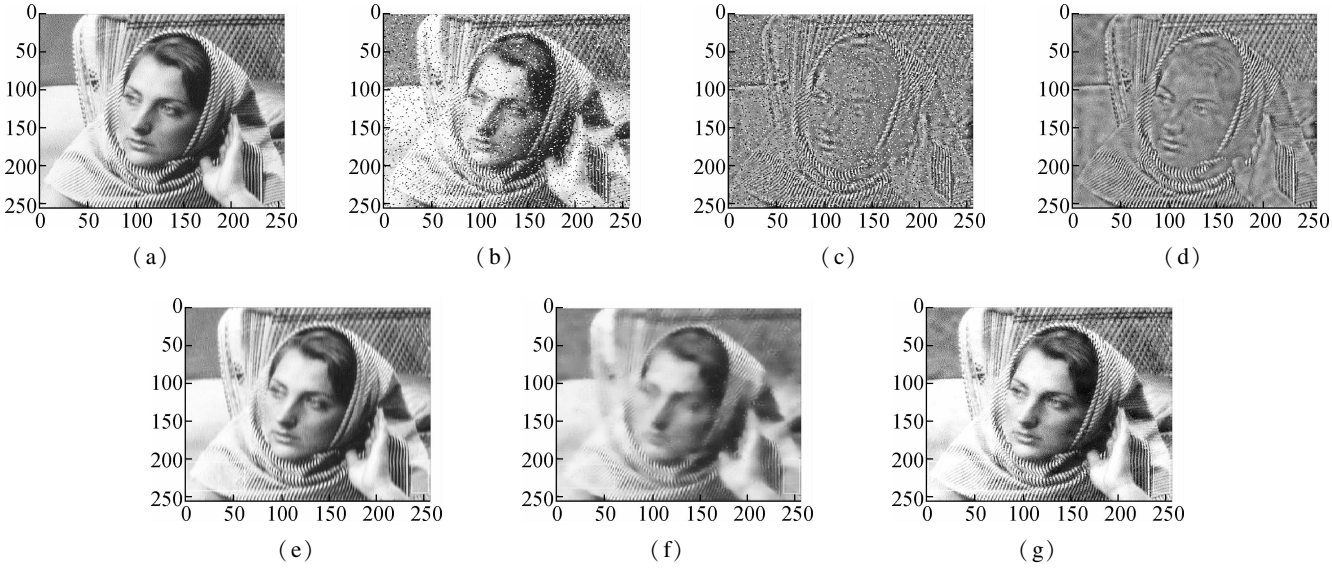

Fig. 9 Salt and pepper noise removal by Algorithm 2. (a) Original image (256×256); (b) Image contaminated by salt and pepper noise (noise intensity $d = 10\%$); (c) High-frequency component of (b); (d) Restored high-frequency component; (e) Restored image by balanced frame-based model; (f) Restored image by the non-local mean model; (g) Restored image by Algorithm 2

5 Conclusions

1) Based on the wavelet characterization of the LSD, we find that the l_1 norm of the LSD matrix is equivalent to a $l_{2,1}$ mixed norm of the wavelet coefficients. This mixed norm introduces a coupling between wavelet coefficients and determines a variable group scheme.

2) We find that enforcing the sparsity of the LSD can efficiently equalize the LSD distribution. Thus, we propose a novel group sparse optimization model in the wavelet domain for image denoising. Encoding the group information in addition to sparsity leads to better signal feature selection.

3) The group sparse optimization problem is more difficult to solve than the conventional l_1 norm regularized problem. We applied the ADMM and FISTA to solve the group sparse model efficiently based on different cases.

4) Several experiments reveal that the proposed group sparse model outperforms traditional wavelet-based model and frame-based model, and has competitive performance with TV and non-local mean image restoration models.

References

[1] Tikhonov A N. Regularization of incorrectly posed problems [J]. *Soviet Mathematics Doklady*, 1963, **4**: 1624 – 1627.

[2] Rudin L I, Osher S, Fatemi E. Nonlinear total variation based noise removal algorithms[J]. *Physica D: Nonlinear Phenomena*, 1992, **60**(1/2/3/4): 259 – 268. DOI:10.1016/0167-2789(92)90242-f.

[3] Buades A, Coll B, Morel J M. Nonlocal image and mov-

ie denoising[J]. *International Journal of Computer Vision*, 2008, **76**(2): 123 - 139. DOI:10.1007/s11263-007 - 0052-1.

[4] Ding D, Ram S, Rodriguez J J. Image inpainting using nonlocal texture matching and nonlinear filtering[J]. *IEEE Transactions on Image Processing*, 2019, **28**(4): 1705 - 1719. DOI:10.1109/tip.2018.2880681.

[5] Wang W, Li F, Ng M K. Structural similarity-based nonlocal variational models for image restoration[J]. *IEEE Transactions on Image Processing*, 2019, **28**(9): 4260 - 4272. DOI:10.1109/tip.2019.2906491.

[6] Zhang K, Zuo W, Chen Y J, et al. Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising[J]. *IEEE Transactions on Image Processing*, 2017, **26**(7): 3142 - 3155. DOI:10.1109/tip.2017.2662206.

[7] Jin K H, McCann M T, Froustey E, et al. Deep convolutional neural network for inverse problems in imaging[J]. *IEEE Transactions on Image Processing*, 2017, **26**(9): 4509 - 4522. DOI:10.1109/tip.2017.2713099.

[8] Starck J L, Elad M, Donoho D L. Image decomposition via the combination of sparse representations and a variational approach[J]. *IEEE Transactions on Image Processing*, 2005, **14**(10): 1570 - 1582. DOI:10.1109/tip.2005.852206.

[9] Cai J F, Osher S, Shen Z. Split bregman methods and frame based image restoration[J]. *Multiscale Modeling & Simulation*, 2010, **8**(2): 337 - 369. DOI:10.1137/090753504.

[10] Dong W S, Shi G M, Ma Y, et al. Image restoration via simultaneous sparse coding: Where structured sparsity meets Gaussian scale mixture[J]. *International Journal of Computer Vision*, 2015, **114**(2/3): 217 - 232. DOI:10.1007/s11263-015 - 0808-y.

[11] Zhang L, Zuo W. Image restoration: From sparse and low-rank priors to deep priors[J]. *IEEE Signal Processing Magazine*, 2017, **34**(5): 172 - 179. DOI:10.1109/msp.2017.2717489.

[12] Zhang J, Zhao D, Gao W. Group-based sparse representation for image restoration[J]. *IEEE Transactions on Image Processing*, 2014, **23**(8): 3336 - 3351. DOI:10.1109/tip.2014.2323127.

[13] Kowalski M. Sparse regression using mixed norms[J]. *Applied and Computational Harmonic Analysis*, 2009, **27**(3): 303 - 324. DOI:10.1016/j.acha.2009.05.006.

[14] Yang J Q, Zhu G P, Shi Y Q. Analyzing the effect of JPEG compression on local variance of image intensity[J]. *IEEE Transactions on Image Processing*, 2016, **25**(6): 2647 - 2656. DOI:10.1109/tip.2016.2553521.

[15] Li Z G, Zeng L, Xu Y F, et al. Level set method for image segmentation based on local variance and improved intensity inhomogeneity model[J]. *IET Image Processing*, 2016, **10**(12): 1007 - 1016. DOI:10.1049/iet-ipr.2016.0352.

[16] Meyer Y. *Wavelet and operators*[M]. Cambridge: Cambridge University Pressing, 1992: 21 - 25.

[17] Zhang T, Fan Q B. Wavelet characterization of dyadic BMO norm and its application in image decomposition for distinguishing between texture and noise[J]. *International Journal of Wavelets, Multiresolution and Information Processing*, 2011, **9**(3): 445 - 457. DOI:10.1142/s0219691311004183.

[18] Boyd S. Distributed optimization and statistical learning via the alternating direction method of multipliers[J]. *Foundations and Trends© in Machine Learning*, 2010, **3**(1): 1 - 122. DOI:10.1561/2200000016.

[19] Beck A, Teboulle M. Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems[J]. *IEEE Transactions on Image Processing*, 2009, **18**(11): 2419 - 2434. DOI:10.1109/tip.2009.2028250.

基于小波系数树状结构的组稀疏图像去噪方法

张 涛 魏海广 莫绪涛

(安徽工业大学数理科学与工程学院,马鞍山 243000)

摘要:为了提高图像对比度以及图像质量,受图像的局部均方差变化范围会随噪声强度的增强而变窄这一有趣的现象启发,在小波域中提出了一种基于小波系数树状结构的组稀疏图像去噪模型.由噪声导致的图像局部均方差变化范围的压缩会引起图像对比度以及图像质量的下降.为了平衡图像的局部均方差分布,引入一种混合范数作为图像去噪模型的正则项,以达到对局部均方差矩阵进行稀疏约束的目的.该混合范数引入了小波系数之间的耦合,并且给出了一种小波系数的树状分组方式.利用交替方向乘子法(ADMM)以及快速迭代阈值收缩算法(FISTA)研究模型在不同情况下的求解方法.最后,通过多组实验验证了所提模型的有效性.实验结果表明,提出的组稀疏优化模型能够有效地平衡图像的局部均方差分布,从而提高图像的对比度和图像质量.

关键词:局部均方差;组稀疏; 图像去噪; 混合范数; 纹理;

中图分类号:TP751