Regularization with Multilevel Non-stationary Tight Framelets for Image Restoration

Yan-Ran Li *   Raymond H. F. Chan †   Lixin Shen‡   Xiaosheng Zhuang§

February 7, 2021

Abstract

Variational regularization models are one of the popular and efficient approaches for image restoration. The regularization functional in the model carries prior knowledge about the image to be restored. The prior knowledge, in particular for natural images, are the first-order (i.e. variance in luminance) and second-order (i.e. contrast and texture) information. In this paper, we propose a model for image restoration, using a multilevel non-stationary tight framelet system that can capture the image's first-order and second-order information. We develop an algorithm to solve the proposed model and the numerical experiments show that the model is effective and efficient as compared to other higher-order models.

1   Introduction

The restoration of a degraded image may be modeled as

$$z = Ku + \epsilon, \tag{1}$$

where u denotes the unknown image to be recovered, K a blurring matrix, z an observed blurred image, and ϵ the noise. In general, K is a singular or near-singular matrix and hence the problem of finding the solution u from model (1) is ill-posed. To overcome the difficulties caused by the ill-posedness, regularization techniques such as total-variation regularization and multiscale regularization are often adopted, see [4, 6, 12, 14, 29] and the references therein. The resulting regularized image models have the following generic form

$$\min_u \{\mathcal{F}(u) + \alpha\mathcal{G}(u)\}, \quad \alpha > 0 \tag{2}$$

where α is the regularization parameter, $\mathcal{F}$ represents the data fidelity term and $\mathcal{G}$ the regularization term. The fidelity term measures the closeness of the estimate obtained from (2) to the data z

1

while the regularization term is used to arrive at a sensible solution. Generally speaking, model (2) integrates knowledge about how data is generated in the fidelity term $\mathcal{F}$ with the regularization functional $\mathcal{G}$ that carries prior knowledge about the image to be restored.

Our main focus of this paper is to choose a proper regularization $\mathcal{G}$ in (2) for image restoration. Here, a proper regularization means a regularization functional that encodes prior knowledge about the image to be restored. Prior knowledge about images, in particular for natural images, includes first-order (i.e. variance in luminance) and second-order (i.e. contrast and texture) information [19]. One commonly used regularization term that exploits the first-order information is the bounded variation semi-norm [29]

$$\mathcal{G}(u) := \int_{\Omega} |\nabla u|, \tag{3}$$

where the image u is defined on the bounded set $\Omega \subset \mathbb{R}^d$. The corresponding model (2), referred as the total variation (TV) based image restoration model, performs incredibly well especially if the image to be reconstructed is piecewise constant. The total variation functional does not penalize discontinuities in images and thus allows us to recover the edges of the original image. However, it does not distinguish between jumps and smooth transitions, therefore it tends to give piecewise constant images with staircase artifacts. Due to this notably staircase phenomenon, the TV-based model is not suited for reconstructing images that are not nearly piecewise constant. It was pointed out in [13] that whereas the reconstruction generated with the first-order model will display jumps, the basic geometric structure of the original intensity surfaces is missing, even if it appears in the data. It was further mentioned that using higher order models, these artifacts from the first-order model can be eliminated and some of the fine geometric structures, particularly planar and quadric patches, of the original image can be recovered.

One of the earliest models using higher derivatives was proposed in [5] where the infimal convolution of the first and second order derivatives was proposed as regularizer

$$\mathcal{G}(u) := \inf_{v} \int_{\Omega} |\nabla u - \nabla v| + \alpha |\nabla(\nabla v)|. \tag{4}$$

It approximates locally the gradient of the function u by ∇v, that itself has a low total variation. Different second-order functionals for staircase reduction have been considered in other papers, for example, see [8, 26]. Based on tensor algebra, the regularizer with derivatives of arbitrary order was introduced in [2]. The corresponding regularizer was called total generalized variation (TGV). In particular, the TGV of second-order is

$$\mathcal{G}(u) := \min_{v} \int_{\Omega} (\alpha_1 |\nabla u - v| + \alpha_2 |\mathcal{E}(v)|), \tag{5}$$

where the parameters α_1, α_2 are positive, and

$$\mathcal{E}(v) = \begin{bmatrix} \partial_1 v_1 & \frac{1}{2}(\partial_1 v_2 + \partial_2 v_1) \\ \frac{1}{2}(\partial_1 v_2 + \partial_2 v_1) & \partial_2 v_2 \end{bmatrix}$$

with v_1 and v_2 being the components of v. Note that for twice differentiable u, $\mathcal{E}(\nabla u)$ is the Hessian of u. We note that the TGV of second-order (5) is similar to, but structural different from, the regularizer (4). The use of TGV and its variants in a plethora of applications has been reported in [1, 30] and the references therein.

Motivated from the fact that an image/signal naturally has a hierarchical structure and allows to be represented in a multiscale structure, we exploit this structure to formulate a regularization term $\mathcal{G}$ in (2) that is different from the aforementioned ones. To this end, we first construct a two-level

2

non-stationary tight framelet system that is suitable for representing images to be restored. More specifically, the tight framelet system in the first level is the directional Haar framelet (DHF) system introduced in our recent work [21] while the one in the second level is constructed from the discrete cosine transform (DCT). We then use the framelet coefficients of an image under this two-level non-stationary tight framelet system to formulate the regularization term $\mathcal{G}$. More precisely, the framelet coefficients of the image with the DHF consist of the first-order information of the image in the vertical, horizontal, and $\pm 45^\circ$ directions. As a result, the regularization term $\mathcal{G}$ contains not only the TV term but also ameliorates it by including the diagonal information. The coarse approximation to this image resulting from the low-pass filter of the DHF, considered as a smooth version of this image, will facilitate the extraction of the second-order information of the image. As shown in our previous work [24], the second-order information of the image can be reliably extracted from the DCT-based tight framelet coefficients of this smoothed image. Our proposed regularization term $\mathcal{G}$ also includes these second-order information. We remark that the success of tight framelets have been proven to be useful in image processing, see, e.g., [3, 7, 22, 23, 24, 31] and the references therein. However, despite that our two-level non-stationary tight framelet system is new, the proposed regularization is also different from the existing ones in the following perspectives:

- Due to the DHF, our regularization assimilates the advantages of both the total variation regularization and other framelet regularizations, and remedies their drawbacks. On the one hand, the filters associated with DHF have the shortest support among all tight framelet systems, therefore, it can suppress ringing artifacts arising from other framelet regularizations. In comparison, the filters associated with the 2-dimensional orthogonal Haar wavelet have the shortest support only among all compactly supported orthogonal wavelets. On the other hand, the diagonal first-order information provided by the DHF can reduce the staircase artifacts (or block effect) arising from the classical TV regularization. In comparison, the 2-dimensional orthogonal Haar wavelet only provide first-order information in the vertical and horizontal directions.

- We exploit the second-order information of the underlying image from its smoothed version rather than from the image itself. The main idea behind it is that the high frequency spatial information of the image will be suppressed in its smoothed one and therefore the second-order information of the image will be faithfully computed, in particularly, for images with high degree of noise.

- Finally, the properties of the tight framelet can be easily exploited to develop algorithms with computational efficiency and to analyze the convergence of the resulting algorithms.

To summarize, the proposed regularizer $\mathcal{G}$ contains the first and second order information of the image to be constructed for (2). The resulting optimization problem (2) can be efficiently solved and the efficiency and accuracy of this regularizer will be confirmed for image restoration.

The rest of this paper is organized as follows. In Section 2 we first briefly review the tight framelet systems, we then propose an image restoration model regularized by a two-level non-stationary tight framelet system and develop an algorithm to solve this model. The performance of the proposed model for image restoration is presented in Section 3.

2 Model and Algorithm with Multi-Level Non-Stationary Tight Framelets

This section consists of three parts. In the first part, we briefly review the multi-level non-stationary tight framelet systems. In the second part, we propose our image restoration model using a two-level non-stationary tight framelet system. In the last part, we propose an algorithm to solve the resulting optimization problem.

2.1 Multi-Level Non-Stationary Tight Framelets

Tight framelets are closely related to filter banks. A tight framelet filter bank can be used to (sparsely) represent data sequences through its associated discrete framelet transforms as well as its underlying discrete affine system [17]. Before proceeding to their connections, let us recall some definitions and notation first.

By $l(\mathbb{Z}^d)$ we denote the set of all sequences and $l_0(\mathbb{Z}^d)$ the set of all finitely supported sequences. A *filter* or *mask* $h = \{h(k)\}_{k \in \mathbb{Z}^d} : \mathbb{Z}^d \to \mathbb{C}$ on $\mathbb{Z}^d$ is a sequence in $l_0(\mathbb{Z}^d)$. For a filter $h \in l_0(\mathbb{Z}^d)$, its *Fourier series* is defined to be $\widehat{h}(\xi) := \sum_{k \in \mathbb{Z}^d} h(k) e^{-ik \cdot \xi}$ for $\xi \in \mathbb{R}^d$, which is a $2\pi \mathbb{Z}^d$-periodic trigonometric polynomial. In particular, by $\boldsymbol{\delta}$ we denote *the Dirac sequence* such that $\boldsymbol{\delta}(0) = 1$ and $\boldsymbol{\delta}(k) = 0$ for all $k \in \mathbb{Z}^d \backslash \{0\}$. Throughout the paper, we assume the tight framelets are *dyadic dilated*, that is, the *dilation matrix* is $2I_d$ with I_d the $d \times d$ identity matrix.

For filters $\tau_0, \tau_1, \ldots, \tau_s \in l_0(\mathbb{Z}^d)$, we say that a filter bank $\{\tau_0; \tau_1, \ldots, \tau_s\}$ is *a (d-dimension dyadic) tight framelet filter bank* if

$$\sum_{\ell=0}^{s} \widehat{\tau_\ell}(\xi) \overline{\widehat{\tau_\ell}(\xi + \pi \omega)} = \boldsymbol{\delta}(\omega), \qquad \forall \xi \in \mathbb{R}^d, \omega \in \{0, 1\}^d, \tag{6}$$

where for a number $x \in \mathbb{C}$, $\bar{x}$ denotes its complex conjugate. Equation (6) is equivalent to the *perfect reconstruction property* of the discrete framelet transforms associated with the filter bank $\{\tau_0; \tau_1, \ldots, \tau_s\}$ ([18, Theorems 1.1.1 and 1.1.4]). The filter τ_0 is usually a *low-pass filter* satisfying $\widehat{\tau_0}(0) = 1$ while τ_ℓ's are the *high-pass filters* satisfying $\widehat{\tau_\ell}(0) = 0$ for $\ell \geqslant 1$.

In practice, multi-level decomposition and reconstruction of data using *discrete framelet transform* associated with tight framelet filter banks are commonly used in order to exploit the sparse property of the data. Moreover, in signal/image processing, translation invariance property of a discrete framelet transform is desirable especially in the scenario of signal denoising/inpainting. To preserve the translation invariance property, one usually considers the redundant version of discrete framelet transform, that is, the *undecimated discrete framelet transform* (UDFmT). More precisely, denote a filter bank at level j as $\boldsymbol{\eta}_j := \{\tau_0^j; \tau_1^j, \ldots, \tau_{s_j}^j\}$ and consider a sequence $\{\boldsymbol{\eta}_j : j = 1, \ldots, J\} = \cup_{j=1}^{J} \{\tau_0^j; \tau_1^j, \ldots, \tau_{s_j}^j\}$ of J filter banks with $j = J \geqslant 1$ being the finest level and $j = 1$ being the coarsest level. Let the *convolution* operation $*$ be defined by $[h * v](\gamma) := \sum_{k \in \mathbb{Z}^d} h(\gamma - k) v(k)$, for $v \in l(\mathbb{Z}^d)$, $h \in l_0(\mathbb{Z}^d)$, $\gamma \in \mathbb{Z}^d$, and the *upsampling operator* $\uparrow m$ with $m \in \mathbb{N}$ be given by

$$[v \uparrow m](\gamma) := \begin{cases} v(m^{-1}\gamma), & \text{if } m^{-1}\gamma \in \mathbb{Z}^d; \\ 0, & \text{otherwise.} \end{cases}$$

For a filter h, let $h^\star$ be a filter defined by $h^\star(k) = \overline{h(-k)}$, $k \in \mathbb{Z}^d$. Then, for a given input data sequence $v = v_J$, the UDFmT includes (i) Decomposition:

$$v_{j-1} = v_j * ((\tau_0^j)^\star \uparrow 2^{J-j}), \quad w_{j-1;\ell} = v_j * ((\tau_\ell^j)^\star \uparrow 2^{J-j}), \quad \ell = 1, \ldots, s_j, \quad j = J, \ldots, 1, \tag{7}$$

4

and (ii) Reconstruction:

$$v_j = v_{j-1} * (\tau_0^j \uparrow 2^{J-j}) + \sum_{\ell=1}^{s} w_{j-1;\ell} * (\tau_\ell^j \uparrow 2^{J-j}), \quad j = 1, \ldots, J. \tag{8}$$

One can show that if each filter bank $\{\tau_0^j; \tau_1^j, \ldots, \tau_{s_j}^j\}$ satisfies the *partition of unity condition*: $\sum_{\ell=0}^{s_j} |\widehat{\tau_\ell^j}(\xi)|^2 = 1$, $\xi \in \mathbb{R}^d$, then any input data sequence $v \in l(\mathbb{Z}^d)$ can be perfectly reconstructed via (8) from its framelet coefficient sequences $\{v_0\} \cup \{w_{j;\ell} : \ell = 1, \ldots, s_j\}_{j=1}^J$ decomposed from (7). The framelet system associated with such a sequence $\{\boldsymbol{\eta}_j : j = 1, \ldots, J\}$ is then called a *multi-level non-stationary tight framelet system*.

In this paper, we consider $J = 2$, that is, two-level non-stationary tight framelet system. One can of course consider $J > 2$. However, in terms of efficiency and simplicity, $J = 2$ is the best choice for the development of this paper.

2.2 Regularization with a Two-level Non-stationary Tight Framelet System

In this subsection, we integrate two different tight framelet systems as a two-level non-stationary tight framelet system which will be exploited for the optimization problem (2).

The tight framelets in the first level is the directional Haar framelet (DHF) system proposed in [21]. The filters associated with this DHF are

$$\tau_0 = \tfrac{1}{4} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \quad \tau_1 = \tfrac{1}{4} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \quad \tau_2 = \tfrac{1}{4} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \quad \tau_3 = \tfrac{1}{4} \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix},$$
$$\tau_4 = \tfrac{1}{4} \begin{bmatrix} 1 & 0 \\ -1 & 0 \end{bmatrix}, \quad \tau_5 = \tfrac{1}{4} \begin{bmatrix} 0 & 0 \\ 1 & -1 \end{bmatrix}, \quad \tau_6 = \tfrac{1}{4} \begin{bmatrix} 0 & 1 \\ 0 & -1 \end{bmatrix}.$$

As two-dimensional filters, the indices of the entries (top-left, top-right, bottom-left, and bottom-right) in each filter are $(0,0)$, $(0,1)$, $(1,0)$, and $(1,1)$, respectively. The first filter τ_0 is a low-pass filter and the rest are high-pass filters that have the ability to provide directional information of an image when these filters are applied to the image. More precisely, the filters τ_1 and τ_2 act as the first-order difference operators in the $45°$ and $135°$ directions, respectively. The results of these two filters convolving with an image will highlight changes in intensity of the image in these two diagonal directions. The filters τ_3 and τ_5 are the first-order difference operators in the horizontal direction while the filters τ_4 and τ_6 are the first-order difference operators in the vertical direction. The convolutions of these filters with the underlying image are the coefficients of the image under the corresponding filters, which are the multiplications of some associated transformation matrices with the image.

Now, we propose a generic regularization term based on DHF. Let $u \in \mathbb{R}^n$ be the vector representing the column-stacked version of an image. We denote by M_κ the associated matrix representation of the filters τ_κ, $\kappa = 0, 1, \ldots, 6$, under a proper boundary condition. We further denote

$$B_{1\ell} := M_0 \quad \text{and} \quad B_{1h} := [M_1^\top, \ldots, M_6^\top]^\top. \tag{9}$$

By the tight frame property of $\{\tau_0; \tau_1, \ldots, \tau_s\}$, these two matrices satisfy the following perfect reconstruction condition

$$B_{1\ell}^\top B_{1\ell} + B_{1h}^\top B_{1h} = I.$$

Let $\Phi_{1\Lambda} : \mathbb{R}^{6n} \to \mathbb{R}$ be defined through a function $\varphi_1 : \mathbb{R}^6 \to \mathbb{R}$ and a non-negative parameter vector $\Lambda = [\lambda_1, \lambda_2, \ldots, \lambda_n]$ as follows

$$\Phi_{1\Lambda}(v) := \sum_{i=1}^{n} \lambda_i \varphi_1(v_i, v_{i+n}, \ldots, v_{i+5n}). \tag{10}$$

With this function $\Phi_{1\Lambda}$, we propose a functional based on DHF in the following form

$$\mathcal{G}_1(u) := \Phi_{1\Lambda}(B_{1h}u), \tag{11}$$

from which the TV regularization and its variants can be derived by properly chosen φ_1 in (10). For example, if we choose $\varphi_1(x_1, x_2, x_3, x_4, x_5, x_6) = |x_3| + |x_4|$, the regularization in (11) is reduced to the so-called anisotropic TV; If we choose $\varphi_1(x_1, x_2, x_3, x_4, x_5, x_6) = \sqrt{|x_3|^2 + |x_4|^2}$, the regularization in (11) is reduced to the so-called isotropic TV.

We choose, in this paper,

$$\varphi_1(x_1, x_2, x_3, x_4, x_5, x_6) = \sqrt{|x_1|^2 + |x_2|^2} + \sqrt{|x_3|^2 + |x_4|^2}. \tag{12}$$

One of the advantages of the regularization $\mathcal{G}_1$ with φ_1 given in (12) is that it assimilates the advantages of both total variation and wavelet regularizations and remedies their drawbacks. The way of avoiding or suppressing ringing artifacts arising from wavelet regularizations is to choose a wavelet system whose filters have small supports. The filters associated with the 2-dimensional orthogonal Haar wavelet have the shortest support among all compactly supported orthogonal wavelets, but the staircase artifacts (or blocky effect) will appear in the neighborhoods of edges in the directions about $\pm 45°$. Since φ_1 in (12) includes the diagonal first-order information from the filters τ_1 and τ_2, the staircase artifacts can be reduced.

The tight framelet in the second level is generated from the standard 3×3 DCT-II orthogonal matrix whose three rows are $c_0 = \frac{\sqrt{3}}{3}[1, 1, 1]$, $c_1 = \frac{\sqrt{2}}{2}[1, 0, -1]$, and $c_2 = \frac{\sqrt{6}}{6}[1, -2, 1]$. In the sequel, this system is referred to as the DCT-based tight framelet system. The filters of the DCT-based tight framelet system are $\tau_{3i+j} = \frac{1}{3} c_i^\top c_j$ with $i, j \in \{0, 1, 2\}$, where τ_0 is the low-pass filter and the others are high-pass filters. Here, for simplicity of notation, we use τ_κ to denote the filters associated with both the DHF or DCT-based tight framelet. The expansions of these filters are

$$\tau_0 = \frac{1}{9} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \quad \tau_1 = \frac{\sqrt{6}}{18} \begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix}, \quad \tau_2 = \frac{\sqrt{2}}{18} \begin{bmatrix} 1 & -2 & 1 \\ 1 & -2 & 1 \\ 1 & -2 & 1 \end{bmatrix},$$

$$\tau_3 = \frac{\sqrt{6}}{18} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix}, \quad \tau_4 = \frac{1}{6} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix}, \quad \tau_5 = \frac{\sqrt{3}}{18} \begin{bmatrix} 1 & -2 & 1 \\ 0 & 0 & 0 \\ -1 & 2 & -1 \end{bmatrix},$$

$$\tau_6 = \frac{\sqrt{2}}{18} \begin{bmatrix} 1 & 1 & 1 \\ -2 & -2 & -2 \\ 1 & 1 & 1 \end{bmatrix}, \quad \tau_7 = \frac{\sqrt{3}}{18} \begin{bmatrix} 1 & 0 & -1 \\ -2 & 0 & 2 \\ 1 & 0 & -1 \end{bmatrix}, \quad \tau_8 = \frac{1}{18} \begin{bmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{bmatrix}.$$

The filters τ_1 and τ_3, known as the Prewitt operator in image processing, are used to compute an approximation of the gradient (i.e., the first-order information) of the image intensity function. The convolution of τ_1 (resp. τ_3) with an image gives the horizontal (resp. vertical) changes of the image intensity and they compute changes of intensity with smoothing due to $\tau_1 = \frac{1}{3} c_0^\top c_1$ and $\tau_3 = \frac{1}{3} c_1^\top c_0$. The filter τ_2 (resp. τ_6) computes the discrete second-order difference in vertical (resp. horizontal) direction with smoothing due to $\tau_2 = \frac{1}{3} c_0^\top c_2$ and $\tau_6 = \frac{1}{3} c_2^\top c_0$. The other filters τ_4, τ_5, τ_7, and τ_8 perform like discrete high-order difference operators.

We should note that the first-order derivative operators exaggerate the effects of noise while the second-order derivatives will exaggerated noise twice as much [15]. Therefore, the applicability of the second-order derivatives is limited to images with low noise level. Motivated from the Laplacian of a Gaussian (LOG) and difference of Gaussian (DOG) operators in computer vision, see, for example, [25, 27], we propose to take the second-order derivatives on the blurred or smoothed images in order to reduce the effect of the presence of noise in an image. To this end, we denote by P_κ the matrix representation of the filters τ_κ, $\kappa = 0, 1, \ldots 8$, under a proper boundary condition. Let us define

$$B_{2\ell} := P_0 \quad \text{and} \quad B_{2h} := [P_1^\top, \ldots, P_8^\top]^\top. \tag{13}$$

We have that

$$B_{2\ell}^\top B_{2\ell} + B_{2h}^\top B_{2h} = I.$$

Let $\Phi_{2\Theta} : \mathbb{R}^{8n} \to \mathbb{R}$ be defined through a nonnegative parameter sequence $\Theta = \{\theta_i = (\theta_{i1}, \theta_{i2}, \ldots, \theta_{i8}) \in \mathbb{R}^8 : 1 \leqslant i \leqslant n\}$ with non-negative elements as follows

$$\Phi_{2\Theta}(v) := \sum_{i=1}^{n} \|[\theta_{i1}v_i, \theta_{i2}v_{i+n}, \ldots, \theta_{i8}v_{i+7n}]\|_1, \tag{14}$$

where $\| \cdot \|_1$ denotes the ℓ_1 norm. With this function $\Phi_{2\Theta}$, we propose a functional based on the DCT-based tight framelet system in the following form

$$\mathcal{G}_2(u) := \Phi_{2\Theta}(B_{2h}B_{1\ell}u), \tag{15}$$

where $B_{1\ell}u$ is viewed as the smooth version of u.

All together, our proposed image restoration model is

$$\min_u \{\mathcal{F}(u) + \mathcal{G}_1(u) + \mathcal{G}_2(u)\}. \tag{16}$$

The efficiency of the regularization functional $\mathcal{G}_1(u) + \mathcal{G}_2(u)$ in (16) will be presented in Section 3 when it is compared with several possible regularization functionals formulated from the DHF and DCT-based tight framelet, and with other existing higher-order regularization functionals.

2.3   Algorithm

In this subsection, we specify the data fidelity $\mathcal{F}$ in (16). For Gaussian noise, the natural choice for $\mathcal{F}$ is $\mathcal{F}(u) = \frac{1}{2}\|Ku - z\|^2$ where $\| \cdot \|$ denotes either the vector 2-norm or matrix 2-norm. That is, the optimization problem we consider here is

$$\min_{u \in [0,1]^n} \frac{1}{2}\|Ku - z\|^2 + \Phi_{1\Lambda}(B_{1h}u) + \Phi_{2\Theta}(B_{2h}B_{1\ell}u), \tag{17}$$

where $\Phi_{1\Lambda}$ is given in (10) and $\Phi_{2\Theta}$ is given in (14). Here, we assume that all pixel values of an image are in $[0, 1]$.

We next introduce our notation and recall some necessary background from convex analysis. The class of all lower semicontinuous convex functions $f : \mathbb{R}^d \to (-\infty, +\infty]$ such that dom $f := \{x \in \mathbb{R}^d : f(x) < +\infty\} \neq \emptyset$ is denoted by $\Gamma_0(\mathbb{R}^d)$. The indicator function of a closed convex set C in $\mathbb{R}^d$ is defined, at $u \in \mathbb{R}^d$, as

$$\iota_C(u) := \begin{cases} 0, & \text{if } u \in C, \\ +\infty, & \text{otherwise.} \end{cases}$$

Clearly, the indicator function ι_C is in $\Gamma_0(\mathbb{R}^d)$ for any closed nonempty convex set C.

For a function $f \in \Gamma_0(\mathbb{R}^d)$, the proximity operator of f with parameter λ, denoted by $\mathrm{prox}_{\lambda f}$, is a mapping from $\mathbb{R}^d$ to itself, defined for a given point $x \in \mathbb{R}^d$ by

$$\mathrm{prox}_{\lambda f}(x) := \mathrm{argmin}\left\{ \frac{1}{2}\|u - x\|^2 + \lambda f(u) : u \in \mathbb{R}^d \right\}.$$

We also need the notation of conjugate. The conjugate of $f \in \Gamma_0(\mathbb{R}^d)$ is the function $f^* \in \Gamma_0(\mathbb{R}^d)$ defined at $x \in \mathbb{R}^d$ by $f^*(x) := \sup\{\langle u, x \rangle - f(u) : u \in \mathbb{R}^d\}$. A key property of the proximity operators of f and its conjugate is

$$\mathrm{prox}_{\lambda f}(x) + \lambda \mathrm{prox}_{\lambda^{-1} f^*}(x/\lambda) = x, \tag{18}$$

which holds for all $x \in \mathbb{R}^n$ and any $\lambda > 0$.

Now, we turn to the optimization problem (17). Define

$$f(u) = \frac{1}{2}\|Ku - z\|^2, \quad g(u) = \iota_{[0,1]^n}, \quad p(s) = \Phi_{1\Lambda}(s_1) + \Phi_{2\Theta}(s_2), \quad \text{and} \quad A = \begin{bmatrix} B_{1h} \\ B_{2h}B_{1\ell} \end{bmatrix}, \tag{19}$$

where $u \in \mathbb{R}^n$ and $s = (s_1, s_2)$ with $s_1 \in \mathbb{R}^{6n}$ and $s_2 \in \mathbb{R}^{8n}$. Then, our optimization problem (17) can be viewed as a special case of the optimization problem whose objective function is the sum of three lower semicontinuous convex functions in the form of

$$\min_{u \in \mathbb{R}^n} f(u) + g(u) + p(Au), \tag{20}$$

where A is a $d \times n$ matrix, $f \in \Gamma_0(\mathbb{R}^n)$ is differentiable, $g \in \Gamma_0(\mathbb{R}^n)$, and $p \in \Gamma_0(\mathbb{R}^d)$.

Several algorithms have been developed for the optimization problem (20), see, for example, [10, 11, 20, 33]. We adopt the algorithm given in [33] for problem (20) since it converges under a much weaker condition and can choose a larger step-size, yielding a faster convergence. This algorithm, named as Primal-Dual Three-Operator splitting (PD3O), has the following iteration:

$$u^k = \mathrm{prox}_{\gamma g}(v^k) \tag{21a}$$

$$s^{k+1} = \mathrm{prox}_{\delta p^*}\left((I - \gamma\delta AA^\top)s^k + \delta A(2u^k - v^k - \gamma\nabla f(u^k)) \right) \tag{21b}$$

$$v^{k+1} = u^k - \gamma\nabla f(u^k) - \gamma A^\top s^{k+1} \tag{21c}$$

One PD3O iteration can be viewed as an operator $\mathrm{T}_{\mathrm{PD3O}}$ such that $(v^{k+1}, s^{k+1}) = \mathrm{T}_{\mathrm{PD3O}}(v^k, s^k)$. The convergence analysis of PD3O is given in the following lemma.

Lemma 1 (Sublinear convergence rate [33]). *Let $f \in \Gamma_0(\mathbb{R}^n)$ and its gradient be Lipschitz continuous with constant L, let $g \in \Gamma_0(\mathbb{R}^n)$, and $p \in \Gamma_0(\mathbb{R}^d)$. Choose γ and δ such that $\gamma < 2/L$ and $M = \frac{\gamma}{\delta}(I - \gamma\delta AA^\top)$ is positive definite. Let (v^*, s^*) be any fixed point of $\mathrm{T}_{\mathrm{PD3O}}$, and $\{(v^k, s^k)\}_{k \geqslant 0}$ be the sequence generated by PD3O. Define $\|(v, s)\|_M := \sqrt{\|v\|^2 + \langle s, Ms \rangle}$. Then, the following statements hold.*

(i) *The sequence $\{(\|(v^k, s^k) - (v^*, s^*)\|_M)\}_{k \geqslant 0}$ is monotonically nonincreasing.*

(ii) *The sequence $\{(\|(v^{k+1}, s^{k+1}) - (v^k, s^k)\|_M)\}_{k \geqslant 0}$ is monotonically nonincreasing. Moreover, $\|(v^{k+1}, s^{k+1}) - (v^k, s^k)\|_M^2 = o\left(\frac{1}{k+1}\right)$.*

To adapt PD3O for our optimization problem (17) with f, g and p, and the matrix A given in (19), some preparations are provided in the following lemmas.

Lemma 2. *Let $\delta > 0$ and $\Phi_{1\Lambda}$ be given in (10). For any $v \in \mathbb{R}^{6n}$, if $y = \mathrm{prox}_{\delta^{-1}\Phi_{1\Lambda}}(v)$, then*

$$y_{(i)} = \mathrm{prox}_{\delta^{-1}\lambda_i\varphi_1}(v_{(i)}), \tag{22}$$

where $y_{(i)} = \begin{bmatrix} y_i & y_{i+n} & \cdots & y_{i+5n} \end{bmatrix}^\top$ and $v_{(i)} = \begin{bmatrix} v_i & v_{i+n} & \cdots & v_{i+5n} \end{bmatrix}^\top$. Furthermore, let $y_{(ij)}$ and $v_{(ij)}$ be $y_{i+(j-1)n}$ and $v_{i+(j-1)n}$, respectively, for $i = 1, \ldots, n$ and $j = 1, \ldots, 6$, then

$$\begin{bmatrix} y_{(i1)} \\ y_{(i2)} \end{bmatrix} = \left(1 - \frac{\lambda_i\delta^{-1}}{\max\{\| \begin{bmatrix} v_{(i1)} & v_{(i2)} \end{bmatrix} \|, \lambda_i\delta^{-1}\}} \right) \begin{bmatrix} v_{(i1)} \\ v_{(i2)} \end{bmatrix},$$

where the pair $(y_{(i3)}, y_{(i4)})$ is obtained by simply replacing $(v_{(i1)}, v_{(i2)})$ in the right hand side of the above formula by $(v_{(i3)}, v_{(i4)})$, and $y_{(i5)} = v_{(i5)}$, $y_{(i6)} = v_{(i6)}$.

Proof. The proof is based on the block separable property of $\Phi_{1\Lambda}$ in (10). By the definition of proximity operator and equations (10) and (12),

$$\begin{aligned} \mathrm{prox}_{\delta^{-1}\Phi_{1\Lambda}}(v) &= \mathrm{argmin}\left\{ \frac{1}{2}\|u - v\|^2 + \delta^{-1}\Phi_{1\Lambda}(u) : u \in \mathbb{R}^{6n} \right\} \\ &= \mathrm{argmin}\left\{ \sum_{i=1}^{n} \frac{1}{2}\|u_{(i)} - v_{(i)}\|^2 + \delta^{-1}\lambda_i\varphi_1(u_{(i)}) : u_{(i)} \in \mathbb{R}^6, i = 1, \ldots, n \right\}. \end{aligned}$$

Hence, equation (22) holds. Notice that φ_1 is also a block separable function. By using the definition of proximity again and the proximity operator of the ℓ_2 norm (see, for example, [9, 28]), we obtain the explicit expression for $y_{(i)}$ as given above. $\qquad \square$

Lemma 3. *Let $\delta > 0$ and $\Phi_{2\Theta}$ be given in (14). For any $v \in \mathbb{R}^{8n}$, if $y = \mathrm{prox}_{\delta^{-1}\Phi_{2\Theta}}(v)$, then*

$$y_{(i)} = \mathrm{prox}_{\delta^{-1}\|\cdot\|_1 \circ \mathrm{diag}(\theta_i)}(\mathrm{diag}(v_{(i)})). \tag{23}$$

where $y_{(i)} = \begin{bmatrix} y_i & y_{i+n} & \cdots & y_{i+7n} \end{bmatrix}^\top$ and $v_{(i)} = \begin{bmatrix} v_i & v_{i+n} & \cdots & v_{i+7n} \end{bmatrix}^\top$. Furthermore, let $y_{(ij)}$ and $v_{(ij)}$ be $y_{i+(j-1)n}$ and $v_{i+(j-1)n}$, respectively, for $i = 1, \ldots n$ and $j = 1, \ldots, 8$, then

$$y_{(ij)} = \max\{|v_{(ij)}| - \delta^{-1}\theta_{ij}, 0\}\mathrm{sgn}(v_{(ij)}).$$

Proof. The proof is based on the block separable property of $\Phi_{2\Theta}$ in (14). By the definition of proximity operator,

$$\begin{aligned} \mathrm{prox}_{\delta^{-1}\Phi_{2\Theta}}(v) &= \mathrm{argmin}\left\{ \frac{1}{2}\|u - v\|_2^2 + \delta^{-1}\Phi_{2\Theta}(u) : u \in \mathbb{R}^{8n} \right\} \\ &= \mathrm{argmin}\left\{ \sum_{i=1}^{n} \frac{1}{2}\|u_{(i)} - v_{(i)}\|_2^2 + \delta^{-1}\|\mathrm{diag}(\theta_i)u_{(i)}\|_1 : u_{(i)} \in \mathbb{R}^8, i = 1, \ldots, n \right\}. \end{aligned}$$

Hence, equation (23) holds. Furthermore, notice that $\mathrm{prox}_{\delta^{-1}\|\cdot\|_1 \circ \mathrm{diag}(\theta_i)}$ is the well-known soft thresholding operator, the rest of result holds. $\qquad \square$

Lemma 4. *Let $\Phi_{1\Lambda}$ be given in (10) and $\Phi_{2\Theta}$ be given in (14). For any $v \in \mathbb{R}^{14n}$, write $v = (v_1, v_2)$ with $v_1 \in \mathbb{R}^{6n}$ and $v_2 \in \mathbb{R}^{8n}$, and define $p(v) = \Phi_{1\Lambda}(v_1) + \Phi_{2\Theta}(v_2)$. Then, for any $\delta > 0$,*

$$\mathrm{prox}_{\delta^{-1}p}(v) = \mathrm{prox}_{\delta^{-1}\Phi_{1\Lambda}}(v_1) \times \mathrm{prox}_{\delta^{-1}\Phi_{2\Theta}}(v_2).$$

9

The result in the above lemma comes from the block separability of the function p. Therefore, we omit its proof here.

To apply Lemma 1 to problem (17), we verify all the requirements listed in Lemma 1. First, for the function f in (19), we have that $\nabla f(u) = K^\top(Ku - z)$, the gradient of f is $\|K\|^2$-Lipschitz continuous. Next, we discuss the positive definiteness of the matrix $I - \gamma\delta AA^\top$.

Lemma 5. *Let A be given in (19). Then, for positive numbers γ and δ, the matrix $I - \gamma\delta AA^\top$ is positive semidefinite (or definite) if and only if $\gamma\delta \leqslant 1$ (or $\gamma\delta < 1$).*

Proof. First, we show that $\|A\| = 1$. For any $u \in \mathbb{R}^n$, we have that

$$u^\top A^\top Au = u^\top B_{1h}^\top B_{1h} u + u^\top B_{1\ell}^\top B_{2h}^\top B_{2h} B_{1\ell} u.$$

Since $B_{2h}^\top B_{2h} + B_{2\ell}^\top B_{2\ell} = I$ and $B_{1h}^\top B_{1h} + B_{1\ell}^\top B_{1\ell} = I$, from the above we have that

$$u^\top A^\top Au \leqslant u^\top B_{1h}^\top B_{1h} u + u^\top B_{1\ell}^\top B_{1\ell} u = u^\top u.$$

Hence $\|A\| \leqslant 1$. Further, since the null space of B_{1h} is non-empty, therefore, $\|A\| = 1$.

Next, since $AA^\top$ is positive semi-definite and its largest eigenvalue is 1, hence, $I - \gamma\delta AA^\top$ is positive semidefinite (or definite) if and only if $\gamma\delta \leqslant 1$ (or $\gamma\delta < 1$). $\qquad\square$

The explicit form of $\text{prox}_{\delta^{-1}p}$ is given in Lemma 4 with the help of Lemmas 2 and 3. Therefore, the proximity operator $\text{prox}_{\delta p^*}$ can be computed via (18). With the above preparation, the complete procedure for solving (17) based on (21a)-(21c) is described in Algorithm 1. This algorithm is refereed to as TNTF (two-level non-stationary tight framelet) algorithm.

Algorithm 1 Two-level Non-stationary Tight Framelet (TNTF) Algorithm

---

1: Set parameters $\gamma < \frac{2}{\|K\|^2}$, $\gamma\delta < 1$; pre-given parameters Λ and Θ.
2: Initialize $v^0 = 0$ and $s^0 = 0$
3: Auxiliary variable x^k and write $s^k = (s_1^k, s_2^k)$
4: **for** $k = 1, 2, \ldots$ **do**
5:

$$u^k = \text{Proj}_{[0,1]}(v^k) \tag{24a}$$

$$x^k = \gamma(B_{1h}^\top s_1^k + B_{1\ell}^\top B_{2h}^\top s_2^k) - (2u^k - v^k) + \gamma K^\top(Ku^k - z) \tag{24b}$$

$$s_1^{k+1} = (s_1^k - \delta B_{1h} x^k) - \delta \cdot \text{prox}_{\delta^{-1}\Phi_{1\Lambda}}(\delta^{-1}(s_1^k - \delta B_{1h} x^k)) \tag{24c}$$

$$s_1^{k+2} = (s_2^k - \delta B_{2h} B_{1\ell} x^k) - \delta \cdot \text{prox}_{\delta^{-1}\Phi_{2\Theta}}(\delta^{-1}(s_2^k - \delta B_{2h} B_{1\ell} x^k)) \tag{24d}$$

$$v^{k+1} = u^k - \gamma K^\top(Ku^k - z) - \gamma(B_{1h}^\top s_1^{k+1} + B_{1\ell}^\top B_{2h}^\top s_2^{k+2}) \tag{24e}$$

6: **end for**

The convergence analysis for Algorithm 1 is as follows.

Theorem 1. *Let (v^*, s^*) be any fixed point of T_{PD3O} with f, g, p and A given in (19). Let $\{(v^k, s^k)\}_{k\geqslant 0}$ be the sequence generated by Algorithm 1, where $s^k = (s_1^k, s_2^k)$. Choose γ and δ such that $\gamma < 2/\|K\|^2$ and $\gamma\delta < 1$. Define $M = \frac{\gamma}{\delta}(I - \gamma\delta AA^\top)$. Then, the following statements hold.*

(i) The sequence $\{(\|(v^k, s^k) - (v^*, s^*)\|_M)\}_{k\geqslant 0}$ is monotonically nonincreasing.

(ii) *The sequence* $\{(\|(v^{k+1}, s^{k+1}) - (v^k, s^k)\|_M)\}_{k \geqslant 0}$ *is monotonically nonincreasing. Moreover,*
$$\|(v^{k+1}, s^{k+1}) - (v^k, s^k)\|_M^2 = o\left(\frac{1}{k+1}\right).$$

Proof. We know that the gradient of f in (19) is $\|K\|^2$-Lipschitz continuous. By Lemma 5, the matrix M is positive definite if $\gamma\delta < 1$, the result of this theorem follows immediately from Lemma 1. □

Remark: the computational cost of Algorithm 1 depends on mainly two factors: the UDFmT used in steps (24b-e) and the total number of iterations for k. The UDFmT decompositions include $B_{1h}x^k$, $B_{2h}B_{1\ell}x^k$ in (24c-d) while the UDFmT reconstructions include $B_{1h}^{\top}s_1^k$, $B_{1\ell}^{T}B_{2h}^{\top}s_2^k$ in (24b) and $B_{1h}^{\top}s_1^{k+1}$ and $B_{1\ell}^{\top}B_{2h}^{\top}s_2^{k+2}$ in (24e). Since UDFmT in Algorithm 1 uses convolutions with 7 DHF filters in the first level and 9 DCT-based filters in the second level, the UDFmT can be implemented with computational cost $O(n)$, where n is the number of pixels in u. For the total number of iterations k in Algorithm 1, it depends on when the algorithm converges and the maximum number $\mathbb{K}$ of iterations set manually. Consequently, the total computational cost for Algorithm 1 is $O(\mathbb{K}n)$.

Finally, we discuss how to choose the parameters in the algorithm. In our tests below we choose $\gamma = 1.99, \delta = 0.5$ to ensure $\gamma\delta < 1$. Recall from (12) that we only use the first four subband coefficients of DHF in the first level. The corresponding regularization parameters λ_i (defined in (10)) are chosen to adaptively adjust to local variations. Let $\mathcal{I}(i)$ be the set containing all indices in the neighborhood at the ith pixel. Then λ_i is set as

$$\lambda_i = \frac{\lambda \times |\mathcal{I}(i)|}{\max\{\sum_{p \in \mathcal{I}(i)} \|w_p\|, 10^{-10}\}}, \tag{25}$$

where $w_p = [v_{(p1)}, v_{(p2)}]^{\top}$ or $w_p = [v_{(p3)}, v_{(p4)}]^{\top}$ are defined as in (22). In our tests, we choose the neighborhood of window size 3×3 and the parameter λ is set by hand.

For the regularization parameters θ_i associated with the DCT-based tight framelet coefficients (see (14)), they are all automatically estimated and updated using the approach in our previous work [24]. More precisely, for the regularization parameters $\theta_{i\kappa}$, $\kappa = 1, \ldots, 8$, used in (14), they are automatically estimated according to the local variations of framelet coefficients and noise level. Suppose the ϵ in model (1) is the Gaussian noise with the standard deviation σ. As it was done in our previous work [24], σ_κ^2 the noise variance of the framelet coefficients coming from the filter τ_κ at the second decomposed level is estimated as $\sigma_\kappa^2 = \frac{\sigma^2}{4}\|\tau_\kappa\|_F^2$, where $\|\tau_\kappa\|_F$ is the Frobenius norm of τ_κ; $(\sigma_i^\kappa)^2$, the local signal variance of the ith framelet coefficients coming from the filter κth, is computed as $(\sigma_i^\kappa)^2 = \max\{(\sum_{p \in \mathcal{I}^\kappa(i)} |v_p|/|\mathcal{I}^\kappa(i)|)^2 - \sigma_\kappa^2, 10^{-10}\}$, where $\mathcal{I}^\kappa(i)$ is the set containing all indices in the neighborhood at the ith framelet coefficients from the filter κ. With them, the regularization parameters are estimated as

$$\theta_{i\kappa} = \frac{\sqrt{2}\sigma_\kappa^2}{\sigma_i^\kappa}, \quad \kappa = 1, \ldots, 8. \tag{26}$$

To save computational cost of estimating parameters λ_i in (25) and $\theta_{i\kappa}$ in (26), we only update these parameters when the iteration k is a multiple of 30 and fix them after the 200th iteration in our numerical experiments.

3   Experiments

In this section, we present numerical experiments to illustrate the effectiveness and efficiency of our proposed model (17) for image restoration. We use the images "Square Circle", "Cameraman",

and "Montage" of size 256×256 as the original images u in our experiments, see Figure 1. The pixel values of these images are normalized to the interval $[0, 1]$. The quality of the restored image, say $\widetilde{u}$, is evaluated in terms of the peak-signal-to-noise ratio (PSNR) that is defined by

$$\text{PSNR} := 10 \log_{10} \frac{255^2 n}{\|\widetilde{u} - u\|^2},$$

where n is the number of pixels in u. To incorporate structural information in image comparisons, the metric of structural similarity (SSIM) [32] of $\widetilde{u}$ to u is reported as well. The higher the PSNR and SSIM, the better the quality of the restored image.
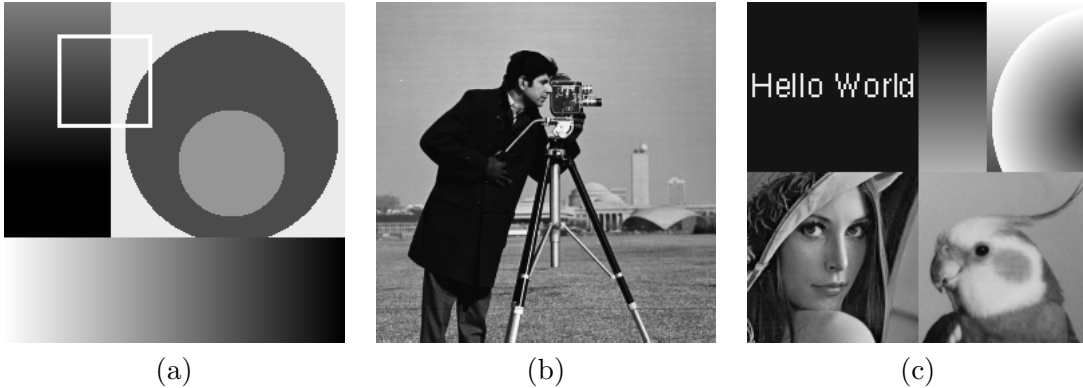

(a)　　　　　　　　　　　(b)　　　　　　　　　　　(c)

Figure 1: Original image: (a) Square Circle; (b) Cameraman; (c) Montage.

Two sets of comparisons for image restoration will be conducted in this section. The first set is to compare with other tight frame regularizers. The second set is to compare with some derivative-based models.

3.1 Comparison with Tight Frame Regularizers

Here we compare the proposed regularization functional (16) with two other tight frame regularization functionals $\mathcal{G}_{\text{DCT}}$ and $\mathcal{G}_{\text{DHF+DCT}}$ while using the classical TV regularizer $\mathcal{G}_{\text{TV}}$ (see (3)) as a benchmark. The $\mathcal{G}_{\text{DCT}}$ is defined as $\mathcal{G}_{\text{DCT}}(u) = \Phi_{2\Theta}(B_{2h}u)$ which only uses the DCT-based tight framelet and takes the first- and second-order information on the image u, where $\Phi_{2\Theta}$ is given in (14). The $\mathcal{G}_{\text{DHF+DCT}}$ is defined as $\mathcal{G}_{\text{DHF+DCT}}(u) = \Phi_{1\Lambda}(B_{1h}u) + \Phi_{2\Theta}(B_{2h}u)$, where $\Phi_{1\Lambda}$ is given in (11) with φ_1 in (12). The main difference between our proposed regularization functional (16) and $\mathcal{G}_{\text{DHF+DCT}}$ is that the action B_{2h} takes on the smoothed image $B_{1\ell}u$ for our regularization functional while the action B_{2h} takes directly on the image u for $\mathcal{G}_{\text{DHF+DCT}}$.

In our experiment, the image of "Square Circle" in Figure 1(a) (which is the same as Figure 2(a)) is blurred by a 5×5 average kernel (using the Matlab command FSPECIAL('AVERAGE', [5:5]), followed by adding Gaussian noise of mean zero and standard deviation $\sigma = 0.04$. The values of the pair of (PSNR, SSIM) of these restored images by $\mathcal{G}_{\text{TV}}$, $\mathcal{G}_{\text{DCT}}$, $\mathcal{G}_{\text{DHF+DCT}}$, and the proposed regularization functional (16) are (33.66dB, 0.962), (32.54dB, 0.970), (33.14dB, 0.980), and (35.00dB, 0.980), respectively. To view the visual quality of the restored images, the square portion marked in the image 1(a) is displayed in Figure 2. For the regions pointed by two arrows, we can conclude that the proposed regularization functional (16) leads to the restored images having better visual quality than the others. The results clearly show that our combined tight frame model is better than other intuitive tight frame models.
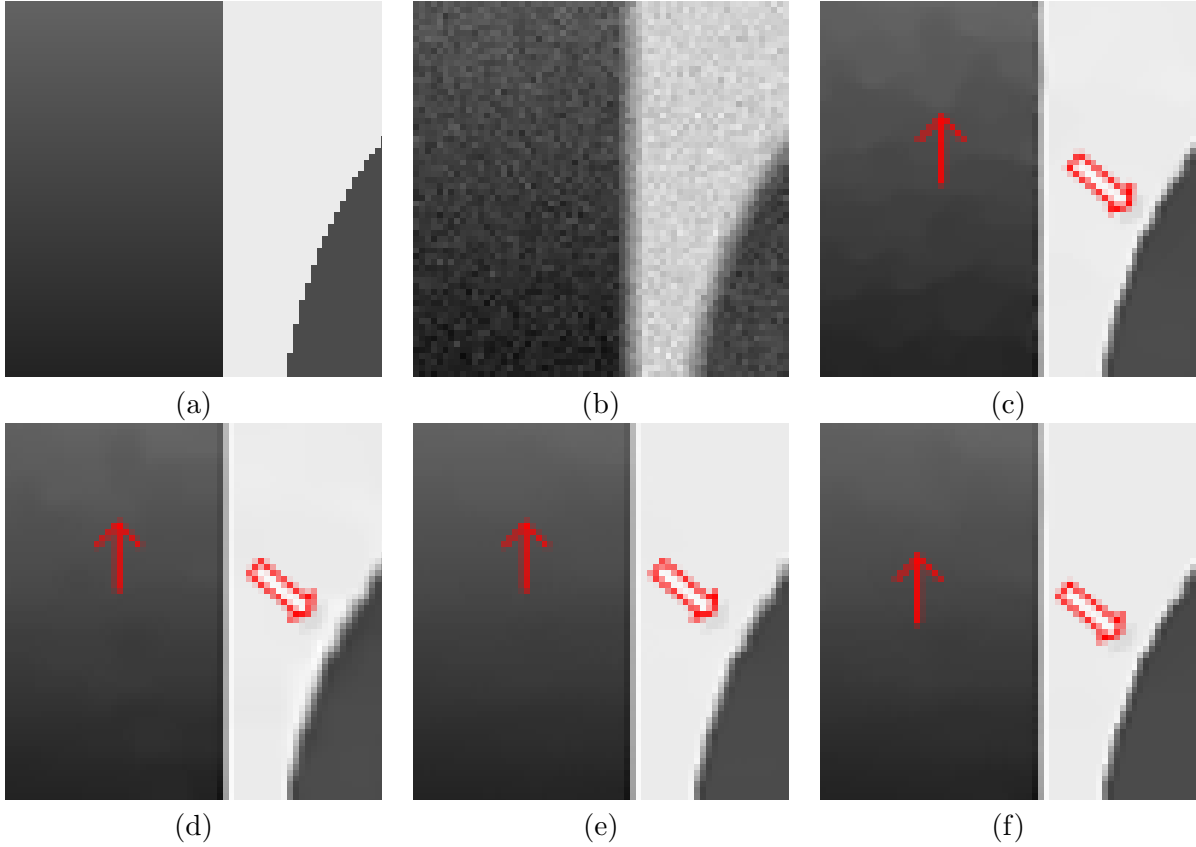
Figure 2: (a) A region of the image of "Square Circle"; (b) The blurred image by the kernel FSPECIAL('AVERAGE', [5:5]) with Gaussian noise of mean zero and variance $\sigma = 0.04$; The restored images with regularization (c) TV with $\alpha = 0.04$ (see (2) and (3)); (d) $\mathcal{G}_{\mathrm{DCT}}$; (e) $\mathcal{G}_{\mathrm{DHF+DCT}}$; and (f) the proposed regularization functional (16) with $\lambda = 0.00035$ (see (25)) in the first level, respectively.

13

3.2 Comparison with Derivative-based Regularizers

Now we give a comprehensive comparison between our model (17) and the TV and TGV models. The TV model uses $\mathcal{G}(u)$ in (3) as its regularization term while the TGV model uses $\mathcal{G}(u)$ in (5) as its regularization term. The software of TGV model was provided by the authors in [16]. All algorithms are carried out until the stopping condition $\|u^{(k+1)} - u^{(k)}\|_2 / \|u^{(k)}\|_2 < 10^{-9}$ is satisfied or the maximal iterations is 400.

In our experiments, the test images in Figure 1 are blurred by a 5×5 average kernel (using the Matlab command FSPECIAL('AVERAGE', [5:5]), followed by adding Gaussian noise of mean zero and standard deviation σ. For different values of σ, the PSNR and SSIM values of the restored images by TV, TGV, and our TNTF are reported in Table 1. The highest values of PSNR and SSIM for each σ in each test image are highlighted. It clearly shows that our proposed TNTF performs the best in terms of both PSNR and SSIM values. We remark that the regularization parameters in the second level for our TNTF are automatically estimated based on the approach in our work [24].

Table 1: The PSNR (dB) and SSIM for the restored results of each algorithm with blurred images contaminated by Gaussian noise. The test images are blurred by the blurring kernel FSPECIAL('AVERAGE',[5:5]).

Algorithm	" Square Circle"		"Cameraman"		"Montage"		Case
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	
TV	35.40dB	0.980	26.43dB	0.815	28.21dB	0.907	
TGV	35.58dB	0.976	26.27dB	0.811	28.84dB	0.910	STD σ=0.02
TNTF	**38.19**dB	**0.992**	**27.06**dB	**0.821**	**30.19**dB	**0.924**	
TV	34.51dB	0.969	25.64dB	0.791	26.92dB	0.886	
TGV	34.33dB	0.967	25.58dB	0.788	26.93dB	0.884	STD σ=0.03
TNTF	**36.14**dB	**0.985**	**26.01**dB	**0.800**	**28.91**dB	**0.910**	
TV	33.66dB	0.962	25.10dB	0.774	25.94dB	0.867	
TGV	33.41dB	0.955	24.94dB	0.770	26.17dB	0.875	STD σ=0.04
TNTF	**35.00**dB	**0.980**	**25.31**dB	**0.784**	**27.88**dB	**0.898**	

In the rest of this section, we provide qualitative results of the restored images from the above three algorithms. We first show the case for the blurred image of "Square Circle" with Gaussian noise of STD $\sigma = 0.03$ in Figure 3. The noisy and blurry image is shown in Figure 3(a). The regularization parameter $\alpha = 0.02$ (see (2) and (3)) is used for the TV model and $(\alpha_1, \alpha_2) = (0.0105, 0.026)$ (see (5)) is used for the TGV model based on the best achievable PSNR values. The regularization parameter $\lambda = 0.0002$ in the first level (see (25)) is used in our proposed model. We can observe that Figure 3(b) produced by the TV has lots of staircase artifacts even without zooming in. As we can see from Figure 3(c) and (d), this kind of staircase artifacts is significantly reduced by the TGV and TNTF. To have a closer look at the visual quality of the restored images by various algorithms, two parts of Figure 3 are zoomed in and displayed in the first column of Figure 4. The corresponding parts in the restored images by TV, TGV, and TNTF are shown in Figure 4(b), (c), and (d), respectively. We can conclude that the horizontal line in the image is well preserved by the TNTF.

Figure 5(a) is the blurred image of "Cameraman" corrupted by Gaussian noise of STD σ=0.02. The restored images by TV, TGV, and TNTF are displayed in Figure 5(b), (c), and (d), respectively. The regularization parameters for TV, TGV, and TNGV are $\alpha = 0.006$, $(\alpha_1, \alpha_2) = (0.0035, 0.0095)$, and $\lambda = 0.0004$, respectively. The structures of the building as well as the man are well preserved in the restored image by our TNTF. Block artifacts are clearly observed in the sky of the restored
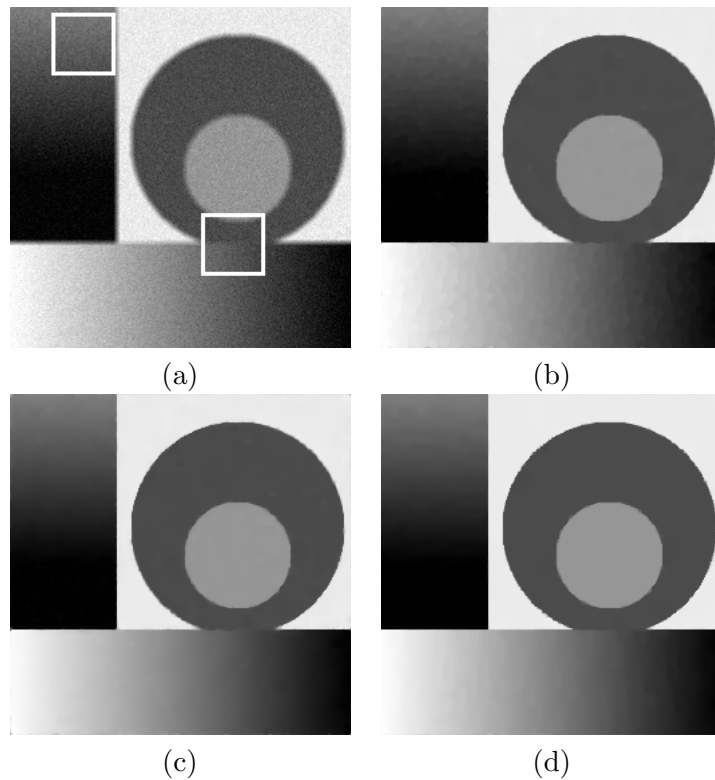
Figure 3: (a) Image blurred by kernel FSPECIAL('AVERAGE', [5:5]) and added Gaussian noise with $\sigma = 0.03$; Images reconstructed by (b) TV with $\alpha = 0.02$ (see (2) and (3)), (c) TGV with $(\alpha_1, \alpha_2) = (0.0105, 0.026)$ (see (5)), and (d) TNTF with $\lambda = 0.0002$ (see (25)), respectively.
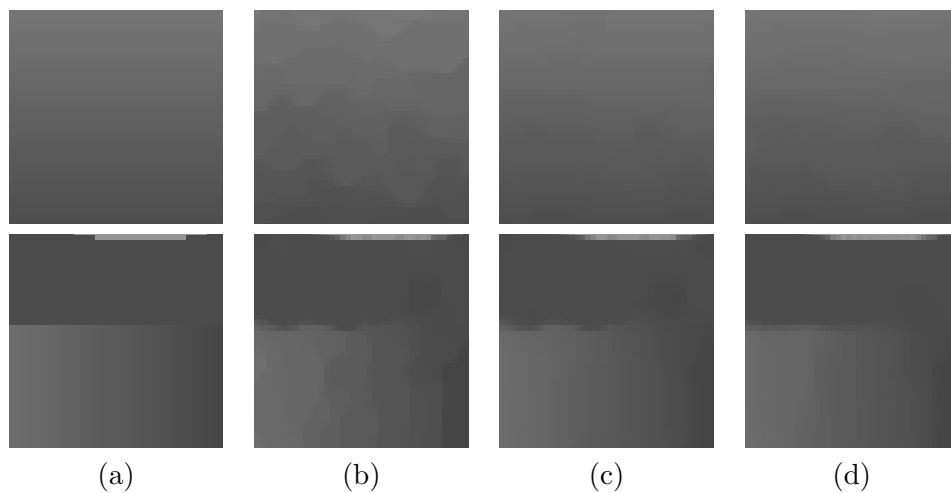

Figure 4: Two zoom-in parts of Fig. 3: (a) Original image; images reconstructed by (b) TV, (c) TGV, and (d) TNTF, respectively.

images by TV and TGV (see Figure 5(b), (c)), but not in Figure 5(d). The zoom-in part of Figure 5 is displayed in Figure 6. The shape of the camera lens in the restored image by TNTF is more closer to the original one than that by TV and TGV.
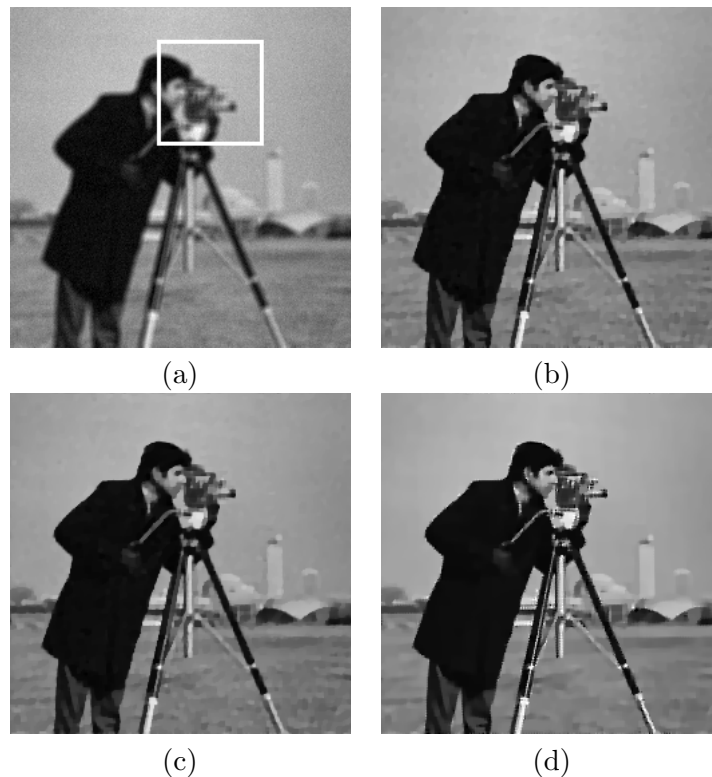

(a)   (b)

(c)   (d)

Figure 5: (a) Image blurred by kernel FSPECIAL('AVERAGE', [5:5]) and added Gaussian noise with $\sigma = 0.02$; images reconstructed by (b) TV with $\alpha = 0.006$; (c) TGV with $(\alpha_1, \alpha_2) = (0.0035, 0.0095)$, and (d) TNTF with $\lambda = 0.0004$, respectively.

Figure 7(a) is the blurred image of "Montage" corrupted by Gaussian noise of STD σ=0.04. The restored images by TV, TGV, and TNTF are displayed in Figure 7(b), (c), and (d), respectively. The regularization parameters for TV, TGV, and TNGV are $\alpha = 0.019$, $(\alpha_1, \alpha_2) = (0.0085, 0.0155)$, and $\lambda = 0.00015$, respectively. Severe artifacts appeared in Figure 7(b) by TV, are significantly suppressed in Figure 7(c) and (d) by TGV and TNTF. Two zoom-in parts of the results are shown in Figure 8. It is evident that the lines are well preserved in Figure 8(d) by TNTF.

4 Conclusion

In this paper, we have designed a two-level non-stationary tight framelet system and utilized it in a regularization model for image restoration. This framelet system has the ability to capture the first and second order information of the image to be reconstructed. We developed an algorithm to solve the resulting optimization problem. The numerical experiments show the effectiveness of the proposed image restoration model and the corresponding algorithm.
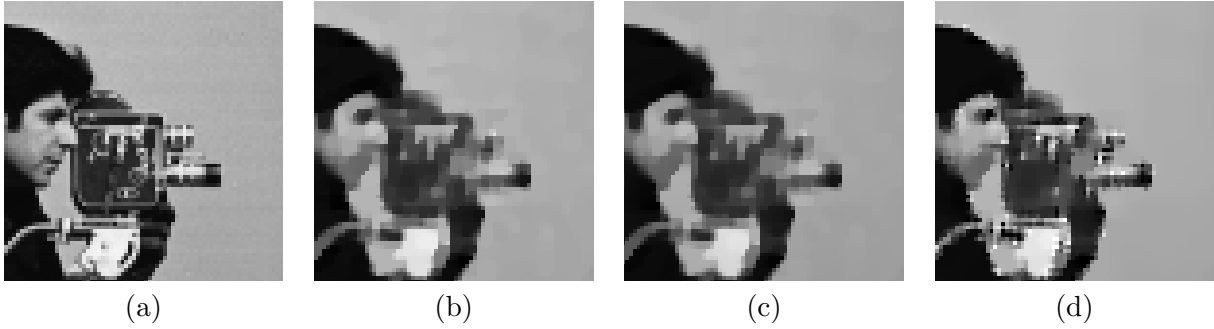
Figure 6: Zoom-in parts of Figure 5: (a) Original image; images reconstructed by (b) TV, (c) TGV, and (d) TNTF.
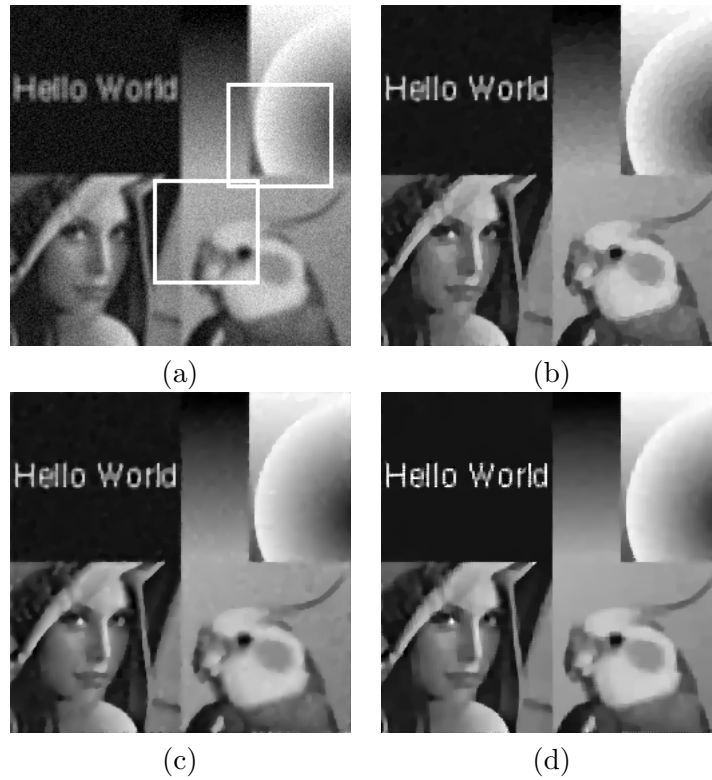

Figure 7: (a) Image blurred by kernel FSPECIAL('AVERAGE', [5:5]) and added Gaussian noise with $\sigma = 0.04$; images reconstructed by (b) TV with $\alpha = 0.019$, (c) TGV with $(\alpha_1, \alpha_2) = (0.0035, 0.0095)$, and (d) TNTF with $\lambda = 0.00015$.
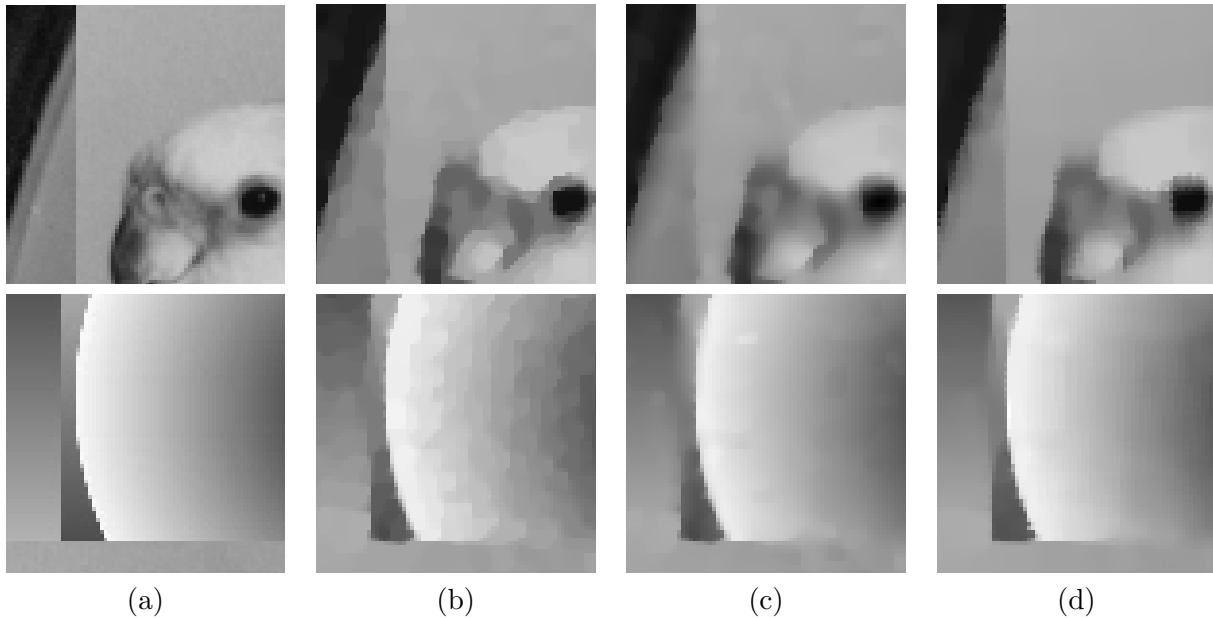
Figure 8: Two zoom-in parts of Fig. 7: (a) Original image; images reconstructed by (b) TV, (c) TGV, and (d) TNTF.

References

[1] K. Bredies and M. Holler, *A TGV-based framework for variational image decompression, zooming, and reconstruction. Part II: Numerics*, SIAM Journal on Imaging Sciences, 8 (2015), pp. 2851–2886.

[2] K. Bredies, K. Kunisch, and T. Pock, *Total generalized variation*, SIAM Journal on Imaging Sciences, 3 (2010), pp. 492–526.

[3] J.-F. Cai, B. Dong, S. Osher, and Z. Shen, *Image restorations: total variation, wavelet frames and beyond*, Journal of the American Mathematical Society, 25 (2012), pp. 1033–1089.

[4] A. Chambolle, R. DeVore, N.-Y. Lee, and B. Lucier, *Nonlinear wavelet image processing: Variational problems, compression, and noise removal through wavelet shrinkage*, IEEE Transactions on Image Processing, 7 (1998), pp. 319–335.

[5] A. Chambolle and P. L. Lions, *Image recovery via total variational minimization and related problems*, Numerische Mathematik, 76 (1997), pp. 167–188.

[6] R. Chan, T. Chan, L. Shen, and Z. Shen, *Wavelet algorithms for high-resolution image reconstruction*, SIAM Journal on Scientific Computing, 24 (2003), pp. 1408–1432.

[7] R. Chan, S. D. Riemenschneider, L. Shen, and Z. Shen, *Tight frame: The efficient way for high-resolution image reconstruction*, Applied and Computational Harmonic Analysis, 17 (2004), pp. 91–115.

[8] T. Chan, A. Marquina, and P. Mulet, *High-order total variation-based image restoration*, SIAM Journal on Scientific Computing, 22 (2000), pp. 503–516.

[9] P. Combettes and V. Wajs, *Signal recovery by proximal forward-backward splitting*, Multiscale Modeling and Simulation: A SIAM Interdisciplinary Journal, 4 (2005), pp. 1168–1200.

[10] P. L. Combettes and J.-C. Pesquet, *Primal-dual splitting algorithm for solving inclusions with mixtures of composite, lipschitzian, and parallel-sum type monotone operators*, Set-Valued and Variational Analysis, 20 (2012), pp. 307–330.

[11] L. Condat, *A primal-dual splitting method for convex optimization involving lipschitzian, proximable and linear composite terms*, Journal of Optimization Theory and Applications, 158 (2013), pp. 460–479.

[12] M. Figueiredo and R. D. Nowak, *An EM algorithm for wavelet-based image restoration*, IEEE Transactions on Image Processing, 12 (2003), pp. 906–916.

[13] D. Geman and G. Reynolds, *Constrained restoration and the recovery of discontinuities*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 14 (1992), pp. 367–383.

[14] S. Geman and D. Geman, *Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 6 (1984), pp. 721–741.

[15] R. Gonzalez and R. Woods, *Digital Image Processing*, Addison-Wesley, Boston, MA, 1993.

[16] W. Guo, J. Qin, and W. Yin, *A new detail-preserving regularization scheme*, SIAM Journal on Imaging Sciences, 7 (2014), pp. 1309–1334.

[17] B. Han, *Properties of discrete framelet transforms*, Mathematical Modelling of Natural Phenomena, 8 (2013), pp. 18–47.

[18] ——, *Framelets and Wavelets: Algorithms, Analysis, and Applications*, Springer International Publishing, 2018.

[19] A. P. Johnson and C. L. Baker, *First- and second-order information in natural images: a filter-based approach to image statistics*, Journal of the Optical Society of America A, 21 (2004), pp. 913–925.

[20] Q. Li and N. Zhang, *Fast proximity-gradient algorithms for structured convex optimization problems*, Applied and Computational Harmonic Analysis, 41 (2016), pp. 491 – 517.

[21] Y.-R. Li, R. H. Chan, L. Shen, Y.-C. Hsu, and W.-Y. I. Tseng, *An adaptive directional haar framelet-based reconstruction algorithm for parallel magnetic resonance imaging*, SIAM Journal on Imaging Sciences, 9 (2016), pp. 794–821.

[22] Y.-R. Li, D. Dai, and L. Shen, *Multiframe super-resolution reconstruction using sparse directional regularization*, IEEE Transactions on Circuits and Systems for Video Technology, 20 (2010), pp. 945–956.

[23] Y.-R. Li, L. Shen, D. Q. Dai, and B. W. Suter, *Framelet algorithms for de-blurring images corrupted by impulse plus gaussian noise*, IEEE Transactions on Image Processing, 20 (2011), pp. 1822–1837.

[24] Y.-R. Li, L. Shen, and B. W. Suter, *Adaptive inpainting algorithm based on DCT induced wavelet regularization*, IEEE Transactions on Image Processing, 22 (2013), pp. 752–763.

[25] D. G. Lowe, *Distinctive image features from scale-invariant keypoints*, International Journal of Computer Vision, 60 (2004), p. 91–110.

[26] M. Lysaker, A. Lundervold, and Xue-Cheng Tai, *Noise removal using fourth-order partial differential equation with applications to medical magnetic resonance images in space and time*, IEEE Transactions on Image Processing, 12 (2003), pp. 1579–1590.

[27] D. Marr and E. Hildreth, *Theory of edge detection*, Proceedings of the Royal Society of London, B-207 (1980), p. 187–217.

[28] C. A. Micchelli, L. Shen, and Y. Xu, *Proximity algorithms for image models: Denoising*, Inverse Problems, 27 (2011), p. 045009(30pp).

[29] L. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D, 60 (1992), pp. 259–268.

[30] S. Setzer, G. Steidl, and T. Teuber, *Infimal convolution regularizations with discrete ℓ_1-type functionals*, Communications in Mathematical Sciences, 9 (2011), pp. 797–827.

[31] L. Shen, I. Kakadiaris, M. Papadakis, I. Konstantinidis, D. Kouri, and D. Hoffman, *Image denoising using a tight frame*, IEEE Transactions on Image Processing, 15 (2006), pp. 1254–1263.

[32] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, *Image quality assessment: From error visibility to structural similarity*, IEEE Transactions on Image Processing, 13 (2004), pp. 600–612.

[33] M. Yan, *A new primal-dual algorithm for minimizing the sum of three functions with a linear operator*, Journal of Scientific Computing, 76 (2018), pp. 1698–1717.