RESEARCH ARTICLE

Simultaneous Cartoon and Texture Reconstruction For Image Restoration By Bivariate Function

You-Wei Wen[a,b*], Raymond H. Chan[c] and Wai-Ki Ching[d]

[a]*Department of Mathematics, South China Agricultural University, P. R. China*;
[b]*Center for Wavelets, Approximation and Information Processing, National University of Singapore, Singapore*;
[c]*Department of Mathematics, The Chinese University of Hong Kong, Hong Kong*;
[d]*Advanced Modeling and Applied Computing Laboratory, Department of Mathematics, The University of Hong Kong, Hong Kong.*

We study the simultaneous cartoon and texture reconstruction problem. We propose a new model to approximate the cartoon and texture part by a sparse linear combination of some bases. A bivariate function is employed as the cost function. One of the variables is the decomposition image and the other is the sparse representation of the decomposition image. An alternating minimization algorithm is used to solve the minimization problem. We prove that the algorithm converges for both the l_1-norm and the l_0-norm. Numerical simulations are given to illustrate the efficiency of our method.

Keywords: image

1. Simultaneous Model

Image restoration is one of the major goals in image processing. The observed model is given by

$$\boldsymbol{g} = \mathcal{M}\boldsymbol{f} + \boldsymbol{n}, \tag{1}$$

where $\boldsymbol{g}, \boldsymbol{f}, \boldsymbol{n}$ are the observed image, the original image, and the Gaussian noise respectively. The operator $\mathcal{M}$ will be the identity matrix in image decomposition models, a mask matrix in inpainting problems, or a Toeplitz-like matrix for deblurring problems. A classical approach is to decompose the original image $\boldsymbol{f}$ into two components $\boldsymbol{f}_u + \boldsymbol{f}_v$. The cartoon component $\boldsymbol{f}_u$ is well-structured and models the homogeneous regions with sharp boundaries. The texture component $\boldsymbol{f}_v$ contains some repeated pattern in small scales.

There are two main approaches to cartoon and texture decomposition: total variational (TV) approach and sparse-representation based approach. The popular TV model was proposed by Rudin, Osher and Fatemi (ROF) [42]. Meyer [37] proposed

2

an improvement on the TV model. Using the G-norm

$$\|\boldsymbol{f}_v\|_G = \inf\{\sqrt{\boldsymbol{v}_1^2 + \boldsymbol{v}_2^2} : \boldsymbol{f}_v = \partial_x \boldsymbol{v}_1 + \partial_y \boldsymbol{v}_2\},$$

he defined the G-space as the dual space of functions of bounded variation in the G-norm. Then the texture component $\boldsymbol{f}_v$ is represented by two subcomponents $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ such that $\boldsymbol{f}_v = \partial_x \boldsymbol{v}_1 + \partial_y \boldsymbol{v}_2$ in the G-space. However in practice, this model is difficult to implement due to the definition of the G-norm. Vese and Osher proposed to use L^p-norm to replace the G-norm [47]. Later, Osher, Sole, and Vese [43] considered the case $p = 2$ which approximates the G-norm by the semi-norm in H^{-1}. Aujol et $al.$ [1] proposed to approximately solve Meyer's original model by adding extra regularization terms. In [18], Daubechies and Teschke represented the cartoon component in a Besov space rather than in the bounded variational space [1, 2, 37, 43, 47].

In [36], an image is treated as a superposition of coherent layers, each layers can be sparsely represented using the wavelet packed transform. In [30], a primal sketch model based on the matching pursuit [35] algorithm and an MRF modeling is proposed. In the sense of morphological component analysis [26, 29, 46], the cartoon part and the texture part can be represented by some sparse linear combinations of basis such that $\boldsymbol{f}_u = \mathcal{D}_u \boldsymbol{\alpha}_u$ and $\boldsymbol{f}_v = \mathcal{D}_v \boldsymbol{\alpha}_v$ for some dictionaries $\mathcal{D}_u$ and $\mathcal{D}_v$, and $\boldsymbol{\alpha}_u$ and $\boldsymbol{\alpha}_v$ are expected to be sparse. The dictionaries are chosen such that each leads to a sparse representation over the content part it is serving, while yielding a non-sparse representation for the other content part. Examples of dictionaries that can sparsely approximate piecewise smooth contents in images are curvelets [9, 10], orthonormal wavelets [17], and wavelet tight frames constructed by the unitary extension principle in [44]. Examples of dictionaries that can sparsely approximate textures are the local discrete cosine transform and the Gabor transform, see [26, 34, 37]. Hence the task of image decomposition is to seek the sparsest representations from the dictionaries which can sparsely approximate both piecewise smooth contents and textures.

Under the assumption that $\mathcal{D}_u \boldsymbol{\alpha}_u$ is mostly piecewise smooth and $\mathcal{D}_v \boldsymbol{\alpha}_v$ is mostly texture, the image decomposition model can be expressed as

$$\min_{\boldsymbol{\alpha}_u, \boldsymbol{\alpha}_v} \left\{ \|\Lambda_u \boldsymbol{\alpha}_u\|_0 + \|\Lambda_v \boldsymbol{\alpha}_v\|_0 + \frac{\eta}{2} \|\boldsymbol{g} - \mathcal{M}(\mathcal{D}_u \boldsymbol{\alpha}_u - \mathcal{D}_v \boldsymbol{\alpha}_v)\|_2^2 \right\}. \tag{2}$$

Here $\Lambda_i, i = u, v$ are the diagonal matrices with the diagonal entries $\lambda_i[s]$, $\|\boldsymbol{x}\|_0$ is the number of non-zero entries in $\boldsymbol{x}$, and η is a regularization parameter. However, the l_0-norm is non-convex and therefore (2) becomes a combinatorial problem and thus it is NP-complete [19, 38]. The convex relaxation of l_1-norm often replaces the l_0-norm because the l_1-norm is convex and leads to a linear programming.

Clearly, one hopes that the solutions for the l_0-norm and the l_1-norm coincide [13, 24]. Since the dictionary $(\mathcal{D}_u, \mathcal{D}_v)$ is a redundant system, there may be other sequences $(\boldsymbol{\alpha}_u^T, \boldsymbol{\alpha}_v^T)^T$ such that $\boldsymbol{f} = \mathcal{D}_u \boldsymbol{\alpha}_u + \mathcal{D}_v \boldsymbol{\alpha}_v$. Therefore, one can restrict $\mathcal{D}_u \boldsymbol{\alpha}_u$ to be piecewise smooth and add an edge-preserving regularization term ϕ with a regularization parameter γ. Hence, In [26, 29, 46], the authors proposed to solve the cartoon and texture reconstruction problem by

$$\{\boldsymbol{\alpha}_u, \boldsymbol{\alpha}_v\} = \operatorname{argmin}_{\boldsymbol{\alpha}_u, \boldsymbol{\alpha}_v} \{\|\Lambda_u \boldsymbol{\alpha}_u\|_1 + \|\Lambda_v \boldsymbol{\alpha}_v\|_1$$
$$+ \frac{\eta}{2} \|\boldsymbol{g} - \mathcal{M}(\mathcal{D}_u \boldsymbol{\alpha}_u - \mathcal{D}_v \boldsymbol{\alpha}_v)\|_2^2 + \gamma \phi(\mathcal{D}_u \boldsymbol{\alpha}_u) \right\}. \tag{3}$$

We call this model Model 1. A modification of Model 1 can be found in [7]. An advantage of changing the l_0-norm to the l_1-norm is that the l_1 sparsity penalty term leads to a soft thresholding scheme. However, images can be decomposed neatly into the cartoon and texture parts only when the chosen dictionaries are appropriate; and this is not always true for arbitrary images. Therefore Model 1 may give a non-sparse pair $\{\boldsymbol{\alpha}_u, \boldsymbol{\alpha}_v\}$. Moreover, if the l_0-norm is replaced by the l_1-norm, the separation of the cartoon and texture will fail since the matrix in Model 1, i.e., $[\mathcal{M}\mathcal{D}_u \ \mathcal{M}\mathcal{D}_v]$ doesn't satisfy the restrict isometry property (RIP) [8].

Our objective in this paper is to propose a new model for the cartoon and texture image decomposition. In this model, the cartoon part and the texture part can be approximately represented by some sparse linear combination of basis such that

$$\|\boldsymbol{f}_u - \mathcal{D}_u\boldsymbol{\alpha}_u\|_2^2 \le \epsilon, \qquad \|\boldsymbol{f}_v - \mathcal{D}_v\boldsymbol{\alpha}_v\|_2^2 \le \epsilon, \qquad \text{and} \qquad \boldsymbol{f} = \boldsymbol{f}_u + \boldsymbol{f}_v.$$

Then we solve the following optimization problem:

$$\min_{\boldsymbol{\alpha}_u, \boldsymbol{\alpha}_v, \boldsymbol{f}_u, \boldsymbol{f}_v} \mathcal{J} = \|\Lambda_u\boldsymbol{\alpha}_u\|_p + \|\Lambda_v\boldsymbol{\alpha}_v\|_p + \frac{1}{2}\|\boldsymbol{f}_u - \mathcal{D}_u\boldsymbol{\alpha}_u\|_2^2 + \frac{1}{2}\|\boldsymbol{f}_v - \mathcal{D}_v\boldsymbol{\alpha}_v\|_2^2$$

$$+ \frac{\eta}{2}\|\boldsymbol{g} - \mathcal{M}(\boldsymbol{f}_u + \boldsymbol{f}_v)\|_2^2 + \gamma\phi(\boldsymbol{f}_u). \qquad (4)$$

Here $p = 0$ or $p = 1$. We call this model Model 2..

To simplify the discussions, we denote

$$A = \begin{bmatrix} I \ I \end{bmatrix}, \quad \boldsymbol{z} = \begin{pmatrix} \boldsymbol{\alpha}_u \\ \boldsymbol{\alpha}_v \end{pmatrix}, \quad \text{and} \quad \boldsymbol{x} = \begin{pmatrix} \boldsymbol{f}_u \\ \boldsymbol{f}_v \end{pmatrix},$$

and hence the objective function (4) can be reformulated as

$$\mathcal{Q}(\boldsymbol{x}, \boldsymbol{z}) = \|\Lambda\boldsymbol{z}\|_p + \frac{1}{2}\|\boldsymbol{x} - \mathcal{D}\boldsymbol{z}\|_2^2 + \gamma\Phi(\boldsymbol{x}) + \frac{\eta}{2}\|\boldsymbol{g} - \mathcal{M}A\boldsymbol{x}\|_2^2. \qquad (5)$$

Here

$$\Lambda = \begin{pmatrix} \Lambda_u & 0 \\ 0 & \Lambda_v \end{pmatrix}, \quad \mathcal{D} = \begin{pmatrix} \mathcal{D}_u & 0 \\ 0 & \mathcal{D}_v \end{pmatrix}, \quad \text{and} \quad \Phi(\boldsymbol{x}) = \phi(\boldsymbol{f}_u).$$

The minimization problem (5) is solved alternatively for $\boldsymbol{x}$ and $\boldsymbol{z}$. We will see in the next section that, if we fix $\boldsymbol{x}$, the minimizer $\boldsymbol{z}$ can be obtained by using shrinkage when $\mathcal{D}_u$ and $\mathcal{D}_v$ are unitary transforms. If we fix $\boldsymbol{z}$, the reduced nonlinear equation in $\boldsymbol{x}$ can be solved by a fixed point method. In this paper, we show that the sequence obtained by the alternating minimization algorithm converges to a fixed point. We also prove that the number of non-zero entries in the sequence $\boldsymbol{z}_k$ does not change when k is large enough.

The outline of the paper is as follows. In Section 2, we introduce our alternating iterative method to find the minimizer of (5). In Section 3, we study the convergence of the algorithm. In Section 4, numerical examples are given to demonstrate the effectiveness of the proposed model. Finally, a summary is given in Section 5.

2.   Iterative Algorithm

There are two variables in the objective function (5). One is the cartoon and texture image $\boldsymbol{x}$, and the other is the sparse representation $\boldsymbol{z}$ of the cartoon and texture

4

image $\boldsymbol{x}$. We use an alternating minimization algorithm to find them. Starting from an initial guess $\boldsymbol{x}_0$, this method computes $\boldsymbol{z}_k$ and $\boldsymbol{x}_k$ alternatively using

$$\mathcal{S}_d(\boldsymbol{x}_{k-1}) := \boldsymbol{z}_k = \arg\min_{\boldsymbol{z}} \mathcal{Q}(\boldsymbol{x}_{k-1}, \boldsymbol{z}), \quad k = 1, 2, \cdots, \tag{6}$$

$$\mathcal{S}_h(\boldsymbol{z}_k) := \boldsymbol{x}_k = \arg\min_{\boldsymbol{x}} \mathcal{Q}(\boldsymbol{x}, \boldsymbol{z}_k), \quad k = 1, 2, \cdots. \tag{7}$$

Note that we have

$$\boldsymbol{x}_k = \mathcal{S}_h(\mathcal{S}_d(\boldsymbol{x}_{k-1})), \quad k = 1, 2, \cdots. \tag{8}$$

In (6), we find $\boldsymbol{z}_k$ of the cartoon and texture image with $\boldsymbol{x}_{k-1}$ given. The dictionaries $\mathcal{D}_u$ and $\mathcal{D}_v$ in $\mathcal{Q}$ are usually built by taking the union of one or several transforms. Generally each corresponds to an orthogonal basis or a tight frame. For simplicity, one can assume that the dictionaries involved are unitary such that $\mathcal{D}_i^T \mathcal{D}_i = I$, $i = u, v$. Recall that $\mathcal{D} = \mathrm{diag}(\mathcal{D}_u, \mathcal{D}_v)$, we have $\mathcal{D}^T \mathcal{D} = I$. We denote $\boldsymbol{\beta}_{k-1} = \mathcal{D}^T \boldsymbol{x}_{k-1}$. By using the unitary invariance property of the 2-norm, the minimizer of the optimization problem (6) can be formulated as

$$\min_{\boldsymbol{z}} \frac{1}{2}\|\boldsymbol{x}_{k-1} - \mathcal{D}\boldsymbol{z}\|_2^2 + \|\Lambda \boldsymbol{z}\|_p = \min_{\boldsymbol{z}} \sum_s \left\{ \frac{1}{2}(\boldsymbol{\beta}_{k-1}[s] - z[s])^2 + |\lambda[s]z[s]|^p \right\}. \tag{9}$$

It is easy to see that each coefficient $z[s]$ can be solved independently as a scalar optimization problem. In particular, when $p = 1$, the solution of (9) is just the popular soft thresholding scheme [23]:

$$z[s] = \begin{cases} \boldsymbol{\beta}_{k-1}[s] - \lambda[s], & \text{if } \boldsymbol{\beta}_{k-1}[s] \geq \lambda[s], \\ 0, & \text{if } |\boldsymbol{\beta}_{k-1}[s]| < \lambda[s], \\ \boldsymbol{\beta}_{k-1}[s] + \lambda[s], & \text{if } \boldsymbol{\beta}_{k-1}[s] \leq -\lambda[s]. \end{cases} \tag{10}$$

When $p = 0$, if we define $0^0 = 0$, then the solution in (9) is just the popular hard thresholding scheme:

$$z[s] = \begin{cases} \boldsymbol{\beta}_{k-1}[s], & \text{if } |\boldsymbol{\beta}_{k-1}[s]| \geq \sqrt{2\lambda[s]}, \\ 0, & \text{if } |\boldsymbol{\beta}_{k-1}[s]| < \sqrt{2\lambda[s]}. \end{cases} \tag{11}$$

Non-unitary or redundant representations is widely used because of its superiority to non-redundant cases [5, 16, 20–22, 27, 28, 32, 45]. We remark that when the dictionaries $\mathcal{D}_i$ are non-unitary or redundant, thresholding is still practiced, see [25]. We can apply block coordinate relaxation method [6] to generate a similar result for the non-unitary case.

Next we consider the iteration (7) where we have to find $\boldsymbol{x}_k$ with $\boldsymbol{z}_k$ given. Note that

$$\frac{\partial}{\partial \boldsymbol{x}} \mathcal{Q}(\boldsymbol{x}, \boldsymbol{z}_k) = (\boldsymbol{x} - \mathcal{D}\boldsymbol{z}_k) + \eta A^T \mathcal{M}^T (\mathcal{M} A \boldsymbol{x} - \boldsymbol{g}) + \gamma B(\boldsymbol{x}) \boldsymbol{x} = 0. \tag{12}$$

Here $B(\boldsymbol{x})$ is the Hessian matrix of $\Phi(\boldsymbol{x})$. If $\phi(\boldsymbol{f}_u) = \|\boldsymbol{f}_u\|_{TV}$ is used, (12) is no longer linear with respect to $\boldsymbol{x}$. Several numerical methods have been proposed for solving (12) in this case: partial differential equation methods such as the explicit [42], the semi-implicit [31], the operator splitting [33] and the lagged diffusivity

fixed-point iterations [48]. Lagged diffusivity fixed-point iterations can be interpreted within the framework of generalized Weiszfeld's methods or majorization minimization algorithms [4]. As proved in [12], this method is monotonically convergent, i.e. the values of the objective function in the iterations is monotonically decreasing, and that the convergence rate is linear.

Here we solve (12) by a simple fixed-point iteration: given $\boldsymbol{x}_{k,p-1}$, we get $\boldsymbol{x}_p$ by solving

$$(\boldsymbol{x}_{k,p} - \mathcal{D}\boldsymbol{z}_k) + \eta A^T \mathcal{M}^T (\mathcal{M}A\boldsymbol{x}_{k,p} - \boldsymbol{g}) + \gamma B(\boldsymbol{x}_{k,p-1})\boldsymbol{x}_{k,p} = 0.$$

Hence we have

$$\boldsymbol{x}_{k,p} = (I + \eta A^T \mathcal{M}^T \mathcal{M}A + \gamma B(\boldsymbol{x}_{k,p-1}))^{-1}(\eta A^T \mathcal{M}^T \boldsymbol{g} + \mathcal{D}\boldsymbol{z}_k), \qquad (13)$$

and the solution of (12) is given by $\boldsymbol{x}_k = \lim_{p\to\infty} \boldsymbol{x}_{k,p}$. Conjugate gradient methods can be used to solve (13) at each iteration. Convergence rate can be improved by using preconditioning techniques such as the transform-based preconditioning techniques [11, 39].

If $\phi(\boldsymbol{f}_u) = \frac{1}{2}\|L\boldsymbol{f}_u\|_2^2$, where L is a regularization matrix, then $B(\boldsymbol{x})$ is independent of $\boldsymbol{x}$. In this case, (12) becomes a linear equation that one can solve directly, see (13) with $B(\boldsymbol{x}_{k,p-1}) \equiv B$. Conjugate gradient methods or other direct methods can be used to solve the linear system. For instance, if $\mathcal{M}$ is a blurring matrix generated by a symmetric point spread function, then $\mathcal{M}$ can be diagonalized by a fast transform matrix, and (13) can be solved by using three fast transforms in $O(n^2 \log n)$ operations for an n-by-n image, see for instance [40].

3. Convergence Analysis

In this section, we study the convergence of our alternating minimization algorithm (8). We use the following Opial theorem to show that it converges to a minimizer of the cost function $\mathcal{Q}$ given in (5).

Theorem 3.1 : *(Opial [41]) Let the mapping $\mathcal{T}$ from $\mathbb{R}^{n^2}$ to $\mathbb{R}^{n^2}$ satisfy the following:*

    (1) $\mathcal{T}$ is asymptotically regular;
    (2) $\mathcal{T}$ is non-expansive;
    (3) the set $\mathbf{F}$ of the fixed-points of $\mathcal{T}$ is not empty.

Then for any $\boldsymbol{f} \in \mathbb{R}^{n^2}$, the sequence $(\mathcal{T}^n \boldsymbol{f})_{n\in\mathcal{N}}$ converges to a fixed-point in $\mathbf{F}$.

We will show that the mapping $\mathcal{T} := \mathcal{S}_h \circ \mathcal{S}_d$ defined in (8) satisfies these assumptions. For the first assumption, we need the following lemma.

Lemma 3.2: *Assume that $\Phi(\boldsymbol{x})$ is a convex function. Let $\boldsymbol{x}_k$ be the sequence generated by (8). Then*

$$\sum_{j=0}^{\infty} \|\boldsymbol{x}_j - \boldsymbol{x}_{j+1}\|_2^2$$

is bounded and convergent.

Proof: Denote $\mathcal{Q}_1(\boldsymbol{x}, \boldsymbol{z}_k) = \frac{1}{2}\|\boldsymbol{x} - \mathcal{D}\boldsymbol{z}_k\|_2^2 + \frac{\eta}{2}\|\boldsymbol{g} - \mathcal{M}A\boldsymbol{x}\|_2^2$. Notice that

$$\mathcal{Q}(\boldsymbol{x}, \boldsymbol{z}_k) = \mathcal{Q}_1(\boldsymbol{x}, \boldsymbol{z}_k) + \|\Lambda \boldsymbol{z}_k\|_p + \gamma \Phi(\boldsymbol{x}),$$

we have

$$\begin{aligned}
&\mathcal{Q}(\boldsymbol{x}_{k-1}, \boldsymbol{z}_k) - \mathcal{Q}(\boldsymbol{x}_k, \boldsymbol{z}_k) \\
&= \mathcal{Q}_1(\boldsymbol{x}_{k-1}, \boldsymbol{z}_k) - \mathcal{Q}_1(\boldsymbol{x}_k, \boldsymbol{z}_k) + \gamma \Phi(\boldsymbol{x}_{k-1}) - \gamma \Phi(\boldsymbol{x}_k).
\end{aligned} \tag{14}$$

Since $\Phi(\boldsymbol{x})$ is a convex function, we deduce that

$$\Phi(\boldsymbol{x}_{k-1}) - \Phi(\boldsymbol{x}_k) \geq (\boldsymbol{x}_{k-1} - \boldsymbol{x}_k)^T \frac{\partial}{\partial \boldsymbol{x}} \Phi(\boldsymbol{x}_k). \tag{15}$$

Next we consider the lower bound of $\mathcal{Q}_1(\boldsymbol{x}_{k-1}, \boldsymbol{z}_k) - \mathcal{Q}_1(\boldsymbol{x}_k, \boldsymbol{z}_k)$. The Taylor expansion of $\mathcal{Q}_1(\boldsymbol{x}, \boldsymbol{z}_k)$ is

$$\begin{aligned}
\mathcal{Q}_1(\boldsymbol{x}, \boldsymbol{z}_k) &= \mathcal{Q}_1(\boldsymbol{x}_k, \boldsymbol{z}_k) + (\boldsymbol{x} - \boldsymbol{x}_k)^T \frac{\partial}{\partial \boldsymbol{x}} \mathcal{Q}_1(\boldsymbol{x}_k, \boldsymbol{z}_k) \\
&\quad + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}_k)^T \frac{\partial^2}{\partial \boldsymbol{x}^2} \mathcal{Q}_1(\boldsymbol{x}_k, \boldsymbol{z}_k)(\boldsymbol{x} - \boldsymbol{x}_k).
\end{aligned}$$

Since $\mathcal{Q}_1(\boldsymbol{x}, \boldsymbol{z}_k)$ is a quadratic function and

$$\frac{\partial^2}{\partial \boldsymbol{x}^2} \mathcal{Q}_1(\boldsymbol{x}, \boldsymbol{z}_k) = (I + \eta A^T \mathcal{M}^T \mathcal{M} A),$$

we have

$$\begin{aligned}
&\mathcal{Q}_1(\boldsymbol{x}, \boldsymbol{z}_k) - \mathcal{Q}_1(\boldsymbol{x}_k, \boldsymbol{z}_k) \\
&= (\boldsymbol{x} - \boldsymbol{x}_k)^T \frac{\partial}{\partial \boldsymbol{x}} \mathcal{Q}_1(\boldsymbol{x}_k, \boldsymbol{z}_k) + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}_k)^T (I + \eta A^T \mathcal{M}^T \mathcal{M} A)(\boldsymbol{x} - \boldsymbol{x}_k). \\
&\geq (\boldsymbol{x} - \boldsymbol{x}_k)^T \frac{\partial}{\partial \boldsymbol{x}} \mathcal{Q}_1(\boldsymbol{x}_k, \boldsymbol{z}_k) + \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{x}_k\|_2^2.
\end{aligned}$$

Let $\boldsymbol{x} = \boldsymbol{x}_{k-1}$ in the above inequality, then with the use of (14) and (15), we obtain

$$\begin{aligned}
&\mathcal{Q}(\boldsymbol{x}_{k-1}, \boldsymbol{z}_k) - \mathcal{Q}(\boldsymbol{x}_k, \boldsymbol{z}_k) \\
&\geq (\boldsymbol{x}_{k-1} - \boldsymbol{x}_k)^T \frac{\partial}{\partial \boldsymbol{x}} \mathcal{Q}_1(\boldsymbol{x}_k, \boldsymbol{z}_k) + \gamma(\boldsymbol{x}_{k-1} - \boldsymbol{x}_k)^T \frac{\partial}{\partial \boldsymbol{x}} \Phi(\boldsymbol{x}_k) + \frac{1}{2}\|\boldsymbol{x}_{k-1} - \boldsymbol{x}_k\|_2^2.
\end{aligned}$$

Since

$$\frac{\partial}{\partial \boldsymbol{x}} \mathcal{Q}(\boldsymbol{x}_k, \boldsymbol{z}_k) = \frac{\partial}{\partial \boldsymbol{x}} \mathcal{Q}_1(\boldsymbol{x}_k, \boldsymbol{z}_k) + \gamma \frac{\partial}{\partial \boldsymbol{x}} \Phi(\boldsymbol{x}_k)$$

and $\boldsymbol{x}_k$ is the minimizer of the cost function $\mathcal{Q}(\boldsymbol{x}, \boldsymbol{z}_k)$, we have $\frac{\partial}{\partial \boldsymbol{x}} \mathcal{Q}(\boldsymbol{x}_k, \boldsymbol{z}_k) = 0$. Therefore we obtain

$$\mathcal{Q}(\boldsymbol{x}_{k-1}, \boldsymbol{z}_k) - \mathcal{Q}(\boldsymbol{x}_k, \boldsymbol{z}_k) \geq \frac{1}{2}\|\boldsymbol{x}_{k-1} - \boldsymbol{x}_k\|_2^2.$$

Notice that $\boldsymbol{z}_k = \mathrm{argmin}_{\boldsymbol{z}} \mathcal{Q}(\boldsymbol{x}_{k-1}, \boldsymbol{z})$, we deduce $\mathcal{Q}(\boldsymbol{x}_{k-1}, \boldsymbol{z}_{k-1}) \geq \mathcal{Q}(\boldsymbol{x}_{k-1}, \boldsymbol{z}_k)$, and hence

$$\mathcal{Q}(\boldsymbol{x}_{k-1}, \boldsymbol{z}_{k-1}) - \mathcal{Q}(\boldsymbol{x}_k, \boldsymbol{z}_k) \geq \mathcal{Q}(\boldsymbol{x}_{k-1}, \boldsymbol{z}_k) - \mathcal{Q}(\boldsymbol{x}_k, \boldsymbol{z}_k)$$
$$\geq \frac{1}{2} \|\boldsymbol{x}_{k-1} - \boldsymbol{x}_k\|_2^2.$$

Summing the inequalities from $k = 1$ to p, we obtain

$$\mathcal{Q}(\boldsymbol{x}_0, \boldsymbol{z}_0) - \mathcal{Q}(\boldsymbol{x}_p, \boldsymbol{z}_p) = \sum_{j=0}^{p-1} \left(\mathcal{Q}(\boldsymbol{x}_j, \boldsymbol{z}_j) - \mathcal{Q}(\boldsymbol{x}_{j+1}, \boldsymbol{z}_{j+1}) \right)$$
$$\geq \frac{1}{2} \sum_{j=0}^{p-1} \|\boldsymbol{x}_j - \boldsymbol{x}_{j+1}\|_2^2.$$

Let $p \to \infty$, then we see that $\sum_{j=0}^{\infty} \|\boldsymbol{x}_j - \boldsymbol{x}_{j+1}\|_2^2$ is bounded and convergent. $\square$

We have following lemma immediately.

Lemma 3.3: *For a given $\boldsymbol{x}_0$, the sequence $\boldsymbol{x}_k$ generated by (8) satisfies*

$$\lim_{k \to \infty} \|\boldsymbol{x}_k - \boldsymbol{x}_{k-1}\|_2 = \lim_{k \to \infty} \|\mathcal{T}^k(\boldsymbol{x}_0) - \mathcal{T}^{k-1}(\boldsymbol{x}_0)\|_2 = 0,$$

where $\mathcal{T}(\boldsymbol{x}) := \mathcal{S}_h(\mathcal{S}_d(\boldsymbol{x}))$. Therefore the operator $\mathcal{T}(\cdot) = \mathcal{S}_h(\mathcal{S}_d(\cdot))$ is asymptotically regular.

Next we consider the non-expansiveness of the operator $\mathcal{S}_h(\cdot)$.

Definition 3.4: [14] An operator $\mathcal{F} : \mathbb{R}^{n^2} \to \mathbb{R}^{n^2}$ is called *non-expansive* if for any $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{R}^{n^2}$, we have $\|\mathcal{F}(\boldsymbol{x}_1) - \mathcal{F}(\boldsymbol{x}_2)\|_2 \leq \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2$. If there exists a number $\kappa \in (0, 1)$ and a non-expansive operator $\mathcal{E} : \mathbb{R}^{n^2} \to \mathbb{R}^{n^2}$ such that $\mathcal{F} = (1 - \kappa)\mathcal{I} + \kappa\mathcal{E}$ is non-expansive, then $\mathcal{F}$ is called *κ-averaged*. In particular, when $\kappa = 1/2$, $\mathcal{F}$ is called a *firmly non-expansive operator*.

We remark that an equivalent definition of firmly non-expansive operator is:

$$\|\mathcal{F}(\boldsymbol{x}_1) - \mathcal{F}(\boldsymbol{x}_2)\|_2^2 \leq [\mathcal{F}(\boldsymbol{x}_1) - \mathcal{F}(\boldsymbol{x}_2)]^T (\boldsymbol{x}_1 - \boldsymbol{x}_2), \tag{16}$$

see for instance [15].

Lemma 3.5: *[15, Lemma 2.4] Let φ be convex and semi-continuous, $\alpha > 0$ and*

$$\widehat{\boldsymbol{x}} := \arg\min_{\boldsymbol{x}} \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{x}\|_2^2 + \alpha\varphi(\boldsymbol{x}). \tag{17}$$

Then the operator $\mathcal{S}$ defined by $\mathcal{S}(\boldsymbol{y}) = \widehat{\boldsymbol{x}}$ is $\frac{1}{2}$-averaged non-expansive.

For the operator $\mathcal{S}_h(\boldsymbol{z})$, we have

$$\mathcal{S}_h(\boldsymbol{z}) = \arg\min_{\boldsymbol{x}} \frac{1}{2} \|\boldsymbol{x} - \mathcal{D}\boldsymbol{z}\|_2^2 + \gamma\Phi(\boldsymbol{x}) + \frac{\eta}{2} \|\boldsymbol{g} - \mathcal{M}HA\boldsymbol{x}\|_2^2.$$

Since $\gamma\Phi(\boldsymbol{x}) + \eta\|\boldsymbol{g} - \mathcal{M}HA\boldsymbol{x}\|_2^2$ is convex and semi-continuous, we know that $\mathcal{S}_h(\boldsymbol{z})$ is also $\frac{1}{2}$-averaged non-expansive.

8

Next, we consider the non-expansiveness of the operator $\mathcal{S}_d(\boldsymbol{x})$. When the l_1 penalty is used, $\mathcal{S}_d(\boldsymbol{x})$ is just the soft thresholding operator given in (10) which is known to be non-expansive. When the l_0 penalty is used, $\mathcal{S}_d(\boldsymbol{x})$ becomes the hard thresholding operator in (11). Hard thresholding operators are not non-expansive in general. We however will show that the hard thresholding operator in our alternating minimization approach is equivalent to a projection under some assumptions, and therefore is non-expansive. We consider the l_1 penalty and the l_0 penalty separately.

3.1.   *The l_1 Penalty*

When the l_1 penalty is used, as $\sum_s |\lambda[s]z[s]|$ is convex and semi-continuous, we know that the operator $\mathcal{S}_d(\cdot)$ is non-expansive. The product of two averaged non-expansive operators is again averaged non-expansive, and hence the operator $\mathcal{S}_h(\mathcal{S}_d(\cdot))$ is averaged non-expansive. According to the Opial theorem (Theorem 3.1), the sequence $\{\boldsymbol{x}_k\}$ with $\boldsymbol{x}_{k+1} = \mathcal{S}_h(\mathcal{S}_d(\boldsymbol{x}_k))$ converges should it satisfy the three conditions mentioned there. We have already shown that $\mathcal{S}_h(\mathcal{S}_d(\cdot))$ is asymptotically regular and non-expansive. Hence we only need to show that the set of fixed-points of $\mathcal{S}_h(\mathcal{S}_d(\cdot))$ is not empty. To do this, we first prove that the minimizer of the cost function $\mathcal{Q}(\boldsymbol{x}, \boldsymbol{z})$ is a fixed-point of the operator $\mathcal{S}_h(\mathcal{S}_d(\cdot))$, and then we show that the cost function $\mathcal{Q}(\boldsymbol{x}, \boldsymbol{z})$ is coercive. It then guarantees that the set of minimizers $(\boldsymbol{x}, \boldsymbol{z})$ of $\mathcal{Q}(\boldsymbol{x}, \boldsymbol{z})$ is non-empty.

Definition 3.6:   A function $\phi : \mathbb{R}^{n^2} \to \mathbb{R}$ is *proper* over a set $\boldsymbol{X} \subset \mathbb{R}^{n^2}$ if $\phi(\boldsymbol{x}) < \infty$ for at least one $\boldsymbol{x} \in \boldsymbol{X}$ and $\phi(\boldsymbol{x}) > -\infty$ for all $\boldsymbol{x} \in \boldsymbol{X}$. A function $\phi : \mathbb{R}^{n^2} \to \mathbb{R}$ is *coercive* over a set $\boldsymbol{X} \subseteq \mathbb{R}^{n^2}$ if for every sequence $\{\boldsymbol{x}_k\} \subset \boldsymbol{X}$ such that $\|\boldsymbol{x}_k\|_2 \to \infty$, we have $\lim_{k\to\infty} \phi(\boldsymbol{x}_k) = \infty$.

Lemma 3.7:   *[3, Proposition 2.1.1] Let $\phi : \mathbb{R}^{n^2} \to \mathbb{R}$ be a closed, proper, and coercive function. Then the set of minimizers of ϕ over $\mathbb{R}^{n^2}$ is nonempty and compact.*

Note that for any $(\boldsymbol{x}^T, \boldsymbol{z}^T)^T \to \infty$, we have $\|\Lambda\boldsymbol{z}\|_1 + \frac{1}{2}\|\boldsymbol{x} - \mathcal{D}\boldsymbol{z}\|_2^2 \to \infty$. Therefore the cost function $\mathcal{Q}(\boldsymbol{x}, \boldsymbol{z})$ is coercive and the set of minimizers for $\mathcal{Q}(\boldsymbol{x}, \boldsymbol{z})$ is nonempty.

We now relate the set of minimizers of $\mathcal{Q}(\boldsymbol{x}, \boldsymbol{z})$ to the set of fixed-points of the operator $\mathcal{S}_h(\mathcal{S}_d(\cdot))$. Assume that $(\boldsymbol{x}, \boldsymbol{z})$ is a minimizer of $\mathcal{Q}(\boldsymbol{x}, \boldsymbol{z})$, i.e.,

$$\begin{pmatrix} \frac{\partial \mathcal{Q}}{\partial \boldsymbol{x}}(\boldsymbol{x}, \boldsymbol{z}) \\ \frac{\partial \mathcal{Q}}{\partial \boldsymbol{z}}(\boldsymbol{x}, \boldsymbol{z}) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

It implies that

$$\begin{cases} \boldsymbol{z} = \operatorname{argmin} \mathcal{Q}(\boldsymbol{x}, \cdot) \\ \boldsymbol{x} = \operatorname{argmin} \mathcal{Q}(\cdot, \boldsymbol{z}) \end{cases}$$

If $(\boldsymbol{x}, \boldsymbol{z})$ is a minimizer of $\mathcal{Q}(\boldsymbol{x}, \boldsymbol{z})$, we deduce that $\boldsymbol{x}$ is the fixed-point of $\mathcal{S}_h(\mathcal{S}_d(\cdot))$. Hence the set of fixed-points of $\mathcal{S}_h(\mathcal{S}_d(\cdot))$ is non-empty. On the other hand, if $(\boldsymbol{x}, \boldsymbol{z})$ is a fixed-point of $\mathcal{S}_h(\mathcal{S}_d(\cdot))$, we have $\boldsymbol{z} = \mathcal{S}_d(\boldsymbol{x})$ and $\boldsymbol{x} = \mathcal{S}_h(\boldsymbol{z})$. This means that $(\boldsymbol{x}, \boldsymbol{z})$ is a minimizer of $\mathcal{Q}(\boldsymbol{x}, \boldsymbol{z})$.

Since the set of minimizers $(\boldsymbol{x}, \boldsymbol{z})$ of $\mathcal{Q}(\boldsymbol{x}, \boldsymbol{z})$ is non-empty, the set of fixed-points of $\boldsymbol{x} = \mathcal{S}_h(\mathcal{S}_d(\boldsymbol{x}))$ is also non-empty. Moreover the operator $\mathcal{S}_h(\mathcal{S}_d(\cdot))$ is non-expansive and asymptotically regular. Hence we have the following theorem.

Theorem 3.8: *The sequence $\{\boldsymbol{x}_k\}$ with $\boldsymbol{x}_{k+1} = \mathcal{S}_h(\mathcal{S}_d(\boldsymbol{x}_k))$ converges to a fixed-point $\boldsymbol{x}$, and the corresponding point $(\boldsymbol{x}, \mathcal{S}_d(\boldsymbol{x}))$ is a minimizer of $\mathcal{Q}(\boldsymbol{x}, \boldsymbol{z})$.*

3.2. The l_0 Penalty

To show the convergence when the l_0-norm is used, we need the following lemma.

Lemma 3.9: *Let $\phi(\boldsymbol{f}) = \|L\boldsymbol{f}\|_2^2$. Then for any given $\epsilon > 0$, there exists a constant M such that for all $k > M$, we have $\|\boldsymbol{z}_{k-1} - \boldsymbol{z}_k\|_2^2 < \epsilon$.*

Proof: Since $\phi(\boldsymbol{f}) = \|L\boldsymbol{f}\|_2^2$, the Hessian matrix of $\phi(\boldsymbol{f})$ is independent on $\boldsymbol{f}$. Hence B is independent on $\boldsymbol{x}$. Assume that the minimum eigenvalue of the matrix $(I + \eta A^T \mathcal{M}^T \mathcal{M} A + \gamma B)^{-1}$ is ρ. From Lemma 3.3, we know that given $\epsilon > 0$, there exists a constant M such that for any $k > M$, we have $\|\boldsymbol{x}_k - \boldsymbol{x}_{k+1}\|_2^2 < \rho^2 \epsilon$. Since

$$\boldsymbol{x}_k = (I + \eta A^T \mathcal{M}^T \mathcal{M} A + \gamma B)^{-1}(\eta A^T \mathcal{M}^T \boldsymbol{g} + \mathcal{D}\boldsymbol{z}_{k-1})$$

We obtain

$$\|\boldsymbol{x}_k - \boldsymbol{x}_{k+1}\|_2^2 = \|(I + \eta A^T \mathcal{M}^T \mathcal{M} A + \gamma B)^{-1}\mathcal{D}(\boldsymbol{z}_{k-1} - \boldsymbol{z}_k)\|_2^2$$
$$\geq \|\rho \mathcal{D}(\boldsymbol{z}_{k-1} - \boldsymbol{z}_k)\|_2^2 = \rho^2 \|\boldsymbol{z}_{k-1} - \boldsymbol{z}_k\|_2^2.$$

Therefore, the result holds. □

The following lemma states that the number of non-zero entries in the vector $\boldsymbol{z}_k$ does not change when k is large enough.

Lemma 3.10: *Let $\phi(\boldsymbol{f}) = \|L\boldsymbol{f}\|_2^2$. Give any ϵ such that $0 < \epsilon < \min_s 2\lambda[s]$, then there exists a constant M such that for all $k > M$, we have $|\boldsymbol{z}_k[s]| \geq \sqrt{2\lambda[s]}$ if $|\boldsymbol{z}_{k-1}[s]| \geq \sqrt{2\lambda[s]}$ and $\boldsymbol{z}_k[s] = 0$ if $\boldsymbol{z}_{k-1}[s] = 0$.*

Proof: Assume $|\boldsymbol{z}_{k-1}[s]| \geq \sqrt{2\lambda[s]}$ while $\boldsymbol{z}_k[s] = 0$, or $\boldsymbol{z}_{k-1}[s] = 0$ while $|\boldsymbol{z}_k[s]| \geq \sqrt{2\lambda[s]}$. Then we have $|\boldsymbol{z}_k[s] - \boldsymbol{z}_{k-1}[s]|^2 \geq 2\lambda[s]$. Hence

$$\|\boldsymbol{z}_k - \boldsymbol{z}_{k-1}\|_2^2 = \sum_s |\boldsymbol{z}_k[s] - \boldsymbol{z}_{k-1}[s]|^2 \geq \min_s 2\lambda[s] > \epsilon.$$

This contradicts Lemma 3.9. Therefore the result holds. □

Lemma 3.11: *Let $\phi(\boldsymbol{f}) = \|L\boldsymbol{f}\|_2^2$ and assume the operator $\mathcal{S}_h(\mathcal{S}_d(\cdot))$ has a fixed point. Then the sequence $\boldsymbol{x}_{k+1} = \mathcal{S}_h(\mathcal{S}_d(\boldsymbol{x}_k))$ converges.*

Proof: There exists a constant M such that for all $k > M$, $\|\boldsymbol{z}_k - \boldsymbol{z}_{k-1}\|_2^2 < \epsilon$. Let $\Gamma = \{s : |\boldsymbol{z}_M[s]| \geq \sqrt{2\lambda[s]}\}$ and define the projection $\mathcal{P}(\boldsymbol{x})$ such that $\mathcal{P}(\boldsymbol{x})[s] = \boldsymbol{x}[s]$ if $s \in \Gamma$ and $\mathcal{P}(\boldsymbol{x})[s] = 0$ if $s \notin \Gamma$. Immediately we see that $\mathcal{P}(\cdot)$ is non-expansive.

Now we show that $\mathcal{P}(\boldsymbol{x}_{k-1}) = \mathcal{S}(\boldsymbol{x}_{k-1})$ if $k > M$. For any $k > M$ and the vector $\boldsymbol{\beta}_{k-1} = \mathcal{D}^T \boldsymbol{x}_{k-1}$, we consider two cases: 1) $|\boldsymbol{\beta}_{k-1}[s]| \geq \sqrt{2\lambda[s]}$ and 2) $|\boldsymbol{\beta}_{k-1}[s]| < \sqrt{2\lambda[s]}$.

(1) $|\boldsymbol{\beta}_{k-1}[s]| \geq \sqrt{2\lambda[s]}$: By (11), we have $|\boldsymbol{z}_k[s]| = |\boldsymbol{\beta}_{k-1}[s]| \geq \sqrt{2\lambda[s]}$. Assume that $s \notin \Gamma$, i.e., $|\boldsymbol{z}_M[s]| < \sqrt{2\lambda[s]}$. Then according to Lemma 3.10, we have $|\boldsymbol{z}_k[s]| < \sqrt{2\lambda[s]}$ for all $k > M$. This leads to a contradiction. Therefore, $s \in \Gamma$ and $\mathcal{P}(\boldsymbol{x})[s] = \boldsymbol{x}[s] = \mathcal{S}(\boldsymbol{x})[s]$.

(2) $|\boldsymbol{\beta}_{k-1}[s]| < \sqrt{2\lambda[s]}$: By (11), we have $\boldsymbol{z}_k[s] = 0$. Assume that $s \in \Gamma$, i.e. $|\boldsymbol{z}_M[s]| \geq \sqrt{2\lambda[s]}$. Using Lemma 3.10, we have $|\boldsymbol{z}_k[s]| \geq \sqrt{2\lambda[s]}$ for all

$k > M$. This also leads to a contradiction. Therefore, $s \notin \Gamma$ and $\mathcal{P}(\boldsymbol{x})[s] = 0 = \mathcal{S}(\boldsymbol{x})[s]$.

Therefore, when $k > M$, we have $\boldsymbol{x}_{k+1} = \mathcal{S}_h(\mathcal{S}_d(\boldsymbol{x}_k)) = \mathcal{S}_h(\mathcal{P}(\boldsymbol{x}_k))$. Since $\mathcal{P}$ and $\mathcal{S}_h$ are non-expansive and the sequence $\boldsymbol{x}_k$ is asymptotically regular, by applying Opial's theorem (Theorem 3.1), we know that $\boldsymbol{x}_k$ converges. $\qquad\square$

4. Numerical Results

In this section, we test the performance of our proposed algorithm. Peak Signal-to-Noise Ratio (PSNR) is used to measure the quality of the restoration, it is defined as follows:

$$\mathrm{PSNR} = 10 \log \frac{255^2}{\frac{1}{mn}\|\boldsymbol{f}_u + \boldsymbol{f}_v - \boldsymbol{f}\|_2^2}.$$

Here $\boldsymbol{f}_u$ and $\boldsymbol{f}_v$ are the restored cartoon and texture parts respectively and $\boldsymbol{f}$ is the $m \times n$ original image. The stopping criterion of both Model 1 (see (3)) and our proposed Model 2 (see (4)) is when the relative difference between the successive iterates satisfies:

$$\frac{\|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\|_2}{\|\boldsymbol{x}_k\|_2} < 10^{-4},$$

where $\boldsymbol{x}_k = \begin{pmatrix} \boldsymbol{f}_{u,k} \\ \boldsymbol{f}_{v,k} \end{pmatrix}$. The initial guess is $\boldsymbol{f}_{u,0} = \boldsymbol{g}$ with $\boldsymbol{g}$ being the recorded image and $\boldsymbol{f}_v = 0$.

The first example is depicted in Fig. 1 where the Barbara image was superimposed by a textual mask covering 20% of its original area. The dictionaries used are the fast curvelet transform and the local DCT with block size 32×32. Here we give the results for the l_1-norm only as the results for the l_0-norm are similar. The cartoon part and texture part restored by using Model 1 and Model 2 are shown in middle row and bottom row of Fig. 1 respectively. We see that our method not only restores the corrupted image, but also separates it nicely into the cartoon part and texture part.

We note that it is an open problem for choosing optimal parameters. Comparing with Model 1, our Model 2 introduces an additional parameter η. To illustrate that the parameter η is not difficult to tune, we fixed the other parameters and only changed the parameter η. The relationship between PSNR and η is shown in Fig. 2. We notice that only for extremely small values of η would the reconstruction of the cartoon and texture parts fail. In our experiments, we found that when the values of η is large enough, for example $\eta > 100$, PSNR is not sensitive to the change of η. The other parameters are determined by fixing all but one parameter and adjusting it to give the best restoration measured in PSNR. All the adjusted parameters are tuned one by one.

The second test is a zooming problem, see Fig. 3, where the image of Barbara was regularly down-sampled with 75% missing values. Again the curvelet transform and the local DCT are used. The result obtained by our Model 2 outperforms Model 1.

In the last example, we used a random mask. Fig. 4 shows the Barbara image with 50% pixels randomly filled in. 5 presents the results obtained by using the fast curvelet transform and the local DCT. We notice that in both cases, the results by our model are better than those from Model 1.
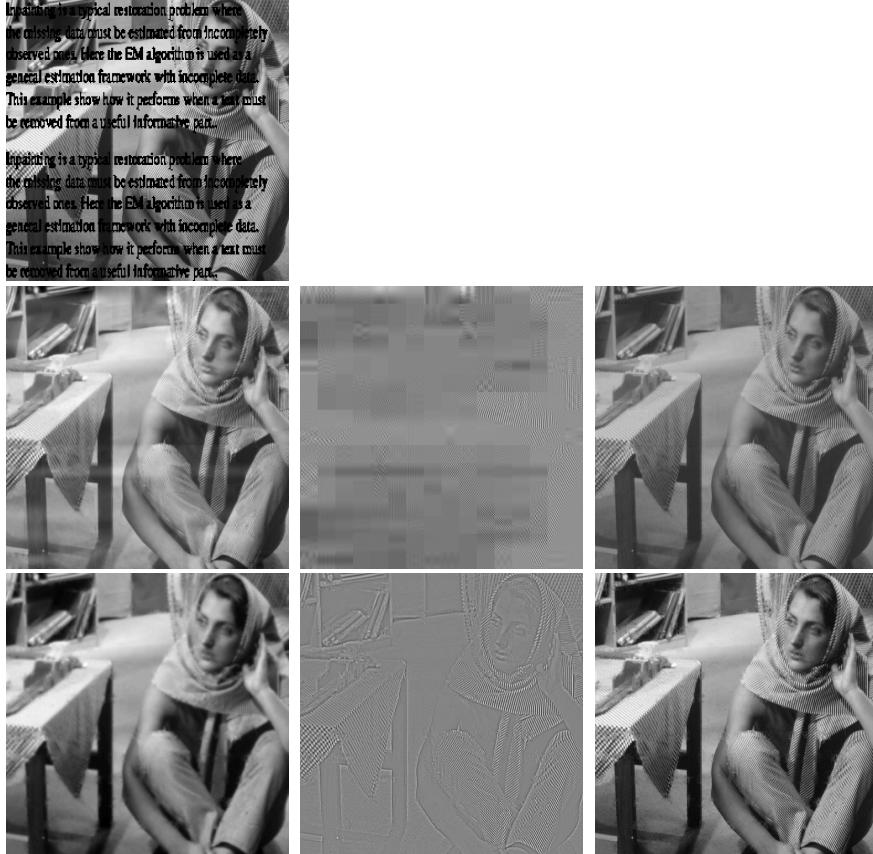
Figure 1.   Example with textual mask. Top: Observed image. Middle: the cartoon part, the texture part and the restored image (PSNR=30.44) by using Model 1. Bottom: the cartoon part, the texture part and the restored image (PSNR=30.64) by using Model 2.
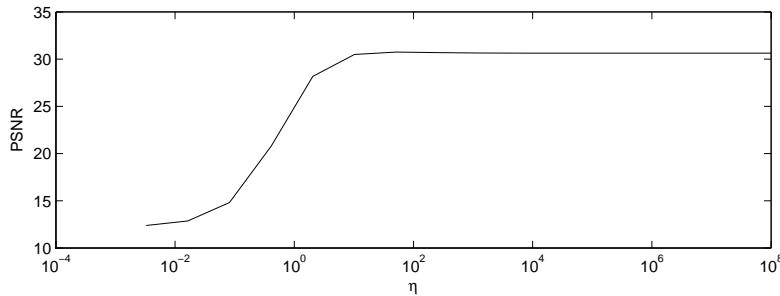

Figure 2. The parameter η versus PSNR (with $\gamma = 0.05$).

5.   Conclusion

We proposed a new model for simultaneous cartoon and texture reconstruction. In the new model, we assume that the cartoon part and the texture part can be approximately represented by some sparse linear combination of basis. An alternating minimization algorithm is used to find the minimum of the cost function. We give the convergence analysis for the proposed algorithm. Our experiments show that our restored images are better than the model 1.

12
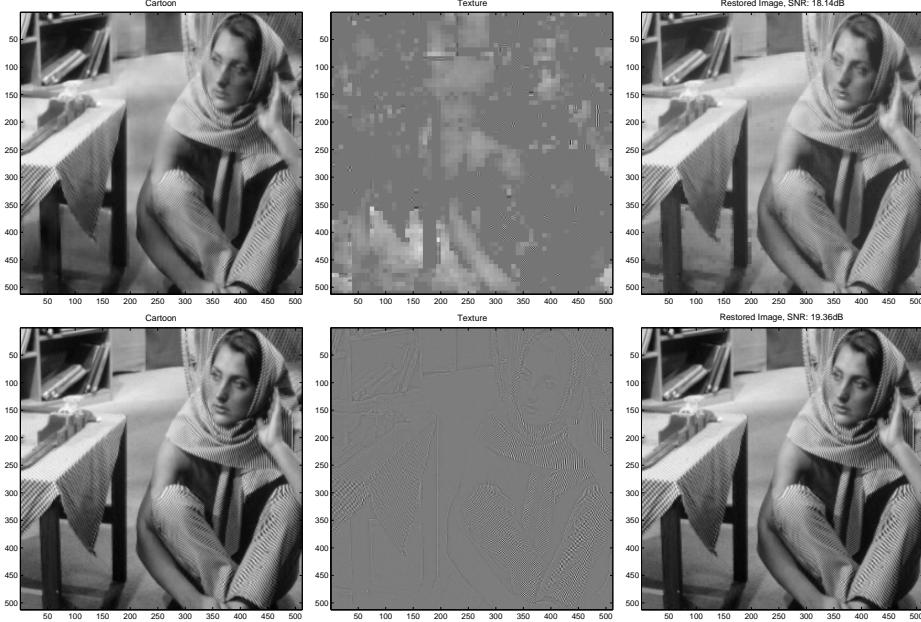
Figure 3. Zooming. The reconstruction image for Model 1 (top row) and Model 2 (bottom row).
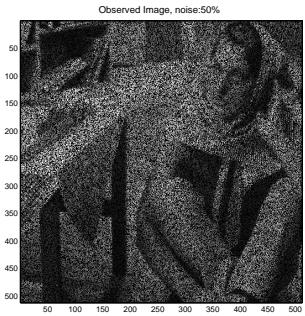

Figure 4. The corrupted image with 50% pixels randomly filled in.
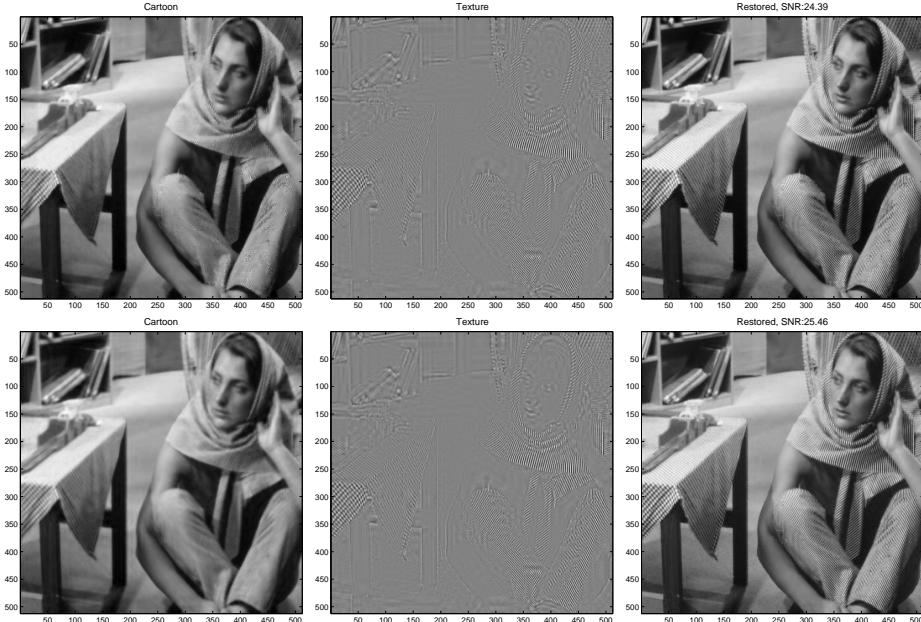
Figure 5. Curvelet transform and local DCT for Model 1 (top row) and Model 2 (bottom row).

References

[1] J. Aujol, G. Aubert, L. Blanc-Feraud, and A. Chambolle. Image decomposition application to SAR images. *Lecture Notes in Computer Science*, 2695:297–312, 2003.

[2] J. Aujol and A. Chambolle. Dual norms and image decomposition models. *International Journal of Computer Vision*, 63(1):85–104, 2005.

[3] D. Bertsekas, A. Nedic, and E. Ozdaglar. *Convex Analysis and Optimization*. Athena Scientific, 2003.

[4] J. Bioucas-Dias, M. Figueiredo, and R. Nowak, Total Variation-based image deconvolution: a majorization-minimization approach. *IEEE International Conference on Acoustics, Speech, and Signal Processing - ICASSP'2006*,2:II-II, 2006.

[5] T. Blu and F. Luisier. The sure-let approach to image denoising. *IEEE Transactions on Image Processing*, 16(11):2778–2786, 2007.

[6] A. Bruce, S. Sardy, and P. Tseng. Block coordinate relaxation methods for nonparametric signal de-noising. *Proceedings of SPIE*, 3391:75–86, 1998.

[7] J. Cai, S. Osher, and Z. Shen. Split Bregman methods and frame based image restoration. *Multiscal Model. Sumul.*, 8(2):337–369, 2009.

[8] E. Candès and T. Tao. Near-optimal signal recovery from random projections: universal encoding strategies. *IEEETrans.Inform. Theory*,52(12):5406-5425, 2006.

[9] E. Candès and D. Donoho. Curvelets: A surprisingly effective nonadaptive representation of objects with edges. Technical report, 1999.

[10] E. Candès and D. Donoho. New tight frames of curvelets and optimal representation of objects with piecewise C2 singularities. *Communications on Pure and Applied Mathematics*, 57:219–266, 2004.

[11] R. Chan and X. Jin. *An Introduction to Iterative Toeplitz Solvers*, SIAM, Philadelphia, PA, 2007.

[12] T. Chan and P. Mulet. On the convergence of the lagged diffusivity fixed point method in total variation image restoration. *SIAM Journal on Numerical Analysis*, 36:354–367, 1999.

[13] S. Chen, D. Donoho, and M. Saunders. Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, 20(1):33–61, 1999.

[14] P. Combettes. Solving monotone inclusions via compositions of nonexpansive averaged operators. *Optimization*, 53(5–6):475–504, 2004.

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

[16] A. Da Cunha, J. Zhou, and M. Do. The nonsubsampled contourlet transform: Theory, design, and applications. *IEEE Transactions on Image Processing*, 15(10):3089–3101, 2006.

[17] I. Daubechies. *Ten Lectures on Wavelets*. SIAM, Philadelphia, PA, 1992.

[18] I. Daubechies and G. Teschkeb. Variational image restoration by means of wavelets: Simultaneous decomposition, deblurring, and denoising. *Applied Compuational Harmonic Analysis.*, 19:1–16, 2005.

[19] G. Davis, S. Mallat, and M. Avellaneda. Adaptive greedy approximations. *Constructive Approximation*, 13(1):57–98, 1997.

[20] M. Do and M. Vetterli. Rotation invariant texture characterization and retrieval using steerable wavelet-domain hidden markov models. *IEEE Transactions on Multimedia*, 4(4):517–527, 2002.

[21] M. Do and M. Vetterli. The finite ridgelet transform for image representation. *IEEE Transactions on Image Processing*, 12(1):16–28, 2003.

[22] M. Do and M. Vetterli. Framing pyramids. *IEEE Transactions on Signal Processing*, 51(9):2329–2342, 2003.

[23] D. Donoho. Denoising by soft-thresholding. *IEEE Transactions on Information Theory*, 41, 1995.

[24] D. Donoho, M. Elad, and V. Temlyakov. Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Transactions on Information Theory*, 52(1):6–18, 2006.

[25] M. Elad. Why simple shrinkage is still relevant for redundant representation. *IEEE Transactions on Information Theory*, 52(12):5559 – 5569, 2006.

[26] M. Elad, J. Starck, P. Querre, and D. Donoho. Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA). *Applied and Computational Harmonic Analysis*, 19(3):340 – 358, 2005.

[27] R. Eslami and H. Radha. The contourlet transform for image denoising using cycle spinning. In H. Radha, editor, *Conference Record of the Thirty-Seventh Asilomar Conference on Signals, Systems and Computers*, volume 2, pages 1982–1986 Vol.2, 2003.

[28] R. Eslami and H. Radha. Translation-invariant contourlet transform and its application to image denoising. *IEEE Transactions on Image Processing*, 15(11):3362–3374, 2006.

[29] M. Fadili, J. Starck, and F. Murtagh. Inpainting and zooming using sparse represntations. *The Computer Journal*, 2007.

[30] C. Guo, S. Zhu, and Y. Wu. Towards a mathematical theory of primal sketch and sketchability. In Song-Chun Zhu, editor, *Proceedings of Ninth IEEE International Conference on Computer Vision, 2003.*, pages 1228–1235 vol.2, 2003.

[31] D. Krishnan, P. Lin, and X. Tai. An efficient operator splitting method for noise removal in images. *Communications in Computational Physics*, 1:847–858, 2006.

[32] M. Lang, H. Guo, J. Odegard, C. Burrus, and R. Wells. Noise reduction using an undecimated discrete wavelet transform. *IEEE Signal Processing Letters*, 3(1), 1996.

[33] M. Lysaker and X. Tai. Noise removal using smoothed normals and surface fitting. *IEEE Transactions on Image Processing*, 13:1345–1357, 2004.

[34] S. Mallat. *A Wavelet Tour of Signal Processing*. 2nd edition. Academic Press: San Diego, 1999.

[35] S. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 41(12):3397–3415, 1993.

[36] F. Meyer, A. Averbuch, and R. Coifman. Multilayered image representation: application to image compression. *IEEE Transactions on Image Processing*, 11(9):1072–1080, 2002.

[37] Y. Meyer. Oscillating patterns in linage processing and nonlinear evolution equations. *University Lecture Series*, 22, 2001.

[38] B. Natarajan. Sparse approximate solutions to linear systems. *SIAM Journal on Computing*, 24:227–234, 1995.

[39] M. Ng. *Iterative Methods for Toeplitz Systems*. Oxford University Press, 2004.

[40] M. Ng, R. Chan, and W. Tang. A fast algorithm for deblurring models with Neumann boundary conditions. *SIAM Journal on Scientific Computing*, 21(3):851–866, 2000.

[41] Z. Opial. Weak convergence of the sequence of successive approximations for nonexpansive mappings. *Bulletin of American Mathematical Society*, 73:591–597, 1967.

[42] S. Osher, L. I. Rudin, and E. Fatem. Nonlinear total variation based noise removal algorithms. *Physics D*, 60:259–268, 1992.

[43] S. Osher, A. Sole, and L. Vese. Image decomposition and restoration using total variation minimization and the H^{-1} norm. *Multiscale Modeling & Simulation*, 1(3):349–370, 2003.

[44] A. Ron and Z. Shen. Affine systems in $L_2(\mathbf{R}^d)$: the analysis of the analysis operator. *Journal of Functional Analysis*, 148(2):408–447, 1997.

[45] J. Starck, E. Candes, and D. Donoho. The curvelet transform for image denoising. *IEEE Transactions on Image Processing*, 11(6):670–684, 2002.

[46] J. Starck, M. Elad, and D. Donoho. Image decomposition via the combination of sparse representations and a variational approach. *IEEE Transactions on Image Processing*, 14(10):1570 – 1582, 2005.

[47] L. Vese and S. Osher. Modeling textures with total variation minimization and oscillating patterns in image processing. *Journal of Scientific Computing*, 19(1–3):553–572, 2002.

[48] C. Vogel and M. Oman. Iterative method for total variation denoising. *SIAM Journal on Scientific Computing*, 17:227–238, 1996.