Journal of Zhejiang University-SCIENCE C (Computers & Electronics)
ISSN 1869-1951 (Print); ISSN 1869-196X (Online)
www.zju.edu.cn/jzus; www.springerlink.com
E-mail: jzus@zju.edu.cn

Adaptive contourlet-wavelet iterative shrinkage/thresholding for remote sensing image restoration*

Nu WEN†[1,2,3], Shi-zhi YANG[1,2], Cheng-jie ZHU[1,2,3], Sheng-cheng CUI[1,2]

([1]*Anhui Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Hefei 230031, China*)

([2]*Key Laboratory of Optical Calibration and Characterization, Chinese Academy of Sciences, Hefei 230031, China*)

([3]*University of Chinese Academy of Sciences, Beijing 100049, China*)

†E-mail: wennu89@mail.ustc.edu.cn

Received Dec. 26, 2013; Revision accepted Apr. 14, 2014; Crosschecked July 16, 2014

Abstract: In this paper, we present an adaptive two-step contourlet-wavelet iterative shrinkage/thresholding (TcwIST) algorithm for remote sensing image restoration. This algorithm can be used to deal with various linear inverse problems (LIPs), including image deconvolution and reconstruction. This algorithm is a new version of the famous two-step iterative shrinkage/thresholding (TwIST) algorithm. First, we use the split Bregman Rudin-Osher-Fatemi (ROF) model, based on a sparse dictionary, to decompose the image into cartoon and texture parts, which are represented by wavelet and contourlet, respectively. Second, we use an adaptive method to estimate the regularization parameter and the shrinkage threshold. Finally, we use a linear search method to find a step length and a fast method to accelerate convergence. Results show that our method can achieve a signal-to-noise ratio improvement (ISNR) for image restoration and high convergence speed.

Key words: Image restoration, Adaptive, Cartoon-texture decomposition, Linear search, Iterative shrinkage/thresholding
doi:10.1631/jzus.C1300377 **Document code:** A **CLC number:** TP7

1 Introduction

In remote sensing image processes, image restoration is an important and practical problem for researchers. A remote sensing image is subjected to degradation caused by different kinds of atmospheric turbulence effects, aerosol scattering, and physical limitations of the sensors. Image restoration is one of the earliest classical linear inverse problems (LIPs) in imaging. Many methods were proposed to deal with these types of inverse problems, which mainly contain regularization, total variation (Bioucas-Dias *et al.*, 2006), and expectation-maximization (EM) (Figueiredo and Nowak, 2003).

An LIP can be solved by using an optimization method to minimize the sum of two convex functions:

$$\min_{\boldsymbol{x}} f(\boldsymbol{x}), \tag{1a}$$

$$f(\boldsymbol{x})=F(\boldsymbol{x})+G(\boldsymbol{x}), \tag{1b}$$

where $F(\boldsymbol{x})$ is a smooth convex function defined by a black-box oracle, and $G(\boldsymbol{x})$ is a general closed convex function that is possibly non-smooth. The image restoration problem is usually described by a linear space invariant convolution (blurring) operator $\boldsymbol{R}$ and additive Gaussian noise $\boldsymbol{n}$ in the literature (Figueiredo and Nowak, 2003; Daubechies *et al.*, 2004; Bioucas-Dias *et al.*, 2006; Figueiredo *et al.*, 2007; Beck and Teboulle, 2009a). We use $\boldsymbol{x}$ to represent the original image and $\boldsymbol{y}$ to represent the observed image. The image degradation model can be formulated as $\boldsymbol{y}=\boldsymbol{R}\boldsymbol{x}+\boldsymbol{n}$. Thus, the original image $\boldsymbol{x}$ can be estimated by minimizing the objective function

* Project supported by the National Science & Technology Pillar Program (No. 2011BAB01B03), the National Natural Science Foundation of China (No. 41305019), and the Anhui Provincial Natural Science Foundation (No. 1308085QD70)

$$f(\boldsymbol{x})=\frac{1}{2}\|\boldsymbol{y}-\boldsymbol{R}\boldsymbol{x}\|_2^2+\tau\phi(\boldsymbol{x}), \tag{2}$$

where $\boldsymbol{R}$ represents a linear operator that is usually singular or ill-conditioned, $\|\cdot\|_2$ is a norm in some given vector space (such as real Hilbert space), and $F(\boldsymbol{x})=\frac{1}{2}\|\boldsymbol{y}-\boldsymbol{R}\boldsymbol{x}\|_2^2$ is the data misfit that represents the error energy between the observed image and the image obtained by the image degradation model, the purpose of which is to insure that the recovered solution is close to the true one. τ is the regularization parameter. $\phi(\boldsymbol{x})$ is finite for all $\boldsymbol{x}$ and is called the regularizer or regularization function, the purpose of which is to overcome the ill-posedness of image restoration.

In this paper, the regularization terms used are l_1-norm ($\phi(\boldsymbol{x})=\|\boldsymbol{x}\|_1$) and TV-norm ($\phi(\boldsymbol{x})=\|\boldsymbol{x}\|_{\mathrm{TV}}$). So, Eq. (2) can be rewritten as

$$f(\boldsymbol{x})=\frac{1}{2}\|\boldsymbol{y}-\boldsymbol{R}\boldsymbol{x}\|_2^2+\tau\|\boldsymbol{x}\|_1, \tag{3a}$$

or

$$f(\boldsymbol{x})=\frac{1}{2}\|\boldsymbol{y}-\boldsymbol{R}\boldsymbol{x}\|_2^2+\tau\|\boldsymbol{x}\|_{\mathrm{TV}}. \tag{3b}$$

This paper is concerned mainly with algorithms for minimizing (3a). All of the numerical experiments are used to solve l_1-norm except for the experiment in Section 5.3. The relative merits of a more general objective function and the choice of $\phi(\boldsymbol{x})$, however, will not be discussed here.

A new algorithm called the iterative shrinkage/thresholding (IST) algorithm was recently proposed to solve LIPs of the form (2). To solve such LIPs, some state-of-the-art algorithms belonging to the IST family were proposed by different researchers. These algorithms include a fast iterative shrinkage/thresholding algorithm (FISTA) (Beck and Teboulle, 2009b), a two-step iterative shrinkage/thresholding (TwIST) algorithm (Bioucas-Dias and Figueiredo, 2007a), and the two-step algorithm (Bioucas-Dias and Figueiredo, 2007b). The EM algorithm for wavelet-based deconvolution was proposed by Nowak and Figueiredo (2001) and Figueiredo and Nowak (2003). The majorization minimization is an optimization technique used in different algorithms, including IST/FISTA algorithms (Hunter and Lange, 2004; Bioucas-Dias *et al.*, 2006; Figueiredo *et al.*, 2007). Recently, Combettes and Wajs (2005) brought important contributions to the understanding of the forward-backward splitting algorithm which includes IST. Another class of IST algorithms called iterative reweighed shrinkage (IRS) was proposed by Bioucas-Dias (2006).

The original IST and TwIST use the single wavelet transform to represent the regularization $\phi(\boldsymbol{x})$. In this paper, we propose the adaptive contourlet-wavelet iterative shrinkage/thresholding algorithm (TcwIST) to solve optimization problems as given in Eq. (2). The proposed algorithm is more effective than the former methods in that it can achieve a better signal-to-noise ratio (SNR) improvement and can more appropriately adapt to the remote sensing image with plenty of texture.

2 Image decomposition model

The cartoon+texture decomposition model has recently received considerable interest in the field of image processing. The algorithm was first proposed by Buades *et al.* (2010), and the theory was proposed by Meyer (2001). Meyer proposed that any image $\boldsymbol{x}$ can be decomposed into the sum of a cartoon/geometric part, $\boldsymbol{u}$, which is represented as a function of bounded variation, and a textural/oscillatory part $\boldsymbol{v}$ with oscillating patterns. By using this kind of decomposition method, we can deal with the different image parts to achieve good image restoration results.

The Rudin-Osher-Fatemi (ROF) model (Rudin *et al.*, 1992) is used to solve a minimization problem of the form

$$\hat{\boldsymbol{x}}=\min_{\boldsymbol{x}\in\mathrm{BV}(\Omega)}\int_{\Omega}\|\nabla\boldsymbol{x}\|+\frac{\tau}{2}\int_{\Omega}\|\boldsymbol{y}-\boldsymbol{x}\|^2,\tau>0, \tag{4}$$

where Ω is an open and bounded subset of $\mathbb{R}^2$, and $\mathrm{BV}(\Omega)$ is a bounded variation in Ω. $\boldsymbol{y}$ and $\boldsymbol{x}$ are noisy and denoised images, respectively. $\|\nabla\boldsymbol{x}\|$ is the regularized function/regularizer, and the second term denotes data fidelity. This algorithm is widely used for image restoration and denoising for its capability to preserve piecewise smooth functions.

The variable splitting idea and split Bregman technique can be found in Bioucas-Dias and Figueiredo (2008), Goldstein and Osher (2009), and Figueiredo *et al.* (2009). This technique is a good formulation choice for l_1-norm regularized problems. The split Bregman is used to solve unconstrained convex minimization problems of the form

$$\hat{\boldsymbol{x}} = \min_{\boldsymbol{x},\boldsymbol{c}} J(\boldsymbol{c}) + E(\boldsymbol{x}) + \frac{\tau}{2} \| \boldsymbol{y} - \boldsymbol{x} - \boldsymbol{c} \|_2^2, \quad (5)$$

where $J(\boldsymbol{c})=\|\boldsymbol{c}\|_1$ and $E(\boldsymbol{x}) = \frac{1}{2} \| \boldsymbol{y} - \boldsymbol{x} \|_2^2$. The key to the split Bregman is that we will 'decouple' the l_1 and l_2 terms to form different sub-problems which can be solved more easily than the original composite objective function. By combining the split Bregman and ROF model, Gilles and Osher (2011) proposed Bregman Meyer's G-norm to separate $\boldsymbol{x}$ into $\boldsymbol{u}+\boldsymbol{v}$. The method works well in separating texture and cartoon parts and requires only a few iterations to reach convergence. The combination of ROF with split Bregman iteration is shown in Algorithm 1, where P is the project operator, λ and μ represent the cartoon and texture parameters, respectively, and $P_{\mathrm{ROF}}(a, b)$ is defined as the output of the ROF model applied to a with a coefficient b and is efficiently implemented by the split Bregman iterations.

Algorithm 1 Combination of ROF with split Bregman iteration

Initialization: $\boldsymbol{u}_0=\boldsymbol{v}_0=\boldsymbol{0}$
while 'not converged' **do**
  Update $\boldsymbol{u}$ by $\boldsymbol{u}_{n+1}=P_{\mathrm{ROF}}(\boldsymbol{x}-\boldsymbol{v}_n, \lambda)$
  Update $\boldsymbol{v}$ by $\boldsymbol{v}_{n+1}=\boldsymbol{x}-\boldsymbol{u}_{n+1}-P_{\mathrm{ROF}}(\boldsymbol{x}-\boldsymbol{u}_{n+1}, 1/\mu)$
end **while**

In this study an image is decomposed in the transform domain to achieve a good result. The key factor is to choose two appropriate dictionaries, one for the representation of texture, and the other for the natural scene parts. This expectation relies on the assumptions made earlier about dictionary $\boldsymbol{T}_{\boldsymbol{u}}$ and $\boldsymbol{T}_{\boldsymbol{v}}$ being very efficient in representing one content type and highly non-effective in representing the other. We use the curvelet dictionary $\boldsymbol{T}_{\boldsymbol{u}}$ and discrete cosine transform (DCT) dictionary $\boldsymbol{T}_{\boldsymbol{v}}$ to represent cartoon and texture parts, respectively. Thus, Algorithm 1 can be modified to Algorithm 2 using a dictionary for representation, where $\boldsymbol{\theta}_{\boldsymbol{u}}$ and $\boldsymbol{\theta}_{\boldsymbol{v}}$ are the curvelet coefficient and DCT coefficient, respectively.

Algorithm 2 Image decomposition model represented by curvelet and DCT

Initialization: $\boldsymbol{u}=\boldsymbol{v}=\boldsymbol{0}$, $\boldsymbol{\theta}_{\boldsymbol{u}}=\boldsymbol{\theta}_{\boldsymbol{v}}=\boldsymbol{0}$
while 'not converged' **do**
  Update $\boldsymbol{u}$ by $\boldsymbol{u}^{n+1} = P_{\mathrm{ROF}}(\boldsymbol{x} - \boldsymbol{T}_{\boldsymbol{v}}\boldsymbol{\theta}_{\boldsymbol{v}}^n, \lambda)$
  Update $\boldsymbol{v}$ by
  $\boldsymbol{v}^{n+1} = \boldsymbol{x} - \boldsymbol{T}_{\boldsymbol{u}}\boldsymbol{\theta}_{\boldsymbol{u}}^n - P_{\mathrm{ROF}}(\boldsymbol{x} - \boldsymbol{T}_{\boldsymbol{u}}\boldsymbol{\theta}_{\boldsymbol{u}}^n, 1/\mu)$
end while

Algorithm 2 is used to numerically solve the cartoon+texture decomposition model based on Meyer's G-norm.

3 Gradient-based shrinkage/thresholding algorithm

3.1 Optimization problem

The formulation (2) is usually called the analysis unconstrained approach, which applies a regularizer directly to the unknown image. The other form can be represented as

$$f(\boldsymbol{\theta}) = \frac{1}{2} \| \boldsymbol{y} - \boldsymbol{R}\boldsymbol{W}^{-1}\boldsymbol{\theta} \|_2^2 + \tau\phi(\boldsymbol{\theta}). \quad (6)$$

We refer to this technique as the synthesis approach. The formulation applies a regularization function to the image representation coefficients ($\boldsymbol{\theta}=\boldsymbol{W}\boldsymbol{x}$). The operator $\boldsymbol{W}$ is a wavelet-like frame (Afonso *et al.*, 2010), such as 'wavelet', 'contourlet', 'Ridgelet', and 'Beamlet'; $\boldsymbol{W}$ is a forward transform and $\boldsymbol{W}^{-1}$ is an inverse transform. Thus, the analysis and synthesis constrained versions can be written, respectively, as

$$\min_{\boldsymbol{x}} \phi(\boldsymbol{x}) \quad \text{subject to } \| \boldsymbol{y} - \boldsymbol{R}\boldsymbol{x} \|_2^2 \le \varepsilon, \quad (7)$$

and

$$\min_{\boldsymbol{\theta}} \phi(\boldsymbol{\theta}) \quad \text{subject to } \| \boldsymbol{y} - \boldsymbol{R}\boldsymbol{W}^{-1}\boldsymbol{\theta} \|_2^2 \le \varepsilon. \quad (8)$$

The parameter ε usually has a clear meaning (it is proportional to the noise variance) and is more natural to determine than the parameter τ, but the constrained version may be more difficult to solve than the unconstrained version. In this study we focus on solving the unconstrained problem (6) in the contourlet and wavelet domain.

3.2 Existing gradient-based algorithms

To be consistent with the former algorithms, we replace image coefficient $\boldsymbol{\theta}$ with $\boldsymbol{x}$ and denote the operator as $\boldsymbol{A}=\boldsymbol{R}\boldsymbol{W}^{-1}$. Thus, objective function (6) can be rewritten as

$$f(\boldsymbol{x})=\frac{1}{2}\|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}\|_2^2+\tau\phi(\boldsymbol{x}). \tag{9}$$

In the TwIST algorithm, we have a general formulation

$$\boldsymbol{x}_{k+1}=(1-\alpha)\boldsymbol{x}_{k-1}+(\alpha-\beta)\boldsymbol{x}_k+\beta\Gamma_\tau(\boldsymbol{x}_k), \tag{10}$$

and the relaxed parameters can be taken as

$$\alpha=1+\rho^2,\ \beta=2\alpha/(1+\xi).$$

The variable ξ is a small value, and ρ is determined by the ith eigenvalue of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$. The operator $\Gamma_\tau(\boldsymbol{x})$ is defined as

$$\Gamma_\tau(\boldsymbol{x})=\Upsilon_\tau(\boldsymbol{x}-\gamma\nabla f(\boldsymbol{x})),$$

where Υ_τ is a denoising operator, given by a soft shrinkage function

$$\Upsilon_\tau(\boldsymbol{x},\delta)=\begin{cases}x(i,\,j)-\delta, & x(i,\,j)\geq\delta,\\ 0, & |x(i,\,j)|<\delta,\\ x(i,\,j)+\delta, & x(i,\,j)\leq-\delta,\end{cases}$$

where $x(i, j)$ represents the (i, j)th entry of matrix $\boldsymbol{x}$. Using a shrinkage function, Eq. (10) can be written as

$$\boldsymbol{x}_{k+1}=(1-\alpha)\boldsymbol{x}_{k-1}+(\alpha-\beta)\boldsymbol{x}_k+\beta\Upsilon_\tau(\boldsymbol{x}_k-\gamma_k\nabla f(\boldsymbol{x}_k),\,\gamma_k\delta). \tag{11}$$

We also provide a compromise shrinkage function

$$\Upsilon_\tau(\boldsymbol{x},\delta,\delta_0)=\begin{cases}(1-m)x(i,\,j)+m\,\mathrm{sgn}(x(i,\,j))(|x(i,\,j)|-\delta), & |x(i,\,j)|\geq\delta,\\ 0, & |x(i,\,j)|<\delta_0,\\ n\,\mathrm{sgn}(x(i,\,j))(|x(i,\,j)|-(\delta-\delta_0)), & \delta_0\leq|x(i,\,j)|\leq\delta.\end{cases}$$

Herein parameters m and n are used to determine the shrinkage rule: m=0 denotes a hard shrinkage function and m=1 denotes a soft shrinkage function; n deals with the contourlet and wavelet coefficients to reserve the significant components. The compromise shrinkage function (CSF) is a compromise of hard and soft shrinkage. We can adjust or shrink the coefficients by changing these parameters. The CSF can give a better SNR than hard shrinkage, especially for some m less than zero. While the coefficients are between δ and δ_0, important image information can be retained by adjusting the parameter n. We have verified the difference in shrinkage operators (Section 5.3), but the results are not shown in this paper. Fig. 1 shows a plot of the shrinkage function.

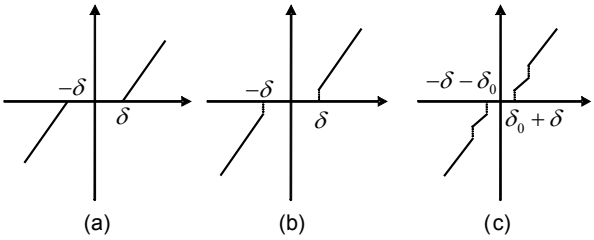


Fig. 1 Shrinkage function

(a) Hard shrinkage function; (b) Soft shrinkage function; (c) Compromise shrinkage function

4 Contourlet-wavelet iterative shrinage/thresholding algorithm

4.1 Description of the proposed algorithm

In this study the proposed TcwIST algorithm is used to solve Eq. (9). We use Algorithm 2 to decompose the degraded image into cartoon and texture parts. The regularization $\phi(\boldsymbol{x})$ is represented by wavelet and contourlet. Thus, Eq. (9) can be written as

$$f(\boldsymbol{x})=\frac{1}{2}\|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}\|_2^2+\tau_{\mathrm{c}}\|\boldsymbol{\Psi}_{\mathrm{c}}(\boldsymbol{x})\|_p+\tau_{\mathrm{w}}\|\boldsymbol{\Psi}_{\mathrm{w}}(\boldsymbol{x})\|_p, \tag{12}$$

where $\boldsymbol{\Psi}_{\mathrm{c}}$ and $\boldsymbol{\Psi}_{\mathrm{w}}$ are two different transform matrices, which represent the contourlet and wavelet frame/basis, respectively, and $\|\cdot\|_p$ represents the corresponding l_1-norm or TV-norm. We assume that these matrices correspond to orthogonal bases or the Parseval frame. Thus, we can obtain

$$\boldsymbol{\Psi}_{\mathrm{c}}\in\mathbb{R}^{N_1\times N},\ \boldsymbol{\Psi}_{\mathrm{w}}\in\mathbb{R}^{N_2\times N}\ \text{with}\ \boldsymbol{\Psi}_{\mathrm{c}}^{\mathrm{T}}\boldsymbol{\Psi}_{\mathrm{c}}=\boldsymbol{\Psi}_{\mathrm{w}}^{\mathrm{T}}\boldsymbol{\Psi}_{\mathrm{w}}=\boldsymbol{I}_N, \tag{13}$$

where $N_1>N$, $N_2>N$, $\boldsymbol{\Psi}$ is the forward-transform and its transpose is $\boldsymbol{\Psi}^{\mathrm{T}}$, and $\boldsymbol{I}_N$ is an identity matrix with size N. The parameter pairs τ_{c}, τ_{w} of objective function (12) are determined by noise standard deviation σ to some extent. By using decomposition on observation, we can solve objective function (12) as the optimization problem

$$\begin{aligned}\left\langle\hat{\boldsymbol{x}}_{\boldsymbol{u}},\hat{\boldsymbol{x}}_{\boldsymbol{v}}\right\rangle=\arg\min_{\boldsymbol{x}_{\boldsymbol{u}},\boldsymbol{x}_{\boldsymbol{v}}}\{f(\boldsymbol{x}_{\boldsymbol{u}},\boldsymbol{x}_{\boldsymbol{v}})\equiv\frac{1}{2}\|\boldsymbol{y}-\boldsymbol{A}_{\boldsymbol{u}}\boldsymbol{x}_{\boldsymbol{u}}-\boldsymbol{A}_{\boldsymbol{v}}\boldsymbol{x}_{\boldsymbol{v}}\|_2^2\\+\tau_{\mathrm{c}}\|\boldsymbol{x}_{\boldsymbol{u}}\|_p+\tau_{\mathrm{w}}\|\boldsymbol{x}_{\boldsymbol{v}}\|_p\},\end{aligned} \tag{14}$$

where $\boldsymbol{x}_{\boldsymbol{u}}$ represents the wavelet coefficients, and $\boldsymbol{x}_{\boldsymbol{v}}$ represents the contourlet coefficients. We obtain the following formula (see Section 3.1 for details):

$$\boldsymbol{A}\boldsymbol{x}=\boldsymbol{R}\boldsymbol{W}^{-1}\boldsymbol{x}=\boldsymbol{R}(\boldsymbol{W}_{\mathrm{w}}^{-1}\boldsymbol{x}_{\boldsymbol{u}}+\boldsymbol{W}_{\mathrm{c}}^{-1}\boldsymbol{x}_{\boldsymbol{v}})=\boldsymbol{A}_{\boldsymbol{u}}\boldsymbol{x}_{\boldsymbol{u}}+\boldsymbol{A}_{\boldsymbol{v}}\boldsymbol{x}_{\boldsymbol{v}}, \tag{15}$$

where $\boldsymbol{W}_{\mathrm{w}}^{-1}\boldsymbol{x}_{\boldsymbol{u}}$ and $\boldsymbol{W}_{\mathrm{c}}^{-1}\boldsymbol{x}_{\boldsymbol{v}}$ represent the cartoon and texture parts, respectively.

In this study, the wavelet transform is a discrete wavelet transform (DWT) and the contourlet transform is a pyramidal directional filter bank (that is, discrete contourlet). We use two different shrinkage thresholds for the smooth and texture parts. In the texture part, we provide a larger threshold because the smooth part contains more image details and the texture part contains more image noise.

4.2 Parameter selection

In this section, we discuss the parameter property. The main parameters are the shrinkage threshold, regularization parameter, and step length.

First, we discuss the shrinkage threshold. Shrinkage threshold tuning can generally be selected to match the noise standard deviation for a good denoising result. The value of the initial threshold is proportional to noise standard deviation σ; that is, higher noise has a larger initial threshold. We can empirically estimate the threshold using an exponential formulation $\delta=0.0779\times\sigma^{1.4972}$. In the compromise shrinkage function, we have $\delta_0=n\delta$. Although the quality of the recovery image depends on the choice of δ, the rate of algorithm convergence does not depend on the choice of the shrinkage threshold because the shrinkage function mainly decreases the noise of the degraded image without accelerating the algorithm.

Second, the regularization parameter is usually hand tuned to achieve SNR improvement and good convergence speed. In this paper we use an adaptive and decreasing regularization parameter, which was mentioned by Pan and Blu (2011; 2013). The adaptive regularization parameter is given by

$$\tau_{\mathrm{c}},\ \tau_{\mathrm{w}}\propto E_{\mathrm{n}}/K, \tag{16}$$

where E_{n} and K are the noise energy and blurred level, respectively. The advantage of TcwIST is that we can estimate the parameter without knowing the image prior information, which is usually difficult to obtain. The regularization parameter is chosen based on noise energy and blurred level. When an observed image is contaminated by high noise, we have a large parameter to restrain noise. When the image was deserved by a high blur operator, we have a small parameter to focus on image deconvolution. The experimental results have verified that the adaptive regularization parameter can be a better choice for SNR improvement than a fixed parameter.

Finally, the effect of the iterative step length γ_k is twofold. A larger γ_k makes the algorithm more efficient, but the result will be worse if the algorithm exceeds the optimal solution. The algorithm converges slowly when using a small γ_k. A tradeoff exists between accuracy and efficiency in choosing γ_k. In the TcwIST algorithm, we use a linear search method (Algorithm 3) to identify the step length.

In the linear search method, Eq. (9) can be simplified as a quadratic function

$$f(\boldsymbol{x})=\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x}-\boldsymbol{b}^{\mathrm{T}}\boldsymbol{x}, \tag{17}$$

where $\boldsymbol{Q}$ ($\boldsymbol{Q}$ is consistent to $\boldsymbol{A}$) is symmetric and positive definite, and $\boldsymbol{b}=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y}$. The constant terms of $\boldsymbol{y}$ and the regularization are neglected because their differentials are equal to zero. We can obtain the step length by differentiating

$$f(\boldsymbol{x}_k-\gamma_k\nabla f_k)=\frac{1}{2}(\boldsymbol{x}_k-\gamma_k\nabla f_k)^{\mathrm{T}}\boldsymbol{Q}(\boldsymbol{x}_k-\gamma_k\nabla f_k)-\boldsymbol{b}^{\mathrm{T}}(\boldsymbol{x}_k-\gamma_k\nabla f_k)$$

with respect to γ_k, such that we obtain

$$\gamma_k=\frac{\nabla f_k^{\mathrm{T}}\nabla f_k}{\nabla f_k^{\mathrm{T}}\boldsymbol{Q}\nabla f_k}. \tag{18}$$

Algorithm 3 Backtracking line search

Choose $\gamma>0$, $\eta>1$, $\kappa\in(0, 1)$ and constants $\gamma_{\min}$, $\gamma_{\max}$ ($0<\gamma_{\min}<\gamma_{\max}$)
Repeat
  Choose $\gamma_k\in[\gamma_{\min}, \gamma_{\max}]$
  Repeat
    $f(\boldsymbol{x}_k+\gamma_k p_k)\leq f(\boldsymbol{x}_k)+\kappa\gamma_k\nabla f^{\mathrm{T}}(\boldsymbol{x}_k)p_k$
    with $p_k=-\nabla f(\boldsymbol{x}_k)$
    $\gamma_k\leftarrow\eta\gamma_k$
  Until x_{k+1} satisfies an acceptance criterion
  $k\leftarrow k+1$
Until the stopping criterion is satisfied

4.3 Convergence of TcwIST and stopping criterion

In FISTA (Beck and Teboulle, 2009b), the information from the two previous iterations is combined in a special manner that is significantly faster than standard gradient-based methods, such as IST algorithms. This scheme has a global convergence rate $O(1/k^2)$, where k is the iteration counter. The fast convergence idea is also used in this paper. For the purpose of our analysis, we consider formulation (1). In the IST, the solution can be represented as

$$\boldsymbol{x}_{k+1}=\boldsymbol{x}_k-\gamma_k\nabla F(\boldsymbol{x}_k). \tag{19}$$

This term can be expanded to obtain a new formulation

$$\boldsymbol{x}_{k+1}=F(\boldsymbol{x}_k)+\left\langle(\boldsymbol{x}-\boldsymbol{x}_k),\nabla F(\boldsymbol{x}_k)\right\rangle+\frac{1}{2\gamma_k}\|\boldsymbol{x}-\boldsymbol{x}_k\|_2^2, \tag{20}$$

where the right-hand side of the equation contains constant, first-order, and two-order terms. By extending this gradient updating mechanism to problem (9) and neglecting the constant term, Eq. (20) can be rewritten as

$$\boldsymbol{x}_{k+1}=\frac{1}{2\gamma_k}\left\|\boldsymbol{x}-\left(\boldsymbol{x}_k-\gamma_k\nabla F(\boldsymbol{x}_k)\right)\right\|_2^2+G(\boldsymbol{x}). \tag{21}$$

Thus, we can obtain an inequality

$$F(\boldsymbol{x})+G(\boldsymbol{x})\leq F(\boldsymbol{x}_k)+\left\langle(\boldsymbol{x}-\boldsymbol{x}_k),\nabla F(\boldsymbol{x}_k)\right\rangle+\frac{L}{2}\|\boldsymbol{x}-\boldsymbol{x}_k\|_2^2+G(\boldsymbol{x}), \tag{22}$$

where L ($L=1/\gamma_k$) is the Lipschitz constant. The solution can be easily proven to satisfy an inequality

$$F(\boldsymbol{x}_k)-F(\boldsymbol{x}^*)\leq\frac{L\|\boldsymbol{x}_0-\boldsymbol{x}^*\|_2^2}{2k}, \tag{23}$$

where $\boldsymbol{x}^*$ is a global optimal solution. $L\|\boldsymbol{x}_0-\boldsymbol{x}^*\|_2^2$ is a constant, such that the convergence of IST is determined by k. In FISTA, the solution is represented as

$$F(\boldsymbol{x}_k)-F(\boldsymbol{x}^*)\leq\frac{2L\|\boldsymbol{x}_0-\boldsymbol{x}^*\|_2^2}{(k+1)^2}. \tag{24}$$

Compared with the convergence of IST ($O(1/k)$), the convergence of FISTA is $O(1/k^2)$. More details can be found in Nesterov (1983) and Beck and Teboulle (2009b).

In this paper we use the relative error of the objective function as a stopping criterion. This error is also used in IST and TwIST.

$$\mathrm{error}_{\mathrm{obj}}=\frac{|f(\boldsymbol{x}_{k+1})-f(\boldsymbol{x}_k)|}{f(\boldsymbol{x}_k)}. \tag{25}$$

A good option is for a user to tune the stopping criterion based on application requirements. Further details on the stopping criteria for formulation (9) can be found in Wright *et al.* (2009) (Section II-H).

5 Numerical experiments

In this section, we present results from numerical experiments to compare our method with state-of-the-art methods. We perform experiments on three cover types of remote sensing images (other images show the same trends): airport (512×512), vegetated surface (512×512), and water body (512×512). To compare other methods, a 256×256 Cameraman image was used in our experiments. In TcwIST, we solved the optimization problem (14) and the choice of different parameters was discussed in Section 4.2. The image restoration process can be described as follows: (1) Use Algorithm 2 to decompose the observed image and obtain the cartoon and texture parts; (2) Use the TcwIST algorithm to solve Eq. (14), and the regularization parameters and shrinkage threshold are computed by the adaptive method and experiential fomula; and (3) A backtracking method is adopted in the iteration to find the step length.

5.1 Cartoon-texture decomposition

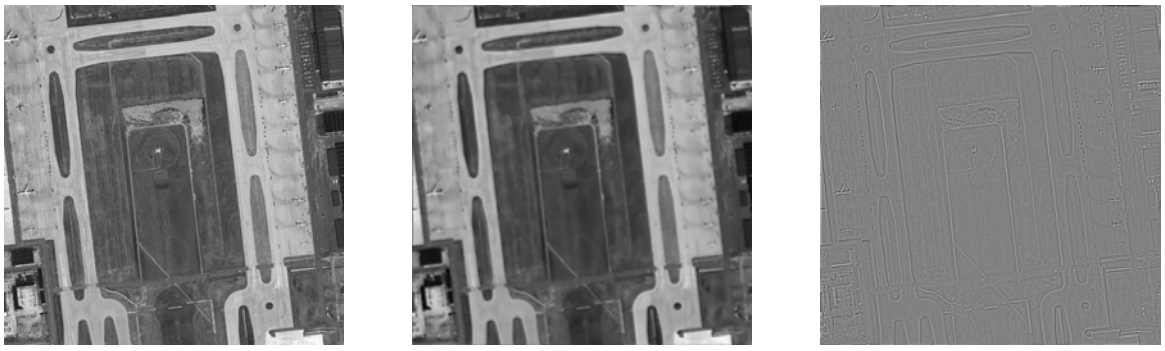
In the first test, we apply Algorithm 2 to show the cartoon-texture decomposition results for the image depicted in Fig. 2. Figs. 2b and 2c show the corresponding cartoon and texture parts obtained by the algorithm, respectively. In our experiment we set λ=1000, μ=1000.

5.2 Image deconvolution with l_1-norm

We use the TcwIST algorithm to deal with image restoration in Experiments 2, 3, and 4. Figs. 3–5 show the image restoration results for the vegetated surface, Beijing airport, and water body, respectively.

5.3 Image deconvolution with TV-norm

Fig. 6 shows the image deconvolution results with TV-norm in Experiment 3. TV-norm can achieve better restoration performance than l_1-norm.

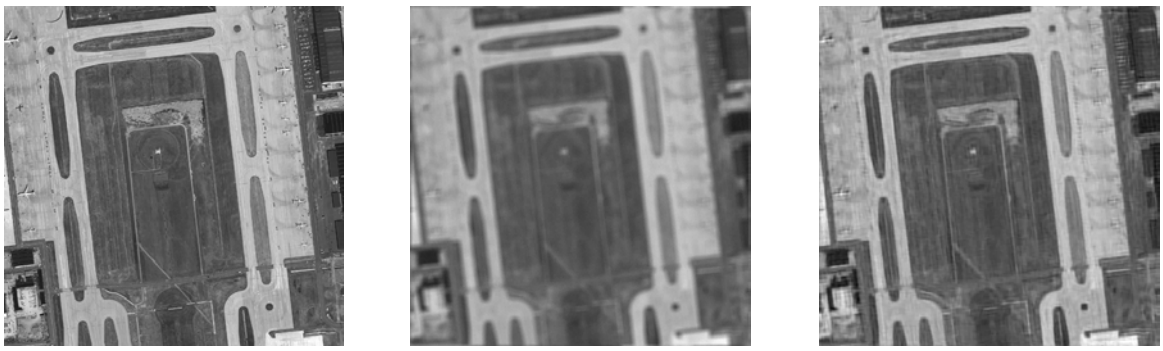


(a) (b) (c)

Fig. 2 Image decomposition with Algorithm 2
(a) Original airport image; (b) Cartoon part; (c) Texture part

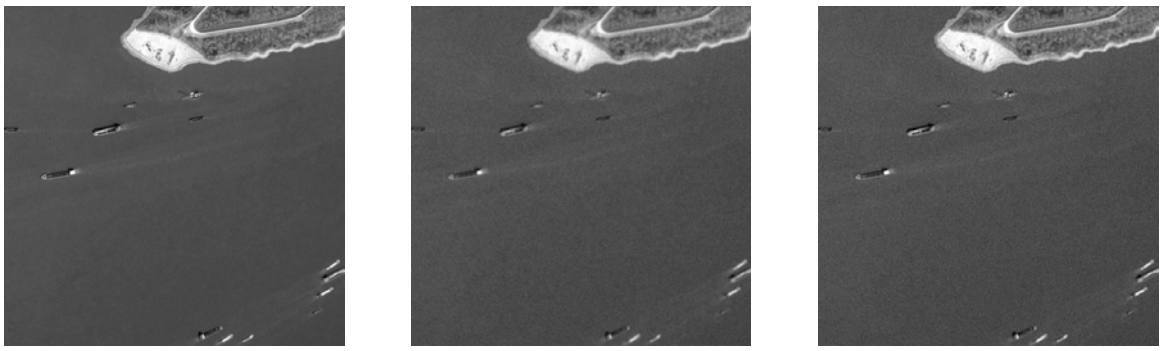
(a) (b) (c)

Fig. 3 Image deconvolution in Experiment 2
(a) Original vegetated surface image; (b) Blurred image; (c) Restoration image



(a) (b) (c)

Fig. 4 Image deconvolution in Experiment 3
(a) Original Beijing airport image; (b) Blurred image; (c) Restoration image



(a) (b) (c)

Fig. 5 Image deconvolution in Experiment 4
(a) Original water body image; (b) Blurred image; (c) Restoration image

(a) (b) (c)

Fig. 6 Image deconvolution with TV-norm
(a) Original Cameraman image; (b) Blurred image; (c) Restoration image

5.4 Experiment setting and analysis

Table 1 shows the settings used in the four experiments: Gaussian blur with low noise, blur operator $1/(1+i^2+j^2)$ with medium noise, uniform blur with medium noise, and blur operator $[1, 4, 6, 4, 1]^{\mathrm{T}}[1, 4, 6, 4, 1]/256$ with high noise. In this section we present the experimental results for image restoration using both l_1-norm and TV-norm regularization. For comparison, we provide two criteria: (1) SNR improvement (ISNR=10lg($\|\boldsymbol{y}-\boldsymbol{x}\|/\|\boldsymbol{x}_k-\boldsymbol{x}\|$), where $\boldsymbol{x}$ is the original image) of restoration, and (2) algorithm convergence speed.

First, we compare the methods based on ISNR. The ISNR along the iterations is shown in Fig. 7. To provide a comprehensive analysis, we show the image restoration results in Table 2. A distinct improvement is observed when the proposed method is used compared to the TwIST algorithm, especially in the medium- and high-noise situations.

Table 1 Experimental settings

Experiment	Blur operator	σ
1	H_1: Gaussian	0.56
2	H_2: $1/(1+i^2+j^2)$	$\sqrt{2}$
3	H_3: uniform	$\sqrt{8}$
4	H_4: $[1, 4, 6, 4, 1]^T[1, 4, 6, 4, 1]/256$	7

Second, we discuss the convergence speed. The objective function $f(\boldsymbol{x}_u, \boldsymbol{x}_v)$ along the iterations is shown in Fig. 8. In Table 3, we compare the convergence speed of our method and those of the state-of-the-art algorithms: IST, TwIST, and sparse reconstruction by separable approximation (SpaRSA).

The objective function value c^{1000}_{IST} at convergence is obtained by running the IST algorithm with 1000 iterations, and the other methods are required to reach the same objective function value of $1.001\times c^{1000}_{\mathrm{IST}}$ for comparison. In the test we give four different blur operators with different noise levels. Our method is faster than the other methods in most cases, especially in the low- and medium-noise cases. However, each iteration of TcwIST is more expensive

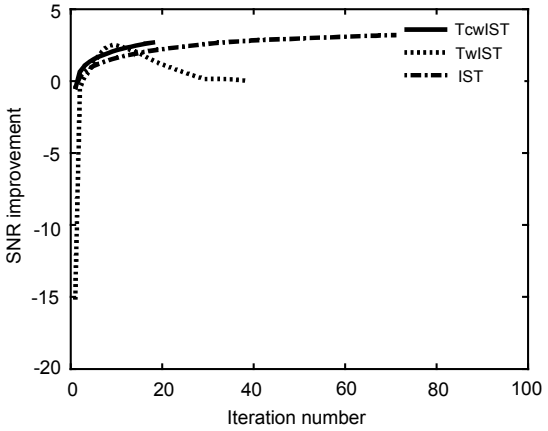


Fig. 7 SNR improvement produced by TcwIST, TwIST, and IST for the first 100 iterations of image deconvolution of the Cameraman image in Experiment 3

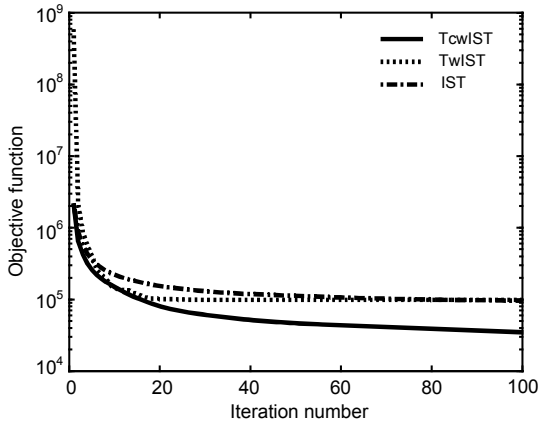


Fig. 8 Convergence speed of TcwIST, TwIST, and IST for the first 100 iterations on the Cameraman image in Experiment 2

Table 2 Experimental results for the comparison of different methods

Image	Method	ISNR				Number of iterations			
		Exp 1	Exp 2	Exp 3	Exp 4	Exp 1	Exp 2	Exp 3	Exp 4
Vegetated surface	IST	12.0117	5.8891	2.5235	−1.9004	30	127	90	72
	FISTA	12.6477	5.4166	2.4206	−3.8022	10	97	77	69
	TwIST	11.6069	0.0631	2.2701	−16.3308	19	51	24	70
	TcwIST	12.6298	6.4389	2.8450	1.9142	7	29	21	19
Beijing airport	IST	12.2667	4.3813	3.2175	−2.7629	14	110	83	70
	FISTA	12.2489	4.0793	2.9580	−4.6344	10	87	71	68
	TwIST	12.0773	−1.1992	−4.8401	−17.2974	10	39	47	73
	TcwIST	12.2865	5.1783	4.1338	2.2261	8	15	42	10
Water body	IST	5.7042	1.7555	1.5322	−5.7278	13	86	58	68
	FISTA	5.6216	0.3155	0.2627	−7.9396	10	77	57	67
	TwIST	5.5557	−7.1274	−12.6817	−21.1395	9	44	53	80
	TcwIST	5.6931	4.0123	3.1182	3.6309	7	24	16	7

because of the contourlet transform, the total runtime of which is 0.2709 s (FISTA, 0.0519 s). Thus, two iterations of the transform consume 0.2166 s because we have to transform a cell matrix into a normal one for computation. Therefore, it is necessary to find a fast and convenient method to solve this problem.

All experiments are conducted using MATLAB 2011b for Windows XP system on a desktop computer equipped with an Intel Pentium Dual-Core 3.2 GHz CPU and 2 GB RAM.

Table 3 Number of iterations required to reach convergence for 256×256 Cameraman image under different blurred SNR (from 10 dB to 40 dB)

Blur operator	Method	Number of iterations						
		40	35	30	25	20	15	10
H_1	IST	29	21	22	25	28	32	35
	TwIST	18	14	14	15	15	17	18
	SpaRSA	13	10	10	11	11	11	12
	TcwIST	**7**	**7**	**7**	**7**	**7**	**7**	**7**
H_2	IST	865	916	976	992	997	998	1001
	TwIST	59	72	67	48	**50**	**46**	**44**
	SpaRSA	108	90	110	14	88	115	112
	TcwIST	**15**	**16**	**22**	**72**	178	278	370
H_3	IST	950	946	954	975	987	992	994
	TwIST	64	62	56	48	**44**	**41**	**35**
	SpaRSA	91	90	99	120	78	97	55
	TcwIST	**17**	**17**	**19**	**35**	125	252	345
H_4	IST	911	906	913	946	968	980	984
	TwIST	67	65	72	67	61	**57**	**53**
	SpaRSA	101	127	115	118	98	132	118
	TcwIST	**9**	**10**	**11**	**16**	**59**	159	267

6 Conclusions

In this paper, we propose a new class of iterative shrinkage threshold algorithm called TcwIST for remote sensing image restoration. The remote sensing image contains different cover types of terrain, and we can obtain good image restoration results by using an image decomposition model. An adaptive method and an experiential method are used to estimate the regularization parameter and shrinkage threshold to obtain a better ISNR than that using the original parameters, especially in severe linear inverse problems and high-noise situations. We also use a fast algorithm and a linear search method in the TcwIST algorithm to achieve fast convergence.

References

Afonso, M.V., Bioucas-Dias, J.M., Figueiredo, M.A.T., 2010. Fast image recovery using variable splitting and constrained optimization. *IEEE Trans. Image Process.*, **19**(9):2345-2356. [doi:10.1109/TIP.2010.2047910]

Beck, A., Teboulle, M., 2009a. Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. *IEEE Trans. Image Process.*, **18**(11): 2419-2434. [doi:10.1109/TIP.2009.2028250]

Beck, A., Teboulle, M., 2009b. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imag. Sci.*, **2**(1):183-202. [doi:10.1137/080716542]

Bioucas-Dias, J.M., 2006. Bayesian wavelet-based image deconvolution: a GEM algorithm exploiting a class of heavy-tailed priors. *IEEE Trans. Image Process.*, **15**(4): 937-951. [doi:10.1109/TIP.2005.863972]

Bioucas-Dias, J.M., Figueiredo, M.A.T., 2007a. A new TwIST: two-step iterative shrinkage/thresholding algorithms for image restoration. *IEEE Trans. Image Process.*, **16**(12): 2992-3004. [doi:10.1109/TIP.2007.909319]

Bioucas-Dias, J.M., Figueiredo, M.A.T., 2007b. Two-step algorithms for linear inverse problems with non-quadratic regularization. Proc. IEEE Int. Conf. on Image Processing, p.I-105-I-108. [doi:10.1109/ICIP.2007.4378902]

Bioucas-Dias, J.M., Figueiredo, M.A.T., 2008. An iterative algorithm for linear inverse problems with compound regularizers. Proc. 15th IEEE Int. Conf. on Image Processing, p.685-688. [doi:10.1109/ICIP.2008.4711847]

Bioucas-Dias, J.M., Figueiredo, M.A.T., Oliveira, J.P., 2006. Total variation-based image deconvolution: a majorization-minimization approach. Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing, p.II. [doi:10.1109/ICASSP.2006.1660479]

Buades, A., Le, T.M., Morel, J.M., *et al*., 2010. Fast cartoon+texture image filters. *IEEE Trans. Image Process.*, **19**(8): 1978-1986. [doi:10.1109/TIP.2010.2046605]

Combettes, P.L., Wajs, V.R., 2005. Signal recovery by proximal forward-backward splitting. *Multiscale Model. Simul.*, **4**(4):1168-1200. [doi:10.1137/050626090]

Daubechies, I., Defrise, M., De Mol, C., 2004. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Commun. Pure Appl. Math.*, **57**(11): 1413-1457. [doi:10.1002/cpa.20042]

Figueiredo, M.A.T., Nowak, R.D., 2003. An EM algorithm for wavelet-based image restoration. *IEEE Trans. Image Process.*, **12**(8):906-916. [doi:10.1109/TIP.2003.814255]

Figueiredo, M.A.T., Bioucas-Dias, J.M., Nowak, R.D., 2007. Majorization-minimization algorithms for wavelet-based image restoration. *IEEE Trans. Image Process.*, **16**(12): 2980-2991. [doi:10.1109/TIP.2007.909318]

Figueiredo, M.A.T., Bioucas-Dias, J.M., Afonso, M.V., 2009. Fast frame-based image deconvolution using variable splitting and constrained optimization. Proc. IEEE/SP 15th Workshop on Statistical Signal Processing, p.109-112. [doi:10.1109/SSP.2009.5278628]

Gilles, J., Osher, S., 2011. Bregman Implementation of Meyer's *G*-Norm for Cartoon+Textures Decomposition. UCLA CAM Report.

Goldstein, T., Osher, S., 2009. The split Bregman method for L1-regularized problems. *SIAM J. Imag. Sci.*, **2**(2):323-343. [doi:10.1137/080725891]

Hunter, D.R., Lange, K., 2004. A tutorial on MM algorithms. *Am. Stat.*, **58**(1):30-37. [doi:10.1198/0003130042836]

Meyer, Y., 2001. Oscillating Patterns in Image Processing and Nonlinear Evolution Equations: the Fifteenth Dean Jacqueline B. Lewis Memorial Lectures. American Mathematical Society Boston, MA, USA.

Nesterov, Y., 1983. A method of solving a convex programming problem with convergence rate $O(1/k^2)$. *Sov. Math. Doklady*, **27**(2):372-376.

Nowak, R.D., Figueiredo, M.A.T., 2001. Fast wavelet-based image deconvolution using the EM algorithm. Proc. 35th Asilomar Conf. on Signals, Systems and Computers, p.371-375. [doi:10.1109/ACSSC.2001.986953]

Pan, H.J., Blu, T., 2011. Sparse image restoration using iterated linear expansion of thresholds. Proc. 18th IEEE Int. Conf. on Image Processing, p.1905-1908. [doi:10.1109/ICIP.2011.6115842]

Pan, H.J., Blu, T., 2013. An iterative linear expansion of thresholds for l_1-based image restoration. *IEEE Trans. Image Process.*, **22**(9):3715-3728. [doi:10.1109/TIP.2013.2270109]

Rudin, L.I., Osher, S., Fatemi, E., 1992. Nonlinear total variation based noise removal algorithms. *Phys. D*, **60**(1-4):259-268. [doi:10.1016/0167-2789(92)90242-F]

Wright, S.J., Nowak, R.D., Figueiredo, M.A.T., 2009. Sparse reconstruction by separable approximation. *IEEE Trans. Signal Process.*, **57**(7):2479-2493. [doi:10.1109/TSP.2009.2016892]

2013 JCR of Thomson Reuters for *JZUS (A/B/C)*

ISI Web of Knowledge℠

Journal Citation Reports®

WELCOME HELP RETURN TO LIST

2013 JCR Science Edition

Journal: Journal of Zhejiang University-SCIENCE A

Mark	Journal Title	ISSN	Total Cites	Impact Factor	5-Year Impact Factor	Immediacy Index	Citable Items	Cited Half-life	Citing Half-life
☐	J ZHEJIANG UNIV-SC A	1673-565X	744	0.608	0.607	0.047	85	5.2	8.0

Journal: Journal of Zhejiang University-SCIENCE B

Mark	Journal Title	ISSN	Total Cites	Impact Factor	5-Year Impact Factor	Immediacy Index	Citable Items	Cited Half-life	Citing Half-life
☐	J ZHEJIANG UNIV-SC B	1673-1581	1393	1.293	1.424	0.147	129	5.0	8.0

Journal: Journal of Zhejiang University-SCIENCE C-Computers & Electronics

Mark	Journal Title	ISSN	Total Cites	Impact Factor	5-Year Impact Factor	Immediacy Index	Citable Items	Cited Half-life	Citing Half-life
☐	J ZHEJIANG U-SCI C	1869-1951	132	0.380	0.399	0.157	89	2.5	6.6