Neurocomputing 129 (2014) 362–377

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Framelet based pan-sharpening via a variational method

Faming Fang [a], Guixu Zhang [a,c,*], Fang Li [b], Chaomin Shen [a]

[a] Department of Computer Science, East China Normal University, Shanghai, China
[b] Department of Mathematics, East China Normal University, Shanghai, China
[c] Key Laboratory of Geographic Information Science, Ministry of Education, East China Normal University

ARTICLE INFO

Article history:
Received 26 January 2013
Received in revised form
25 July 2013
Accepted 10 September 2013
Communicated by Qi Li
Available online 1 November 2013

Keywords:
Pan-sharpening
Framelet
Sparsity
Variational method
Split Bregman iteration

ABSTRACT

Pan-sharpening is a process of combining a low resolution multi-spectral (MS) image and a high resolution panchromatic (PAN) image to obtain a single high resolution MS image. In this paper, we propose two pan-sharpening methods based on the framelet framework. The first method, as a basic work, is called a framelet-based pan-sharpening (FP) method. In the FP method, we first decompose the MS and PAN images into framelet coefficients, then obtain a new set of coefficients by choosing the approximation coefficients in MS and detail coefficients in PAN, and finally construct the pan-sharpened image from the new set of coefficients. To overcome the inflexibility of FP, in the second method, by combining FP and other three fusion requirements, i.e., geometry keeping, spectral preserving and the sparsity of the image in the framelet domain, four assumptions are established. Based on these assumptions, a framelet based variational energy functional, whose minimizer is related to the final pan-sharpened result, is then formulated. To improve the numerical efficiency, the split Bregman iteration is further introduced, and the result of FP method is set as an initial value. We refer this method as the variational framelet pan-sharpening (VFP) method. To verify the effectiveness of our methods, we present the results of the two methods on the QuickBird and IKONOS images, compare them with five existing methods qualitatively and quantitatively, analyze the influence of parameters of VFP, and extend the VFP to hyperspectral data as well as comparison study. The experimental results demonstrate the superiority of our methods.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

With the increasing amount of images acquired from the earth observing satellites, the demand of high quality representation is also increasing. However, due to technical reasons, many sensors do not provide imagery with both high spatial and spectral resolutions. Instead, dual images, a multi-spectral image (MS) and a panchromatic image (PAN), are often captured. For example, as a high resolution satellite, QuickBird provides a single band PAN image (spatial resolutions: 0.7 m), and a four-band MS image: red (R), green (G), blue (B) and near-infrared (NIR) (spatial resolutions: 2.8 m). The MS image contains high spectral quality with low spatial resolution, whereas the PAN image collects fine spatial details with less spectral information [1,2]. In order to make better use of these satellite data, there is always a demand to obtain a high resolution MS image by using an image processing technique [3]. This technique, generally called fusion, integration or pan-sharpening, has become an important direction, and it is very useful for many remote sensing problems such as target recognition, change detection, spectrum matching, vegetation mapping and hazard monitoring [4–6].

More specifically, pan-sharpening is a process of integrating a low resolution MS with a PAN image to construct a higher resolution MS result. For effective pan-sharpening, the following conditions must be established [7,8]: (1) the time interval between acquiring MS and PAN images should be short; and (2) the MS and PAN images should be registered. The result of pan-sharpening should be more suitable for various types of applications. Thus, we should try to construct a pan-sharpened image which contains both high spectral and spatial qualities [2].

In the past few decades, a variety of pan-sharpening methods have been developed. Some of them, such as the intensity-hue-saturation (IHS) [9–11], Brovey [12], principal-component-analysis [9,12] and Gram–Schmidt (GS) [13,14], can be classified as component substitution (CS) methods [15]. The CS methods have access to high spatial quality but suffer noticeable spectral distortion. Another family of pan-sharpening methods, e.g., wavelet [16,17] and Laplacian pyramids [18,19], is the multi-resolution analysis (MRA) methods, in which each image is decomposed into a set of bandpass channels in the spatial frequency domain [15]. MRA methods can overcome the spectral distortion problem of the CS methods. However, in the case of high-pass detail injection, their results sustain some spatial degradation, such as ringing or aliasing effects [20].

* Corresponding author. Tel./fax: +86 21 54345108.
E-mail addresses: fmfang@cs.ecnu.edu.cn (F. Fang), gxzhang@cs.ecnu.edu.cn (G. Zhang), fli@math.ecnu.edu.cn (F. Li), cmshen@cs.ecnu.edu.cn (C. Shen).

0925-2312/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2013.09.022

Recently, the variational methods are widely used in image processing. A variational method generally presents an energy functional for a certain problem based on some properties, and tries to search local or global minima for this energy, where the minima correspond to the desired result. Compared with other schemes, a variational method has remarkable advantages in both theory and numerical implementation [21]. For a detailed description of variational methods, one may refer to [21–23]. In 2006, based on the main assumption that the topographic map of the PAN image should contain the geometry of the spectral channels, Ballester et al. presented the first variational pan-sharpening method named P+XS [24]. The P+XS can produce a promising result, but it faces some spatial blurring. Following the idea of combining the advantages of wavelet and P+XS algorithms, Moeller et al. proposed a variational wavelet pan-sharpening method (VWP) [25] in 2009. This method, to some extent, can overcome the spatial degradation. However, its time complexity is high due to the space conversion between the wavelet and the spatial domain. To address this problem, an alternate VWP (AVWP) was then deduced to reduce the running time at the expense of the image quality [25].

In all the above methods, the wavelet-based fusion methods are prominent due to the well balancing of preserving the high spectral and spatial information [26]. The main idea of these methods is as follows. First, each image is decomposed into wavelet coefficients; second, with a proper pan-sharpening rule, a new set of coefficients is composed; finally the pan-sharpened result is obtained from the new set of coefficients by an inverse wavelet transform [27,28]. In terms of the pan-sharpening rule, many schemes have been proposed. Among them, a remarkable idea is introduced by Zhou et al. [16], in which the detail coefficients of the PAN image and the approximation coefficients of each MS band are taken to compose the new coefficients. The choice of a wavelet transform also has a great influence on the fused image. As mentioned in [29], the orthogonal wavelet transform may suffer some aliasing effects. To avoid this drawback, the stationary wavelet transform, which leads to many redundant information because it does not subsample after each decomposition, is raised. This transform obviously outperforms the orthogonal wavelet transform [25,30].

As mentioned in many literatures, such as [31,32], the framelet transform is more precise than wavelet in extracting the spectral and spatial information. Thus, in this paper, we will introduce the framelet transform into pan-sharpening work, and will develop two new methods for pan-sharpening. The first method, which is similar to the aforementioned wavelet fusion [16], is named the framelet-based pan-sharpening method (FP), in which each image is first decomposed into framelet coefficients; then a new set of coefficients is composed by a certain rule; finally the pan-sharpening result is acquired from the new set of coefficients using an inverse framelet transform. To overcome the inflexibility of the FP method, by combining FP with other three reasonable assumptions, a variational pan-sharpening method in the framelet domain is then performed. We refer this method as the variational framelet pan-sharpening method (VFP). Furthermore, we compare our methods with many state-of-the-art schemes qualitatively and quantitatively using the QuickBird and IKONOS data, analyze the influence of parameters of VFP, and extend the VFP to hyperspectral data as well as comparison study.

The rest of the paper is organized as follows. In Section 2, we introduce the framelet theory and image representation. In Section 3, the FP method based on framelet system is developed. In Section 4, we detail the VFP method and present its numerical scheme using the split Bregman iterator. In Section 5, we perform our pan-sharpened results as well as comparison study. Finally the conclusion is given in Section 6.

2. Framelets and image representation

In this section, we give a brief introduction to the framelet system. For simplicity, we only show the univariate framelets, and the framelets in the bivariate setting can be obtained by a tensor product of univariate framelets. Tight frames in a finite dimensional space, which is derived from framelets and their matrix forms, are also performed. One who is interested in the framelets can refer to [33–35] for more details.

2.1. Framelets in $L^2(\mathbb{R})$

First, we give the definition of the tight frame. A countable function subset $X \subset L^2(\mathbb{R})$ is called a tight frame of $L^2(\mathbb{R})$, if

$$f = \sum_{h \in X} \langle f, h\rangle h, \quad \forall f \in L^2(\mathbb{R}).$$

This is equivalent to

$$\|f\|^2 = \sum_{h \in X} |\langle f, h\rangle|^2, \quad \forall f \in L^2(\mathbb{R}),$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product of $L^2(\mathbb{R})$, and $\| \cdot \| = \langle \cdot, \cdot \rangle^{1/2}$ is the norm of $L^2(\mathbb{R})$.

As a generalization of the orthogonal basis, the tight frame relaxes the requirements of the orthogonality and linear independence to bring in redundancy. This redundancy has been verified to be useful in many applications such as deblurring [35].

Given a finite set $\Phi := \{\phi^1, \ldots, \phi^r\} \subset L^2(\mathbb{R})$, a wavelet system $X(\Phi)$ is defined as the collection of dilations and shifts of Φ, i.e.,

$$X(\Phi) := \{2^{k/2}\phi^j(2^k x - l) : 1 \le j \le r; k, l \in \mathbb{Z}\}.$$

When $X(\Phi)$ forms a tight frame, it is named as a wavelet tight frame, and ϕ^j is named as a (tight) framelet.

In order to construct compactly supported wavelet tight frames $X(\Phi)$, one generally first obtains a compactly supported refinable function $\psi \in L^2(\mathbb{R})$ with a refinement mask (low-pass filter) g_0 such that

$$\psi(x) = \sum_l g_0(l)\psi(2x - l).$$

Then a wavelet tight frame can be constructed by finding an appropriate set of framelets $\Phi := \{\phi^1, \ldots, \phi^r\} \subset L^2(\mathbb{R})$, which is defined as

$$\phi^j = \sum_l g_j(l)\psi(2x - l), \quad j = 1, \ldots, r,$$

where g_j is a high-pass filter. Thus, the construction of framelets Φ essentially is to design the filters $g_0, g_1, \ldots, g_r$. The unitary extension principle (UEP) in [36] asserts that the system $X(\Phi)$ forms a tight frame in $L^2(\mathbb{R})$ if the filters $g_0, g_1, \ldots, g_r$ satisfy

$$\zeta_{g_0}(\omega)\overline{\zeta_{g_0}(\omega + \gamma\pi)} + \sum_{j=1}^{r} \zeta_{g_j}(\omega)\overline{\zeta_{g_j}(\omega + \gamma\pi)} = \delta(\gamma), \quad \gamma = 0, 1,$$

for almost all $\omega \in \mathbb{R}$. Here $\zeta_g(\omega) = \sum_l g(l)e^{il\omega}$ and $\delta(\gamma)$ is a delta function.

As a conventional application of the UEP, a piecewise linear B-spline is used as the refinable function ψ. The corresponding filters are

$$g_0 = \left[\frac{1}{4}, \frac{1}{2}, \frac{1}{4}\right], \quad g_1 = \left[-\frac{1}{4}, \frac{1}{2}, -\frac{1}{4}\right], \quad g_2 = \left[\frac{\sqrt{2}}{4}, 0, -\frac{\sqrt{2}}{4}\right].$$

2.2. Frames in $\mathbb{R}^N$

Since images are finite dimensional, we are concentrated to obtain the framelet coefficients of a finite dimension vector in a certain frame. This task can be realized by matrix operation

directly. We will describe how to convert a vector of $\mathbb{R}^N$ to coefficients in a certain frame.

Let $\mathcal{A}$ be a matrix with size of $K \times N (K \geq N)$. We denote the system, which consists of all the rows of $\mathcal{A}$, as $\mathcal{A}$ again. Then, $\mathcal{A}$ is a tight frame in $\mathbb{R}^N$ if

$$\|x\|_2^2 = \sum_{y \in \mathcal{A}} |\langle x, y\rangle|^2, \quad \forall x \in \mathbb{R}^N,$$

which is equivalent to the perfect reconstruction formula:

$$x = \sum_{y \in \mathcal{A}} \langle x, y\rangle y,$$

where $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$ are inner product and norm in $\mathbb{R}^N$ respectively. The matrix $\mathcal{A}$ is named as the decomposition (analysis) operator and $\mathcal{A}^T$ is named as the reconstruction (synthesis) operator. Then, the perfect decomposition and reconstruction process can be formulated as

$$x = \mathcal{A}^T \mathcal{A} x.$$

Thus, $\mathcal{A}$ is a tight frame in $\mathbb{R}^N$ if and only if $\mathcal{A}^T\mathcal{A} = \mathcal{I}$. Here in general $\mathcal{A}\mathcal{A}^T \neq \mathcal{I}$, unless in the orthogonal case.

Next, we construct $\mathcal{A}$ from the filters related to any certain framelet system. For a given filter $g = \{g(j)\}_{j=-J}^{J}$, under the Neumann boundary condition, let $\mathcal{S}(g)$ be the convolution operator with filter g, then

$$\mathcal{S}(g) = \mathcal{T}(g) + \mathcal{H}(g),$$

where $\mathcal{T}(g)$ and $\mathcal{H}(g)$ are Toeplitz and Hankel matrices respectively, which are defined as [36,37]

$$\mathcal{T}(g) = \begin{bmatrix} g(0) & \cdots & g(-J) & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ g(J) & \ddots & \ddots & \ddots & g(-J) \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & g(J) & \cdots & g(0) \end{bmatrix},$$

$$\mathcal{H}(g) = \begin{bmatrix} g(1) & g(2) & \cdots & g(J) & 0 \\ g(2) & \ddots & \ddots & \ddots & g(-J) \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ g(J) & \ddots & \ddots & \ddots & g(-2) \\ 0 & g(-J) & \cdots & g(-2) & g(-1) \end{bmatrix}.$$

Many applications tend to use a multi-level tight frame system relating to the tight framelet decomposition without downsampling. To introduce it, recalling the aforementioned filter $g = \{g(j)\}_{j=-J}^{J}$, we define the filter g^l at level l corresponding to the decomposition without downsampling as

$$g^l = \Big\{ g(-J), \underbrace{0, \ldots, 0}_{2^{l-1}-1}, g(-J+1), 0, \ldots, 0, g(-1), \underbrace{0, \ldots, 0}_{2^{l-1}-1}, g(0), \underbrace{0, \ldots, 0}_{2^{l-1}-1}, g(1), 0, \ldots, 0, h(J-1), \underbrace{0, \ldots, 0}_{2^{l-1}-1}, h(J) \Big\}.$$

For the given filters $\{g_j\}_{j=0}^{r}$, by denoting $\mathcal{Z}_j^l = \mathcal{S}(g_j^l)$, then the decomposition matrix $\mathcal{A}$ with L level is given by

$$\mathcal{A} = \begin{bmatrix} \prod_{l=0}^{L-1} \mathcal{Z}_0^{L-l} \\ \mathcal{Z}_1^L \prod_{l=1}^{L-1} \mathcal{Z}_0^{L-l} \\ \vdots \\ \mathcal{Z}_r^L \prod_{l=1}^{L-1} \mathcal{Z}_0^{L-l} \\ \vdots \\ \mathcal{Z}_1^1 \\ \vdots \\ \mathcal{Z}_r^1 \end{bmatrix} := \begin{bmatrix} \mathcal{A}_0 \\ \mathcal{A}_1 \end{bmatrix}$$

where $\mathcal{A}_0 = \prod_{l=0}^{L-1} \mathcal{Z}_0^{L-l}$. Thanks to UEP, the following equation holds

$$\mathcal{A}^T\mathcal{A} = \mathcal{A}_0^T\mathcal{A}_0 + \mathcal{A}_1^T\mathcal{A}_1 = \mathcal{I}.$$

With matrix $\mathcal{A}$, the frame transformation process can be easily described. Let $\boldsymbol{f}$ be a vector, the frame coefficient vector $\boldsymbol{u}$ is given by

$$\boldsymbol{u} = \mathcal{A}\boldsymbol{f}.$$

Since $\mathcal{A}_0$ denotes the low pass filter, $\mathcal{A}_0\boldsymbol{f}$ are the approximation coefficients at a given approximation level L. $\mathcal{A}_1$ denotes the band-pass and high pass filters, thus $\mathcal{A}_1\boldsymbol{f}$ are the detail coefficients. Besides, the frame reconstruction process can be expressed as,

$$\boldsymbol{f} = \mathcal{A}^T\boldsymbol{u}.$$

Finally, it should be noted, in this paper, that we use the same decomposition and reconstruction algorithms as [35]. And we work in the bivariate case for our pan-sharpening task. This corresponding transform matrix, still denoted as $\mathcal{A}$, can be easily obtained by using the Kronecker product of the matrix corresponding to the univariate frame transform.

As an example, Fig. 1 shows a result of one level framelet decomposition, in which the original image presented in (a) is acquired from QuickBird satellite [38]. Fig. 1(b) performs the framelet coefficients. Particularly, the approximation coefficients are shown in

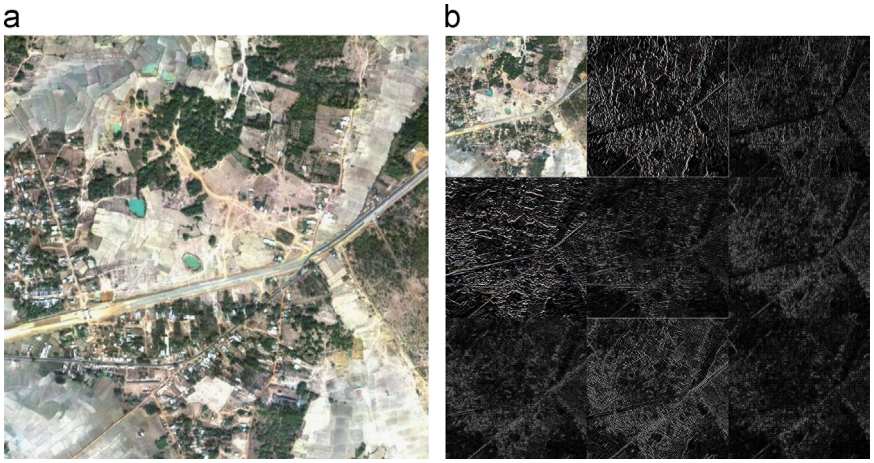


Fig. 1. An example of frame decomposition (with one level): (a) the input image (RGB bands, 512×512, acquired from QuickBird satellite); and (b) the result of decomposition.

the top left corner of Fig. 1(b), and the detail coefficients are shown in the other part of Fig. 1(b). Note that the image size in Fig. 1(b) is nine times as that in Fig. 1(a).

3. Framelet-based pan-sharpening

In this section, we will introduce our pan-sharpening algorithm based on the framelet system [31,36]. We refer this method as FP.

First of all, we clarify some useful mathematical formulations and notations for our task. Let $\boldsymbol{p}$ denote the original PAN image; $\boldsymbol{m}=(\boldsymbol{m}_1,\ldots,\boldsymbol{m}_N)$ denote the low resolution MS image which has been upsampled to be the same size as $\boldsymbol{p}$ by nearest neighbor interpolation, where N is the number of bands of MS image; $\boldsymbol{f}=(\boldsymbol{f}_1,\ldots,\boldsymbol{f}_N)$ denote the desired high resolution MS image; and $\mathcal{A}=[\mathcal{A}_0;\mathcal{A}_1]$ denote the transform matrix with a prescribed level L which corresponding to the framelet system, where $\mathcal{A}_0$ represents the low-pass filter and $\mathcal{A}_1$ represents band-pass and high-pass filters (see Section 2 for details).

As mentioned in Section 2, a functional (in $L^2(\mathbb{R}^2)$) can be decomposed into a set of framelet coefficients, and also it can be reconstructed from the coefficients using the inverse transform. This conclusion is also satisfied in a discrete form (for example, an image).

Due to the nature of the framelet transform, while decomposing an image into framelet coefficients, its spectral and spatial information can be sufficiently reflected in the approximation coefficients (generated by $\mathcal{A}_0$) and the detail coefficients (generated by $\mathcal{A}_1$), respectively. Fig. 1 is a good explanation of this nature. On the other hand, in the pan-sharpening problem, the MS image $\boldsymbol{m}$ includes most of the spectral information, while the PAN image $\boldsymbol{p}$ contains the majority of the spatial information. Therefore, we can use the approximation coefficients of $\boldsymbol{m}_n$ and detail coefficients of $\boldsymbol{p}$ as the approximation and detail coefficients of the $\boldsymbol{u}_n$, respectively. Here, $\boldsymbol{u}_n$ represents the framelet coefficients of the pan-sharpened image $\boldsymbol{f}_n$.

Mathematically, given the framelet coefficients of $\boldsymbol{p},\boldsymbol{m}_n$ which are defined by

$$\mathcal{A}\boldsymbol{p}=\begin{bmatrix}\mathcal{A}_0\boldsymbol{p}\\ \mathcal{A}_1\boldsymbol{p}\end{bmatrix},\quad \mathcal{A}\boldsymbol{m}_n=\begin{bmatrix}\mathcal{A}_0\boldsymbol{m}_n\\ \mathcal{A}_1\boldsymbol{m}_n\end{bmatrix}.$$

$\boldsymbol{u}_n$ can be formalized as

$$\boldsymbol{u}_n=\begin{bmatrix}\mathcal{A}_0\boldsymbol{m}_n\\ \mathcal{A}_1\boldsymbol{p}\end{bmatrix}. \tag{1}$$

Then the pan-sharpened image $\boldsymbol{f}_n$ can be directly acquired by $\boldsymbol{f}_n=\mathcal{A}^T\boldsymbol{u}_n$.

On the whole, the overall procedure of our FP method can be divided into three steps, and the all steps are performed in Algorithm 1. In order to facilitate understanding, as a supplement, Fig. 2 shows a visual statement for the entire process of the FP method. Note that we set the decomposition level $L=2$ for simplicity.

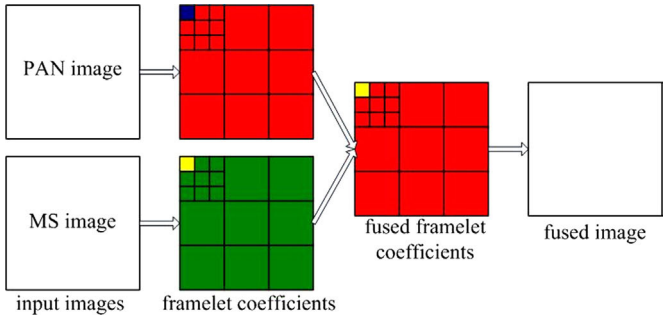


Fig. 2. Process of the FP method (note that we use the decomposition algorithm without downsampling, and set the decomposition level $L=2$ for simplicity).

With respect to the decomposition level L, in practice, we find it is promising for balancing the computational complexity and quality when $L=2$. Thus we use 2 level framelet decomposition in all our implementations.

Algorithm 1. The main process of the FP method.

Input: the low-resolution MS image $\boldsymbol{m}$ and the PAN image $\boldsymbol{p}$

step 1: decomposition:

$$\mathcal{A}\boldsymbol{p}=\begin{bmatrix}\mathcal{A}_0\boldsymbol{p}\\ \mathcal{A}_1\boldsymbol{p}\end{bmatrix},\quad \mathcal{A}\boldsymbol{m}_n=\begin{bmatrix}\mathcal{A}_0\boldsymbol{m}_n\\ \mathcal{A}_1\boldsymbol{m}_n\end{bmatrix},$$

step 2: coefficient choosing:

$$\boldsymbol{u}_n=\begin{bmatrix}\mathcal{A}_0\boldsymbol{m}_n\\ \mathcal{A}_1\boldsymbol{p}\end{bmatrix}.$$

step 3: construction:

$$\boldsymbol{f}_n=\mathcal{A}^T\boldsymbol{u}_n.$$

Output: the pan-sharpened image $\boldsymbol{f}$.

4. Variational framelet pan-sharpening

The result of the FP method is promising due to its balancing on preserving the spectral and spatial information from the original images. However, its result is unique because there is no adjustable parameter. Since some tasks, such as segmentation and feature extraction, tend to require more high spatial information; while others, e.g. target recognition and spectrum matching, are apt to desire more high spectral information; the unique result may not meet various practical applications. As introduced in [21–23], the variational method can deal with this drawback due to its flexibility. To hold the advantage and avoid the disadvantage of the FP method, in what follows, by combining FP (could be regarded as a term) and other three assumptions, we will develop a new variational pan-sharpening method based on the framelet system, of which the result is adjustable according to the different requirements. We refer this method as VFP.

Particularly, our proposed energy functional consists of four terms, where each term is built based on a certain assumption. In what follows, we will introduce these terms in detail. Note that we adopt the same notations as Section 3. Besides, we set $\Omega\subset\mathbb{R}^2$ as our image domain, where Ω is an open, bounded domain with the Lipschitz boundary.

4.1. Geometry information preserving term

As has been stated, the PAN image contains the vast majority of the spatial information compared to the MS bands. Therefore, we can assume that a linear combination of the desired pan-sharpened bands should be approximate to the PAN image in the spatial information.

The expression form of the spatial information is a foundational issue. A universal and an effective idea is to express it using the gradient field [39]. That is, given the PAN image $\boldsymbol{p}$ and the desired pan-sharpened image $\boldsymbol{f}$, their spatial information can be expressed by the gradient fields $\nabla\boldsymbol{p}$ and $\nabla\boldsymbol{f}_n$ respectively. Then the above assumption can be formalized as

$$\nabla\left(\sum_{n=1}^{N}\alpha_n\boldsymbol{f}_n\right)=\nabla\boldsymbol{p}, \tag{2}$$

where α_n is a mixing coefficient.

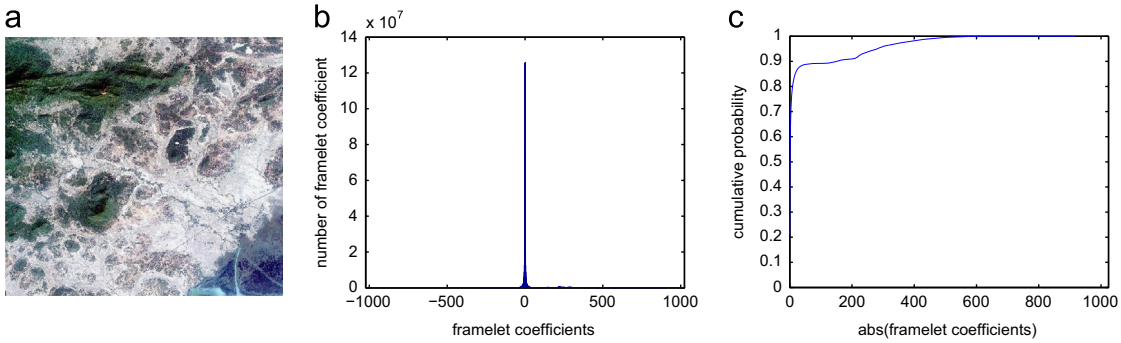


Fig. 3. Statistics of the framelet coefficients. (a) The thumbnail of an example image (acquired from the Chilka Lake area in India at February 23, 2005, the original image is at size of $4096 \times 4096 \times 4$); (b) histogram of the framelet coefficients; and (c) cumulative probability distribution corresponding to (b).

To help impose our assumption into a variational framework, we present the following energy functional:

$$E_G(\boldsymbol{f}) = \left\| \nabla\left(\sum_{n=1}^{N} \alpha_n \boldsymbol{f}_n \right) - \nabla \boldsymbol{p} \right\|_2^2 = \left\| \sum_{n=1}^{N} \alpha_n \nabla \boldsymbol{f}_n - \nabla \boldsymbol{p} \right\|_2^2. \tag{3}$$

Choosing α_n is a challenging problem. Many related methods set $\alpha_n = 1/N$ or use the experimentally determined coefficients [10]. Since $\alpha_n = 1/N$ is certainly inaccurate, and empirical value is difficult to acquire, inspired by [10], we obtain α_n by another method.

As introduced in [40], the contrast inversion phenomenon is a common situation occurring in a remote sensing image. To avoid this phenomenon, we can add the following constraint when choosing α_n:

$$\alpha_n \nabla \boldsymbol{f}_n \le \nabla \boldsymbol{p}, \quad n = 1, \ldots, N.$$

Then, α_n can be obtained by minimizing the following energy:

$$\min_{\alpha_n} H(\alpha_n) = \left\| \sum_{n=1}^{N} \alpha_n \nabla \boldsymbol{f}_n - \nabla \boldsymbol{p} \right\|_2^2 + \| \alpha_n \nabla \boldsymbol{f}_n - \nabla \boldsymbol{p} \|_2^2, \tag{4}$$

where the first term is used to keep the linear relationship and the second term is used to avoid the contrast inversion phenomenon.

The first variation of (4) is

$$\frac{\delta E}{\delta \alpha_n} = \nabla \boldsymbol{f}_n \left(\sum_{i=1}^{N} \alpha_i \nabla \boldsymbol{f}_i - \nabla \boldsymbol{p} \right) + \nabla \boldsymbol{f}_n (\alpha_n \nabla \boldsymbol{f}_n - \nabla \boldsymbol{p}).$$

Thus we can obtain α_n by the gradient descent method:

$$\alpha_n^{j+1} = \alpha_n^j - dt * \frac{\delta E}{\delta \alpha_n},$$

where dt is a time variable. We let $\alpha_n = \alpha_n^{j+1}$ when $(\alpha_n^{j+1} - \alpha_n^j)/\alpha_n^{j+1} < 10^{-6}$ arrived.

4.2. Spectral information preserving term

Observing on the upsampled low resolution MS image $\boldsymbol{m}$, one may find that it appears blurry and ambiguous. While we aim to obtain a high spectral quality result $\boldsymbol{f}$, which is clear and distinct and contain the same spectral information as $\boldsymbol{m}$, it is safe to treat $\boldsymbol{m}$ as a degradation of $\boldsymbol{f}$ in each band. That is, $\boldsymbol{m}_n$ can be acquired by $\boldsymbol{f}_n$ with a low-pass filter (kernel) $\boldsymbol{k}_n$. Mathematically, we have

$$\boldsymbol{m}_n = \boldsymbol{k}_n * \boldsymbol{f}_n, \quad n = 1, \ldots, N, \tag{5}$$

where '$*$' denotes the convolution operation.

Note that a similar assumption is also introduced in [24], in which (5) is also adopted with a prescribed kernel $\boldsymbol{k}_n$. Compared to [24], we assume that the kernels differ from image to image, i.e., $\boldsymbol{k}_n$ is an unknown kernel.

While both $\boldsymbol{f}_n$ and $\boldsymbol{k}_n$ are unknown, (5) is an under-constraint problem. To effectively solve this problem, we need to add some constraints to reduce the uncertainty. As mentioned in many literatures such as [35], a kernel can be regarded as a special type of image which is sparse in the framelet domain. We use this sparsity property as a regularization term to constrain (5). Then, to impose this regularization term and (5) into a variational framework, we present the following energy functional:

$$E_S(\boldsymbol{f}_n, \boldsymbol{k}_n) = \eta \| \boldsymbol{k}_n * \boldsymbol{f}_n - \boldsymbol{m}_n \|_2^2 + \| \mathcal{A} \boldsymbol{k}_n \|_1, \quad n = 1, \ldots, N \tag{6}$$

where η is a balancing parameter. Here $\| \mathcal{A} \boldsymbol{k}_n \|_1$, namely the analysis sparsity regularization term [41], is used to control framelet coefficients from too large. That is, $\| \mathcal{A} \boldsymbol{k}_n \|_1$ can decrease the number of large discontinuities pixels in $\boldsymbol{k}_n$.

4.3. Sparsity prior

As introduced in [35,42], the images usually have sparse representations or approximations in the framelet domain. To examine whether the theory is applicable to the remote sensing image, we first acquire an example remote sensing image from QuickBird satellite (size: 4096×4096, band: 4, data type: unit16) whose thumbnail is shown in Fig. 3(a), then decompose it into framelet coefficients and calculate the number of coefficients in each value. The statistical result is performed in Fig. 3(b) and (c), where Fig. 3(b) is the histogram value over all $4096 \times 4096 \times 4 \times 9$ framelet coefficients, and Fig. 3(c) is the corresponding cumulative probability distribution. Observing on the figure, we find that most of the coefficients have zero values (see (b)), about 80% of the coefficients is less than 10, and almost 90% of the coefficients is below 50 (see (c)). That is, most of the framelet coefficients are equal to or very close to 0. Thus the sparse representations in the framelet domain is also adopted for remote sensing images.

We then formulate the sparse prior into a variational framework

$$E_P(\boldsymbol{f}_n) = \| \mathcal{A} \boldsymbol{f}_n \|_1, \quad n = 1, \ldots, N. \tag{7}$$

Here, minimizing E_P is to seek a solution with sparse framelet coefficients. That is, the result tends to be a smooth image. Thus, min E_P is robust to the noise.

4.4. Framelet match term

For more effective pan-sharpening, we add the FP method (see Section 3) as a match term to our variational formula. To formalize this idea, we impose (1) into the following energy functional:

$$E_F(\boldsymbol{f}_n) = \left\| \boldsymbol{u}_n - \begin{bmatrix} \mathcal{A}_0 \boldsymbol{m}_n \\ \mathcal{A}_1 \boldsymbol{p} \end{bmatrix} \right\|_2^2$$

$$= \left\| \mathcal{A}\boldsymbol{f}_n - \begin{bmatrix} \mathcal{A}_0\boldsymbol{m}_n \\ \mathcal{A}_1\boldsymbol{p} \end{bmatrix} \right\|_2^2, \quad n = 1, \ldots, N. \tag{8}$$

Since many applications may in favor of the image with high spectral quality, whereas others tend to have the image with more spatial information, in order to better meet the practical application, we introduce two parameters c_0 and c_1 to control the emphasis of the two aspects. In other words, we have the following modified formula:

$$E_F'(\boldsymbol{f}_n) = c_0\|\mathcal{A}_0\boldsymbol{f}_n - \mathcal{A}_0\boldsymbol{m}_n\|_2^2 + c_1\|\mathcal{A}_1\boldsymbol{f}_n - \mathcal{A}_1\boldsymbol{p}\|_2^2, \quad n = 1, \ldots, N, \tag{9}$$

The choice of c_0 and c_1 depends on the practical application: for high spectral quality we can set a larger value for c_0, vice versa we increase c_1 for higher spatial quality.

4.5. The proposed total energy

Taking all the above terms (3), (6), (7) and (9) into account, we can obtain the total energy functional for our pan-sharpening task

$$\begin{aligned} E(\boldsymbol{f}, \boldsymbol{k}) &= \sum_{n=1}^{N} E_P(\boldsymbol{f}_n) + \frac{\lambda}{2} E_G(\boldsymbol{f}) \\ &\quad + \frac{1}{2}\sum_{n=1}^{N} E_S(\boldsymbol{f}_n, \boldsymbol{k}_n) + \frac{1}{2}\sum_{n=1}^{N} E_F'(\boldsymbol{f}_n) \\ &= \sum_{n=1}^{N} \|\mathcal{A}\boldsymbol{f}_n\|_1 + \frac{\lambda}{2}\left\| \sum_{n=1}^{N} \alpha_n \nabla \boldsymbol{f}_n - \nabla \boldsymbol{p} \right\|_2^2 \\ &\quad + \frac{1}{2}\sum_{n=1}^{N} (\eta\|\boldsymbol{k}_n * \boldsymbol{f}_n - \boldsymbol{m}_n\|_2^2 + \|\mathcal{A}\boldsymbol{k}_n\|_1) \\ &\quad + \frac{1}{2}\sum_{n=1}^{N} (c_0\|\mathcal{A}_0\boldsymbol{f}_n - \mathcal{A}_0\boldsymbol{m}_n\|_2^2 + c_1\|\mathcal{A}_1\boldsymbol{f}_n - \mathcal{A}_1\boldsymbol{p}\|_2^2). \end{aligned} \tag{10}$$

where λ is the nonnegative parameter with the larger value indicating the more importance of its corresponding term.

Solving our pan-sharpening problem (10) is equivalent to minimizing the following energy:

$$(\boldsymbol{f}_n, \boldsymbol{k}_n) = \underset{\boldsymbol{f}_n, \boldsymbol{k}_n \in \mathbb{R}^N}{\operatorname{argmin}} E(\boldsymbol{f}, \boldsymbol{k}), \quad n = 1, \ldots, N. \tag{11}$$

We note that (11), the so-called analysis-sparsity-based minimization [41], is searching for the sparsest solution from the coefficients which are decomposed from images. Its result tends to have better performance than the synthesis-sparsity-based minimization [42] which is looking for the most sparse solution in all coefficients space [35]. Besides, since $\min E_P(\boldsymbol{f}_n)$ has robustness to noise, (11) also robust to noise obviously.

Although (10) is not global convex, it is convex with respect to $\boldsymbol{f}_n$ and $\boldsymbol{k}_n$. Therefore, we can use the alternating iteration algorithm to minimize (11). Given $\boldsymbol{k}^{(0)} = (\boldsymbol{k}_1^{(0)}, \ldots, \boldsymbol{k}_N^{(0)})$ and $\boldsymbol{f}^{(0)} = (\boldsymbol{f}_1^{(0)}, \ldots, \boldsymbol{f}_N^{(0)})$ as the initial values of $\boldsymbol{k}$ and $\boldsymbol{f}$. The alternating iteration algorithm is performed in Algorithm 2.

Algorithm 2. Scheme of the alternative iteration algorithm for VFP method.

For $i = 0, 1, 2, \ldots$
  For $n = 0, 1, 2, \ldots, N$
  1. Given the kernel $\boldsymbol{k}^{(i)}$ and the pan-sharpened image $\boldsymbol{f}^{(i)}$, calculate the pan-sharpened image band $\boldsymbol{f}_n^{(i+1)}$:

$$\boldsymbol{f}_n^{(i+1)} = \underset{\boldsymbol{f}_n}{\operatorname{argmin}} \|\mathcal{A}\boldsymbol{f}_n\|_1 + R(\boldsymbol{f}_n), \tag{12}$$

  where

$$R(\boldsymbol{f}_n) = \tfrac{\lambda}{2}\|\alpha_n \nabla \boldsymbol{f}_n + \sum_{j=1, j\neq n}^{N} \alpha_j \nabla \boldsymbol{f}_j^{(i)} - \nabla \boldsymbol{p}\|_2^2 + \tfrac{\eta}{2}\|\boldsymbol{k}_n^{(i)} * \boldsymbol{f}_n - \boldsymbol{m}_n\|_2^2 + \frac{1}{2}(c_0\|\mathcal{A}_0\boldsymbol{f}_n - \mathcal{A}_0\boldsymbol{m}_n\|_2^2 + c_1\|\mathcal{A}_1\boldsymbol{f}_n - \mathcal{A}_1\boldsymbol{p}\|_2^2).$$

  2. Given the pan-sharpened image $\boldsymbol{f}^{(i+1)}$, calculate the kernel image $\boldsymbol{k}_n^{(i+1)}$:

$$\boldsymbol{k}_n^{(i+1)} = \underset{\boldsymbol{k}_n}{\operatorname{argmin}} \left(\tfrac{1}{2}\|\boldsymbol{k}_n * \boldsymbol{f}_n^{(i+1)} - \boldsymbol{m}_n\|_2^2 + \|\mathcal{A}\boldsymbol{k}_n\|_1 \right). \tag{13}$$

  end
end

Theoretically, when the initial value is not proper, Algorithm 2 may converge to a local minimum, rather than the global one. However, as verified in practice, it can converge to a high-quality result consistently.

4.6. Numerical scheme

In this subsection, we will detail the numerical algorithm for VFP method schemed in Algorithm 2. Generally, both steps in Algorithm 2 can be solved using the gradient descent flow. However, because of the non-differentiability of the l_1-norm, one needs to add a small regular value to avoid it, yet this may lead to an inexact result. In what follows, we will use a more effective solver, named the split Bregman iteration [43], to solve the VFP method.

Based on the Bregman divergence, [43] introduced the split Bregman iteration which can solve the L^1 norm minimization problems more effectively. The split Bregman iteration extends the utility of the Bregman iteration and the linear Bregman iteration. As has been verified in many literatures such as [43], the split Bregman iteration is a promising method which can significantly reduce the time and space overhead.

In step 1 of Algorithm 2, we need to solve (12) for our purpose. The main idea of the split Bregman algorithm is that we should split the l_1 and l_2 portions. Thus (12) can be rewritten as

$$\min_{\boldsymbol{f}_n, \boldsymbol{d}_1} \|\boldsymbol{d}_1\|_1 + R(\boldsymbol{f}_n) \quad \text{s.t.} \quad \boldsymbol{d}_1 = \mathcal{A}\boldsymbol{f}_n. \tag{14}$$

To solve it, we should transform it into an unconstrained optimization problem:

$$\begin{cases} (\boldsymbol{f}_n^{(i+1,l+1)}, \boldsymbol{d}_1^{(i+1,l+1)}) = \underset{\boldsymbol{f}_n, \boldsymbol{d}_n}{\operatorname{argmin}} \|\boldsymbol{d}_1\|_1 \\ \quad + R(\boldsymbol{f}_n) + \dfrac{\beta}{2}\|\boldsymbol{d}_1 - \mathcal{A}\boldsymbol{f}_n - \boldsymbol{b}_1^{(i+1,l)}\|, \quad (15) \\ \boldsymbol{b}_1^{(i+1,l+1)} = \boldsymbol{b}_1^{(i+1,l)} + (\mathcal{A}\boldsymbol{f}_n^{(i+1,l+1)} - \boldsymbol{d}_1^{(i+1,l+1)}), \quad (16) \end{cases}$$

where β is a prescribed constant and $\boldsymbol{b}_1$ is a proper vector. We can solve (15) efficiently by alternately iterating $\boldsymbol{f}_n$ and $\boldsymbol{d}_1$ separately [44]. The two steps are as follows:

$$\begin{cases} \boldsymbol{f}_n^{(i+1,l+1)} = \underset{\boldsymbol{f}_n}{\operatorname{argmin}} R(\boldsymbol{f}_n) \\ \quad + \dfrac{\beta}{2}\|\boldsymbol{d}_1^{(i+1,l)} - \mathcal{A}\boldsymbol{f}_n - \boldsymbol{b}_1^{(i+1,l)}\|, \quad (17) \\ \boldsymbol{d}_1^{(i+1,l+1)} = \underset{\boldsymbol{d}_1}{\operatorname{argmin}} \|\boldsymbol{d}_1\|_1 \\ \quad + \dfrac{\beta}{2}\|\boldsymbol{d}_1 - \mathcal{A}\boldsymbol{f}_n^{(i+1,l+1)} - \boldsymbol{b}_1^{(i+1,l)}\|. \quad (18) \end{cases}$$

Since the subproblem (17) is differentiable, the optimality conditions of $\boldsymbol{f}_n^{(i+1,l+1)}$ are easily obtained by calculating its first variations, i.e.,

$$K\boldsymbol{f}_n^{(i+1,l+1)} = rhs(\boldsymbol{d}_1^{(i+1,l)}, \boldsymbol{b}_1^{(i+1,l)}),$$

where

$$\begin{cases} K = -\lambda\alpha_n^2\Delta + \eta[\boldsymbol{k}_n^{(i)}]^T[\boldsymbol{k}_n^{(i)}] \\ \quad + c_0\mathcal{A}_0^T\mathcal{A}_0 + c_1\mathcal{A}_1^T\mathcal{A}_1 + \beta, \\ rhs(\boldsymbol{d}_1^{(i+1,l)}, \boldsymbol{b}_1^{(i+1,l)}) \\ \quad = \lambda\alpha_n\left(\sum_{j=1, j\neq n}^{N}\alpha_j\Delta\boldsymbol{f}_j^{(i)} - \Delta\boldsymbol{p}\right) \\ \quad + \eta[\boldsymbol{k}_n^{(i)}]^T\boldsymbol{m}_n + c_0\mathcal{A}_0^T\mathcal{A}_0\boldsymbol{m}_n \\ \quad + c_1\mathcal{A}_1^T\mathcal{A}_1\boldsymbol{p}_n + \beta\mathcal{A}^T(\boldsymbol{d}_1^{(i+1,l)} - \boldsymbol{b}_1^{(i+1,l)}), \end{cases}$$

$[\cdot]$ denotes the matrix form of the convolution operator, i.e., $\boldsymbol{k}_n^{(i)} * \boldsymbol{f}_n = [\boldsymbol{k}_n^{(i)}]\cdot\boldsymbol{f}_n$, and Δ denotes the Laplacian operator. We further convert the above problem to an equivalent form:

$$\mathcal{F}(K)\mathcal{F}(\boldsymbol{f}_n^{(i+1,l+1)}) = \mathcal{F}(rhs(\boldsymbol{d}_1^{(i+1,l)}, \boldsymbol{b}_1^{(i+1,l)})),$$

where $\mathcal{F}$ denotes the Fast Fourier Transform (FFT). As solving the above equation for $\mathcal{F}(\boldsymbol{f}_n^{(i+1,l+1)})$, we take the inverse FFT, denoted as $\mathcal{F}^{-1}$, to obtain the closed form solution of $\boldsymbol{f}_n^{(i+1,l+1)}$, i.e.,

$$\boldsymbol{f}_n^{(i+1,l+1)} = \mathcal{F}^{-1}\left(\frac{\mathcal{F}(rhs(\boldsymbol{d}_1^{(i+1,l)}, \boldsymbol{b}_1^{(i+1,l)}))}{\mathcal{F}(K)}\right). \tag{19}$$

Meanwhile, the subproblem (18) can be solved by the following soft-thresholding formula directly

$$\boldsymbol{d}_1^{(i+1,l+1)} = \text{shrink}\left(\mathcal{A}\boldsymbol{f}_n^{(i+1,l+1)} + \boldsymbol{b}_1^{(i+1,l)}, \frac{1}{\beta}\right), \tag{20}$$

where

$$\text{shrink}(\phi, \beta) = \frac{\phi}{|\phi|}\cdot\max(|\phi| - \beta, 0).$$

Therefore, the iterative numerical algorithm for (12) is summarized as follows:

$$\begin{cases} \boldsymbol{f}_n^{(i+1,l+1)} = \mathcal{F}^{-1}\left(\dfrac{\mathcal{F}(rhs(\boldsymbol{d}_1^{(i+1,l)}, \boldsymbol{b}_1^{(i+1,l)}))}{\mathcal{F}(K)}\right) \\ \boldsymbol{d}_1^{(i+1,l+1)} = \text{shrink}\left(\mathcal{A}\boldsymbol{f}_n^{(i+1,l+1)} + \boldsymbol{b}_1^{(i+1,l)}, \dfrac{1}{\beta}\right) \\ \boldsymbol{b}_1^{(i+1,l+1)} = \boldsymbol{b}_1^{(i+1,l)} + (\mathcal{A}\boldsymbol{f}_n^{(i+1,l+1)} - \boldsymbol{d}_1^{(i+1,l+1)}), \end{cases} \tag{21}$$

where $\boldsymbol{f}_n^{(i+1,0)} = \boldsymbol{f}_n^{(i)}$, $\boldsymbol{d}_1^{(i+1,0)} = \boldsymbol{d}_1^{(i)}$ and $\boldsymbol{b}_1^{(i+1,0)} = \boldsymbol{b}_1^{(i)}$.

Theorem 1. *Assume that $\lambda, \eta, c_0, c_1 > 0$, we assert that (12) has at least one minimizer $\boldsymbol{f}_n^{(i+1,\star)}$, and the iteration (21) satisfies*

$$\lim_{l\to\infty}\|\mathcal{A}\boldsymbol{f}_n^{(i+1,l)}\|_1 + R(\boldsymbol{f}_n^{(i+1,l)}) = \|\mathcal{A}\boldsymbol{f}_n^{(i+1,\star)}\|_1 + R(\boldsymbol{f}_n^{(i+1,\star)}).$$

Furthermore, the following formula holds if (12) has a unique solution:

$$\lim_{l\to\infty}\|\boldsymbol{f}_n^{(i+1,l)} - \boldsymbol{f}_n^{(i+1,*)}\|_2 = 0.$$

Proof. First, since $\mathcal{A}$ is a tight frame, and $R(\boldsymbol{f}_n)$ is a convex functional, by the definition of (12), the existence of a minimizer $\boldsymbol{f}_n^{(i+1,\star)}$ can be directly obtained. Then, applying Theorem 3.2 in [41], Theorem 1 is proved. □

Similarly, subproblem (13) can also be solved using the split Bregman iteration. By making a few modifications from (21), we can obtain the iterative numerical algorithm for (13), i.e.,

$$\begin{cases} \boldsymbol{k}_n^{(i+1,l+1)} = \mathcal{F}^{-1}\left(\dfrac{\mathcal{F}([\boldsymbol{f}_n^{(i+1)}]^T\boldsymbol{m}_n + \gamma\mathcal{A}^T(\boldsymbol{d}_2^{(i+1,l)} - \boldsymbol{b}_2^{(i+1,l)}))}{\mathcal{F}([\boldsymbol{f}_n^{(i+1)}]^T[\boldsymbol{f}_n^{(i+1)}] + \gamma)}\right) \\ \boldsymbol{d}_2^{(i+1,l+1)} = \text{shrink}(\mathcal{A}\boldsymbol{k}_n^{(i+1,l+1)} + \boldsymbol{b}_2^{(i+1,l)}, \dfrac{1}{\gamma}) \\ \boldsymbol{b}_2^{(i+1,l+1)} = \boldsymbol{b}_2^{(i+1,l)} + (\mathcal{A}\boldsymbol{k}_n^{(i+1,l+1)} - \boldsymbol{d}_2^{(i+1,l+1)}), \end{cases} \tag{22}$$

where γ denotes a parameter similar to β, $\boldsymbol{k}_n^{(i+1,0)} = \boldsymbol{k}_n^{(i)}$, $\boldsymbol{d}_2^{(i+1,0)} = \boldsymbol{d}_2^{(i)}$ and $\boldsymbol{b}_2^{(i+1,0)} = \boldsymbol{b}_2^{(i)}$.

Theoretically, at every iteration i, we should choose the exact result of the two subproblems (12) and (13), i.e., $\boldsymbol{f}_n^{(i+1)} = \boldsymbol{f}_n^{(i+1,+\infty)}$ and $\boldsymbol{k}_n^{(i+1)} = \boldsymbol{k}_n^{(i+1,+\infty)}$. However, because $\boldsymbol{f}_n^{(i)}$ and $\boldsymbol{k}_n^{(i)}$ are not precise, this leads to the inaccuracy of $\boldsymbol{f}_n^{(i+1)}$ and $\boldsymbol{k}_n^{(i+1)}$. Thus, the infinite loop will be waste of time. In practice, for effectively dealing with our problem, we only need to choose $\boldsymbol{f}_n^{(i+1)} = \boldsymbol{f}_n^{(i+1,1)}$ and $\boldsymbol{k}_n^{(i+1)} = \boldsymbol{k}_n^{(i+1,1)}$. Besides, since the result of FP method (named $\boldsymbol{f}^{FP}$) is relative reasonable, to improve the convergence speed, we can initialize $\boldsymbol{f}_n^{(0)}$ as $\boldsymbol{f}_n^{FP}$.

Next, we define the stopping criteria for our algorithm. When the relative error between two iterations is less than a given parameter ε for all $n = 1, \ldots, N$, our algorithm can be considered to reach a steady state. That is, our stopping criterion is

$$\max\left(\frac{\|\boldsymbol{f}_n^{(i+1)} - \boldsymbol{f}_n^{(i)}\|}{\|\boldsymbol{f}_n^{(i+1)}\|}\right)_{n=1,\ldots,N} < \varepsilon.$$

A large number of experiments we have done show that $\varepsilon = 10^{-3}$ is reasonable.

On the whole, taking all the above analyses into account, we can obtain the overall numerical algorithm of the VFP method. The detailed description is shown in Algorithm 3.

Algorithm 3. The overall numerical scheme for VFP method.

Input: MS image $\boldsymbol{m}$, PAN image $\boldsymbol{p}$.
Initialization:
$\boldsymbol{b}_1^{(0)} = \boldsymbol{b}_2^{(0)} = 0$, $\boldsymbol{d}_1^{(0)} = \boldsymbol{d}_2^{(0)} = 0$, $\boldsymbol{k}_n^{(0)} = 0$, $\boldsymbol{f}_n^{(0)} = \boldsymbol{f}_n^{FP}$.
While $\max\left(\frac{\|\boldsymbol{f}_n^{(i+1)} - \boldsymbol{f}_n^{(i)}\|}{\|\boldsymbol{f}_n^{(i+1)}\|}\right)_{n=1,\ldots,N} \geq \varepsilon$
For $n = 1, \ldots, N$

$$\begin{cases} \boldsymbol{f}_n^{(i+1)} = \mathcal{F}^{-1}\left(\dfrac{\mathcal{F}(rhs(\boldsymbol{d}_1^{(i)}, \boldsymbol{b}_1^{(i)}))}{\mathcal{F}(K)}\right) \\ \boldsymbol{d}_1^{(i+1)} = \text{shrink}\left(\mathcal{A}\boldsymbol{f}_n^{(i+1)} + \boldsymbol{b}_1^{(i)}, \dfrac{1}{\beta}\right) \\ \boldsymbol{b}_1^{(i+1)} = \boldsymbol{b}_1^{(i)} + (\mathcal{A}\boldsymbol{f}_n^{(i+1)} - \boldsymbol{d}_1^{(i+1)}) \\ \boldsymbol{d}_2^{(i+1,0)} = \boldsymbol{d}_2^{(i)}, \boldsymbol{b}_2^{(i+1,0)} = \boldsymbol{b}_2^{(i)} \\ \boldsymbol{k}_n^{(i+1)} = \mathcal{F}^{-1}\left(\dfrac{\mathcal{F}([\boldsymbol{f}_n^{(i+1)}]^T\boldsymbol{m}_n + \gamma\mathcal{A}^T(\boldsymbol{d}_2^{(i)} - \boldsymbol{b}_2^{(i)}))}{\mathcal{F}([\boldsymbol{f}_n^{(i+1)}]^T[\boldsymbol{f}_n^{(i+1)}] + \gamma)}\right) \\ \boldsymbol{d}_2^{(i+1)} = \text{shrink}(\mathcal{A}\boldsymbol{k}_n^{(i+1)} + \boldsymbol{b}_2^{(i)}, \frac{1}{\gamma}) \\ \boldsymbol{b}_2^{(i+1)} = \boldsymbol{b}_2^{(i)} + (\mathcal{A}\boldsymbol{k}_n^{(i+1)} - \boldsymbol{d}_2^{(i+1)}). \end{cases}$$

end
end
Output: The pan-sharpened image $\boldsymbol{f}$.

5. Experimental results and comparison

In this section, to verify the effectiveness of our methods, i.e., FP and VFP, we describe and analyze them on the QuickBird and IKONOS satellites data provided by the University of Maryland (see

http://glcf.umiacs.umd.edu/data/). Since the MS image of QuickBird and IKONOS have four bands, $N=4$.

All of our experiments are executed by Matlab 7.12 and run on a windows personal computer with an Intel(R) 2.33 GHz CPU and 4-GB RAM. For an image with a size of 256×256, the execution time of FP and VFP methods is about 0.5 and 4 s respectively. Besides, the parameters of split Bregman are selected according to the analysis of [43], i.e., $\beta=10^3$ and $\gamma=10^2$.

5.1. Evaluation methods for assessing pan-sharpening

Evaluation of fusion is a challenging task. Up to now, many evaluation methods have been performed [45–47]. These methods can be roughly divided into two types, i.e., qualitative methods and quantitative methods [48]. Qualitative methods evaluate pan-sharpened results via the human visual system (HVS) directly, while quantitative methods assess the fused images by some evaluation criteria. Since the HVS is not instable and an acknowledged criterion is nonexistent, it might not be very accurate by only using a single type of metrics. For this reason, in our experiments, we take both types of metrics into account to do more reasonable evaluation.

In aspect of the quantitative methods, many criteria have been presented [45,46,49]. The methods can generally be divided into two classes: spectral and spatial quality measures. Spectral quality measures are used to evaluate the degree of spectral information preservation from MS to pan-sharpened image; while spatial quality measures are used to calculate the degree of spatial information preservation from PAN to pan-sharpened image. For a promising pan-sharpening method, the results should be performed well in both the classes.

In what follows, for simplicity, we take the eight representative criteria, in which the first four metrics measure spectral quality and the others measure spatial quality, to do our evaluation task. Note that we have normalized all the images to [0, 1].

1. *Q4 index*: Q4 index is an effective metric for pan-sharpening when the MS image has four spectral bands. It is defined as

Table 1
The choice of parameters in Fig. 4.

Parameter	Fig.4								
	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)
λ	0.5	0.01	0.9	0.5	0.5	0.5	0.5	0.5	0.5
η	0.5	0.5	0.5	0.01	0.9	0.5	0.5	0.5	0.5
c_0	0.1	0.1	0.1	0.1	0.1	0.01	0.5	0.1	0.1
c_1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.01	0.5

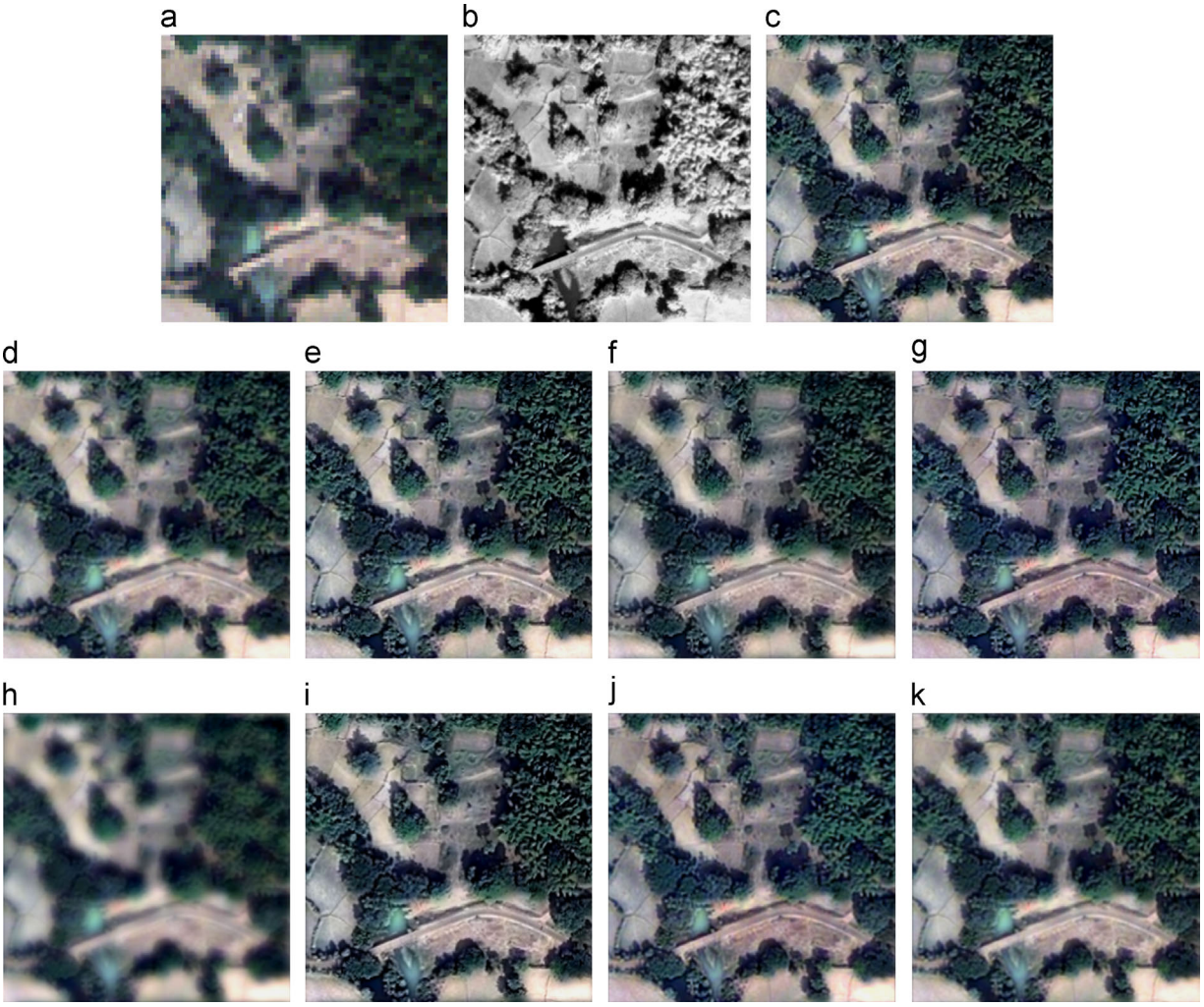


Fig. 4. Source images and the fused results using different parameters. (a) The resampled low-resolution MS image (RGB, 256×256 pixels); (b) the high-resolution PAN images (256×256 pixels); and (c)–(k) the fused RGB bands using different parameters which are shown in the 2nd–10th column of Table 1, respectively. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

[50]

$$Q4=\frac{1}{|\Omega|}\int_{\Omega}\frac{4|\sigma_{yz}|}{\sigma_y^2+\sigma_z^2}\cdot\frac{|\overline{y}|\cdot|\overline{z}|}{|\overline{y}|^2+|\overline{z}|^2}d\boldsymbol{x},$$

where the quaternion $\boldsymbol{y}=\boldsymbol{f}_1(\boldsymbol{x})+i\boldsymbol{f}_2(\boldsymbol{x})+j\boldsymbol{f}_3(\boldsymbol{x})+k\boldsymbol{f}_4(\boldsymbol{x})$ $\boldsymbol{z}=\boldsymbol{m}_1(\boldsymbol{x})+i\boldsymbol{m}_2(\boldsymbol{x})+j\boldsymbol{m}_3(\boldsymbol{x})+k\boldsymbol{m}_4(\boldsymbol{x})$, $\sigma_{\boldsymbol{y}}=E(|\boldsymbol{y}|^2)-|\overline{\boldsymbol{y}}|^2$, $\sigma_{\boldsymbol{yz}}=E(\boldsymbol{yz}^*)-\overline{\boldsymbol{y}}\overline{\boldsymbol{z}}^*$, $\boldsymbol{y}^*=\boldsymbol{f}_1(\boldsymbol{x})-i\boldsymbol{f}_2(\boldsymbol{x})-j\boldsymbol{f}_3(\boldsymbol{x})-k\boldsymbol{f}_4(\boldsymbol{x})$, $|\boldsymbol{y}|=\sqrt{\boldsymbol{y}\boldsymbol{y}^*}=\sqrt{\sum_{n=1}^{4}\boldsymbol{f}_n^2(\boldsymbol{x})}$, and $E(\boldsymbol{y})=\overline{\boldsymbol{y}}$ is a mean of $\boldsymbol{y}$ in a $D\times D$ neighborhood centered at $\boldsymbol{y}$. Obviously, Q4 is easily extended to the case that the MS image has N spectral bands.

The dynamic range of Q4 is [0, 1], and the bigger value indicates the better fusion result.

Table 2
Quantitative analysis of the images in Fig. 4(c)–(k).

Measure	Q4	ERGAS	SAM	SID	FCC	Q^F	AG	H
Reference	1	0	0	0	1	1	$\sqrt{2}$	∞
Fig. 4(c)	0.7606	5.9008	6.8683	0.1518	0.8955	0.4336	0.0733	7.8276
Fig. 4(d)	0.8340	4.3558	5.2511	0.0663	0.7486	0.2451	0.0532	7.8285
Fig. 4(e)	0.7312	6.6078	7.7021	0.2233	0.9044	0.4751	0.0810	7.7948
Fig. 4(f)	0.7006	6.2637	8.1769	0.1697	0.8878	0.3530	0.0627	7.8591
Fig. 4(g)	0.7405	7.2297	7.8872	0.1567	0.8730	0.4226	0.0792	7.7075
Fig. 4(h)	0.7895	4.4549	5.8170	0.0789	0.5787	0.0920	0.0336	7.8382
Fig. 4(i)	0.6967	7.0127	8.6558	0.2516	0.9232	0.4775	0.0783	7.8064
Fig. 4(j)	0.7812	6.4233	6.7214	0.0888	0.8064	0.3576	0.0674	7.7158
Fig. 4(k)	0.8182	4.6623	5.5613	0.0746	0.7977	0.3003	0.0576	7.8372

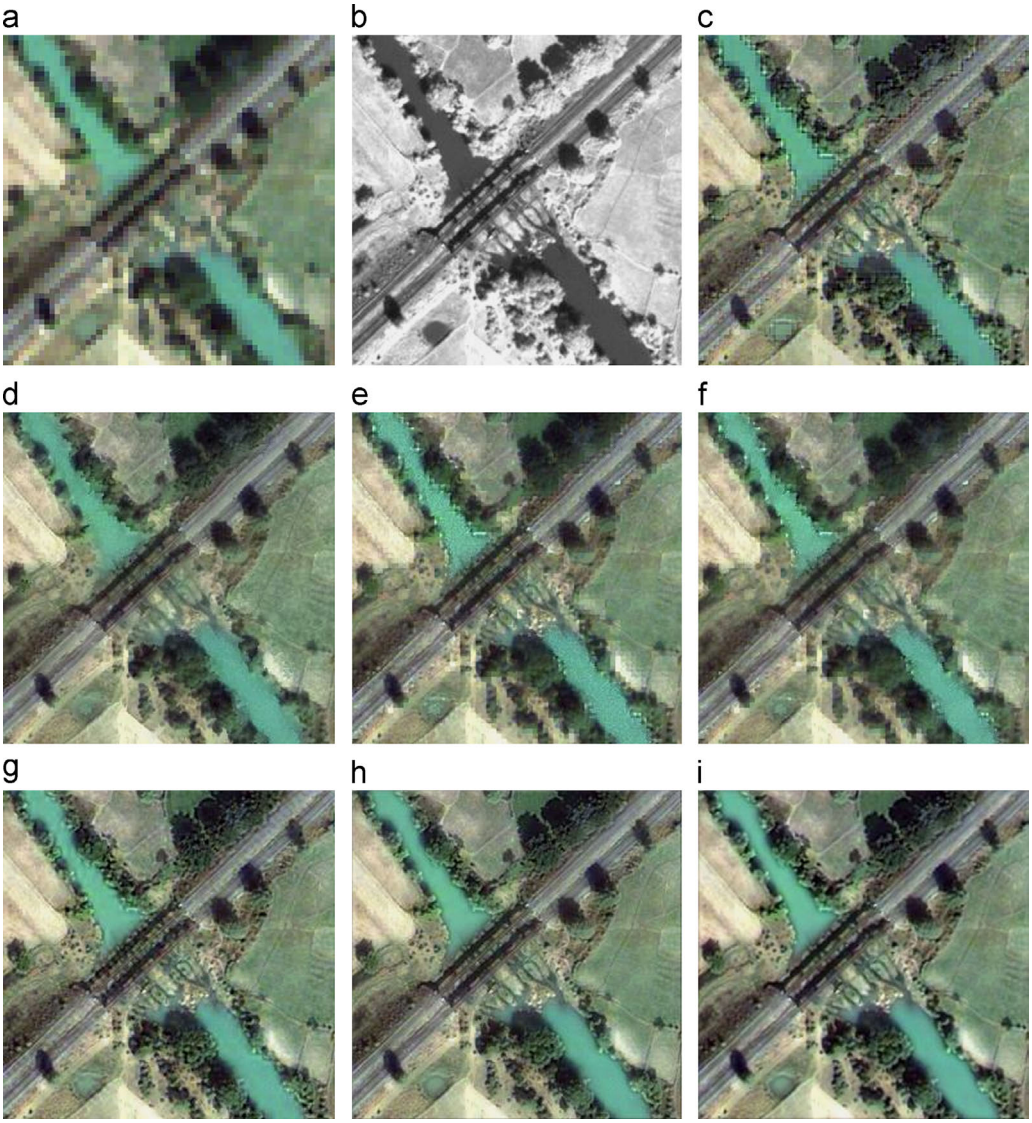


Fig. 5. Qualitative comparison: source QuickBird images (a river and road area in Chilka Lake region of India, acquired on February 23, 2005) and the fused results using different methods. (a) The resampled low-resolution MS image (RGB, 256 × 256 pixels); (b) the high-resolution PAN images (256 × 256 pixels); and (c)–(i) the fused RGB bands by GIHS, P+XS, VWP, AVWP, wavelet, FP and VFP methods, respectively. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

2. *Relative Dimensionless Global Error in Synthesis* (*ERGAS*): ERGAS is a spectral quality metric. It indicates the ratio between RMSE and the mean of each band of $\boldsymbol{f}$. It is sensitive to mean bias and range change. The ERGAS value is defined as [51]

$$\mathrm{RDGES} = 100\frac{h}{l}\sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(\frac{RMSE(\boldsymbol{f}_n,\boldsymbol{m}_n)}{\mu(\boldsymbol{f}_n)}\right)}$$

where h/l is the ratio between the size of original PAN and MS images, and $\mu(\boldsymbol{f}_n)$ is the mean of $\boldsymbol{f}_n$. The closer the ERGAS to 0 is, the more accurate the fusion is.

3. *Spectral angle mapper* (*SAM*): As a spectral correlation quality metric, SAM [20,52] denotes the mean spectral distortion between $\boldsymbol{m}$ and $\boldsymbol{f}$. It is calculated by

$$SAM = \frac{1}{|\Omega|}\int_{\Omega}\arccos\left(\frac{\langle\boldsymbol{f},\boldsymbol{m}\rangle}{\|\boldsymbol{f}\|_2\|\boldsymbol{m}\|_2}\right)dx.$$

The small value of SAM indicates ideal fused image.

4. *Spectral information divergence* (*SID*): The SID [53] describes the degradation degree of spectral correlation information between $\boldsymbol{f}$ an $\boldsymbol{m}$. Given a random vector $\boldsymbol{\theta}(\boldsymbol{x}) = (\boldsymbol{f}_1(\boldsymbol{x}),\ldots,\boldsymbol{f}_N(\boldsymbol{x}))$, by defining its normalized value $\boldsymbol{v}^{\boldsymbol{f}}(x) = (v_1^{\boldsymbol{f}}(x),\ldots,v_N^{\boldsymbol{f}}(x)) = \boldsymbol{\theta}(x)/|\boldsymbol{\theta}(x)|$, SID is given by

$$\mathrm{SID} = \frac{1}{|\Omega|}\int_{\Omega}\sum_{n=1}^{N}\left(v_n^{\boldsymbol{f}}\log\frac{v_n^{\boldsymbol{f}}}{v_n^{\boldsymbol{m}}} + v_n^{\boldsymbol{m}}\log\frac{v_n^{\boldsymbol{m}}}{v_n^{\boldsymbol{f}}}\right)dx.$$

The lower the value of SID the better the preservation degree of the spectral correlation information.

5. *Filtered correlation coefficient* (*FCC*): Based on the idea that the spatial information of PAN image is primarily concentrated in the high frequency domain, [16] presented the FCC as

$$FCC = \frac{1}{N}\sum_{n=1}^{N} Cor(k*\boldsymbol{f}_n, k*\boldsymbol{p}),$$

where $Cor(\cdot)$ is the correlation coefficient and $k = \begin{pmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{pmatrix}$.

The biggest value of FCC indicates ideal fused image.

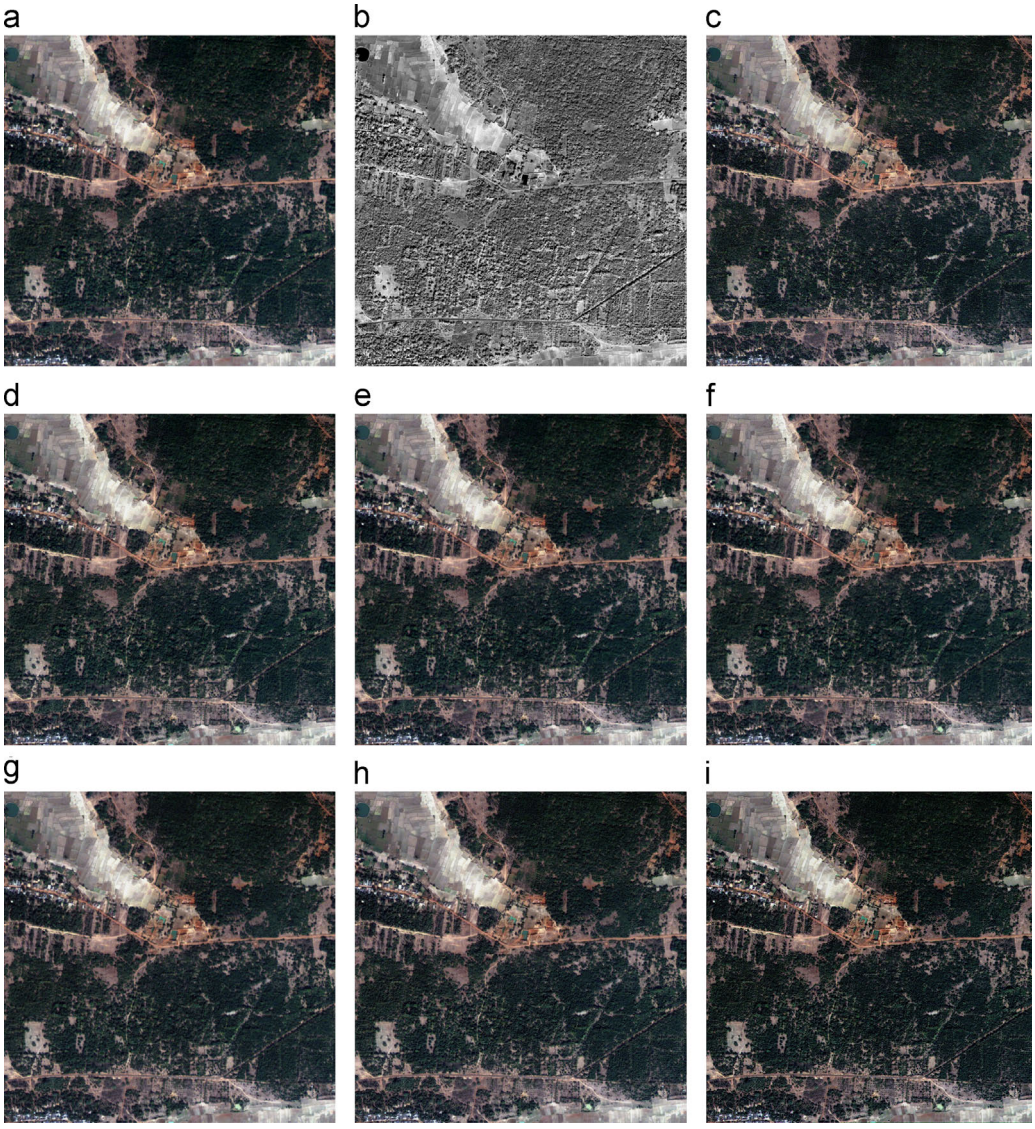


Fig. 6. Qualitative comparison: source QuickBird images (a mountainous area in Chilka Lake region of India, acquired on February 23, 2005) and the fused results using different methods. (a) The resampled low-resolution MS image (RGB, 2000 × 2000 pixels); (b) the high-resolution PAN images (2000 × 2000 pixels); and (c)–(i) the fused RGB bands by GIHS, P+XS, VWP, AVWP, wavelet, FP and VFP methods, respectively. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

6. *Objective image fusion performance measure* (Q^F): The objective image fusion performance [54] measure reflects the degree of the 'edge information preservation' when the pan-sharpening processes. Because the edge information is mainly concentrated in $\boldsymbol{p}$, Q^F can be defined as

$$Q^F = \frac{\sum_{n=1}^{N} \int_{\Omega} Q^{\boldsymbol{f}_n \boldsymbol{p}}(x) \omega^{\boldsymbol{f}_n}(x)\, dx}{\sum_{n=1}^{N} \int_{\Omega} \omega^{\boldsymbol{f}_n(x)} dx},$$

where $\omega^{\boldsymbol{f}_n}(x)$ is a weight and $Q^{\boldsymbol{f}_n \boldsymbol{p}}(x) \in [0,1]$ is an edge information preservation value (see [54] for details). For $0 \le Q^F \le 1$, the bigger the Q^F, the better the fusion result.

7. *Average gradient* (AG): Obviously, gradient is sensitive to subtle details of the image. Thus, its average (i.e., average gradient) can be used to assess the blurring degree of the image. Then, average gradient can be defined as [55]

$$AG = \frac{1}{N|\Omega|} \sum_{n=1}^{N} \int_{\Omega} |\nabla \boldsymbol{f}_n|\, dx.$$

Obviously, $0 \le AG \le \sqrt{2}$, and the bigger value of AG indicates the better fusion result.

8. *Entropy* (H): As an important statistical index, entropy is used to measure the information of an image. The definition of

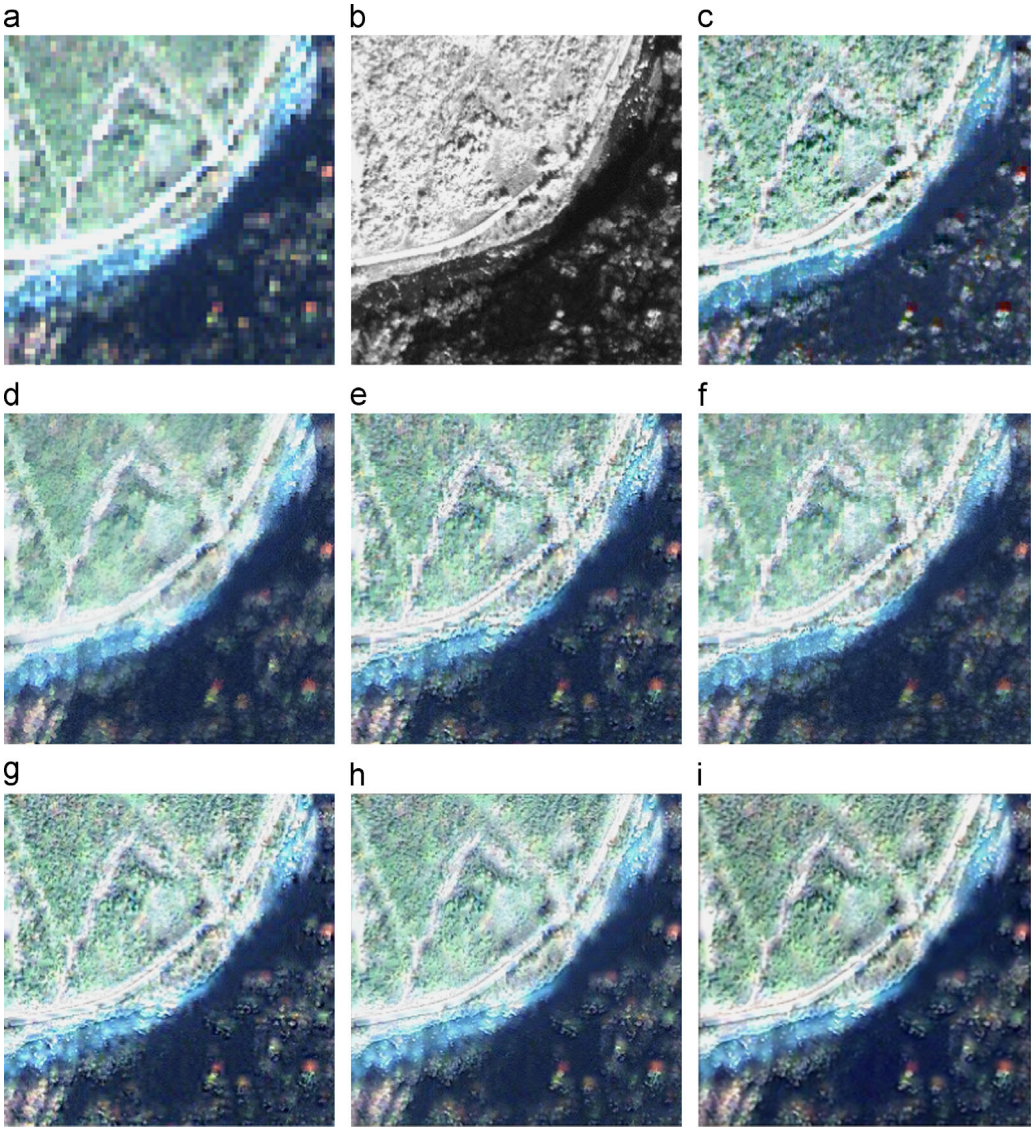


Fig. 7. Qualitative comparison: source IKONOS images (a mountainous area in Sichuan, China, acquired on May 15, 2008) and the fused results using different methods. (a) The resampled low-resolution MS image (RGB, 256 × 256 pixels); (b) the high-resolution PAN images (256 × 256 pixels); and (c)–(i) the fused RGB bands by GIHS, P+XS, VWP, AVWP, wavelet, FP and VFP methods, respectively. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

Table 3
Comparison of the proposed methods with the outstanding methods on QuickBird images shown in Fig. 5.

Measure Reference	Q4 1	ERGAS 0	SAM 0	SID 0	FCC 1	Q^F 1	AG $\sqrt{2}$	H ∞
GIHS	0.7881	3.7377	1.9317	0.0168	0.8084	0.4182	0.0549	7.6949
P+XS	0.7707	3.0993	2.9594	0.0079	0.8499	0.4290	0.0460	7.6850
VWP	0.8137	2.7953	1.4520	**0.0028**	0.8332	0.3372	0.0490	7.7100
AVWP	0.8255	2.6533	1.8105	0.0044	0.7911	0.3617	0.0448	7.7209
Wavelet	0.7352	3.7291	1.8105	0.0210	0.9269	0.3189	0.0567	7.7374
FP	0.8252	3.5017	2.8033	0.0152	0.8900	0.3974	0.0548	7.7404
VFP	**0.8462**	**2.4037**	**1.4207**	0.0125	**0.9290**	**0.4245**	**0.0568**	**7.7664**

entropy is [55]

$$H(x) = -\sum_{i=0}^{L-1} p_i \log_2 p_i, \tag{23}$$

where L is the gray level of the image, for example $L=256$; $p_i = n_i/n$, n_i is the number of pixels in the ith gray-level, and n is the total pixel number of the images.

The larger value of entropy indicates better image quality.

5.2. The influence of the parameters

As mentioned in Section 4, the VFP model consists of four terms, each term has its corresponding parameters. These parameters are λ, η, c_0 and c_1. Many experiments demonstrate that the reasonable ranges of λ, η and c_0, c_1 are [0.01, 1] and [0.001, 0.5], respectively. We now analyze the influence of them.

We first obtain original MS and PAN images from the QuickBird satellite (see Fig. 4(a) and (b)), then use these images to pan-sharpen using different parameters as shown in each column of Table 1. The RGB bands of pan-sharpened results are shown in Fig. 4(c)–(k). Besides, we also calculate the quantitative value for each result (see Table 2).

First, we analyze the influence of λ. Fig. 4(c) (d) and (e) shows the results of different λ, i.e., $\lambda=0.5$, 0.01, 0.9, respectively. Comparing with Fig. 4(c) and (d), we find that (d) is more blurry in detail while more vivid in color, this phenomenon is easily verified by comparing the first two rows of Table 2, in which (d) is better than (c) in Q4, ERGAS, SAM, SID and H metrics while worse than (c) in FCC, Q^F and AG. Similarly, we also find that (e) is clearer than (c) in detail while

Table 4
Comparison of the proposed methods with the outstanding methods on QuickBird images as shown in Fig. 6.

Measure	Q4	ERGAS	SAM	SID	FCC	Q^F	AG	H
Reference	1	0	0	0	1	1	$\sqrt{2}$	∞
GIHS	0.6992	7.3217	6.1171	0.1790	0.7796	0.5187	0.0701	7.6385
P+XS	0.7618	4.7794	5.2545	0.1719	0.8295	0.3770	0.0503	7.6916
VWP	0.7350	5.2009	4.1287	0.1204	0.8388	0.3362	0.0558	7.6829
AVWP	0.7761	4.4990	**4.0876**	0.1524	0.7503	0.3267	0.0467	7.6674
Wavelet	0.6416	7.3240	8.3008	0.1977	0.9510	0.3442	0.0766	7.6454
FP	0.6993	6.4705	7.4369	0.1780	**0.9770**	0.4555	0.0688	7.6610
VFP	**0.7837**	**4.4761**	4.8780	**0.1200**	0.9149	**0.5340**	**0.0767**	**7.7030**

Table 5
Comparison of the proposed methods with the outstanding methods on IKONOS images as shown in Fig. 7.

Measure	Q4	ERGAS	SAM	SID	FCC	Q^F	AG	H
Reference	1	0	0	0	1	1	$\sqrt{2}$	∞
GIHS	0.4142	7.2273	3.3316	0.0460	0.9250	**0.587**2	0.0845	7.7130
P+XS	0.5811	4.1977	3.7007	0.0527	0.8071	0.2981	0.0562	7.8124
VWP	0.5785	5.1789	3.5122	0.0496	0.8724	0.2976	0.0828	7.7263
AVWP	0.5780	4.8613	3.5661	0.0276	0.7606	0.3306	0.0658	7.7517
Wavelet	0.5823	5.2356	3.7576	0.0585	0.9642	0.2594	0.0862	7.7290
FP	0.5749	6.3749	4.9642	**0.0245**	0.8211	0.3511	0.0754	7.8013
VFP	**0.5885**	**3.9484**	**3.2676**	0.0485	**0.9648**	0.3343	**0.0877**	**7.8577**

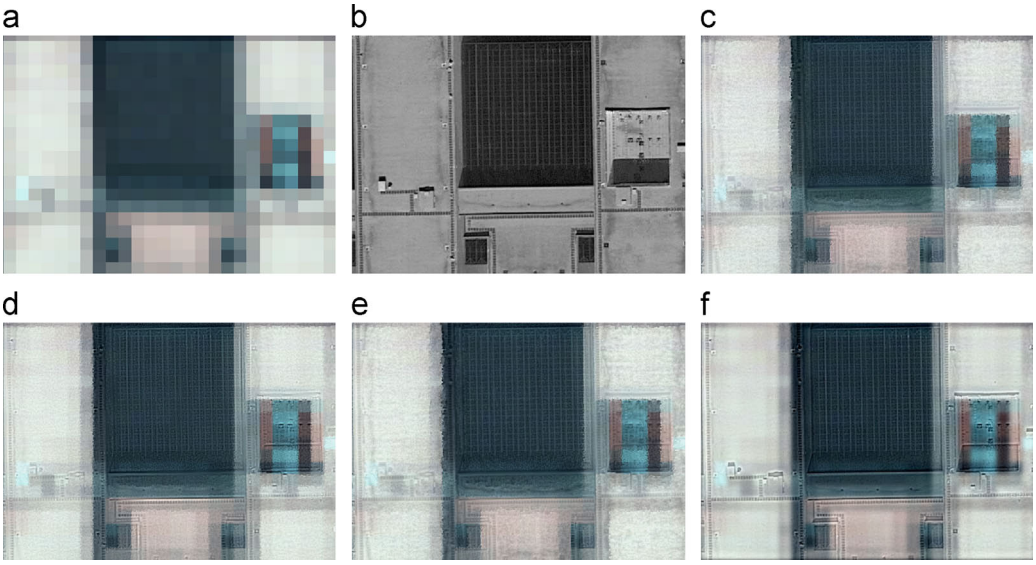


Fig. 8. HYDICE image pan-sharpening. (a) The resampled low-resolution MS image (RGB, 400 × 560 pixels); (b) the high-resolution PAN images (400 × 560 pixels); and (c)–(f) the fused results of P+XS,VWP,AVWP and VFP methods (the blue, green and red colors are bands 17, 27 and 60, respectively). (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

worse than (c) in color. In summary, the larger λ will lead to better spatial resolution and worse spectral quality.

Taking a similar analysis on η (by comparing (c), (f) and (g)), c_0 (by comparing (c), (h) and (i)) and c_1 (by comparing (c), (j) and (k)), we can draw a conclusion that larger λ, c_1 and smaller η, c_0 will result in better spatial resolution and worse spectral quality, in contrast we will obtain an image with worse spatial resolution and better spectral quality.

5.3. Comparative studies

5.3.1. Qualitative analysis

In this part, we present our outcomes and compare them with five state-of-the-art schemes visually. These five schemes contain the Generalized IHS method (GIHS) [9], the P+XS method [24], the variational wavelet pan-sharpening method (VWP) [25], the alternate variational wavelet pan-sharpening method (AVWP) [25], and the wavelet method [16,26], respectively. The comparison results are shown in Figs. 5–7. Note that all the parameters in these five schemes are set based on the authors' suggestions, and 'sym4' is selected as the base in wavelet decomposition. Besides, the balancing parameters in the VFP method are set as $\lambda=0.5$, $\eta=0.5$, $c_0=c_1=0.1$.

In the first two figures, the two original QuickBird data, presented in (a) and (b), are acquired from the Chilka Lake area of India on February 23, 2005. All the results obtained by GIHS, P+XS, VWP, AVWP, Wavelet, FP and VFP are shown in (c), (d), (e), (f), (g), (h), (i), respectively. For visual impact, we only display the first three bands in the MS image, i.e., red, blue and green bands.

First, we can find that all of the results have a significant visual improvement compared with the original images. However, the results obtained by different methods are distinctly different. The GIHS results have great performance in the spatial quality while they suffer some visual degradation. The results of P+XS, VWP and AVWP methods have very rare spectral distortion, yet they are relatively lack of definition and details. The wavelet results have both better spatial and spectral properties, but they suffer some aliasing effects. Instead, the FP and VFP methods seem to produce promising both spatial and spectral results without apparent confusion.

Taking Fig. 5 as an example, first we find that (c)–(i) are clearer than (a). Especially, the outline of the water flow under the bridge in the middle of the image (a) is blurry, while this area is distinct in all (c)–(i). The result of (c) (i.e., GIHS) has sharp outlines, whereas there are some dentate boundaries and little color distortion in both sides of the river and the road. (d)–(f) (i.e., P+XS, VWP and AVWP) enjoy great conservation in color, but they look blurry and over-smoothing in some extent. (f) (i.e., wavelet) seems to be clear both in color and definition, whereas it suffers some aliasing effects. The FP result in (h) looks more vivid than (c)–(f), but it is still not distinct compared to (i) which is produced by VFP. The VFP result seems to have high performance both in hue and in outline. For instance, the trees beside the river in the middle-lower part of the image (i) is more vivid than others. The similar phenomena can also be found in the road and river areas and Fig. 6.

To verify the stability of our methods, in Fig. 7, we further show the IKONOS data and its corresponding results. observing on the results, we can also find that our results outperform other results. Especially, the results of GIHS, P+XS, VWP, AVWP, Wavelet and FP contain much noise, while the result of VFP method looks more vivid and noise-free. This just confirms the aforementioned point that the VFP method is robust to noise.

Therefore, for the visual performance, the FP and VFP methods, especially VFP, seem to be better than other schemes on QuickBird and IKONOS data.

5.3.2. Quantitative analysis

In this part, we quantitatively evaluate our outcomes using eight measures which are mentioned in Section 5.1, and compare the evaluation results with those of the five aforementioned outstanding methods. We still use the results, of which the RGB bands are shown in Figs. 5–7, to do our comparison. Tables 3–5, which are related to the three figure data, are presented to report the quantitative evaluation. It should be noted, in each table, that the best value of each measure is marked in bold, and each value in the reference row is the ideal value of corresponding measure;

Observing these tables, we find that GIHS method generally has good performance in Q4, Q^F, AG and H measures, while it fails in other metrics. P+XS, VWP and AVWP methods are better than the other methods in terms of ERGAS, SAM, SID and FCC measures, but they lose in the other criteria. Nevertheless, wavelet and FP methods are quite balanced in all metrics, but they have few outstanding

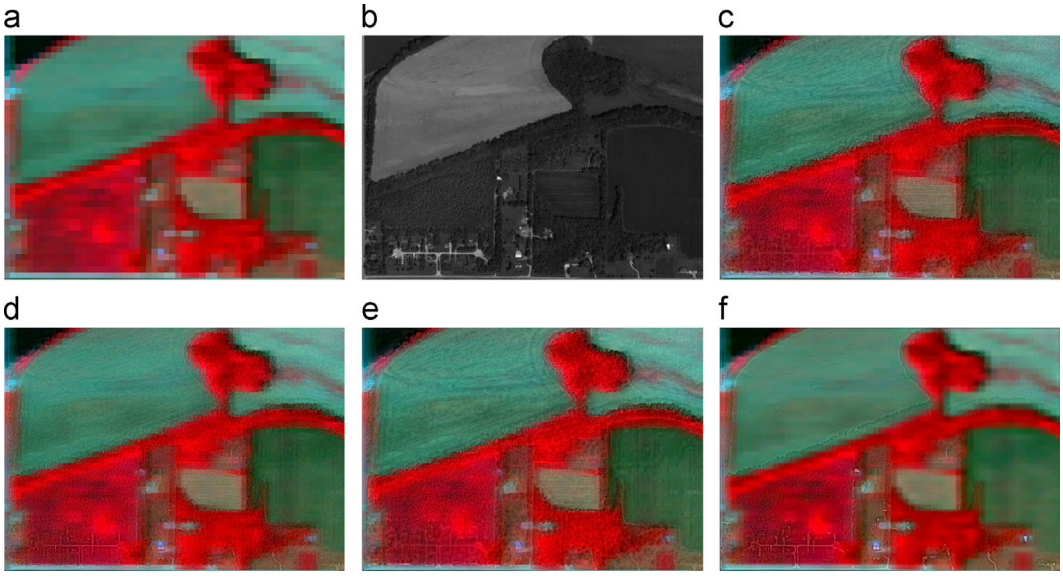


Fig. 9. AVIRIS image pan-sharpening. (a) The resampled low-resolution MS image (RGB, 400 × 560 pixels); (b) the high-resolution PAN images (400 × 560 pixels); and (c)–(f) the fused results of P+XS,VWP,AVWP and VFP methods (the blue, green and red colors are bands 17, 27 and 60, respectively). (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

Table 6
Comparison of the VFP method with the other variational methods on HYDICE images as shown in Fig. 8.

Measure	Q4	ERGAS	SAM	SID	FCC	Q^F	AG	H
Reference	1	0	0	0	1	1	$\sqrt{2}$	∞
P+XS	0.3045	2.2208	1.4286	0.0111	0.7399	0.4689	0.0282	7.6540
VWP	0.3390	2.2888	**0.4786**	0.0104	0.8423	0.4132	0.0307	7.6791
AVWP	0.2792	2.2077	0.5504	0.0100	0.7505	0.4880	0.0294	7.6883
VFP	**0.3400**	**2.1616**	1.0462	**0.0065**	**0.9028**	**0.4943**	**0.0309**	**7.7114**

Table 7
Comparison of the VFP method with the other variational methods on AVIRIS images as shown in Fig. 9.

Measure	Q4	ERGAS	SAM	SID	FCC	Q^F	AG	H
Reference	1	0	0	0	1	1	$\sqrt{2}$	∞
P+XS	0.5067	3.7320	3.3579	0.2717	0.7133	0.3251	0.0264	7.5687
VWP	0.6413	2.6162	**1.6223**	0.1921	0.7647	0.2872	0.0238	7.6331
AVWP	0.5013	3.3388	2.3398	0.1658	0.7019	0.3089	0.0264	**7.6363**
VFP	**0.7435**	**2.3238**	3.8151	**0.0637**	**0.8395**	**0.3330**	**0.0288**	7.3547

performance. Since FCC, Q^F, AG and H criteria measure the spatial resolution, yet Q4 and SID criteria report the spectral quality. This just echoes our aforementioned qualitative comparison.

Compared to other schemes, our methods perform well in all the measures. In terms of the FP method, it is averagely ranked midway in all seven methods. For example, in Table 3, FP is ranked 2 in H criteria, and ranked 3 in Q4 and FCC criteria. Especially, FP is better than the wavelet method in five of eight criteria, that is, FP is generally better than the wavelet method. In the VFP method, we find that it has the best performance compared to the others. Taking Table 3 as an example, VFP method ranked first in 7 of 8 criteria, and performed well in both spatial and spectral quality measures. A similar conclusion can also be found in Table 4 and 5.

Thus, in quantitative aspect, the FP method is generally comparable with the other techniques, and the VFP method performs better than the others. This, to a large extent, has verified our qualitative comparison.

5.4. Extension to hyperspectral imagery

So far, we have studied the techniques for pan-sharpening MS (4–6 bands) and its corresponding PAN images. It could be interesting to pan-sharpen the hyperspectral images. Since the variational pan-sharpening methods are not limited to the number of bands, they are easily extended to hyperspectral data. Thus we study this extension of VFP.

We work with two datasets: HYDICE image of Washington DC Mall (contain 191 bands) and AVIRIS image of Northwest Tippecanoe County (contain 220 bands). For detailed information of these two datasets, one can refer to the website https://engineering.purdue.edu/~biehl/MultiSpec/hyperspectral.html. Since these datasets have no PAN images, we extract the same scenes from Google Maps (which provide high resolution images, see http://ditu.google.cn/) and treat them as the PAN images.

The pan-sharpening results are shown in Figs. 8 and 9. The results of other variational methods, i.e., P+XS, VWP, and AVWP, are also presented for comparison. For simplicity, the images in these figures are made using bands 17, 27 and 60 for blue, green and red colors respectively. Besides, we also show the quantitative results for each figure (see Tables 6 and 7). Note that the image used for quantitative evaluation has no longer three bands, but full, i.e., 191 and 220 bands, respectively.

Observing on these figures and tables, we can find that VFP outperforms the others consistently. For example, in Fig. 8, (f) is better than (c)–(e) in color and detail. Especially, (c)–(e) contain much noise, while (f) is noise-free. This phenomenon can be verified easily from Table 6, in which VFP is better than others in 7 of 8 metrics. Therefore, the VFP method is also remarkable in hyperspectral imagery.

6. Conclusions

We have proposed two pan-sharpening methods based on the framelet framework. The first method, named as FP, first decomposed the MS and PAN images into framelet coefficients, then chose the framelet coefficients to compose new coefficients. Finally, the new coefficients were converted into an image which is exactly the pan-sharpened result. The result of the FP method is promising but inflexible. To overcome this drawback, the second method, termed VFP, was proposed. In VFP method, by combining FP and the fusion requirements of geometry keeping, spectrum preserving and the sparsity of image in the framelet domain, a variational energy functional was presented. To minimize the energy functional more efficiently, the split Bregman iteration was detailed, and the result of FP method was set as the initial value. Furthermore, to verify the effectiveness of our methods, we tested them on QuickBird and IKONOS data, compared them with five outstanding techniques qualitatively and quantitatively, analyzed the influence of parameters of VFP, and extended the VFP to hyperspectral data as well as comparison study. The experimental results demonstrate that the FP method is as good as others, and the VFP method is effective and robust compared with others.

Pan-sharpening is a challenging problem, and there are still many open questions that should be discussed. Our methods are effective, but also could be improved in many aspects. For example, there are several parameters which should be chosen manually. Further research will be extended to develop some automatic methods for parameters choosing. Besides, new pan-sharpening techniques based on the sparsity of the image under

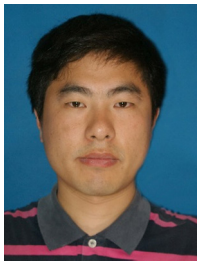
some certain bases have been developed recently. We will combine our variational approach with these techniques in the future.

Acknowledgments

This work is supported by the National Basic Research Program (973 Program, no. 2011CB707104), Program for New Century Excellent Talents in University (no. NCET-08-0193), and the National Science Foundation of China (11001082).

References

[1] S. Ashraf, L. Brabyn, B.J. Hicks, Image data fusion for the remote sensing of freshwater environments, Appl. Geogr. 32 (2012) 619–628.
[2] Yuhendra, I. Alimuddin, J.T.S. Sumantyo, H. Kuze, Assessment of pan-sharpening methods applied to image fusion of remotely sensed multi-band data, Int. J. Appl. Earth Obs. Geoinform. 18 (2012) 165–175.
[3] T. Bretschneider, O.D. Kao, Image fusion in remote sensing, in: First Online Symposium of Electronics Engineers.
[4] V. Vijayaraj, N. Younan, C. Ohara, Concepts of image fusion in remote sensing application, IEEE Trans. Geosci. Remote Sensing 10 (2006) 3781–3784.
[5] G. Andrea, N. Filippo, Pan sharpening of remote sensing images using a multiscale Kalman filter, Pattern Recognit. 40 (2007) 3568–3577.
[6] U. Kumar, C. Mukhopadhyay, T. Ramachandra, Pixel based fusion using IKONOS imagery, Int. J. Recent Trends Eng. 1 (2009) 173–175.
[7] R.A. Schowengerdt, Remote Sensing: Models and Methods for Image Processing, 3rd edition, Academic Press, San Diego, USA, 2007.
[8] L. Fonseca, L. Namikawa, E. Castejon, L. Carvalho, C. Pinho, A. Pagamisse, Image Fusion for Remote Sensing Applications, InTech, 2011.
[9] T. Tu, S. Su, H. Shyn, P. Huang, A new look at IHS-like image fusion methods, Inf. Fusion 2 (2001) 177–186.
[10] S. Rahmani, M. Strait, D. Merkurjev, M. Moeller, T. Wittman, An adaptive IHS pan-sharpening method, IEEE Geosci. Remote Sens. Lett. 7 (2010) 746–750.
[11] T. Carper, T. Lillesand, R. Kiefer, The use of intensity-hue-saturation transformations for merging SPOT panchromatic and multispectral image data, IEEE Geosci. Remote Sens. Lett. 56 (1990) 459–467.
[12] P. Chavez, S. Sides, J. Anderson, Comparison of three different methods to merge multiresolution and multispectral data: LANDSAT TM and SPOT panchromatic, Photogramm. Eng. Remote Sens. 57 (1991) 295–303.
[13] C. Laben, B. Brower, Process for enhancing the spatial resolution of multispectral imagery using pansharpening, 2000, http://www.freepatentsonline.com/6011875.html.
[14] C. Padwick, M. Deskevich, F. Pacifici, S. Smallwood, Worldview-2 pan-sharpening, in: American Society for Photogrammetry and Remote Sensing, 2010.
[15] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, F. Lotti, F. Nencini, M. Selva, Context-sensitive pan-sharpening of multispectral images, in: Proceedings of the Semantic and Digital Media Technologies Second International Conference on Semantic Multimedia, SAMT'07, Springer-Verlag, Berlin, Heidelberg, 2007, pp. 121–125.
[16] J. Zhou, D. Civico, J. Silander, A wavelet transform method to merge landsat TM and SPOT panchromatic data, Int. J. Remote Sens. 19 (1998) 743–757.
[17] J. Nunez, X. Otazu, O. Fors, A. Prades, V. Pala, R. Arbiol, Multiresolution-based image fusion with additive wavelet decomposition, IEEE Trans. Geosci. Remote Sens. 37 (1999) 1204–1211.
[18] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, Context-driven fusion of high spatial and spectral resolution data based on oversampled multiresolution analysis, IEEE Trans. Geosci. Remote Sens. 40 (2002) 2300–2312.
[19] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, M. Selva, Mtf-tailored multiscale fusion of high-resolution MS and pan imagery, Photogramm. Eng. Remote Sens. 5 (2006) 591–596.
[20] L. Alparone, L. Wald, J. Chanussot, C. Thomas, Comparison of pansharpening algorithms: outcome of the 2006 GRS-S data-fusion contest, IEEE Trans. Geosci. Remote Sens. 45 (2007) 3012–3021.
[21] T. Chan, J. Shen, L. Vese, Variational PDE models in image processing, Not. Am. Math. Soc. 50 (2003) 14–26.
[22] G. Sapiro, Geometric Partial Differential Equations and Image Processing, Cambridge University Press, 2001.
[23] G. Aubert, P. Kornprobst, Mathematical problems in image processing: partial differential equations and the calculus of variations, in: Applied Mathematical Sciences, vol. 147, Springer-Verlag, 2nd edition, 2009.
[24] C. Ballester, V. Caselles, L. Igual, J. Verdera, A variational model for P+XS image fusion, Int. J. Comput. Vis. 69 (2006) 43–58.
[25] M. Moeller, T. Wittman, A. Bertozzi, M. Burger, A Variational Approach for Sharpening High Dimensional Images (Ph.D. thesis), Westfalische Wilhelms Universitat Munster, 2009.
[26] R. King, J. Wang, A wavelet based algorithm for pan sharpening landsat 7 imagery, in: IEEE International Geoscience and Remote Sensing Symposium, 2 (2001) 849–851.
[27] P. Hill, N. Canagarajah, D. Bull, Image fusion using complex wavelets, in: The British Machine Vision Conference, 2002.
[28] Y. Song, M. Li, Q. Li, L. Sun, A new wavelet based multi-focus image fusion scheme and its application on optical microscopy, in: 2006 IEEE International Conference on Robotics and Biomimetic, 2006.
[29] Y. Chibani, A. Houacine, Redundant versus orthogonal wavelet decomposition for multisensor image fusion, Pattern Recognit. 36 (2003) 879–887.
[30] X. Otazu, M. González-Audícana, O. Fors, J. Núnez, Introduction of sensor spectral response into image fusion methods. Application to wavelet-based methods, IEEE Trans. Geosci. Remote Sens. 43 (2005) 2376–2385.
[31] I. Daubechies, B. Han, A. Ron, Z. Shen, Framelets: MRA-based constructions of wavelet frames, Appl. Comput. Harmon. Anal. 17 (2003) 1–46.
[32] R. Chan, S. Riemenschneider, L. Shen, Z. Shen, Tight frame: an efficient way for high-resolution image reconstruction, Appl. Comput. Harmon. Anal. 17 (2004) 91–115.
[33] A. Chai, Z. Shen, Deconvolution: a wavelet frame approach, Numer. Math. 106 (2007) 529–587.
[34] J.F. Cai, R. Chan, Z. Shen, A framelet-based image inpainting algorithm, Appl. Comput. Harmon. Anal. 24 (2008) 131–149.
[35] J. Cai, H. Ji, C. Liu, Z. Shen, Framelet-based blind motion deblurring from a single image, IEEE Trans. Image Process. 21 (2012) 562–572.
[36] A. Ron, Z. Shen, Affine system in $l_2(\mathbb{R}^d)$: the analysis of the analysis operator, J. Funct. Anal. 148 (1997) 408–447.
[37] R. Chan, L. Shen, Z. Shen, A Framelet-Based Approach for Image Inpainting, Technical Report, Department of Mathematics, The Chinese University of Hong Kong, 2005.
[38] DigitalGlobe, Quickbird Scene 000000185964_01_P001, Level Standard 2A, DigitalGlobe, Longmont, Colorado, 2004.
[39] G. Piella, Image fusion for enhanced visualization: a variational approach, Int. J. Comput. Vis. 83 (2009) 1–11.
[40] C. Thomas, T. Ranchin, L. Wald, J. Chanussot, Synthesis of multispectral images to high spatial resolution: a critical review of fusion methods based on remote sensing physics, IEEE Trans. Geosci. Remote Sens. (2008) 1301–1312.
[41] J.-F. Cai, S. Osher, Z. Shen, Split Bregman method and frame based image restoration, Not. Am. Math. Soc. 8 (2009) 337–369.
[42] J.-F. Cai, S. Osher, Z. Shen, Linearized Bregman iterations for frame-based image deblurring, SIAM J. Imaging Sci. 2 (2009) 226–252.
[43] T. Goldstein, S. Osher, The split Bregman method for L1-regularized problems, SIAM J. Imaging Sci. 2 (2009) 323–343.
[44] S. Kindermann, S. Osher, P. Jones, Deblurring and denoising of images by nonlocal functionals, SIAM Multiscale Model. Simulation 4 (2005) 1091–1115.
[45] Q. Du, N. Younan, R. King, V. Shah, On the performance evaluation of pan-sharpening techniques, IEEE Geosci. Remote Sens. Lett. 4 (2007) 518–522.
[46] V. Vijayaraj, C. Hara, N. Younan, Quality analysis of pansharpened images, in: International Geoscience and Remote Sensing Symposium, 57 (2004) 85–88.
[47] Q. Du, O. Gungor, J. Shan, Performance evaluation for pansharpening techniques, in: IEEE International Geoscience and Remote Sensing Symposium, 2007 pp. 4264–4266.
[48] S. Zheng, W. Shi, J. Liu, G. Zhu, J. Tian, Multisource image fusion method using support value transform, IEEE Trans. Image Process. 16 (2007) 1831–1839.
[49] I. Parcharidis, L. Tani, Landsat TM and ERS data fusion: a statistical approach evaluation for four different methods, in: International Geoscience and Remote Sensing Symposium, 5 (2000) 2120–2122.
[50] L. Alparone, S. Baronti, A. Garzell, F. Nencini, A global quality measurement of pan-sharpened multispectral imagery, IEEE Geosci. Remote Sens. Lett. 1 (2004) 313–317.
[51] L. Wald, T. Ranchin, M. Mangolini, Fusion of satellite images of different spatial resolutions: assessing the quality of resulting images, Photogramm. Eng. Remote Sens. 63 (1997) 691–699.
[52] S. Li, B. Yang, A new pan-sharpening method using a compressed sensing technique, IEEE Trans. Geosci. Remote Sens. 49 (2011) 738–746.
[53] C. Chang, Spectral information divergence for hyperspectral image analysis, in: International Geoscience and Remote Sensing Symposium, 1 (1999) 509–511.
[54] C. Xydeas, V. Petrovic, Objective image fusion performance measure, Electr. Lett. 36 (2000) 308–309.
[55] W. Liu, J. Huang, Y. Zhao, Image fusion based on PCA and undecimated discrete wavelet transform, in: Neural Information Processing, vol. 4233, Springer, Berlin/Heidelberg, 2006, pp. 481–488.



Faming Fang received the Ph.D. degree in Computer Science from East China Normal University, Shanghai, China, in 2013. He is currently a Lecturer with the Department of Computer Science, East China Normal University, Shanghai, China. His main research area is image processing using mathematical methods.

Guixu Zhang received the Ph.D. degree from Institute of Modern Physics, Chinese Academy of Sciences, Lanzhou, China, in 1998. He is currently a Professor in Department of Computer Science and Technology, East China Normal University, Shanghai, China. His research interests include hyperspectral remote sensing, image processing, and artificial intelligence. He also serves as an Associate Editor of Journal of Applied Remote Sensing.

Fang Li received her MSc degree in Mathematics from the South West China Normal University in 2004 and she received her Ph.D. degree in Mathematics from the East China Normal University in 2007. She is currently an associate professor in the Department of Mathematics, East China Normal University. Her research interests include anisotropic diffusion filtering, the variational methods and PDEs in image processing.

Chaomin Shen is an associate professor of computer science at East China Normal University (ECNU). He received his Master and Ph.D. degrees from the National University of Singapore (NUS) and ECNU, respectively, all in Mathematics. During 1998 and 2004, he was with the Centre for Remote Imaging, Sensing and Processing (CRISP), NUS, as an associate scientist. His research interest includes image processing using mathematical methods and machine learning.