

# An Efficient and High-Quality Pansharpening Model Based on Conditional Random Fields

Yong Yang, Hangyuan Lu, Shuying Huang, Yuming Fang, Wei Tu

**Abstract**—Pansharpening fuses a low spatial resolution multi-spectral (MS) image with the corresponding panchromatic (PAN) image to obtain a high spatial resolution MS (HRMS) image. Traditional fusion methods may easily cause a spectral or spatial distortion when injecting details into an MS image. To preserve the spectral and spatial information, an efficient pansharpening model based on conditional random fields (CRFs) is proposed. With this model, a state feature function is designed to force the HRMS image filtered using a blur function to be consistent with the up-sampled MS image and retain the spectral fidelity. To obtain a proper blur function, a new filter-acquisition method is proposed for the unified CRF-based model. Meanwhile, a transition feature function is defined to enable the transition of HRMS pixels to follow the gradient of a PAN image and ensure the sharpness of the fused image. Considering the characteristics of the gradient domain, a total variation regularization is designed to make the gradient of the HRMS image sparse. Finally, the augmented Lagrangian function of the model is solved by employing an alternating direction method of the multipliers. Experiment results indicate that, compared with previous state-of-the-art pansharpening methods, the proposed method can achieve the best fusion results with high computational efficiency.

**Index Terms**—Pansharpening, CRFs, filter acquisition, feature function, image fusion.

## 1. Introduction

Owing to the payload, cost, and bandwidth limitations and restrictions of a satellite, it is extremely challenging to provide high spatial resolution multispectral (HRMS) images directly [1, 2]. A practical solution is to capture a panchromatic (PAN) image that has a high spatial resolution but contains little spectral information, and then fuse it with an up-sampled multispectral (UPMS) image to achieve an HRMS image through proper fusion technology [3]. This remote sensing image fusion method is also called pansharpening. Pansharpening has been widely applied in many different fields including change detection, object classification, surface mapping, and disaster warning [4].

In general, pansharpening methods can be divided into three classes: component substitute (CS), multi-resolution analysis (MRA), and model-based methods. A CS-based method projects the multispectral (MS) image into another space and replaces the spatial component with the PAN image. Classical CS-based methods include a principle component analysis [5], intensity-hue-saturation (IHS) [6], Gram–Schmidt spectral sharpening [7], and Gram–Schmidt adaptive (GSA) [8] approaches. Although these methods can effectively maintain the spatial fidelity, when the correlation between an MS image and a PAN image is low, they may produce a significant spectral distortion [9].

An MRA-based method decomposes the source images into several scales using a wavelet transform or pyramid transform and fuses them at each scale [10]. Popular MRA-based methods include a decimated wavelet transform [11], an undecimated wavelet transform [12], a Laplacian pyramid [13], and a non-subsampled shearlet transform (NSST) [14]. MRA-based methods can retain the spectral information well, but have certain drawbacks such as an easy spatial distortion and high computational complexity [15].

<sup>T</sup>This work is supported by the National Natural Science Foundation of China (No. 61662026 and No.61862030), and by the Natural Science Foundation of Jiangxi Province (No. 20182BCB22006 and No. 20181BAB202010).

Y. Yang, H. Lu, Y. Fang, and W. Tu are with the School of Information Technology, Jiangxi University of Finance and Economics, Nanchang 330032, China (e-mail: greatyangy@126.com; lhyhzice@163.com; leo.fangyuming@foxmail.com; ncsytuwei@163.com).

S. Huang is with the School of Software and Communication Engineering, Jiangxi University of Finance and Economics, Nanchang 330032, China (e-mail: shuyinghuang2010@126.com).

A model-based method optimizes the variables based on a constructed model to realize pansharpening. Kang *et al.* [16] proposed a matting model for pansharpening. Later, Yang *et al.* [17] improved on this approach by combining a matting model with NSST to retain the improved spectral fidelity. Due to the significant performance of sparse representation (SR) on reconstructing images, an adaptive-weight detail injection method based on SR was also proposed [18]. In addition, to preserve the structural information, a new image fusion method was proposed [19] using image decomposition and SR. However, these methods are computationally expensive. To reduce the computational complexity, a regression model was proposed to exploit the iterative trend of the coefficient in the fused image [20], and a nonlinear IHS model was proposed [21] to obtain a proper intensity component of the image. Nevertheless, balance between the spatial clarity and spectral fidelity remains a problem in pansharpening methods. Some researchers have recently employed a Markov random field (MRF) prior to improve the fusion quality. Khademi *et al.* [22] combined an MRF-based prior model with a Bayesian framework to obtain a fused HRMS image, and Upla *et al.* [23] used an MRF to consider the spatial dependencies among the pixels. These methods consider the blur and down-sampling process in all bands within the MS image, which can improve the fusion quality but increase the computational burden. In addition, proper blur and down-sampling matrices are difficult to obtain. Deep learning has been successfully applied in pansharpening [24, 25] and has obtained impressive fusion results. However, such approaches usually require a significant computational time for the training process or require a large number of training samples.

To overcome these shortcomings, the design of a fast and efficient model with a high fusion quality is essential for practical applications. Injecting details into a UPMS will inevitably cause spectral distortion, and therefore, a way to balance the spectral fidelity, spatial sharpness, and efficiency is a key problem for pansharpening. For remote sensing images, there are two well-known assumptions. One assumption is regarding the relationship between the HRMS and UPMS (RHU): An UPMS image can be considered a blurred version of an HRMS image through a blur function [26]. Another assumption is regarding the relationship between the HRMS and PAN (RHP): The spatial information of the HRMS image is contained in the PAN image, which can be expressed through the measurement of the gradient field [3, 27]. These two assumptions correspond with the goal of retaining the spectral and spatial fidelity, respectively. The conditional random fields (CRFs) consider the local spatial relationship and generate the desired sequence with the maximum likelihood, which can be used to model different types of prior information. Thus, CRFs are highly suitable for modeling the two assumptions.

In this paper, according to the above analysis, an efficient pansharpening model based on the CRFs is proposed, which can successfully balance the spatial sharpness, spectral fidelity, and efficiency, and can thus obtain high-quality fusion results with light computational demand. In the model, a state feature function is designed to satisfy the RHU assumption. To mimic the blur function, a new filter-acquisition algorithm is presented in the unified model. In addition, a transition feature function is defined to satisfy the RHP assumption in the model, and a total variance term is designed to assure sparsity in the gradient domain to reduce artifacts. To improve the efficiency, the MS image is converted into IHS components in the model, and the desired intensity component (I) is obtained by solving the model with an alternating direction method of multipliers (ADMM) [1, 3, 28]. Finally, a new inverse IHS transform method is presented to convert an IHS space into the original space by considering a different injection coefficient for each satellite dataset. Both reduced- and full-scale experiments were conducted in this study. Compared with both classical and recent methods, the proposed method achieves a high quality and high efficiency. The contributions of this study are as follows:

- 1) A new pansharpening model based on CRFs that can acquire high-quality fusion results with a low computational burden is presented.
- 2) A state feature function is designed to satisfy the RHU assumption of the model, and a new filter-acquisition method is simultaneously developed to obtain a proper filter, thereby retaining a higher spectral fidelity.
- 3) A transition feature function is proposed to satisfy the RHP assumption in the model and retain the spatial fidelity, and a total variance term is built to reduce artifacts.

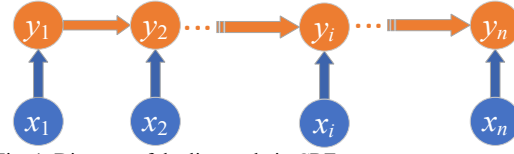


Fig. 1. Diagram of the linear-chain CRFs.

## 2. Basic CRFs Model

Let  $\mathbf{x} = (x_1, x_2, \dots, x_n)$  be a random sequence, and  $\mathbf{y} = (y_1, y_2, \dots, y_n)$  be the vector to be predicted. In general, we assume that  $\mathbf{x}$  and  $\mathbf{y}$  have the same graph structure. A graph of basic linear-chain conditional random fields (CRFs) is shown in Fig. 1. Let  $G = (V, E)$  be the graph, where  $V$  represents the vertices, and  $E$  represents the edges. The vertices include  $x_i$  and  $y_i$ , and the edges indicate the state transition between  $y_i$  and  $y_{i-1}$ . The linear-chain CRFs have a distribution  $P(\mathbf{y} | \mathbf{x})$ , which takes the following form [29]:

$$P(\mathbf{y} | \mathbf{x}) = \frac{1}{Z(\mathbf{x})} \exp(T + S), \quad (1)$$

where

$$T = \sum_{i,k} \lambda_k t_k(y_{i-1}, y_i, x, i), \quad (2)$$

$$S = \sum_{i,l} u_l s_l(y_i, x, i), \quad (3)$$

$$Z(\mathbf{x}) = \sum_{\mathbf{y}} \exp(S + T). \quad (4)$$

Here,  $t_k$  is the  $k$ th transition function and  $\lambda_k$  is the corresponding weight parameter. In addition,  $s_l$  is the  $l$ th state function, and  $u_l$  is the corresponding weight parameter, where  $i$  represents the position in a sequence. Moreover,  $Z(\mathbf{x})$  is a normalization factor and can be considered as a constant;  $T$  is a feature function applied to the edges, which is also called a transition feature because it relies on the current and previous states; and  $S$  is the feature function on the vertices, which is also called a state feature, and relies on the current state. The goal of the CRFs is to maximize the probability distribution  $P(\mathbf{y} | \mathbf{x})$  and obtain the output sequence  $\mathbf{y}$ . Equations (1)–(4) are the basic forms of linear-chain CRFs, which indicate the probability of output sequence  $\mathbf{y}$  under the condition of input sequence  $\mathbf{x}$ .

## 3. Proposed Method

In this section, we present a new CRF-based model to satisfy the two assumptions in remote sensing image fusion. A flowchart of the proposed fusion method is shown in Fig. 2. The intensity (I) component of the UPMS

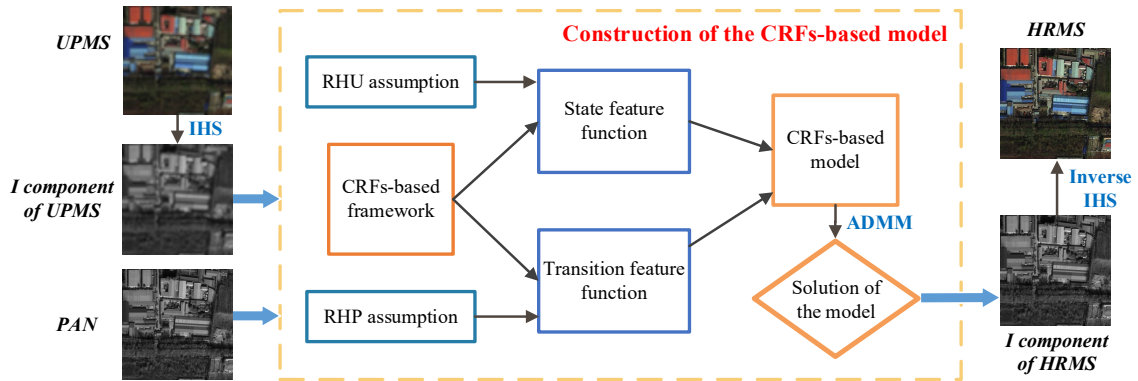


Fig. 2. Flowchart of the CRFs based pansharpening method. The UPMS and HRMS images are shown in RGB bands for better visual effect.

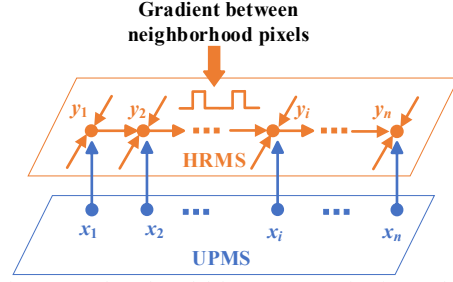


Fig. 3. CRF-based model for remote sensing image fusion.

image and the PAN image are used as inputs. The designed CRF-based model is shown in the orange rectangle. First, according to the assumptions of RHR and RHP, a state feature function and a transition feature function are designed to construct the CRF-based model. Second, the I component of the HRMS image is obtained by solving this model. Finally, the HRMS image is acquired through the inverse IHS transform of the result obtained from the previous step.

The construction of the CRF-based model is our main concern in this study. For consistency between the HRMS and UPMS images in spectral terms and coherence between the HRMS and PAN images in spatial terms, a homogeneous CRF prior that uses a global parameter is preferred in this study. This can also significantly reduce the complexity of the parameter estimation [22]. Taking the pixels  $x_i$  in the UPMS as the observation sequence, and the pixels  $y_i$  in the HRMS as the output sequence, the CRF-based model is shown in Fig. 3. The values of the HRMS pixels depend on the pixels in the UPMS, and the state feature function with a single non-negative parameter  $u$  can be defined as follows:

$$S = \sum_i u s(y_i, x_i). \quad (5)$$

The transition feature function is defined in the gradient domain, which involves the pixels  $y_i$  and the adjacent pixels in the neighborhood  $\Omega$ . Adopting a global non-negative parameter  $\lambda$ , the transition feature function for the pixel  $y_i$  and its nearby pixels  $y_j$  is defined as follows:

$$T = \sum_{i,j \in \Omega} \lambda t(y_i, y_j). \quad (6)$$

### 3.1 Modeling for RHU assumption

The ideal HRMS image should maintain spectral fidelity with the UPMS image. Therefore, after applying a low-pass filter  $H^b$  on the  $b$ th band of the HRMS image ( $M_{HR}^b$ ), the low-resolution version of  $M_{HR}^b$  should coincide with the  $b$ th band of the UPMS image ( $M_{UP}^b$ ), which can be expressed by the following:

$$M_{UP}^b = H^b M_{HR}^b + \varepsilon^b, \quad (7)$$

where  $\varepsilon^b$  is noise or an error. Estimating a proper low-pass filter  $H^b$  is essential to satisfy the RHU assumption. In general, the frequency response of the filter should be a Gaussian-like shape and match the modulation transfer function (MTF) of the MS sensor [26]. The MTF is the frequency domain transform of the sensor's point spread function (PSF). However, there are still certain limitations. On the one hand, modeling the MTF as a Gaussian filter is occasionally not confirmed in practice. On the other hand, an accurate gain of the Nyquist frequency of the MTF is often difficult to acquire and may change with the aging process of the device [30]. Furthermore, the separated filtering process in each band ignores the intrinsic correlation between MS bands, which may lead to a worse performance [31]. Thus, to maintain the consistency of the injection details between each band, the assumption that a unified filter  $H$  affects all of the MS bands is reasonable [31]. In this study, to reduce the computational dimensions from 3D to 2D as well as maintain the spectral information, the MS image is transformed into an IHS space and only the intensity component (I) is altered. Then, applying the proper filter  $H$  on the I component of the HRMS image ( $I_{HR}$ ), the low-resolution version of  $I_{HR}$  should be close to the I component of the UPMS image ( $I_{UP}$ ), which can be expressed as follows:

$$\mathbf{I}_{UP} = \mathbf{H}\mathbf{I}_{HR} + \varepsilon_1, \quad (8)$$

where  $\varepsilon_1$  is the error, and the I component can be obtained by the following [26]:

$$\mathbf{I}_x = \sum_{b=1}^N \alpha^b \mathbf{M}_x^b, \quad (9)$$

where  $N$  is the total bands of MS image,  $\alpha^b$  is the  $b$ th band weight coefficient which is set as an average value of each band generally. The subscript  $x$  represents either UP or HR symbol.

In (8), if the  $\mathbf{I}_{HR}$  is obtained, the final HRMS image can be obtained after inversely transforming the IHS space. In addition, the value of  $\varepsilon_1$ , which indicates the error between  $\mathbf{I}_{UP}$  and  $\mathbf{H}\mathbf{I}_{HR}$ , should be as small as possible in line with the RHU assumption. According to the CRFs model in (1), the pixels in  $\mathbf{I}_{UP}$  are then taken as observation sequence  $\mathbf{x}$ , and the pixels in  $\mathbf{I}_{HR}$  are taken as the output sequence  $\mathbf{y}$ . Maximizing the probability of  $P(\mathbf{y} | \mathbf{x})$  is equal to maximizing  $P(\mathbf{I}_{HR} | \mathbf{I}_{UP})$ . Thus, a state feature function  $S$  in (5) that needs to be maximized is designed as follows:

$$S = -\frac{u}{2} \|\mathbf{I}_{UP} - \mathbf{H}\mathbf{I}_{HR}\|^2. \quad (10)$$

Maximizing function (10) can ensure that the spectral distortion is minimized. According to function (10), if  $\mathbf{I}_{UP}$  and  $\mathbf{I}_{HR}$  are known, then  $\mathbf{H}$  can be obtained by solving (10), and thus we can obtain the filter  $\mathbf{H}$  in a unified model. The filter should be Gaussian-like and smooth [30], and thus we add a regular term to ensure the smooth transition in (10) and constrain the summation of elements in the filter to 1. The state feature function  $S$  with the filter-acquisition method can then be redefined as follows:

$$S = -\left(\frac{u}{2} \|\mathbf{I}_{UP} - \mathbf{H}\mathbf{I}_{HR}\|^2 + \frac{\gamma}{2} \|\nabla \mathbf{H}\|^2\right) \quad s.t. \quad \mathbf{H}^T \mathbf{1} = 1, \quad (11)$$

where  $\gamma$  is a penalty parameter. The filter  $\mathbf{H}$  is normalized to fit the constraint, that is,  $\mathbf{H}^T \mathbf{1} = 1$ . In addition,  $\nabla$  represents the gradient operator. The Laplacian operator is isotropic and can obtain sharpened edges and lines of any direction. Therefore, the classical Laplacian operator is introduced in this study to achieve horizontal and vertical gradients.

By iteratively calculating  $\mathbf{I}_{HR}$  and  $\mathbf{H}$  through the ADMM algorithm,  $\mathbf{I}_{HR}$  and  $\mathbf{H}$  can be simultaneously obtained. In this study, the initial  $\mathbf{H}$  is set as a Gaussian filter and is then optimized by model (11).

### 3.2 Modeling for RHP assumption

The RHP assumption assumes that the gradient of an HRMS image should follow the gradient trend in a PAN image. The gradient of an image depends on the current and adjacent pixels, which is in accordance with the transition feature function  $T$  that relies on the current and adjacent vertices. To acquire a gradient map of the image, the Laplacian operator is utilized in this study. In the IHS space, the gradient of the intensity component should also

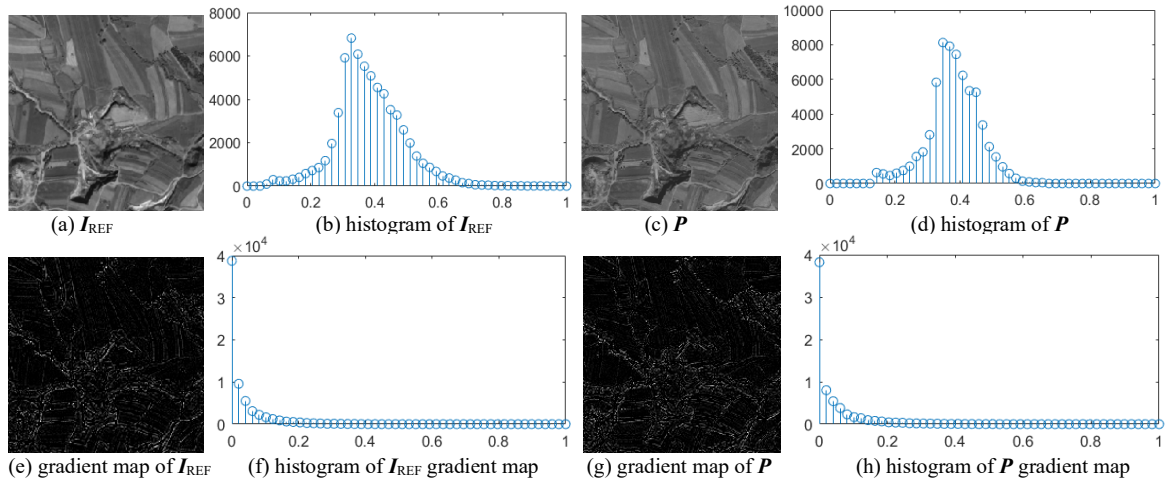


Fig. 4. The comparison of the  $\mathbf{I}_{REF}$  and PAN image as well as the comparison of their gradient maps.

be consistent with that of the PAN image. When taking a remote sensing image from the IKONOS dataset, for example, a comparison of the I component of the reference image ( $\mathbf{I}_{\text{REF}}$ ) and the PAN images, as well as their gradient maps, are shown in Fig. 4. We can see from (a) and (c) and their corresponding histograms in (b) and (d) that there is significant difference in information between  $\mathbf{I}_{\text{REF}}$  and  $\mathbf{P}$ . Their gradient maps, however, are extremely close to each other, which can be seen from (e) and (g) and their corresponding histograms (f) and (h). Thus, to minimize the difference between the gradient of the PAN image and that of the  $\mathbf{I}_{\text{HR}}$  image, we construct a transition feature function  $T$  in (6) based on the CRF theory:

$$T = -\frac{\lambda}{2} \|\nabla \mathbf{P} - \nabla \mathbf{I}_{\text{HR}}\|^2. \quad (12)$$

Maximizing formula (12) can assure the spatial fidelity of a fused image.

In (7) and (8), the noise  $\varepsilon^b$  or  $\varepsilon_l$  should be zero under an ideal situation; however, the artifacts are inevitably generated during the detail injection process. The total variation is an effective constraint that can be applied in image denoising, deblurring, and fusion, as examples [32]. Hence, to ensure the sparsity in the gradient domain, the total variation is employed as a regular term in the CRFs model, and we construct a new transition feature function:

$$T = -\left(\frac{\lambda}{2} \|\nabla \mathbf{P} - \nabla \mathbf{I}_{\text{HR}}\|^2 + \beta \|\nabla \mathbf{I}_{\text{HR}}\|_1\right). \quad (13)$$

The first term is the data-fitting term, and the second term is the total variation regular term. Here,  $\beta$  is a penalty parameter.

### 3.3 CRF-based model

Substituting (11) and (13) into (1), the proposed CRF-based model can be expressed as follows:

$$P(\mathbf{I}_{\text{HR}} | \mathbf{I}_{\text{UP}}, \mathbf{P}) = \frac{1}{Z} \exp \left( -\left(\frac{u}{2} \|\mathbf{I}_{\text{UP}} - \mathbf{H}\mathbf{I}_{\text{HR}}\|^2 + \frac{\gamma}{2} \|\nabla \mathbf{H}\|^2\right) - \left(\frac{\lambda}{2} \|\nabla \mathbf{P} - \nabla \mathbf{I}_{\text{HR}}\|^2 + \beta \|\nabla \mathbf{I}_{\text{HR}}\|_1\right) \right) \quad s.t. \mathbf{H}^T \mathbf{1} = 1, \quad (14)$$

where  $Z$  can be treated as a constant. To maximize the probability in (14), the monotonic logarithm is applied to simplify the calculation. Maximizing  $P(\mathbf{I}_{\text{HR}} | \mathbf{I}_{\text{UP}})$  is then equal to minimizing the energy function  $E(\mathbf{I}_{\text{HR}})$  as follows:

$$E(\mathbf{I}_{\text{HR}}) = \frac{u}{2} \|\mathbf{I}_{\text{UP}} - \mathbf{H}\mathbf{I}_{\text{HR}}\|^2 + \frac{\gamma}{2} \|\nabla \mathbf{H}\|^2 + \frac{\lambda}{2} \|\nabla \mathbf{P} - \nabla \mathbf{I}_{\text{HR}}\|^2 + \beta \|\nabla \mathbf{I}_{\text{HR}}\|_1 \quad s.t. \mathbf{H}^T \mathbf{1} = 1. \quad (15)$$

To reduce the regularization parameters, dividing  $u$  on both sides of the equation, and taking  $E(\mathbf{I}_{\text{HR}}) = E(\mathbf{I}_{\text{HR}})/u$ ,  $\lambda = \lambda/u$ ,  $\gamma = \gamma/u$ , and  $\beta = \beta/u$ , respectively, (15) becomes the following:

$$E(\mathbf{I}_{\text{HR}}) = \frac{1}{2} \|\mathbf{I}_{\text{UP}} - \mathbf{H}\mathbf{I}_{\text{HR}}\|^2 + \frac{\gamma}{2} \|\nabla \mathbf{H}\|^2 + \frac{\lambda}{2} \|\nabla \mathbf{P} - \nabla \mathbf{I}_{\text{HR}}\|^2 + \beta \|\nabla \mathbf{I}_{\text{HR}}\|_1 \quad s.t. \mathbf{H}^T \mathbf{1} = 1 \quad (16)$$

The estimation of  $\mathbf{I}_{\text{HR}}$  can therefore be obtained through the following equation:

$$\hat{\mathbf{I}}_{\text{HR}} = \arg \min_{\mathbf{I}_{\text{HR}}} (E(\mathbf{I}_{\text{HR}})), \quad (17)$$

where  $\hat{\mathbf{I}}_{\text{HR}}$  represents the estimated  $\mathbf{I}_{\text{HR}}$ .

## 4. Solution of CRF-Based Model

In this section, we present the solution of the proposed CRF-based model in formula (17). By replacing  $\nabla \mathbf{I}_{\text{HR}}$  with  $\mathbf{F}$ , the augmented Lagrangian function of (17) can be defined as:

$$E(\mathbf{I}_{\text{HR}}, \mathbf{H}) = \frac{1}{2} \|\mathbf{I}_{\text{UP}} - \mathbf{H}\mathbf{I}_{\text{HR}}\|^2 + \frac{\gamma}{2} \|\nabla \mathbf{H}\|^2 + \frac{\lambda}{2} \|\nabla \mathbf{P} - \nabla \mathbf{I}_{\text{HR}}\|^2 + \beta \|\mathbf{F}\|_1 + \langle \mathbf{M}, \mathbf{F} - \nabla \mathbf{I}_{\text{HR}} \rangle + \frac{\delta}{2} \|\mathbf{F} - \nabla \mathbf{I}_{\text{HR}}\|^2, \quad (18)$$

where  $\mathbf{M}$  is the Lagrangian multiplier, and  $\delta$  is the penalty parameter.

To solve model (18), the ADMM algorithm is employed owing to its successful application in various image processing areas [33]. By alternatively fixing one variant and updating another during the iterative step, the unknown variables  $\mathbf{I}_{\text{HR}}$ ,  $\mathbf{H}$ , and  $\mathbf{\Gamma}$  can be obtained when the algorithm satisfies the stop iteration condition.

#### 4.1 Updating $\mathbf{I}_{\text{HR}}$

The estimation of the  $\mathbf{I}_{\text{HR}}$  is a convex optimization problem. By fixing  $\mathbf{H}$  and  $\mathbf{\Gamma}$ , the  $\mathbf{I}_{\text{HR}}^{t+1}$  in the  $t + 1$  iteration step can be obtained by setting the partial derivation  $\partial E / \partial \mathbf{I}_{\text{HR}}$  to zero, which is given by the following:

$$\begin{aligned} \partial E / \partial \mathbf{I}_{\text{HR}} &= 0 \\ \Rightarrow (\mathbf{H}^T \mathbf{H} + \lambda \nabla^T \nabla + \delta \nabla^T \nabla) \mathbf{I}_{\text{HR}} &= \mathbf{H}^T \mathbf{I}_{\text{UP}} + \lambda \nabla^T \nabla \mathbf{P} + \nabla^T \mathbf{M} + \delta \nabla^T \mathbf{\Gamma}, \end{aligned} \quad (19)$$

For the existence of a gradient operator, the fast Fourier transform (FFT) is introduced to operate in the frequency domain to enhance the computational efficiency. Denoting FFT as  $\mathcal{F}(\bullet)$  and the inverse FFT (IFFT) as  $\mathcal{F}^{-1}(\bullet)$ ,  $\mathbf{I}_{\text{HR}}^{t+1}$  can be obtained by the following:

$$\mathbf{I}_{\text{HR}}^{t+1} = \mathcal{F}^{-1} \left( \frac{\mathcal{F}(\mathbf{H}^T \mathbf{I}_{\text{UP}} + \lambda \nabla^T \nabla \mathbf{P} + \nabla^T \mathbf{M} + \delta \nabla^T \mathbf{\Gamma})}{\mathcal{F}(\mathbf{H}^T \mathbf{H} + \lambda \nabla^T \nabla + \delta \nabla^T \nabla)} \right). \quad (20)$$

#### 4.2 Updating $\mathbf{H}$

The initial value of  $\mathbf{H}$  ( $\mathbf{H}^0$ ) was set as a Gaussian filter to accelerate the convergence. Filter  $\mathbf{H}$  can be obtained by minimizing the energy function (16). Fixing  $\mathbf{I}_{\text{HR}}^{t+1}$  and  $\mathbf{\Gamma}$ , the value of  $\mathbf{H}^{t+1}$  in the  $t + 1$  iteration step can be obtained by setting a partial derivation  $\partial E / \partial \mathbf{H}$  to 0, and a closed-form solution is acquired as follows:

$$\partial E / \partial \mathbf{H} = 0 \Rightarrow ((\mathbf{I}_{\text{HR}}^{t+1})^T \mathbf{I}_{\text{HR}}^{t+1} + \gamma \nabla^T \nabla) \mathbf{H} = (\mathbf{I}_{\text{HR}}^{t+1})^T \mathbf{I}_{\text{UP}}. \quad (21)$$

After applying FFT and IFFT, the  $\mathbf{H}^{t+1}$  can be obtained through the following:

$$\mathbf{H}^{t+1} = \mathcal{F}^{-1} \left( \frac{\mathcal{F}((\mathbf{I}_{\text{HR}}^{t+1})^T) \cdot \mathcal{F}(\mathbf{I}_{\text{UP}})}{\mathcal{F}((\mathbf{I}_{\text{HR}}^{t+1})^T) \cdot \mathcal{F}(\mathbf{I}_{\text{HR}}^{t+1}) + \gamma \mathcal{F}(\nabla^T) \cdot \mathcal{F}(\nabla)} \right), \quad (22)$$

where  $\cdot *$  represents a pixel-wise multiplication, and  $\mathbf{H}^{t+1}$  is normalized such that  $\mathbf{H}^T \mathbf{1} = 1$ .

#### 4.3 Updating $\mathbf{\Gamma}$

The updating of  $\mathbf{\Gamma}$  is unrelated to filter  $\mathbf{H}$ , and thus, by fixing  $\mathbf{I}_{\text{HR}}^{t+1}$ ,  $\mathbf{\Gamma}^{t+1}$  in step  $t + 1$  of the iteration is given by the following:

$$\mathbf{\Gamma}^{t+1} = \arg \min_{\mathbf{\Gamma}} \frac{\beta}{\delta} \|\mathbf{\Gamma}\|_1 + \frac{1}{2} \left\| \mathbf{\Gamma} - (\nabla \mathbf{I}_{\text{HR}}^{t+1} - \frac{1}{\delta} \mathbf{M}) \right\|^2 \quad (23)$$

Equation (23) can be solved using a fast soft-thresholding strategy as follows:

$$\mathbf{\Gamma}^{t+1} = \text{shrink}(\nabla \mathbf{I}_{\text{HR}}^{t+1} - \frac{1}{\delta} \mathbf{M}, \frac{\beta}{\delta}), \quad (24)$$

where

$$\text{shrink}(v, \tau) = \frac{v}{|v|} \cdot \max(|v| - \tau, 0). \quad (25)$$

#### 4.4 Updating $\mathbf{M}$

Using the gradient descent method, the Lagrangian multiplier  $\mathbf{M}$  can be easily updated as

$$\mathbf{M}^{t+1} = \mathbf{M}^t + \delta (\mathbf{\Gamma}^{t+1} - \nabla \mathbf{I}_{\text{HR}}^{t+1}). \quad (26)$$

In each iterative step, we set  $\delta = \rho \delta$ , where  $\rho > 1$ , to accelerate the convergence.

#### 4.5 Summary

In summary, after initializing the parameters, the model of (16) can be solved using the ADMM algorithm, and

---

**Algorithm 1** The algorithm of proposed model
 

---

**Input:** PAN image  $P$ , I component of a UPMS image ( $I_{UP}$ )

**Initialize:**  $\delta=1, M=1, t=0, I^0=0, \rho=1.01$ ,

$H^0$  is set as a Gaussian filter

**While do:**

1. Calculate  $I$  component of the HRMS image: update  $I_{HR}^{t+1}$  by Eq. (20),
2. Estimate the filter  $H$ : update  $H^{t+1}$  by Eq. (22) and normalize  $H^{t+1}$ ,
3. Update  $I^{t+1}$  by Eq. (24),
4. Update Lagrangian multiplier  $M^{t+1}$  by Eq. (26),
5. Update iteration:  $\delta = \rho\delta, t=t+1$

**Until:**  $\frac{\|I_{HR}^{t+1} - I_{HR}^t\|}{\|I_{HR}^t\|} < \zeta$

**Output:**  $\hat{I}_{HR}$  which is the estimated  $I$  component of the HRMS image.

---

$I_{HR}$ ,  $H$ , and  $I$  can be calculated by iteratively updating one argument and fixing the others, as shown in Algorithm 1.

If  $I_{HR}$  is estimated by the inverse transformation of the IHS, the fused image can be acquired. However, owing to the unique characteristics of satellites, the number of desired details extracted from a PAN image varies for the MS images from different satellites. A tuned coefficient  $k$  is therefore defined to determine the number of details, and the final HRMS image is defined as follows:

$$\hat{M}_{HR}^j = M_{UP}^j + k \frac{M_{UP}^j}{(1/N) \sum_{j=1}^N M_{UP}^j} * (\hat{I}_{HR} - I), \quad (27)$$

where  $\hat{M}_{HR}^j$  and  $M_{UP}^j$  represent the  $j$ th band of the estimated HRMS and UPMS images, respectively. The coefficient  $k$  can be obtained using a grid search method, which will be interpreted in the following experiments.

## 5. Experiments and Discussion

### 5.1 Setup of experiments

The remote sensing images were collected from the IKONOS, WorldView-2, and WorldView-3 satellite datasets. All of the datasets were downloaded from <http://www.kosmos-imagesmall.com>. The MS images include four bands, namely, red, green, blue, and near-infrared. Each dataset contains 60 groups of images. To evaluate the fusion quality accurately and objectively, both reduced-scale experiments with the reference images and full-scale experiments without the reference images are conducted. The sizes of the full-scale PAN and MS images were  $1024 \times 1024$  and  $256 \times 256$ , respectively. For the reduced-scale experiment, the PAN and MS images are degraded and decimated by a factor of 4 to  $256 \times 256$  and  $64 \times 64$ , respectively, and the original MS image is considered a reference image, which obeys Wald's protocol [34]. Some classic and recent methods have been compared, such as GSA [8], which is a CS-based method, and bilateral filter luminance proportional (BFLP) [35], which belongs to the MRA. The mixed methods include adaptive IHS with a multiscale guided filter (AIMG) [36], which combines CS with MRA, and full-scale regression-based injection coefficients (FSRIC) [37], which combines an MRA-based approach with a regression-based method. Model-based methods include nonlinear IHS (NIHS) [21], a matting model and multiscale transform (MMMT) [17], and fusion via compressive sensing FCS [1]. In addition, the interpolation method using a polynomial kernel (EXP) [38] is employed as an important baseline. All experiments were conducted using MATLAB (R2016a) on a laptop with 12 GB of RAM.

Some widely used quantitative metrics for reduced-scale fusion results are employed and are briefly introduced in the following:

- (1) The filtered spatial correlation coefficient (SCC) [4] and cross-correlation coefficient (CC) [4] account for the spatial similarity between the reference image and the fused image, and their ideal value is 1.



(2) The universal image quality index (UIQI) [39] combines three quality factors to assess the universal fusion quality. The optimum value is 1.

(3) The root mean square error (RMSE) [17] is widely used to measure the error between the fused and reference images, and a smaller value indicates a better result.

(4) The relative average spectral error RASE [40] is used to assess the average error in the spectral aspect, in which a smaller error indicates a better result.

(5) The SAM [26] reflects a spectral distortion, the ideal value of which is zero.

(6) ERGAS [26] is a global metric that measures both the spectral and spatial quality. Its optimal value is zero.

For a full-scale experimental evaluation, because there are no reference images, the popular metric quality with no reference (QNR) [41] is employed, which is obtained by the following:

$$QNR = (1 - D_\lambda)^p (1 - D_s)^q, \quad (28)$$

where  $D_\lambda$  indicates the spectral distortion and  $D_s$  indicates the spatial distortion. In addition,  $p$  and  $q$  are normally set to 1.

## 5.2 Parameter settings

The parameter  $\lambda$  in (16) indicates the weight of the spatial clarity. The expected fusion result will obtain a clear spatial structure while retaining the spectral fidelity. To determine the best value of  $\lambda$  for each satellite dataset, three quality metrics (SCC, SAM, and ERGAS) were employed to evaluate the spatial and spectral quality. The up-arrow next to the metric indicates that the ideal value is 1, and the down-arrow indicates that the ideal value is zero. For each satellite dataset, 10 images were randomly collected, and the average quality evaluation of 10 fused images under different  $\lambda$  values is shown in Fig. 5. For the IKONOS dataset, we can see that the best  $\lambda$  value is 2. At this value, all metrics can achieve the best results. For the WorldView-2 and WorldView-3 datasets,  $\lambda = 6$  will be optimal.

The parameter  $\beta$  in (16) represents the penalty for unnecessary details. Parameter  $k$  in (27) determines the details needed. Both parameters control the number of details injected into the UPMS; thus, both parameters can be adjusted at the same time through a grid search method. By fixing parameter  $\lambda$ ,  $\beta$  and  $k$  can be obtained by finding the best average fusion quality of 10 images for each dataset. Because each metric emphasizes one aspect of the fusion quality, and the optimal parameter values determined by SCC, SAM, and ERGAS may vary greatly, it is difficult to determine the parameter value by a single metric. To obtain the overall quality in a unified index, a comprehensive index is designed to represent the composite quality, which is expressed as  $Q = SCC +$

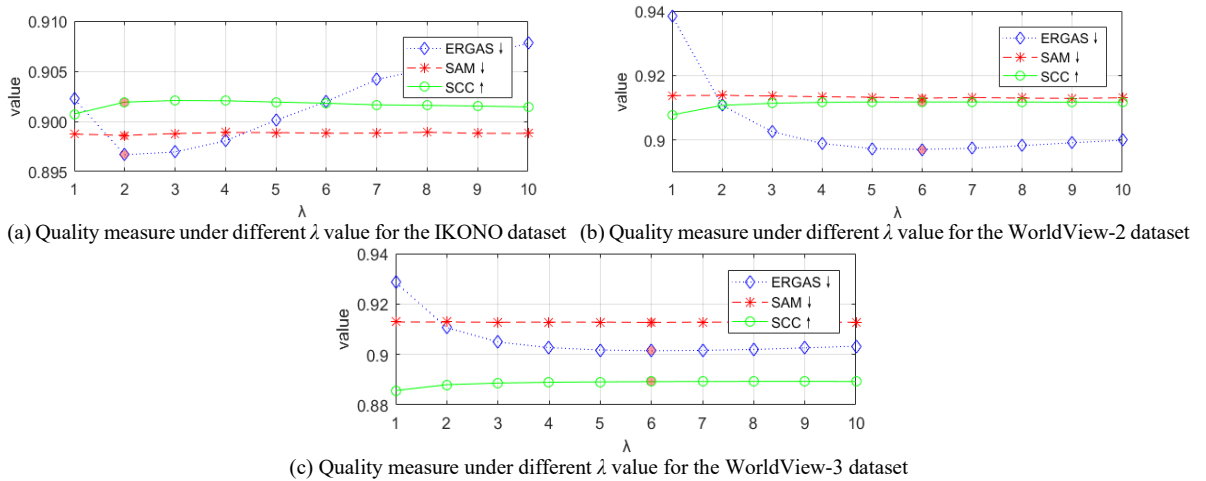


Fig. 5. Average quality measure under different  $\lambda$  value for each dataset. SAM and ERGAS values are normalized. The enhanced red points are the best results.

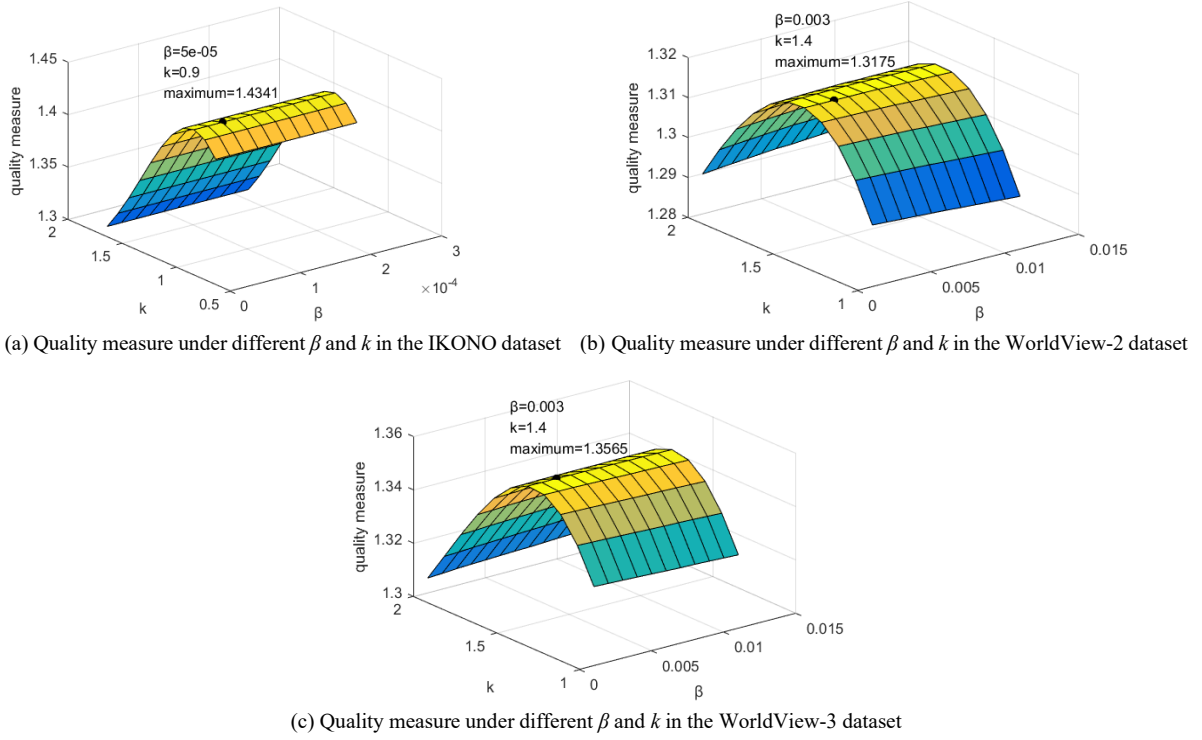


Fig. 6. Average quality measure under different  $\beta$  and  $k$  values for each dataset. The enhanced black points are the best results corresponding to the maximum quality values. The optimal  $\beta$  and  $k$  are displayed above the black points.

1/SAM+1/ERGAS. The maximum Q value indicates the best result. The results are shown in Fig. 6, which demonstrates that, by considering the three quality metrics together, the best results can be obtained when  $\beta = 5 \times 10^{-5}$ ,  $k = 0.9$  for the IKONOS dataset. For the WorldView-2 and WorldView-3 datasets,  $\beta = 0.003$  and  $k = 1.4$  will be optimal.

### 5.3 Iterative results

To verify the effectiveness of the proposed model as well as the ADMM algorithm, taking the WorldView-3 dataset as an example, the fusion results of each iterative step are shown in Fig. 7. We can see that as the iterations proceed, the ERGAS and SCC metrics improve, and the iteration converges at the fifth step. Although the SAM metric inevitably increases as more details are injected, the variation is not significant, and overall, the fusion quality improves after each step. In addition, owing to the proper settings of the initial parameters as well as the efficient solution for the model, the proposed method achieves a rapid convergence, which is a highly efficient advantage of our method.

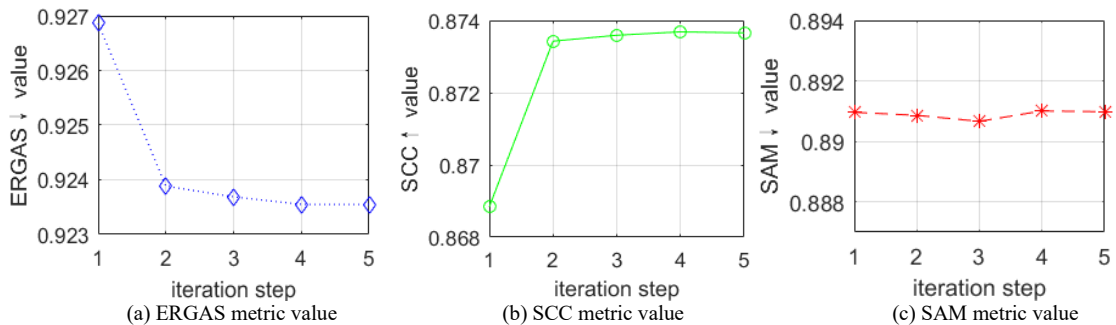


Fig. 7. Fusion quality of each iterative step. SAM and ERGAS values are normalized.

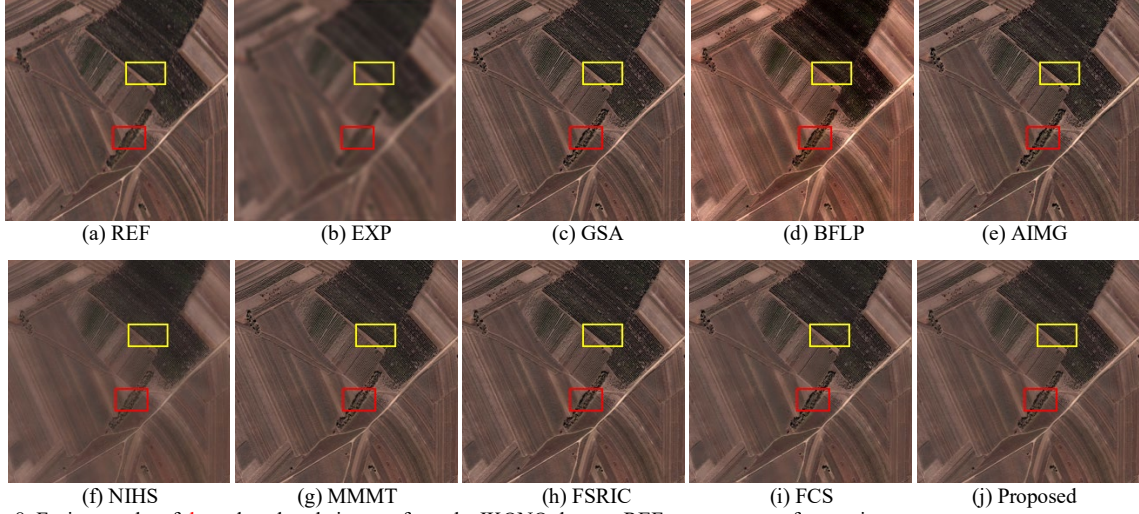


Fig. 8. Fusion results of the reduced scale images from the IKONO dataset. REF represents a reference image.

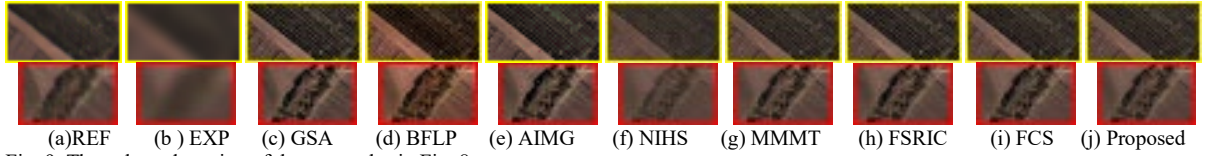


Fig. 9. The enlarged version of the rectangles in Fig. 8.

## 5.4 Reduced-scale experiments

### 5.4.1 IKONOS dataset

The fusion results of the remote sensing images from the IKONOS dataset are shown in Fig. 8. For better visual comparison, the enlarged yellow and red rectangles are shown in Fig. 9. REF represents the reference image. We can see from the images overall that the result of the BFLP apparently produces a spectral distortion as the color of the image is changed, and the result of the NIHS is blurred. The results of the AIMG and FCS introduce extra green artifacts, as shown by the enlarged yellow rectangles. Compared with the REF, an over-injection appeared in the GSA, MMTT, and FSRIC because the forest areas are excessively enhanced, as shown in the red enlarged

TABLE 1  
QUANTITATIVE EVALUATION OF THE FUSION RESULTS IN FIG. 8

Method	CC $\uparrow$	UIQI $\uparrow$	SCC $\uparrow$	RASE $\downarrow$	RMSE $\downarrow$	SAM $\downarrow$	ERGAS $\downarrow$
EXP-02[38]	0.8884	0.6838	0.6969	12.0668	0.0412	2.8315	2.9849
GSA-07[8]	0.9188	0.8132	0.8740	11.1154	0.0379	3.8394	2.9044
BFLP-14[35]	0.9391	0.7598	0.8895	32.9958	0.1126	8.0000	6.0717
AIMG-16[36]	0.9092	0.7948	0.8718	12.2390	0.0417	3.5690	3.1518
NIHS-16[21]	0.9226	0.7586	0.8640	10.4563	0.0357	2.9816	2.6273
MMMT-17[17]	0.9262	0.8105	0.8629	10.0107	0.0341	3.8117	2.5606
FSRIC-18[37]	0.9218	0.8195	0.8799	11.6196	0.0396	3.7886	3.0430
FCS-19[1]	0.9276	0.8159	0.8930	17.3250	0.0591	3.6769	4.7564
Proposed	<b>0.9476</b>	<b>0.8649</b>	<b>0.9168</b>	<b>8.3779</b>	<b>0.0286</b>	<b>2.8214</b>	<b>2.0941</b>

TABLE 2  
AVERAGE QUANTITATIVE EVALUATION ON THE REDUCED-SCALE IKONO DATASETS

Method	CC $\uparrow$	UIQI $\uparrow$	SCC $\uparrow$	RASE $\downarrow$	RMSE	SAM $\downarrow$	ERGAS $\downarrow$	Time(s)
EXP-02[38]	0.8657	0.7105	0.6131	17.7710	0.0542	3.9668	4.3752	-
GSA-07[8]	0.9322	0.8682	0.8505	13.2053	0.0404	5.2420	3.4488	0.07
BFLP-14[35]	0.9394	0.8146	0.8690	31.1994	0.0919	6.2721	6.7567	1.83
AIMG-16[36]	0.9195	0.8444	0.8639	15.9202	0.0485	4.3850	4.0040	0.37
NIHS-16[21]	0.9164	0.8001	0.8496	14.7254	0.0451	4.2731	3.6971	0.95
MMMT-17[17]	0.9349	0.8624	0.8458	12.6455	0.0388	4.9074	3.1605	18.27
FSRIC-18[37]	0.9364	0.8736	0.8637	13.8150	0.0422	5.0906	3.6385	0.17
FCS-19[1]	0.9372	0.8555	0.8754	23.2067	0.0718	6.7845	6.9195	28.60
Proposed	<b>0.9506</b>	<b>0.9004</b>	<b>0.8984</b>	<b>10.7597</b>	<b>0.0322</b>	<b>3.9482</b>	<b>2.7155</b>	0.16

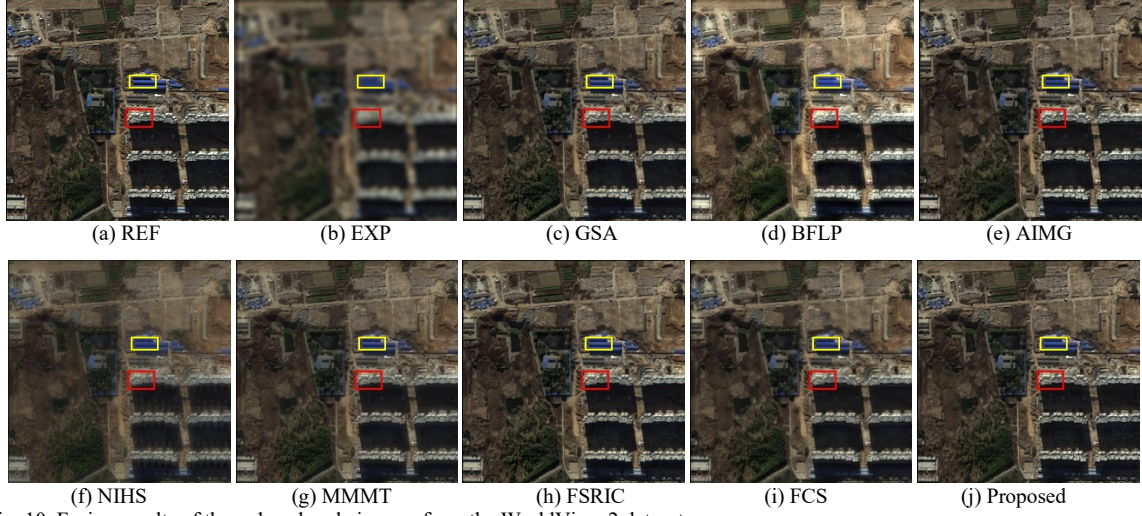


Fig. 10. Fusion results of the reduced scale images from the WorldView-2 dataset.

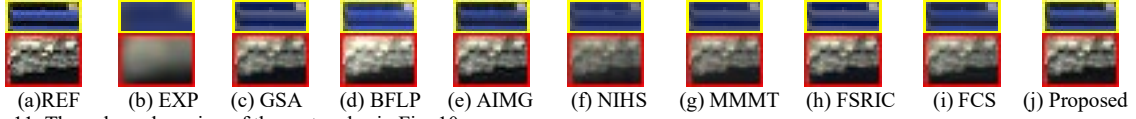


Fig. 11. The enlarged version of the rectangles in Fig. 10

rectangles. The result of the proposed method is closer to the reference image than those of the other methods.

An objective assessment of the fusion results in Fig. 8 is shown in Table 1, with the best results indicated in bold. The names of the comparison methods are followed by the years of publication. We can see that the objective evaluation is in accordance with the subjective evaluation, and the proposed method exceeds the other competing methods in all metrics. In addition, the average objective assessment of the fusion results on the 60 IKONO images is shown in Table 2. The average running time is provided in the last column and indicates that our approach is the best on all metrics, achieving a high efficiency.

#### 5.4.2 WorldView-2 dataset

The fusion results of the remote sensing image from the WorldView-2 dataset are shown in Fig. 10. The enlarged rectangles are shown in Fig. 11. We can see that the result of the BFLP suffers from an obvious spectral distortion

TABLE 3  
QUANTITATIVE EVALUATION OF THE FUSION RESULTS IN FIG. 10

Method	CC $\uparrow$	UIQI $\uparrow$	SCC $\uparrow$	RASE $\downarrow$	RMSE $\downarrow$	SAM $\downarrow$	ERGAS $\downarrow$
EXP-02[38]	0.7910	0.6079	0.5026	36.8846	0.0964	4.6118	9.3126
GSA-07[8]	0.9540	0.9198	0.9169	18.1593	0.0473	5.3057	4.6302
BFLP-14[35]	0.9358	0.8841	0.9165	40.5981	0.1057	5.3868	7.8427
AIMG-16[36]	0.9572	0.9324	0.9272	17.2825	0.0450	4.6726	4.3852
NIHS-16[21]	0.9052	0.7710	0.8756	28.2657	0.0736	4.7281	7.1491
MMMT-17[17]	0.9401	0.8719	0.9098	22.1744	0.0577	5.0875	5.6244
FSRIC-18[37]	0.9519	0.9142	0.9169	18.7710	0.0489	5.1744	4.7744
FCS-19[1]	0.9528	0.9341	0.9125	22.7244	0.0594	6.3776	6.1899
Proposed	<b>0.9614</b>	<b>0.9399</b>	<b>0.9369</b>	<b>16.4295</b>	<b>0.0428</b>	<b>4.5454</b>	<b>4.2109</b>

TABLE 4  
AVERAGE QUANTITATIVE EVALUATION ON THE REDUCED-SCALE WORLDVIEW-2 DATASETS

Method	CC $\uparrow$	UIQI $\uparrow$	SCC $\uparrow$	RASE $\downarrow$	RMSE $\downarrow$	SAM $\downarrow$	ERGAS $\downarrow$	Time(s)
EXP-02[38]	0.7810	0.5995	0.5261	31.4686	0.0791	4.7251	8.0794	-
GSA-07[8]	0.9337	0.8822	0.8917	17.7277	0.0443	5.5989	4.5463	0.08
BFLP-14[35]	0.9329	0.8641	0.9028	32.8643	0.0808	5.9662	6.6014	1.84
AIMG-16[36]	0.9256	0.8752	0.8890	18.8636	0.0462	4.8383	4.6854	0.50
NIHS-16[21]	0.8952	0.7447	0.8557	24.4192	0.0612	4.9002	6.2658	0.73
MMMT-17[17]	0.9217	0.8216	0.8872	20.6743	0.0518	5.2131	5.2581	14.53
FSRIC-18[37]	0.9335	0.8717	0.8916	17.9129	0.0448	5.4222	4.5853	0.17
FCS-19[1]	0.9352	0.8646	0.9016	20.9988	0.0523	6.2885	6.0537	29.2
Proposed	<b>0.9431</b>	<b>0.8956</b>	<b>0.9099</b>	<b>16.5355</b>	<b>0.0415</b>	<b>4.6894</b>	<b>4.2666</b>	0.16



because the image is significantly brighter than that of the REF. The result of the NIHS is blurred and the color is significantly darker than that of the REF. The enlarged yellow and red rectangles indicate that the results of the GSA, MMT, FSRIC, and FCS approaches produce spectral distortions because the roofs are darker than in the REF. The proposed method and the AIMG approach can both preserve the spectral and spatial information well, although it is difficult to detect the distinction visually. An objective evaluation of the images in Fig. 10 is tabulated in Table 3. We can see that the objective results coincide with the subjective evaluation, and that our method obtains the best results under all metrics. In addition, the average quantitative evaluation of the 60 WorldView-2 images is shown in Table 4, and the average running time is tabulated in the last column. We can see that our method also performs the best in all indexes, and the efficiency is higher than that of all other fusion methods with the exception of the GSA.

#### 5.4.3 WorldView-3 dataset

The fusion results of the image from the WorldView-3 dataset are shown in Fig. 12, with the enlarged rectangles shown in Fig. 13. As shown in the enlarged rectangles, the result of the BFLP appears to be affected by an over-injection because the image is excessively bright. The result of the NIHS shows an insufficient injection because the image is blurred and introduces some artifacts. Although the AIMG can preserve the spatial structure well, it also introduces other artifacts. The enlarged yellow rectangles indicate that the GSA, MMT, and FSRIC approaches suffer from spectral distortion because the edges of the red roof become white with the GSA approach, and the light green roofs become red with the MMT and FSRIC methods. The enlarged red rectangles indicate that the FCS introduces extra columns in the middle of the rectangle, the result of the MMT is slightly blurred, and the FSRIC suffers from obvious spectral distortion owing to the overall color change. The proposed method can preserve the spectral and spatial information better and is closer to REF.

To verify the effectiveness of the proposed method, a quantitative evaluation of the images in Fig. 12, and the average quantitative evaluation of the 60 WorldView-3 images, are shown in Tables 5 and 6, respectively. We can see that our approach achieves high efficiency and is the best among all metrics.

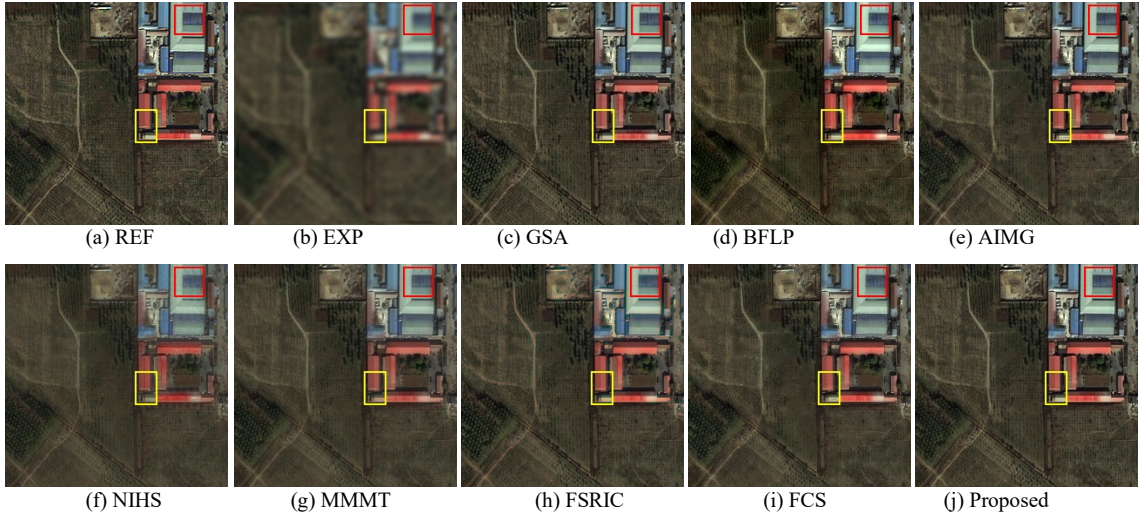


Fig. 12. Fusion results of the reduced scale images from the WorldView-3 dataset.

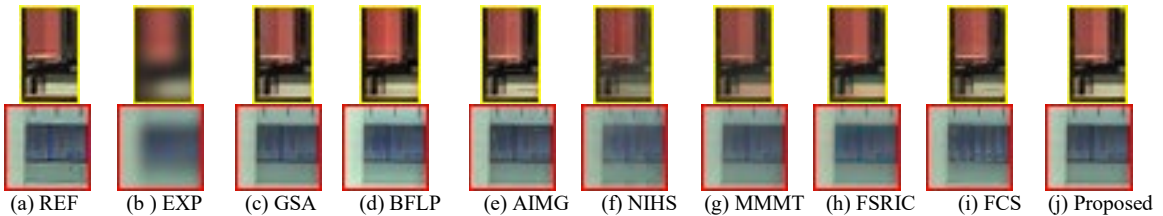


Fig. 13. The enlarged version of the rectangles in Fig. 12.

TABLE 5  
QUANTITATIVE EVALUATION OF THE FUSION RESULTS IN FIG. 12

Method	CC↑	UIQI↑	SCC↑	RASE↓	RMSE↓	SAM↓	ERGAS↓
EXP-02[38]	0.7635	0.5426	0.5156	27.3216	0.078	5.1108	6.8773
GSA-07[8]	0.9254	0.8358	0.8314	17.4033	0.0497	6.8081	4.1270
BFLP-14[35]	0.9388	0.8591	0.8683	19.2977	0.0551	7.5510	5.3359
AIMG-16[36]	0.9413	0.8630	0.8659	15.3945	0.0439	5.0931	3.7523
NIHS-16[21]	0.9058	0.7405	0.8245	21.2395	0.0606	5.2786	5.2519
MMMT-17[17]	0.9260	0.7997	0.8493	18.2887	0.0522	5.5202	4.4628
FSRIC-18[37]	0.9263	0.8342	0.8266	17.3874	0.0496	6.3936	4.1026
FCS-19[1]	0.9317	0.8368	0.8592	18.7790	0.0536	6.2894	4.9750
Proposed	<b>0.9447</b>	<b>0.8711</b>	<b>0.8732</b>	<b>14.9867</b>	<b>0.0425</b>	<b>5.0497</b>	<b>3.6451</b>

TABLE 6  
AVERAGE QUANTITATIVE EVALUATION ON THE WORLDVIEW-3 DATASETS

Method	CC↑	UIQI↑	SCC↑	RASE↓	RMSE↓	SAM↓	ERGAS↓	Time(s)
EXP-02[38]	0.7959	0.5737	0.4936	28.8917	0.0701	4.6726	7.2014	-
GSA-07[8]	0.9392	0.8826	0.8706	16.3735	0.0396	5.1878	4.0729	0.07
BFLP-14[35]	0.9320	0.8378	0.8715	37.1242	0.0844	6.1943	7.2227	1.74
AIMG-16[36]	0.9380	0.8738	0.8682	16.3595	0.0394	4.8494	4.0593	0.53
NIHS-16[21]	0.8958	0.7328	0.8216	22.9085	0.0555	4.7960	5.6964	0.71
MMMT-17[17]	0.9252	0.8200	0.8586	19.0581	0.0462	4.9772	4.7307	15.74
FSRIC-18[37]	0.9365	0.8749	0.8767	16.8704	0.0408	5.0924	4.1836	0.17
FCS-19[1]	0.9371	0.8685	0.8730	17.6705	0.0426	5.7001	4.5867	27.8
Proposed	<b>0.9424</b>	<b>0.8937</b>	<b>0.8818</b>	<b>15.9185</b>	<b>0.0384</b>	<b>4.5930</b>	<b>3.9742</b>	0.31

## 5.5 Full-scale experiments

Taking the IKONOS dataset as an example, the results of the full-scale experiment are shown in Fig. 14. The enlarged red and yellow rectangles are shown in Fig. 15. The result of the EXP was treated as the spectral reference. From the enlarged yellow rectangles, we can see that the black lines are thicker with the BFLP, AIMG, and FCS methods than those in the EXP. In addition, the GSA, BFLP, and AIMG introduce extra noise on the ground area. The NIHS suffers from spectral distortion because the black lines turn gray. The red rectangles indicate that, for the result of the forest area, the proposed method is closer to the EXP than the other methods, which are too bright

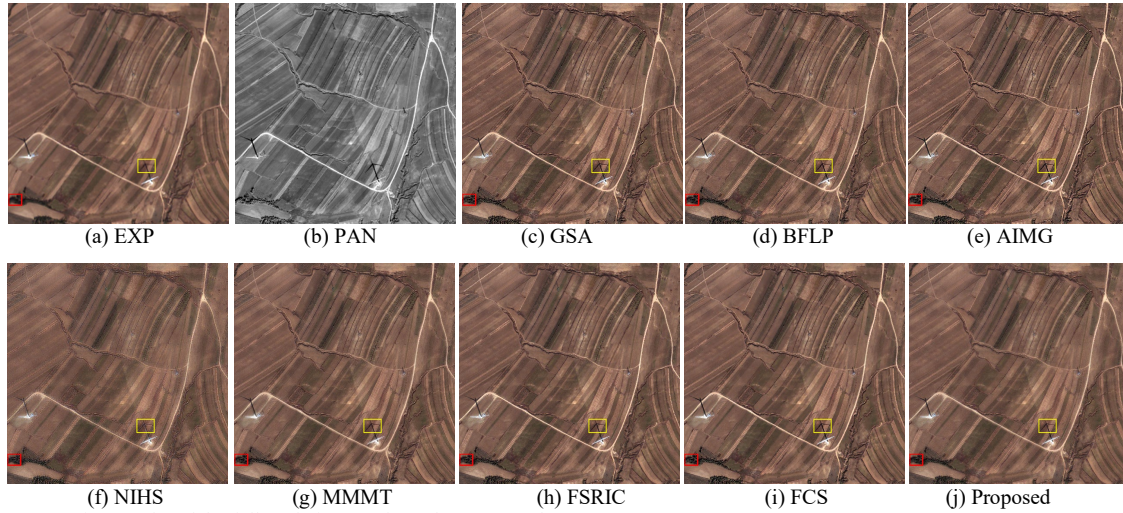


Fig. 14. Fusion results of the full-scale images from the IKONOS dataset.

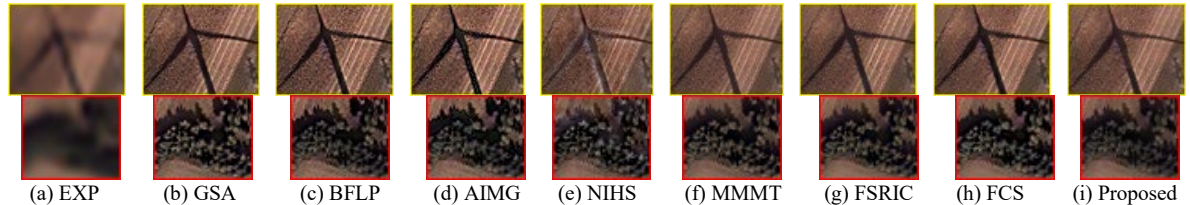


Fig. 15. The enlarged version of the rectangles in Fig. 14.

TABLE 7  
QUANTITATIVE EVALUATION OF THE FUSION RESULTS IN FIG. 14, AND AVERAGE QUANTITATIVE EVALUATION ON THE IKONO DATASETS

METHOD	RESULTS IN FIG. 14			AVERAGE			
	$D_i \downarrow$	$D_s \downarrow$	QNR $\uparrow$	$D_i \downarrow$	$D_s \downarrow$	QNR $\uparrow$	Time(s)
EXP-02[38]	0.0022	0.3163	0.6821	0.0023	0.3259	0.6725	0.32
GSA-07[8]	0.1524	0.1145	0.7506	0.1947	0.1728	0.6678	1.21
BFLP-14[35]	0.1600	0.1073	0.7499	0.2022	0.1607	0.6727	12.43
AIMG-16[36]	0.1516	0.1125	0.7530	0.1962	0.1709	0.6690	1.47
NIHS-16[21]	0.1010	0.0508	0.8534	0.1282	0.0745	0.8096	9.33
MMMT-17[17]	0.1201	<b>0.0290</b>	0.8543	0.1560	0.0658	0.7889	194.24
FSRIC-18[37]	0.1424	0.0803	0.7887	0.1856	0.1338	0.7068	1.00
FCS-19[1]	0.1704	0.0690	0.7723	0.1978	0.1106	0.7150	529.7
Proposed	<b>0.0866</b>	<u>0.0357</u>	<b>0.8808</b>	<b>0.1079</b>	<b>0.0562</b>	<b>0.8417</b>	4.60

compared to the EXP. The objective evaluation of the fusion results in Fig. 14 is shown in the left part of Table 7. We can see that our method performs better in all of the metrics except  $D_s$ , which is the second best (underlined). An evaluation of the average fusion quality of the 60 full-scale images in the IKONO dataset is shown in the right section of Table 7. We can therefore conclude that the proposed method can achieve the best results with relatively high efficiency.

### 5.6 Ablation study

The proposed model mainly consists of a state feature function, a transition feature function, and a total variation regularization term. The filter acquisition algorithm is integrated into the state feature function. A series of ablation experiments were then conducted. Denoting NSF, NTF, and NTV as models with no state feature function, no transition feature function, and no total-variation regular term, respectively, the components of the models after ablation are as shown in Table 8. Ten remote sensing images were randomly selected from the IKONO, WorldView-2, and WorldView-3 datasets. Taking a WorldView-2 image as an example, the fusion results are shown in Fig. 16. We can see from the figure that the result of the NSF method suffers from spectral distortion because there is no state feature function that provides a spectral constraint. The result of the NTF approach is blurred under the absence of a transfer feature function that provides spatial constraints. In addition, the results of the NTV and the proposed model make it difficult to determine the difference visually. The average objective evaluation of the 10 images is shown in Fig. 17. The figure is a radar map, and the metrics are normalized with optimal values of 1. A larger area indicates a better result. We can see from the map that the proposed model achieves the best result. The NTV method is slightly inferior, and the NTF and NSF methods become much worse, which indicates that the designed transition and state feature functions are of critical importance, and the total variation regularization term can also improve the overall performance.

To verify the effectiveness of the proposed iterative filter-acquisition method integrated in terms of the state

TABLE 8  
DETAILS OF THE MODELS AFTER ABLATION

Ablation model	State feature function in formula (11)	Transition feature function in formula (12)	Total variation
NSF	X	✓	✓
NTF	✓	X	✓
NTV	✓	✓	X
Proposed	✓	✓	✓

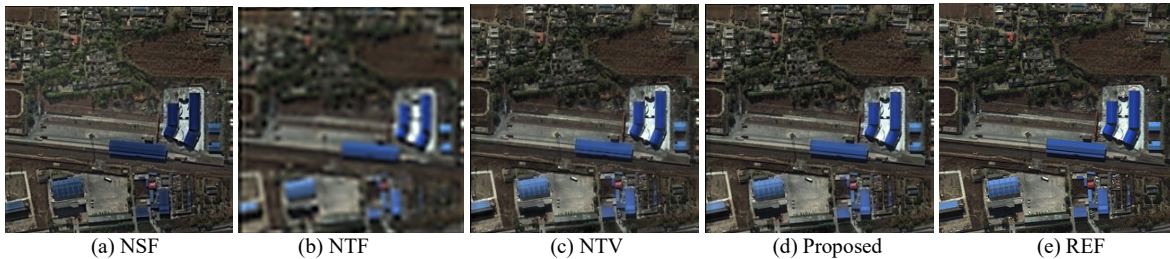


Fig.16. Comparison of the fusion results for different ablation models.



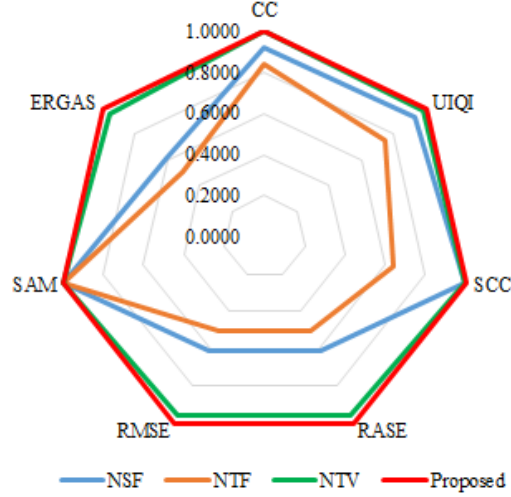


Fig.17. The radar map of the objective evaluation for the ablation models.

feature function, comparison experiments between the method using an MTF-matched filter (the MTF method) and the proposed method were conducted. Because  $HI_{HR}$  should be as close as possible to  $I_{UP}$ , the correlation between  $HI_{HR}$  and  $I_{UP}$  is calculated to represent the similarity between them. The results are shown in Fig. 18, which demonstrates that the correlations of both the MTF method and the proposed method increase as the iteration processes. Although the MTF method converged at the fourth step, which is faster than the proposed method that converges at the fifth step, the correlation of the proposed method is higher than that of the MTF method. This means that the filter acquired by the proposed method can better satisfy the first assumption. In addition, spectral quality metrics are employed to evaluate the spectral fidelity, such as SAM, which indicates a spectral distortion, RASE representing the average spectral error, and ERGAS representing the global quality. Ten remote sensing images were randomly collected, and the average quantitative results are shown in Fig. 19. We can see that the proposed method performs better in all metrics, thereby further verifying the effectiveness of the unified model, which is combined with the filter-acquisition method.

### 5.7 Efficiency analysis

To prove the performance of the proposed pansharpening method, the average running time of different methods for both reduced-scale and full-scale images is shown in the last column of Tables 2, 4, 6, and 7. Owing to the high efficiency of FFT and IFFT, the calculation cost of the proposed method is  $O(m \times n \log_2 n)$ , where  $m$  is the number of iterations and  $n$  is the number of pixels. With a proper initialization, our method has the property of rapid convergence; thus,  $m$  is usually small. MMT with NSST decomposition and FCS based on a sparse representation are computationally expensive. Although the classical CS-based method, GSA, achieves the fastest

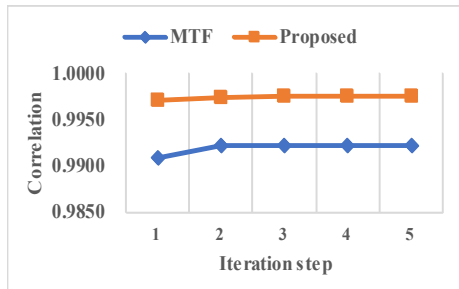


Fig. 18. The correlation between  $HI_{HR}$  and  $I_{UP}$  in each step. The blue line is the result of using an MTF-matched filter, and the orange line is the result of using a proposed filter. The images are from the WorldView-2 dataset.

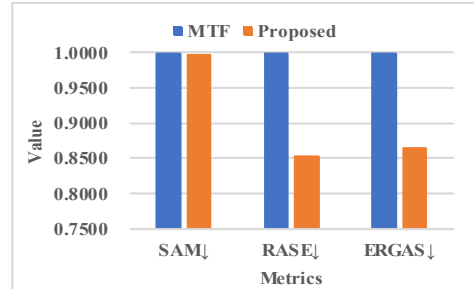


Fig. 19. Average quality evaluation of the MTF and proposed method. The metric values are normalized by dividing the maximum one. The down-arrow means the ideal value is 0.



speed, the fusion quality is not satisfactory, as can be seen from the above experiments. The proposed method shows a preferable performance in terms of both efficiency and quality.

## 6. Conclusion

In this paper, two assumptions corresponding to spectral and spatial preservation were proposed for remote sensing image fusion, and a model based on the CRFs was constructed to satisfy both assumptions. In the model, a state feature function was designed to assure spectral fidelity in line with the first assumption. Owing to the absence of a blur function in the first assumption, a filter-acquisition method was developed in the unified model. Meanwhile, a transition feature function was devised according to the second assumption to ensure spatial fidelity. Owing to the sparsity of the gradient domain, a total variation regular term was built to reduce artifacts.

For the model solution, an efficient optimization algorithm was devised, which employs FFT and IFFT in the ADMM algorithm to iteratively solve the augmented Lagrangian function. Experiments were conducted on the parameter adjustments and iteration processes, which can further verify the effectiveness of the proposed method. In addition, extensive experiments were conducted on both reduced- and full-scale images from the IKONOS, WorldView-2, and WorldView-3 datasets. Compared with both classical and state-of-the-art methods, the results show that the proposed method can obtain the best fusion quality with high efficiency.

## References

- [1] M. Ghahremani, Y. Liu, P. Yuen, A. Behera, Remote sensing image fusion via compressive sensing, *ISPRS Journal of Photogrammetry and Remote Sensing*. 152 (2019) 34–48.
- [2] X. Wang, J. Li, J. Li, H. Yan, Multilevel similarity model for high-resolution remote sensing image registration, *Information Sciences*. 505 (2019) 294–305.
- [3] H. Li, X. He, Z. Yu, J. Luo, Noise-robust image fusion with low-rank sparse decomposition guided by external patch prior, *Information Sciences*. 523 (2020) 14–37.
- [4] T. Wang, F. Fang, F. Li, G. Zhang, High-Quality Bayesian Pansharpening, *IEEE Trans. on Image Process.* 28 (2019) 227–239.
- [5] Zhijun Wang, D. Ziou, C. Armenakis, D. Li, Qingquan Li, A comparative analysis of image fusion methods, *IEEE Trans. Geosci. Remote Sensing*. 43 (2005) 1391–1402.
- [6] J. A. Malpica, Hue adjustment to IHS pan-sharpened IKONOS imagery for vegetation enhancement. *IEEE Geoscience & Remote Sensing Letters*. 4(2007) 27–31.
- [7] C. A. Laben, B. V. Brower. Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening, U.S. Patent 6011875 A (2000).
- [8] B. Aiazzi, S. Baronti, M. Selva, Improving component substitution pansharpening through multivariate regression of MS+Pan data. *IEEE Trans. Geosci. Remote Sensing*. 45 (2007) 3230–3239.
- [9] H. Ghassemian, A review of remote sensing image fusion methods, *Information Fusion*. 32 (2016) 75–89.
- [10] Y. Xu, C. Yang, B. Sun, X. Yan, M. Chen, A novel multi-scale fusion framework for detail-preserving low-light image enhancement, *Information Sciences*. 548 (2021) 378–397.
- [11] S. G. Mallat, A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 11 (1989) 674–693.
- [12] G. P. Nason, B. W. Silverman. The Stationary Wavelet Transform and Some Statistical Applications. *Wavelets and Statistics*. New York, NY: Springer New York. (1995) 281–299.
- [13] P. J. Burt, E. H. Adelson, The Laplacian Pyramid as a Compact Image Code. *IEEE Transactions on Communications*. 31 (1983) 532–540.
- [14] M. Yin, W. Liu, X. Zhao, et al, A novel image fusion algorithm based on nonsubsampling shearlet transform. *Optik International Journal for Light & Electron Optics*. 125 (2014) 2274–2282.
- [15] L.-J. Deng, G. Vivone, W. Guo, M. Dalla Mura, J. Chanussot, A Variational Pansharpening Approach Based on Reproducible Kernel Hilbert Space and Heaviside Function, *IEEE Trans. on Image Process.* 27 (2018) 4330–4344.
- [16] Xudong Kang, Shutao Li, J.A. Benediktsson, Pansharpening With Matting Model, *IEEE Trans. Geosci. Remote Sensing*. 52 (2014) 5088–5099.
- [17] Y. Yang, W. Wan, S. Huang, P. Lin, Y. Que, A Novel Pan-Sharpener Framework Based on Matting Model and Multiscale Transform, *Remote Sensing*. 9 (2017) 391.
- [18] Y. Yang, L. Wu, S. Huang, W. Wan, Y. Que, Remote Sensing Image Fusion Based on Adaptively Weighted Joint Detail Injection, *IEEE Access*. 6 (2018) 6849–6864.
- [19] Z. Zhu, H. Yin, Y. Chai, Y. Li, G. Qi, A novel multi-modality image fusion method based on image decomposition and sparse representation, *Information Sciences*. 432 (2018) 516–529.
- [20] G. Vivone, R. Restaino, J. Chanussot, Full Scale Regression-Based Injection Coefficients for Panchromatic Sharpening, *IEEE Trans. on Image Process.* 27 (2018) 3418–3431.
- [21] M. Ghahremani, H. Ghassemian, Nonlinear IHS: A Promising Method for Pan-Sharpener, *IEEE Geosci. Remote Sensing Lett.* 13 (2016)

- 1606–1610.
- [22] G. Khademi, H. Ghassemian, Incorporating an Adaptive Image Prior Model Into Bayesian Fusion of Multispectral and Panchromatic Images, *IEEE Geosci. Remote Sensing Lett.* 15 (2018) 917–921.
  - [23] K.P. Upla, M.V. Joshi, P.P. Gajjar, An Edge Preserving Multiresolution Fusion: Use of Contourlet Transform and MRF Prior, *IEEE Trans. Geosci. Remote Sensing.* 53 (2015) 3210–3220.
  - [24] J Yang, X Fu, Y Hu, et al, PanNet: A Deep Network Architecture for Pan-Sharpning, 2017 IEEE International Conference on Computer Vision (ICCV). IEEE. (2017) 5449–5457.
  - [25] Y. Xing, M. Wang, S. Yang, L. Jiao, Pan-sharpening via deep metric learning, *ISPRS Journal of Photogrammetry and Remote Sensing.* 145 (2018) 165–183.
  - [26] G. Vivone, L. Alparone, J. Chanussot, M. Dalla Mura, A. Garzelli, G.A. Licciardi, R. Restaino, L. Wald, A Critical Comparison Among Pansharpening Algorithms, *IEEE Trans. Geosci. Remote Sensing.* 53 (2015) 2565–2586.
  - [27] X. Wang, S. Bai, Z. Li, Y. Sui, J. Tao, The PAN and MS image fusion algorithm based on adaptive guided filtering and gradient information regulation, *Information Sciences.* 545 (2021) 381–402.
  - [28] L. Xu, D. Zhao, Y. Yan, S. Kwong, J. Chen, L.-Y. Duan, IDeRs: Iterative dehazing method for single remote sensing image, *Information Sciences.* 489 (2019) 50–62.
  - [29] J. Lafferty, A. McCallum, F.C.N. Pereira, Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data, in *Proc. ICML.* (2001) 282–289.
  - [30] G. Vivone, P. Addesso, R. Restaino, M. Dalla Mura, J. Chanussot, Pansharpening Based on Deconvolution for Multiband Filter Estimation, *IEEE Trans. Geosci. Remote Sensing.* 57 (2019) 540–553.
  - [31] G. Vivone, M. Simoes, M. Dalla Mura, R. Restaino, J.M. Bioucas-Dias, G.A. Licciardi, J. Chanussot, Pansharpening Based on Semiblind Deconvolution, *IEEE Trans. Geosci. Remote Sensing.* 53 (2015) 1997–2010.
  - [32] X. He, L. Condat, J.M. Bioucas-Dias, J. Chanussot, J. Xia, A New Pansharpening Method Based on Spatial and Spectral Sparsity Priors, *IEEE Trans. on Image Process.* 23 (2014) 4160–4174.
  - [33] L.-J. Deng, M. Feng, X.-C. Tai, The fusion of panchromatic and multispectral remote sensing images via tensor-based sparse modeling and hyper-Laplacian prior, *Information Fusion.* 52 (2019) 76–89.
  - [34] L. Wald, T. Ranchin, M. Mangolini, Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images. *Photogrammetric Engineering and Remote Sensing.* 63 (1997) 691–699.
  - [35] N.H. Kaplan, I. Erer, Bilateral Filtering-Based Enhanced Pansharpening of Multispectral Satellite Images, *IEEE Geosci. Remote Sensing Lett.* 11 (2014) 1941–1945.
  - [36] Y. Yang, W. Wan, S. Huang, F. Yuan, S. Yang, Y. Que, Remote Sensing Image Fusion Based on Adaptive IHS and Multiscale Guided Filter, *IEEE Access.* 4 (2016) 4573–4582.
  - [37] G. Vivone, R. Restaino, J. Chanussot, Full Scale Regression-Based Injection Coefficients for Panchromatic Sharpning, *IEEE Trans. on Image Process.* 27 (2018) 3418–3431.
  - [38] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, Context-driven fusion of high spatial and spectral resolution images based on oversampled multiresolution analysis, *IEEE Trans. Geosci. Remote Sensing.* 40 (2002) 2300–2312.
  - [39] Z Wang, A. C. Bovik. A universal image quality index. *IEEE signal processing letters*, 9 (2002) 81–84.
  - [40] L. Wald, T. Ranchin, M. Mangolini, Fusion of high spatial and spectral resolution images: The ARSIS concept and its implementation. *Photogrammetric Engineering & Remote Sensing*, 66 (2000) 49–61.
  - [41] L. Alparone, B. Aiazzi, S. Baronti, A. Garzelli, F. Nencini, M. Selva, Multispectral and Panchromatic Data Fusion Assessment Without Reference, *Photogramm Eng Remote Sensing.* 74 (2008) 193–200.