An Efficient Pansharpening Approach Based on Texture Correction and Detail Refinement

Hangyuan Lu, Yong Yang, *Senior Member, IEEE*, Shuying Huang, *Member, IEEE*, and Wei Tu

***Abstract*—Pansharpening aims at fusing a multispectral (MS) image and panchromatic (PAN) image to obtain a high spatial resolution multispectral (HRMS) image. To obtain accurate details and reduce spectral distortion, this paper proposes an efficient pansharpening approach based on texture correction and detail refinement. First, a texture correction model is constructed based on spatial and spectral fidelity constraints, to obtain a texture image that is highly correlated with the MS image. Second, a detail acquisition model is proposed by the consecutive parameter regression to adaptively refine the extracted details. Finally, the extracted details are injected into the up-sampled MS image to obtain the fused HRMS image. Experimental results demonstrate that the proposed method can obtain the high-quality results with high efficiency.**

***Index Terms*—Pansharpening, texture correction, detail refinement, parameter regression.**

I. Introduction

Remote sensing refers to record the target radiation energy with a distance [1]. With the fast development of satellite technology, remote sensing images become much easier to be acquired. On the one hand, a high spatial resolution multispectral (HRMS) image plays a critical role in the fields of geographic information management on the earth and Mars exploration in outer space [2]–[4]. On the other hand, due to the hardware constraint, current satellite sensors cannot directly provide HRMS images [5]. Instead, they usually carry two kinds of sensors, which provide panchromatic (PAN) images and multispectral (MS) images, respectively. A PAN image has a higher spatial resolution but contains the lower spectral information, while an MS image has the opposite characteristics. Therefore, the fusion of these two kinds of complementary images by a certain fusion technology has become an effective way to obtain an HRMS image [6]. The fusion of MS and PAN images is also called pansharpening.

Traditional pansharpening methods include component substitute (CS) and multi-resolution analysis (MRA) approaches. The CS-based methods usually replace the spatial component of an MS image with the PAN image. Classical CS-based methods include intensity-hue-saturation (IHS), principal component analysis (PCA), Gram-Schmidt adaptive (GSA), and so on [7]. This kind of method usually can obtain high spatial quality with high efficiency, but the spectral quality is not satisfactory [1], [3]. The MRA-based methods decompose the original images into several scales and fuse the images at each scale. Popular MRA-based methods include wavelet transform, pyramid transform, contourlet transform, etc. [8]. Compared with CS-based methods, the MRA-based methods can improve the spectral quality, but the spatial quality and computational efficiency are usually not as good as the CS-based methods [9].

Recently, the model-based methods have attracted more and more attention. Ghahremani *et al.* [10] proposed a new fusion method based on compressive sensing (FCS). This method can improve the fusion quality to a certain extent, but requires relatively high computational cost. To get more accurate details for each band of the MS image, Garzelli *et al.* [11] proposed a model based on band-dependent spatial details (BDSD). Vivone *et al.* [12] improved it by proposing a robust BDSD (RBDSD) method, which can obtain better fusion results for the MS images with 8 bands. Although impressive fusion results have been achieved, the balance among spectral quality, spatial quality and efficiency needs to be further improved [6].

Deep learning-based methods have strong ability in feature learning and have developed quickly in recent years. For example, Jiang *et al.* [13] proposed a pansharpening method based on residual convolution neural network. To overcome the limitation of no reference image, Luo *et al.* [14] proposed an unsupervised iterative network based on PAN-guided strategy. To further exploit the rich spectral characteristics in MS image, Qu *et al.* [15] presented an unsupervised pansharpening network based on the self-attention mechanism. Although these methods can achieve impressive fusion results, they are usually computationally expensive and their generalization performance is usually unsatisfactory [16].

In this paper, to efficiently obtain the fusion results with high spatial and spectral quality, a new pansharpening method is proposed based on texture correction and detail refinement. In our method, a texture correction model is proposed to improve the correlation between source images. Besides, due to the diversity of spatial information in different MS bands, a detail acquisition model is designed through two consecutive parameter-regression models to refine the details. Compared with the traditional and state-of-the-art methods, the experimental results show that the proposed method can efficiently obtain high spatial and spectral quality. The main contributions of this work are as follows:

This work is supported by the National Natural Science Foundation of China (No.62072218 and No.61862030), by the Natural Science Foundation of Jiangxi Province (No.20192ACB20002 and No.20192ACBL21008), and by the Talent project of Jiangxi Thousand Talents Program (No. jxsq2019201056).

H. Lu is with the College of Information Engineering, Jinhua Polytechnic, Jinhua 321007, China (e-mail: lhyhziee@163.com).

Y. Yang, and W. Tu are with the School of Information Technology, Jiangxi University of Finance and Economics, Nanchang 330032, China (e-mail: greatyangy@126.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).

(1) A new pansharpening method is proposed based on texture correction and detail refinement, which can reduce spectral distortion and obtain accurate details for each band.

(2) A texture correction model is constructed by keeping both spectral and spatial fidelity to obtain a texture image that is highly correlated with the MS image.

(3) A detail acquisition model is designed by constructing two consecutive parameter regression models to obtain the precise details of each band.

II. Proposed Method

A. The proposed framework

First of all, for simplicity, we denote $\boldsymbol{P}\in\mathbb{R}^{U\times V}$ as a PAN image, $\boldsymbol{M}\in\mathbb{R}^{u\times v\times B}$ as an MS image, $\boldsymbol{M}_{\mathrm{UP}}\in\mathbb{R}^{U\times V\times B}$ as an up sampled MS image (UPMS), and $\boldsymbol{M}_{\mathrm{HR}}\in\mathbb{R}^{U\times V\times B}$ as an HRMS image. $\boldsymbol{I}_{\mathrm{UP}}\in\mathbb{R}^{U\times V}$ and $\boldsymbol{I}_{\mathrm{HR}}\in\mathbb{R}^{U\times V}$ are the intensity components of UPMS and HRMS images in IHS space, respectively. $\boldsymbol{T}\in\mathbb{R}^{U\times V}$ represents a texture image. Here B is the number of the bands, $U\times V$ denotes the size of the PAN image, and $u=U/4$, $v=V/4$.

The proposed method mainly consists of two key models including texture correction and detail acquisition, as shown in Fig.1. In the former model, two prior terms of the spatial and spectral fidelities are constructed to achieve a new texture image $\boldsymbol{T}$ which is a corrected version of $\boldsymbol{P}$. Here, $\boldsymbol{T}$ contains the texture information similar to $\boldsymbol{P}$, but is corrected to better match the spectral characteristics of $\boldsymbol{M}$. In the latter model, with the source images and $\boldsymbol{T}$, a detail acquisition model is designed to obtain accurate details for each band. Finally, the details are injected into the UPMS image by the injection model to obtain the fused HRMS image.

B. Texture correction model

The radiometric discrepancy between PAN and MS images results in different texture features between them, which leads to the low correlation between the source images and may cause spectral distortion in the fusion process [3], [17]. In order to reduce the spectral distortion, the texture correction of the PAN image is necessary to obtain a new texture image $\boldsymbol{T}$ that is highly correlated with the MS image without losing spatial information of the PAN image. Therefore, we construct a spectral fidelity constraint between $\boldsymbol{T}$ and MS image, and a spatial fidelity constraint between $\boldsymbol{T}$ and PAN image.

According to the prior knowledge of spectral consistency, an UPMS image can be considered as a low-resolution version of an HRMS image (L-HRMS), and a L-HRMS can be obtained by applying a low-pass filter with a Gaussian-like shape on the HRMS image [18]. Vivone *et al.*[19] argued that using a unified filter $\boldsymbol{H}$ on each band of the MS image can better preserve the coherence among the bands. Therefore, to make sure that $\boldsymbol{M}_{\mathrm{UP}}$ is close to $\boldsymbol{H}\boldsymbol{M}_{\mathrm{HR}}$, a spectral-fidelity function E_{spec} is designed as follows:

$$E_{\mathrm{spec}}=\frac{1}{2}\left\|\boldsymbol{M}_{\mathrm{UP}}-\boldsymbol{H}\boldsymbol{M}_{\mathrm{HR}}\right\|_2^2 \tag{1}$$

To improve the efficiency and keep the inter-band relationship, the intensity components of $\boldsymbol{M}_{\mathrm{UP}}$ and $\boldsymbol{M}_{\mathrm{HR}}$ are extracted respectively by the following equation:

$$\boldsymbol{I}_{\mathrm{X}}=\sum_{b=1}^{B}\alpha_b\boldsymbol{M}_{\mathrm{X}}^{b} \tag{2}$$

where b represents the bth band, $\boldsymbol{I}_{\mathrm{X}}$ is either $\boldsymbol{I}_{\mathrm{UP}}$ or $\boldsymbol{I}_{\mathrm{HR}}$, α_b is the weight coefficient of the bth band. According to (1) and (2), we can obtain $E_{\mathrm{spec}}^{'}$ as follows:

$$E_{\mathrm{spec}}^{'}=\frac{1}{2}\left\|\sum_{b=1}^{B}\alpha_b\boldsymbol{M}_{\mathrm{UP}}^{b}-\boldsymbol{H}\sum_{b=1}^{B}\alpha_b\boldsymbol{M}_{\mathrm{HR}}^{b}\right\|_2^2=\frac{1}{2}\left\|\boldsymbol{I}_{\mathrm{UP}}-\boldsymbol{H}\boldsymbol{I}_{\mathrm{HR}}\right\|_2^2 \tag{3}$$

where the weight coefficient α_b is set as $1/B$ for simplicity. Here, the filter $\boldsymbol{H}$ is still uncertain. Because $\boldsymbol{H}$ should be a filter with a Gaussian-like shape [18], and the Nyquist gain of modulation transfer function (MTF) of MS sensor is hard to obtain [18], [19], we thus define a filter with the distribution of $N(0,\sigma^2)$, where σ represents standard deviation and needs to be determined. Since the spatial details of $\boldsymbol{I}_{\mathrm{HR}}$ are similar to those of the PAN image, the low-resolution version of $\boldsymbol{P}$ obtained by applying $\boldsymbol{H}$ on $\boldsymbol{P}$ also needs to be close to $\boldsymbol{I}_{\mathrm{UP}}$. Based on this analysis, we employ a correlation function to represent the degree of the similarity between $\boldsymbol{H}\boldsymbol{P}$ and $\boldsymbol{I}_{\mathrm{UP}}$, and design a filter acquisition algorithm to obtain an optimal σ when the correlation function reaches the maximum, *i.e.*,

$$\hat{\sigma}=\arg\max_{\sigma} corr(\boldsymbol{H}_{\sim N(0,\sigma)}\boldsymbol{P},\boldsymbol{I}_{\mathrm{UP}}) \tag{4}$$

where $corr(\cdot)$ represents the correlation function, $\hat{\sigma}$ represents the estimated σ. Iteratively updating σ, the optimal σ can be obtained.

In the ideal situation, $\boldsymbol{T}$ should be equal to $\boldsymbol{I}_{\mathrm{HR}}$. Replacing $\boldsymbol{I}_{\mathrm{HR}}$ with $\boldsymbol{T}$ in Eq. (3), and combining Eq. (4), we can deduce the spectral fidelity constraint as follows:

$$E_{\mathrm{spec}}^{'}=\frac{1}{2}\left\|\boldsymbol{I}_{\mathrm{UP}}-\boldsymbol{H}_{\sim N(0,\hat{\sigma})}\boldsymbol{T}\right\|_2^2 \tag{5}$$

As for the spatial fidelity, the spatial structure of $\boldsymbol{T}$ should be similar to that of the PAN image, and it can be expressed by the gradient information [6]. The Laplacian operator is employed to obtain gradient information at any direction [3], [20]. Given that Laplacian operator is denoted as $\boldsymbol{L}$, then $\boldsymbol{L}\boldsymbol{T}$ representing the gradient of $\boldsymbol{T}$ should be similar to $\boldsymbol{L}\boldsymbol{P}$ which represents the gradient of $\boldsymbol{P}$. Thus, the spatial-fidelity function can be designed as follows:

$$E_{\mathrm{spat}}^{'}=\frac{\beta}{2}\left\|\boldsymbol{L}\boldsymbol{P}-\boldsymbol{L}\boldsymbol{T}\right\|_2^2 \tag{6}$$

where β represents a penelty parameter. Combing (5) and (6) together, we can obtain $\boldsymbol{T}$ by minimizing the spectral-spatial fidelity function as follows:

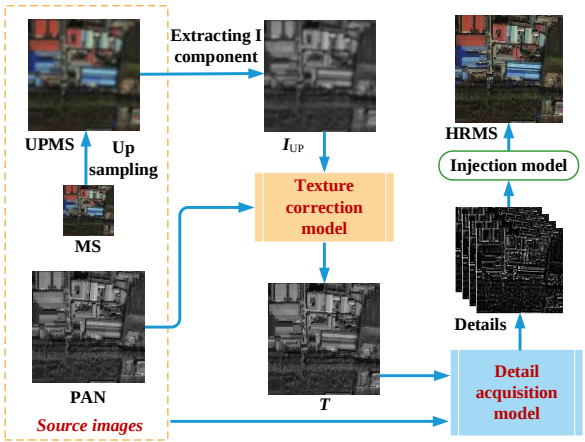

Fig.1. Flowchart of the proposed pansharpening method.

$$\boldsymbol{T}=\arg\min_{\boldsymbol{T}}\left(\frac{1}{2}\left\|\boldsymbol{I}_{\mathrm{UP}}-\boldsymbol{H}_{\sim N(0,\hat{\sigma})}\boldsymbol{T}\right\|_2^2+\frac{\beta}{2}\left\|\boldsymbol{LP}-\boldsymbol{LT}\right\|_2^2\right) \quad (7)$$

A fast solution for the model (7) can be obtained by setting the derivative result to 0, *i.e.*,

$$(\boldsymbol{H}_{\sim N(0,\hat{\sigma})}^{\mathrm{T}}\boldsymbol{H}_{\sim N(0,\hat{\sigma})}+\beta\boldsymbol{L}^{\mathrm{T}}\boldsymbol{L})\boldsymbol{T}=\boldsymbol{H}_{\sim N(0,\hat{\sigma})}^{\mathrm{T}}\boldsymbol{I}_{\mathrm{UP}}+\beta\boldsymbol{L}^{\mathrm{T}}\boldsymbol{LP} \quad (8)$$

Due to the existence of the gradient operation, the fast Fourier transform (FFT, denoted as $\mathcal{F}(\bullet)$) and inverse FFT (denoted as $\mathcal{F}^{-1}(\bullet)$) are employed to quickly solve the Eq.(8), which is shown as follows:

$$\boldsymbol{T}=\mathcal{F}^{-1}\left(\circ\frac{\mathcal{F}(\boldsymbol{H}_{\sim N(0,\hat{\sigma})}^{\mathrm{T}}\boldsymbol{I}_{\mathrm{UP}}+\beta\boldsymbol{L}^{\mathrm{T}}\boldsymbol{LP})}{\mathcal{F}(\boldsymbol{H}_{\sim N(0,\hat{\sigma})}^{\mathrm{T}}\boldsymbol{H}_{\sim N(0,\hat{\sigma})}+\beta\boldsymbol{L}^{\mathrm{T}}\boldsymbol{L})}\circ\right) \quad (9)$$

where $\circ$ represents the pixelwise operation.

C. Detail acquisition model

The details for each band of MS image can be extracted from the obtained texture image. Owing to the significant spatial difference among the bands, extracting the details by subtracting $\boldsymbol{M}_{\mathrm{UP}}$ from $\boldsymbol{T}$ cannot fully satisfy the diversity of the spatial information in different bands. Therefore, the extracted details should be band-adaptive corresponding to the spatial characteristics of each band. Besides, there is some spatial information that exists in the MS image but may be lost in $\boldsymbol{T}$. Thus, not only the details in $\boldsymbol{T}$, but also the details in the MS image should be considered to further refine the details. According to the above analysis, a detail acquisition model is designed based on the two consecutive parameter regressions.

The first-step regression model is constructed based on the band-adaptive detail extraction. Obviously, the details of each band should be consistent with those of the HRMS image, which can be defined as follows:

$$\boldsymbol{D}_{\mathrm{F1}}^{b}=\boldsymbol{T}-\boldsymbol{T}_{\mathrm{L}}^{b}=\boldsymbol{M}_{\mathrm{HR}}^{b}-\boldsymbol{M}_{\mathrm{UP}}^{b}+\varepsilon \quad (10)$$

where $\boldsymbol{D}_{\mathrm{F1}}^{b}$ is the bth band details with respect to the first step, ε is the error and should be as small as possible. $\boldsymbol{T}_{\mathrm{L}}^{b}$ and $\boldsymbol{M}_{\mathrm{HR}}^{b}$ are still unknown. $\boldsymbol{T}_{\mathrm{L}}^{b}$ is the bth band low-resolution version of texture image, which can be determined by two ways. One way is to replace it with $\boldsymbol{I}_{\mathrm{UP}}$, which belongs to the CS-based method. The other way is to apply a Gaussian filter obtained by Eq. (4) on $\boldsymbol{T}$ to get a filtered image $\boldsymbol{T}_{\mathrm{L}}^{\mathrm{G}}$, which belongs to the MRA-based method. To take the advantage of these two methods, we define a band-adaptive algorithm for $\boldsymbol{T}_{\mathrm{L}}^{b}$ by a linear combination of $\boldsymbol{I}_{\mathrm{UP}}$ and $\boldsymbol{T}_{\mathrm{L}}^{\mathrm{G}}$ as follows:

$$\boldsymbol{T}_{\mathrm{L}}^{b}=\omega_{b,1}\boldsymbol{I}_{\mathrm{UP}}+\omega_{b,2}\boldsymbol{T}_{\mathrm{L}}^{\mathrm{G}} \quad (11)$$

where $\omega_{b,1}$ and $\omega_{b,2}$ are the band-adaptive weight coefficients that need to be determined.

Due to the lack of reference image $\boldsymbol{M}_{\mathrm{HR}}^{b}$ in (10), we design a reduced scale algorithm and use the original MS image ($\boldsymbol{M}^{b}$) as a reference image to determine the parameters in Eq. (11). The obtained $\boldsymbol{T}$ and the source images in (10) are degraded and down-sampled by a factor of 4 following Wald protocol [5]. Based on (10) and (11), we have a reduced resolution version of Eq. (10) as follows:

$$\boldsymbol{D}_{\mathrm{R\_F1}}^{b}=\boldsymbol{T}_{\mathrm{R}}-(\omega_{b,1}\boldsymbol{I}_{\mathrm{R\_UP}}+\omega_{b,2}\boldsymbol{T}_{\mathrm{R\_L}}^{\mathrm{G}})=\boldsymbol{M}^{b}-\boldsymbol{M}_{\mathrm{R\_UP}}^{b}+\varepsilon \quad (12)$$

where $\boldsymbol{D}_{\mathrm{R\_F1}}^{b}$, $\boldsymbol{T}_{\mathrm{R}}$, $\boldsymbol{I}_{\mathrm{R\_UP}}$, $\boldsymbol{T}_{\mathrm{R\_L}}^{\mathrm{G}}$ and $\boldsymbol{M}_{\mathrm{R\_UP}}^{b}$ represent the reduced resolution versions of $\boldsymbol{D}_{\mathrm{F1}}^{b}$, $\boldsymbol{T}$, $\boldsymbol{I}_{\mathrm{UP}}$, $\boldsymbol{T}_{\mathrm{L}}^{\mathrm{G}}$ and $\boldsymbol{M}_{\mathrm{UP}}^{b}$, respectively. In (12), $\omega_{b,1}$ and $\omega_{b,1}$ are the only unknown parameters. Defining $\boldsymbol{\omega}_b=[\omega_{b,1},\ \omega_{b,2}]$, $\boldsymbol{Q}=[\ \boldsymbol{I}_{\mathrm{R\_UP}},\ \boldsymbol{T}_{\mathrm{R\_L}}^{\mathrm{G}}]^{\mathrm{T}}$, and making the error ε as small as possible, we design a parameter regression model to estimate $\boldsymbol{\omega}_b$ based on (12), *i.e.*,

$$\begin{cases}\boldsymbol{\omega}_b=\arg\min_{\boldsymbol{\omega}_b}\frac{1}{2}\left\|\boldsymbol{S}_b-\boldsymbol{\omega}_b\boldsymbol{Q}\right\|^2 \quad s.t.\ \boldsymbol{\omega}_b\geq 0\\ \boldsymbol{S}_b=\boldsymbol{T}_{\mathrm{R}}-\boldsymbol{M}^{b}+\boldsymbol{M}_{\mathrm{R\_UP}}^{b}\end{cases} \quad (13)$$

The model in (13) can be solved using the Lagrange multiplier and gradient decent method. The solution is expressed as follows:

$$\begin{aligned}&\boldsymbol{\omega}_b=\arg\min_{\boldsymbol{\omega}_b}\left(\frac{1}{2}\left\|\boldsymbol{S}_b-\boldsymbol{\omega}_b\boldsymbol{Q}\right\|^2+\frac{\lambda}{2}\left\|\max(\boldsymbol{0},-\boldsymbol{\omega}_b)\right\|^2\right)\\ &\Rightarrow\boldsymbol{\omega}_b^{t+1}=\Delta t(\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{S}_b-\boldsymbol{\omega}_b^{t}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q}+\lambda\max(\boldsymbol{0},-\boldsymbol{\omega}_b^{t}))+\boldsymbol{\omega}_b^{t}\end{aligned} \quad (14)$$

where λ is the penalty parameter, t is the iteration step and Δt is the step length. By iterative regression, the final $\boldsymbol{\omega}_b$ can be obtained when the regression meets the end condition. Substituting $\boldsymbol{\omega}_b$ into (12), $\boldsymbol{D}_{\mathrm{R\_F1}}^{b}$ is obtained. Further, substituting $\boldsymbol{\omega}_b$ into (11), $\boldsymbol{D}_{\mathrm{F1}}^{b}$ is obtained by (10) too.

To further refine the details, a band-adaptive fusion algorithm is designed by fusing the details from the UPMS image (denoted as $\boldsymbol{D}_{\mathrm{M}}^{b}$) with $\boldsymbol{D}_{\mathrm{F1}}^{b}$ to refine the details, *i.e.*,

$$\boldsymbol{D}_{\mathrm{F2}}^{b}=\delta_{b,1}\boldsymbol{D}_{\mathrm{F1}}^{b}+\delta_{b,2}\boldsymbol{D}_{\mathrm{M}}^{b} \quad (15)$$

where $\boldsymbol{D}_{\mathrm{F2}}^{b}$ represents the second-step bth band details, $\delta_{b,1}$ and $\delta_{b,2}$ are the weight coefficients. $\boldsymbol{D}_{\mathrm{M}}^{b}$ can be obtained by applying a Gaussian filter of (4) on the UPMS image. $\delta_{b,1}$ and $\delta_{b,2}$ are unknown and can be determined in reduced resolution as Eqs. (10) and (12). Thus, we design a detail extraction algorithm as well as its reduced resolution version as follows:

$$\delta_{b,1}\boldsymbol{D}_{\mathrm{F1}}^{b}+\delta_{b,2}\boldsymbol{D}_{\mathrm{M}}^{b}=\boldsymbol{M}_{\mathrm{HR}}^{b}-\boldsymbol{M}_{\mathrm{UP}}^{b}+\varepsilon \quad (16)$$

$$\delta_{b,1}\boldsymbol{D}_{\mathrm{R\_F1}}^{b}+\delta_{b,2}\boldsymbol{D}_{\mathrm{R\_M}}^{b}=\boldsymbol{M}^{b}-\boldsymbol{M}_{\mathrm{R\_UP}}^{b}+\varepsilon \quad (17)$$

where $\boldsymbol{D}_{\mathrm{R\_M}}^{b}$ represents the reduced resolution version of $\boldsymbol{D}_{\mathrm{M}}^{b}$. Denoting $\boldsymbol{\delta}_b=[\delta_{b,1},\ \delta_{b,2}]$, $\boldsymbol{D}_{\mathrm{R}}^{b}=[\ \boldsymbol{D}_{\mathrm{R\_F1}}^{b},\ \boldsymbol{D}_{\mathrm{R\_M}}^{b}]^{\mathrm{T}}$, the second-step parameter regression model for estimating $\boldsymbol{\delta}_b$ is designed by making the error ε in (17) as small as possible, *i.e.*,

$$\boldsymbol{\delta}_b=\arg\min_{\boldsymbol{\delta}_b}\frac{1}{2}\left\|\boldsymbol{M}^{b}-\boldsymbol{M}_{\mathrm{R\_UP}}^{b}-\boldsymbol{\delta}_b\boldsymbol{D}_{\mathrm{R}}^{b}\right\|^2 \quad s.t.\ \boldsymbol{\delta}_b\geq 0 \quad (18)$$

The solution of (18) is similar to Eq. (14). When $\boldsymbol{\delta}_b$ is acquired, $\boldsymbol{D}_{\mathrm{F2}}^{b}$ can be obtained by substituting $\boldsymbol{\delta}_b$ into (15). The final HRMS image is obtained by the injection model as follows:

$$\boldsymbol{M}_{\mathrm{HR}}^{b}=\boldsymbol{M}_{\mathrm{UP}}^{b}+g\frac{\boldsymbol{M}_{\mathrm{UP}}^{b}}{(1/B)\sum_{b=1}^{B}\boldsymbol{M}_{\mathrm{UP}}^{b}}\boldsymbol{D}_{\mathrm{F2}}^{b} \quad (19)$$

where g is a scale factor.

III. Experimental Results and Analysis

A. Experimental setup

Two remote sensing datasets including WorldView-3 and IKONOS are collected. Each dataset contains 60 remote sensing image pairs, and each pair includes an MS image and a PAN image. All the MS images contain 4 bands including red, green, blue, and near infrared. Both simulation and real experiments are conducted. For the simulation experiments, the image pair is degraded and down-sampled 4 times, and the

original MS image is regarded as a reference image following Wald protocol [5]. For the real experiments, the fusion is performed with the original size of the image. The original size of an MS image is 256×256×4, and that of a PAN image is 1024×1024. There are three parameters that need to be set up, including β in (9), λ in (14), and g in (19). Both β and g are determined by experiments: β is set as 48 and 85, and g is set as 1.2 and 1 for WorldView-3 and IKONOS, respectively. λ is set as 5 according to the reference [3].

Five widely used quality metrics for simulation experiments are employed including the universal image quality index (UIQI) , spectral angle mapper (SAM), erreur relative global adimensionnelle de synthese (ERGAS), spatial correlation coefficient (SCC), and Q4 [1], [5]. For the real experiments, three well-known metrics are introduced, such as D_λ indicating spectral distortion, D_s indicating spatial distortion, and QNR combining D_λ and D_s [5], [21].

The classical fusion methods for comparison include MTF-Generalized Laplacian Pyramid [5], bilateral filtering-based enhanced pansharpening (BFEP) [22], and some latest comparison methods including MMMT [23], RBDSD [12], FCS [10] , and the low-rank fuzzy fusion (LRFF) [9]. The up-sampling method called EXP [5] is treated as spectral benchmark and excluded from the comparison.

B. Ablation study

To verify the effectiveness of the proposed method, the ablation study is first conducted. Our method includes a texture correction (TC) model and a detail acquisition model which is composed of first-step regression (FSR) and the second-step regression (SSR). The ablation experiments with or without these components are conducted. Note that without the proposed models, the original PAN image (OPI) is used for detail extraction. With only TC, the details are obtained by subtracting $\boldsymbol{I}_{\mathrm{UP}}$ from $\boldsymbol{T}$, and the HRMS image is obtained by Eq. (19). Ten image pairs are collected from the two datasets and fused in the ablation study. The average results of Q4 representing the overall quality and SAM representing the spectral distortion are shown in Fig.2. From the figure, we can see that the result of OPI is unsatisfactory. With the proposed models, the results are improved. The intact proposed method including three components performs the best, which proves the effectiveness of the proposed method.

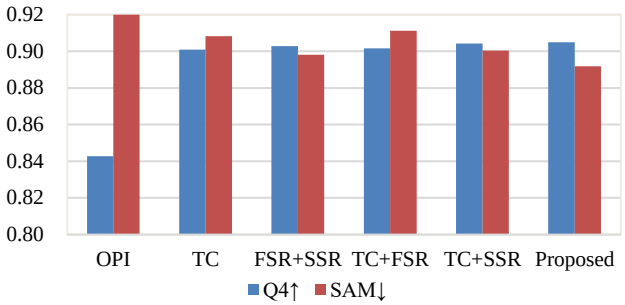

Fig.2. The average Q4 and SAM values of 10 fused images in the ablation experiments. SAM values are normalized proportionally.

C. Simulation and real experiments

Both simulation and real experiments are conducted on the two datasets. Taking an image pair from WorldView-3 as an example for simulation experiment, the closeups of the fusion results are shown in Fig.3. As shown in the red rectangles, the result of FCS obviously suffers from spectral distortion. The results of MMMT and LRFF are blurred and may suffer from insufficient detail injection. The result of BFEP is too bright compared with REF and it may suffer from over detail injection. The result of RBDSD suffers from slight blurry. From the enlarged green rectangles, the results of MTF-GLP, MMMT, FCS, LRFF are too dark and suffer from undersaturation. The proposed method is closer to REF both in spatial and spectral aspects.

Taking an image pair form IKONOS as an example for real experiment, the closeups of the fusion results are shown in Fig.4. From the enlarged red and green rectangles, we can see that the result of BFEP suffers from spatial distortion for the objects are unnatural. The results of RBDSD and MMMT are obviously blurred. The result of FCS suffers from serious spectral distortion for the overall color is changed. The results of MTF-GLP and LRFF suffer from slight spectral distortion for the color is too light compared with that of EXP. The proposed method can keep both spectral and spatial fidelity better.

To further quantify the quality of the fusion results, the

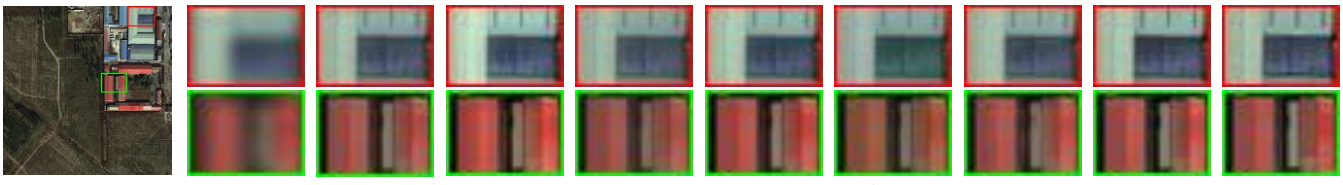

Fig.3. Simulation experiments on an image pair from the WorldView-3 dataset. REF represents a reference image.

Fig.4. Real experiments on an image pair from the IKONO dataset.

TABLE I AVERAGE QUANTITATIVE EVALUATION ON THE FUSION RESULTS

Method	Simulation experiments on WoldView-3 dataset						Real experiments on IKONO dataset			
	UIQI↑	SAM↓	ERGAS↓	SCC↑	Q4↑	Time(s)	D_λ↓	D_s↓	QNR↑	Time(s)
EXP	0.5737	4.6205	7.1010	0.5618	0.5688	-	0.0028	0.2308	0.7670	-
MTF-GLP	0.8578	4.9742	4.3285	0.8702	0.8632	0.22	0.1829	0.1182	0.7222	1.21
BFEP	0.8378	6.1781	8.9122	0.8806	0.7322	1.76	0.2022	0.1607	0.6727	12.49
MMMT	0.8200	4.9792	4.7128	0.8619	0.8229	11.68	0.1560	0.0658	0.7889	200.20
RBDSD	0.8888	5.0968	4.0037	0.8829	0.8912	**0.11**	0.1376	0.0894	0.7866	**0.72**
FCS	0.8685	5.6158	4.3633	0.8776	0.8569	26.57	0.1978	0.1106	0.7150	503.84
LRFF	0.8111	4.7131	4.7836	0.8520	0.8112	7.68	0.1455	**0.0592**	0.8043	75.59
Proposed	**0.8980**	**4.5450**	**3.9061**	**0.8882**	**0.8985**	0.30	**0.0826**	0.0643	**0.8550**	4.78

average objective evaluation of 60 image pairs in each dataset is shown in TABLE I. The metrics of the simulation experiments are shown on the left, while the metrics of the real experiments are shown on the right. The best results are shown in bold red and the second-best results are underlined. The computational time is listed in the table too. From the table, we can see that the proposed method can acquire the best results in all metrics both in the simulation and real experiments except D_s which is the second-best and slightly inferior than the best one. Besides, the computational cost of our method is relatively low, which further verifies that our method has the merits of high fusion quality and efficiency.

IV. CONCLUSION

In this paper, to reduce the spectral distortion and accurately extract details for each band of an MS image, a novel pansharpening approach is proposed based on texture correction and detail refinement. First, a texture correction model is constructed to get a texture image that is highly correlated with the MS image. Second, a detail acquisition algorithm is designed to extract the refined details from the texture image for each band of MS image. In the algorithm, two consecutive parameter regression models are constructed. With the obtained parameters, the details for each band can be accurately extracted. Finally, the HRMS image is obtained by injecting the extracted details into the UPMS image. Both simulation and real experiments show that compared with some classical and the latest methods, the proposed method achieves high fusion quality with high efficiency.

REFERENCES

[1] F. Dadrass Javan, F. Samadzadegan, S. Mehravar, A. Toosi, R. Khatami, and A. Stein, "A review of image fusion techniques for pan-sharpening of high-resolution satellite imagery," *ISPRS J. Photogram. Remote Sens.*, vol. 171, pp. 101–117, Jan. 2021.

[2] X. Meng, H. Shen, H. Li, L. Zhang, and R. Fu, "Review of the pansharpening methods for remote sensing images based on the idea of meta-analysis: Practical discussion and challenges," *Information Fusion*, vol. 46, pp. 102–113, Mar. 2019.

[3] Y. Yang, H. Lu, S. Huang, and W. Tu, "Remote Sensing Image Fusion Based on Fuzzy Logic and Salience Measure," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 11, pp. 1943–1947, Nov. 2020.

[4] C. Kwan, B. Budavari, M. Dao, B. Ayhan, and J. F. Bell, "Pansharpening of Mastcam images," in *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2017, pp. 5117–5120.

[5] G. Vivone, L. Alparone, J. Chanussot, M. Dalla Mura, A. Garzelli, G. A. Licciardi, R. Restaino, and L. Wald, "A Critical Comparison Among Pansharpening Algorithms," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2565–2586, May 2015.

[6] Y. Yang, H. Lu, S. Huang, Y. Fang, and W. Tu, "An efficient and high-quality pansharpening model based on conditional random fields," *Information Sciences*, vol. 553, pp. 1–18, Apr. 2021.

[7] H. Ghassemian, "A review of remote sensing image fusion methods," *Information Fusion*, vol. 32, pp. 75–89, Nov. 2016.

[8] T. Benzenati, Y. Kessentini, A. Kallel, and H. Hallabia, "Generalized Laplacian Pyramid Pan-Sharpening Gain Injection Prediction Based on CNN," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 4, pp. 651–655, Apr. 2020.

[9] Y. Yang, C. Wan, S. Huang, H. Lu, and W. Wan, "Pansharpening Based on Low-Rank Fuzzy Fusion and Detail Supplement," *IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing*, vol. 13, pp. 5466–5479, 2020.

[10] M. Ghahremani, Y. Liu, P. Yuen, and A. Behera, "Remote sensing image fusion via compressive sensing," *ISPRS J. Photogram. Remote Sens.*, vol. 152, pp. 34–48, Jun. 2019.

[11] A. Garzelli, F. Nencini, and L. Capobianco, "Optimal MMSE Pan Sharpening of Very High Resolution Multispectral Images," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 1, pp. 228–236, Jan. 2008.

[12] G. Vivone, "Robust Band-Dependent Spatial-Detail Approaches for Panchromatic Sharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6421–6433, Sep. 2019.

[13] M. Jiang, H. Shen, J. Li, Q. Yuan, and L. Zhang, "A differential information residual convolutional neural network for pansharpening," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 163, pp. 257–271, May 2020.

[14] S. Luo, S. Zhou, Y. Feng, and J. Xie, "Pansharpening via Unsupervised Convolutional Neural Networks," *IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing*, vol. 13, pp. 4295–4310, 2020.

[15] Y. Qu, R. K. Baghbaderani, H. Qi, and C. Kwan, "Unsupervised Pansharpening Based on Self-Attention Mechanism," *IEEE Trans. Geosci. Remote Sensing*, vol. 59, no. 4, pp. 3192–3208, Apr. 2021.

[16] X. Fu, Z. Lin, Y. Huang, and X. Ding, "A Variational Pan-Sharpening With Local Gradient Constraints," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, 2019, pp. 10257–10266.

[17] Y. Yang, H. Lu, S. Huang, and W. Tu, "Pansharpening Based on Joint-Guided Detail Extraction," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 389–401, 2021.

[18] G. Vivone, P. Addesso, R. Restaino, M. Dalla Mura, and J. Chanussot, "Pansharpening Based on Deconvolution for Multiband Filter Estimation," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 540–553, Jan. 2019.

[19] G. Vivone, M. Simoes, M. Dalla Mura, R. Restaino, J. M. Bioucas-Dias, G. A. Licciardi, and J. Chanussot, "Pansharpening Based on Semiblind Deconvolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 1997–2010, Apr. 2015.

[20] Y. Yang, H. Lu, S. Huang, and W. Tu, "An Efficient Pansharpening Method Based On Conditional Random Fields," in *2020 IEEE International Conference on Multimedia and Expo (ICME)*, London, United Kingdom, 2020, pp. 1–6.

[21] C. Kwan, B. Budavari, A. C. Bovik, and G. Marchisio, "Blind Quality Assessment of Fused WorldView-3 Images by Using the Combinations of Pansharpening and Hypersharpening Paradigms," *IEEE Geosci. Remote Sensing Lett.*, vol. 14, no. 10, pp. 1835–1839, Oct. 2017.

[22] N. H. Kaplan and I. Erer, "Bilateral Filtering-Based Enhanced Pansharpening of Multispectral Satellite Images," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 11, pp. 1941–1945, Nov. 2014.

[23] Y. Yang, W. Wan, S. Huang, P. Lin, and Y. Que, "A Novel Pan-Sharpening Framework Based on Matting Model and Multiscale Transform," *Remote Sensing*, vol. 9, no. 4, p. 391, Apr. 2017.