Context adaptive panchromatic band simulation and detail injection for image pansharpening

Dai, Qinling; Luo, Bin; Wang, Leiguang; Tu, Zhigang

2018

Dai, Q., Luo, B., Wang, L., & Tu, Z. (2018). Context adaptive panchromatic band simulation and detail injection for image pansharpening. *Journal of Applied Remote Sensing*, 12(1), 015018-.

<https://hdl.handle.net/10356/86291>

<https://doi.org/10.1117/1.JRS.12.015018>

© 2018 Society of Photo-Optical Instrumentation Engineers (SPIE). This paper was published in *Journal of Applied Remote Sensing* and is made available as an electronic reprint (preprint) with permission of Society of Photo-Optical Instrumentation Engineers (SPIE). The published version is available at: [<http://dx.doi.org/10.1117/1.JRS.12.015018>]. One print or electronic copy may be made for personal use only. Systematic or multiple reproduction, distribution to multiple locations via electronic or other means, duplication of any material in this paper for a fee or for commercial purposes, or modification of the content of the paper is prohibited and is subject to penalties under law.

Downloaded on 26 Apr 2023 15:18:16 SGT

Context adaptive panchromatic band simulation and detail injection for image pansharpening

Qinling Dai,^{a,b,*} Bin Luo,^c Leiguang Wang,^d and Zhigang Tu^e

^aWuhan University, School of Printing and Packaging, Wuhan, China

^bSouthwest Forestry University, School of Design, Kunming, China

^cWuhan University, State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan, China

^dSouthwest Forestry University, School of Forestry, Kunming, China

^eNanyang Technological University, School of Electrical and Electronic, Singapore

Abstract. General component substitution (CS) pansharpening methods establish a global model over the whole image plane and may lead to unexpected spectral distortion. This paper proposes a context adaptive CS pansharpening method, which features a totally local-based processing procedure. The method consists of two processing blocks. The first block is to simulate a low-resolution panchromatic band by a local linear regression model between panchromatic and multispectral bands. The second block extracts spatial details and adds details back to multispectral bands in locally varying ratios. By recasting the local linear regression model into the guided filtering framework and analyzing the implicit statistical assumptions underlying CS methods, the strengths of the local-based pansharpening algorithm are addressed. Experiments test 7 pairs of images acquired from different sensors, such as GF-2, Quickbird, and Worldview-2. Both quantitative and qualitative evaluations reveal that the presented method can better preserve the spectral information than some state-of-the-art methods. © 2018 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: [10.1117/1.JRS.12.015018](https://doi.org/10.1117/1.JRS.12.015018)]

Keywords: image pansharpening; component substitution; linear regression; context adaptive model.

Paper 170820 received Sep. 22, 2017; accepted for publication Feb. 14, 2018; published online Mar. 7, 2018.

1 Introduction

Due to the physical constraint of sensors, remotely sensed images obtained from a single satellite often cannot possess high spatial and spectral resolutions simultaneously. Therefore, the image pansharpening techniques, which can combine both strengths of multispectral image (MS) and panchromatic image (Pan), have become a very hot topic in the remote sensing community.^{1–5}

A large collection of literature has presented all kinds of fusion methods. Some well-known algorithms include Gram–Schmidt (GS),⁶ Brovey,⁷ UNB-pansharpening (UNB),⁸ subtractive resolution merge (SRM),⁹ band-dependent spatial-detail algorithm (BDSD),¹⁰ etc. These methods can be roughly divided into three categories,^{1,4} including ratio enhancement (RE), component substitution (CS), and multiresolution analysis. Critical reviews on pansharpening methods can be found in Refs. 2, 4, and 5. Recently, the deep learning techniques,¹¹ e.g., convolutional neural network-based methods (CNNs), have also been employed to solve low-level vision problems, such as image pansharpening and superresolution,^{12–14} and obtained promising pansharpening results.¹³ However, as training a CNN often requires large amounts of labeled samples and the variety of remote sensing images in spectral and spatial properties, the applications of the techniques in pansharpening are still limited.

The RE category obtains a fused MS band by multiplying a ratio with an upsampled MS band, where the ratio is calculated by the division of the (histogram-adjusted) original Pan band

*Address all correspondence to: Qinling Dai, E-mail: daiqinling@126.com

and the simulated Pan band. Methods, such as Brovey,⁷ UNB-pansharpening,⁸ and the recently developed nearest-neighbor diffusion (NND)-based method,¹⁵ belong to this category. Recent research^{5,16,17} also reveals that the RE category can also be seen as a special subtype of the CS category.

The CS group typically includes GS-based methods^{1,6} and IHS-based methods.^{18,19} These methods involve transformations to project MS bands into new spaces, and then partially or totally substitute one of the obtained components by a simulated Pan band during performing the inverse transformation. These CS methods can also be formulated by two-step processes,^{1,20} including the Pan/intensity simulation and detail injection. Due to their low computation complexity, acceptable spatial quality, many methods belonging to the above two categories have been integrated into some commercial software such as ENVI, ERDAS, and PCI.

However, these methods are also criticized for spectral distortion. And this spectral distortion is generally attributed to the significant difference between substituted and substitution components.^{21,22} Various modifications have been proposed to address this problem.^{8,19,22} For instance, Zhang⁸ utilized both histogram standardization and multivariate linear regression techniques to simulate a low-resolution Pan band; Rahmani et al.¹⁹ included a histogram matching step to ensure that the mean and standard deviation of Pan and MS images are within the same range; Xu et al.²² designed a data fitting method to adjust the mean information of the Pan image.

Although both the histogram adjusting and the linear regression can efficiently alleviate spectral distortions, these techniques are still argued due to some obvious limitations. First, these methods rely on some global statistics. The parameters used for Pan simulation and detail injection (i.e., mean, standard deviation, regression coefficients, and injection weights) are estimated from the entire processing unit. Therefore, the fused results obtained from an entire image or a patch of the image can not be consistent. Second, as pointed out by many works,^{20,22–24} due to the variety of illumination and atmospheric conditions, the simulation and/or injection parameters assessed from the whole scene through global-based techniques cannot accurately represent the relationship between MS and Pan bands in a local area or a specific land class.

Recently, Xie et al.¹⁷ have well revealed the implicit assumptions of the CS methods. One assumption is that the common weight and bias coefficients should be shared by all pixels [see assumptions 4 and related discussion in Ref. 17 (pp. 4, 11, 12)]. Obviously, due to spatial complexity of the earth surface, this assumption is often violated in a large area. Therefore, applying this assumption in a spatial homogeneous area or local small windows is a theoretically better choice. Meanwhile, many works^{1,4,24} have practically improved CS methods by restricting the detail injection model in a local sliding window or distinct block.

Although various local-based techniques have extensively integrated into the CS pansharpening process, the local-based Pan/intensity-simulation model was still less considered in previous works and the performance of various local/global-based simulation and injection models were also not comprehensively compared. In this context, a totally local-based CS method, featuring context adaptive Pan simulation and adaptive spatial detail injection, is proposed. As the models and the related parameters are established and estimated locally, the presented method is not sensitive to the change of the image size, and the local similarity in gray values between substituted and substitution components can be guaranteed. Meanwhile, the adaptive Pan simulation is associated with the guided filtering,²⁵ which casts a new light on this local-based processing strategy.

The remaining sections are organized as follows. Section 2 formulates the problem and describes the proposed local-based pansharpening method. Section 3 presents the experimental results and discussions carried out on multisource datasets. Finally, conclusions are drawn in Sec. 4.

2 Proposed Local-Based CS Method

As shown in Fig. 1, the proposed algorithm consists of four main steps. The first step (A) is to select MS bands whose spectral coverage is overlapped with that of the Pan band. The second (B) and third steps (C) are to simulate a low-resolution Pan band by a local-based regression

Fig. 1 The flowchart of the proposed local-based pansharpening method. The numbers in parentheses index the related equations in the paper.

model. And the last step (D) aims to obtaining a pansharpened MS image by a local-based detail injection model. The problem formulation and description on the algorithm are presented in the following sections.

2.1 Problem Formulation and Notations

Given a Pan image $\mathbf{P}$ and an MS image $\mathbf{M}$ acquired from the same scene, the aim of pansharpening is to obtain a new MS $\hat{\mathbf{M}}$ at the spatial size of the original PAN band, which preserves the spectral property of the original MS bands and possesses the spatial property of the original Pan band.

During fusing of the two kinds of images, the original $\mathbf{M}$ should be first upsampled to the same spatial size of the Pan band $\mathbf{P}$. For brevity, the upsampled spectral image is still denoted as $\mathbf{M}$ in the following sections. Thereafter, in order to prepare spatial details for the detail injection model, a low-resolution version of the panchromatic image $\mathbf{P}$ should be simulated, which is denoted as $\hat{\mathbf{P}}$.

In the following part, the first subscript associated with these above-mentioned matrices (such as $\mathbf{M}$ and $\hat{\mathbf{P}}$) is the position index, and the second one (if it exists) is the band index; d is the number of MS bands; and N_1 and N_2 are the number of pixels of original Pan and MS images, respectively. For instance, $\mathbf{M}_{i,j}$ (with $i = 1, \dots, N_2$ and $j = 1, \dots, d$) denotes the pixel value of the MS image at the position i and the band j .

2.2 Local-Based Pan Simulation

The key assumption underlying the proposed Pan-simulation method is a local linear model. We assume that, in a local window w_k centered at the pixel k , a low-resolution version of the panchromatic image $\mathbf{P}_i$ can be linearly simulated by the upsampled spectral vector $\mathbf{M}_i$ and a residual scalar b_k . This local linear model can be formulated as

$$\hat{\mathbf{P}}_i = \boldsymbol{\beta}_k^T \mathbf{M}_i + b_k, \quad \forall i \in w_k, \quad (1)$$

where both $\boldsymbol{\beta}_k = [\beta_1, \beta_2, \dots, \beta_d]^T$ and b_k are undetermined coefficients, and $\{\boldsymbol{\beta}_k, b_k\}$ are just assumed to be constant in w_k .

For a given window w_k , in order to determine the linear coefficients $\boldsymbol{\beta}_k$ and b_k in Eq. (1), the linear regression cost function can be established as Eq. (2) (see Ref. 26, pp. 44–46)

$$\text{RSS}(\boldsymbol{\beta}_k, b_k) = \sum_{i \in w_k} (\mathbf{P}_i - \boldsymbol{\beta}_k^T \mathbf{M}_i - b_k)^2, \quad (2)$$

where $\mathbf{P}_i$ is the pixel value of the original Pan located i .

By minimizing the residual sum of squares shown in Eq. (2), the coefficients β_k and b_k can be estimated as

$$\beta_k = (\Lambda_k)^{-1} \left(\frac{1}{|w_k|} \sum_{i \in w_k} \mathbf{P}_i \mathbf{M}_i - \bar{\mathbf{P}} \bar{\mathbf{M}} \right), \quad (3)$$

$$b_k = \bar{\mathbf{P}} - \beta_k^T \bar{\mathbf{M}}, \quad (4)$$

where Λ_k is the covariance matrix of upsampled MS bands in w_k , $|w_k|$ is the number of pixels in w_k , $\bar{\mathbf{P}}$ and $\bar{\mathbf{M}}$ are the average vector of the panchromatic and MS bands in w_k , respectively. By introducing Eqs. (3) and (4) into Eq. (1), a synthetic low-resolution panchromatic local patch $\hat{\mathbf{P}}_i$ with $i \in w_k$ can be obtained.

In order to extend the local simulation to the whole image, a sliding-window strategy is considered. As a pixel i is overlapped by multiple windows w_k , the value of $\hat{\mathbf{P}}_i$ in Eq. (1) is different when it is calculated in different windows. Therefore, we average all the possible values of $\hat{\mathbf{P}}_i$. More specifically, the average of all the values of $\hat{\mathbf{P}}_i$ (denoted as $\bar{\hat{\mathbf{P}}}_i$) is defined as the output of the pixel i . Considering the symmetry of windows, this process can be written as

$$\bar{\hat{\mathbf{P}}}_i = \bar{\beta}_i^T \bar{\mathbf{M}}_i + \bar{b}_i \quad \text{with} \quad i = 1, \dots, N_1, \quad (5)$$

where $\bar{\beta}_i = \sum_{k \in w_i} \beta_k / |w_i|$ and $\bar{b}_i = \sum_{k \in w_i} b_k / |w_i|$ are the average coefficients of all windows overlapping i .

In the above Pan-simulation process, one potential problem is that the covariance matrix Λ_k in Eq. (3) is an ill-conditioned or singular matrix. Thereby, the linear regression technique may lead to an unstable estimation of the coefficients (see Ref. 26, pp. 61–69). Analogous to Ref. 24, in this case, the weight coefficient for each MS band is simply calculated from the local image window, and assigned as the ratio of the sum of all pixels of the corresponding MS band over the sum of all Pan pixels.

2.3 Local-Based Detail Injection

In order to yield the final pansharpened bands $\tilde{\mathbf{M}}$, the detail information, obtained by the difference between the original Pan band $\mathbf{P}_i$ and the synthetic Pan bands $\bar{\hat{\mathbf{P}}}_i$, should be added in ratio into the resampled MS bands $\mathbf{M}_i$, respectively. This process can be formulated as

$$\tilde{\mathbf{M}}_i = \bar{\mathbf{g}}_i (\mathbf{P}_i - \bar{\hat{\mathbf{P}}}_i) + \mathbf{M}_i \quad \text{with} \quad i = 1, \dots, N_1, \quad (6)$$

$$\mathbf{g}_k = \frac{\text{Cov}_k(\mathbf{M}, \bar{\mathbf{P}})}{\text{Var}_k(\bar{\mathbf{P}})}, \quad (7)$$

$$\bar{\mathbf{g}}_i = \frac{\sum_{k \in w_i} \mathbf{g}_k}{|w_i|}, \quad (8)$$

where $\mathbf{M}_i$, $\tilde{\mathbf{M}}_i$, and $\bar{\mathbf{g}}_i$ are $d \times 1$ vectors, $\bar{\mathbf{g}}_i$ is calculated by the average of all possible values of $\mathbf{g}_k$ and each $\mathbf{g}_k$ is calculated locally from the window centered at the pixel k by the covariance of each MS band and the simulated Pan band $\text{Cov}_k(\mathbf{M}, \bar{\mathbf{P}})$ over the variance of the simulated Pan $\text{Var}_k(\bar{\mathbf{P}})$, $|w_i|$ is the number of pixels in the window w_i .

In the above pansharpening process, two window sizes for both Pan simulation and CS should be specified, respectively. Empirically, the setting of the window sizes is associated with the spatial resolution ratio r between MS and Pan images.^{10,15,24} The sizes should also be large enough to obtain a stable estimation of the local relationship between bands.

Meanwhile, the sizes are also expected to be small, in order that the estimation can be operated on the neighboring pixels with similar spectra or belonging to the same land class. Therefore, a “trade off” decision has to be made on the two parameters. For instance, the windows size for estimating the injection coefficients is configured as $5 \times r + 1$ in Ref. 24. In our research, the influence of different window sizes was exhaustively tested on an open dataset.⁵ The details are represented in the following Sec. 3.3.1.

2.4 Relationship Between Local-Based Pan Simulation and Guided Filter

It is quite interesting that the local linear assumption underlying the proposed Pan-simulation method is identical to the assumption underlying the guided image filtering. The guided image filtering is implicitly derived from a local linear model, which is based on two key assumptions.²⁵ The first is a local linear model between the guidance image $\mathbf{I}$ and the filtering output q , and the second is the minimization of the difference between the input image p and the output q . The two assumptions can be formulated as

$$q_i = \boldsymbol{\beta}_k^T \mathbf{I}_i + b_k, \quad \forall i \in w_k, \quad (9)$$

$$\text{RSS}(\boldsymbol{\beta}_k, b_k) = \sum_{i \in w_k} [(\boldsymbol{\beta}_k^T \mathbf{I}_i + b_k - p_i)^2 + \varepsilon \|\boldsymbol{\beta}_k\|^2], \quad (10)$$

where $\mathbf{I}_i$ is a spectral vector, $\boldsymbol{\beta}_k$ is a weight vector, p_i , q_i , and b_k are scalars, ε is a prespecified regularization parameter. Equation (10) represents a linear rigid regression model (Ref. 26, p. 61), where ε is to penalize the large $\boldsymbol{\beta}_k$ values.

If the input image and guidance image are assigned as the original Pan image $\mathbf{P}$ and the MS image $\hat{\mathbf{M}}$, respectively, then Eq. (9) is strictly equivalent to the local linear model presented by Eq. (1). Furthermore, by specifying ε as 0, Eq. (10) converts to the objective function expressed by Eq. (2). Therefore, the presented Pan-simulation process can be seen as a special case of the guided image filtering, where the input image and guidance image are assigned as the original Pan image and the MS image, respectively, and the regularization parameter is specified as 0.

This connection also casts a new light on the proposed Pan-simulation process. First, due to the structure-transferring property of the guided filtering (see Sec. 3.6 in Ref. 25), the structure of all MS bands (the guide image) can be well preserved in the simulated Pan band (the filtering output). As shown in Fig. 2, due to the local linear model of $\hat{\mathbf{P}} = \boldsymbol{\beta}_k^T \hat{\mathbf{M}} + b_k$, the output $\hat{\mathbf{P}}$, the simulated low-resolution Pan band, is locally a linear weighted summation of the original MS bands $\hat{\mathbf{M}}$. As $\nabla \hat{\mathbf{P}} = \boldsymbol{\beta}_k^T \nabla \hat{\mathbf{M}}$, the structure in the guidance image $\hat{\mathbf{M}}$ is transferred in ratio $\boldsymbol{\beta}_k$ into $\hat{\mathbf{P}}$. Meanwhile, the guided filtering is an implicit edge-preserving soothing filter,²⁵ which can

Fig. 2 Understanding the proposed pan-simulation model from the view of the guided filtering.

soothe the input panchromatic image $\mathbf{P}$. In this way, the local structure similarity between the substituted component $\widehat{\mathbf{P}}$ and substitution component $\mathbf{P}$ can also be guaranteed. Second, the guided filtering indicates an efficient implementation of the simulation process (see Sec. 4 in Ref. 25). Independent of the processing window size, the simulation process can be implemented through a moving sum method, which has the time complexity of $O(N)$ (N is the number of pixels in a given image).²⁵

2.5 Summarization of Proposed Method

The main difference between the proposed CS model and the traditional CS methods is that all key parameters $\{\beta_i, b_i, \mathbf{g}_i\}$ in Eqs. (1) and (6) are locally estimated. In contrast, the general CS-based method, such as GS,⁶ utilizes the constant parameters $\{\beta, b, \mathbf{g}\}$ for all pixels. It is also worthy noting that the proposed method is not the only method involving a localization strategy. For instance, the BDSD¹⁰ also conducts a CS process on each distinct local windows. A more related algorithm is proposed by Aiazzi et al.,¹ where a context-adaptive (CA) injection model (named as GSA-CA) is achieved by a sliding-window strategy. Although both GSA-CA and the proposed method involve the sliding-window processing step, the differences between them are still twofold. First, as shown in Eq. (8), after computing all potential injection weights through the sliding-window, our model further utilizes an average strategy over all the weights. A potential benefit from this average step is that the fusing result becomes less sensitive to the processing window size. Second, the low-resolution Pan band is also locally simulated, which was not considered by previous works.

Therefore, generally speaking, the novelty of the proposed method lies in the totally localized fusion process, which can provide a more accurate description on the spatial and spectral relationships between MS and Pan bands. The strengths and limitations of various localization strategies will be further discussed in the following experimental section.

3 Performance Evaluations and Discussions

3.1 Datasets

In our experiments, the proposed method is evaluated with different datasets acquired by GF-2 (four pairs), Quickbird (one pair), Worldview-2 (one pair) satellites, and an aerial platform (one pair).

The Pléiades dataset (Fig. 4), which was collected by the aerial platform, covers the urban area of Toulouse in France. This dataset was preprocessed and distributed by CNES, the French Space Agency.^{2,5} The original sizes of the Pan and MS images are 1024×1024 and $256 \times 256 \times 4$, respectively. This dataset is available in Ref. 27. The ground truth with both high spectral and spatial resolutions [Fig. 4(c)] was also provided for evaluation. This dataset is mainly used for evaluating the influence of parameters.

The detailed information on the remaining datasets are listed in Table 1. The datasets are representative, as they were collected from various landscapes with different satellite sensors. Especially, GF-2 is the first submeter civilian optical remote sensing satellite of China. The pansharpening of GF-2 images was also less considered in the previous literature.

Only MS bands, whose spectral range is overlapped with that of Pan band, are utilized for fusion. Therefore, red-edge, red, green, and blue bands were considered for the Worldview-2 dataset; near-infrared (NIR), red, green, and blue bands were considered for both GF-2 and Quickbird datasets. The testing datasets containing various land classes, such as water body, dense and sparse buildings, grass, trees, naked land, etc., are representative for the validation.

3.2 Evaluation Methods

For quantitative assessment, several metrics, including the average spectral angle change of all MS vectors (SAM),²⁸ the relative dimensionless global error in synthesis (ERGAS),²⁹ the quality with no reference (QNR),³⁰ and $Q2^n$,³¹ are calculated.

Table 1 The information of test datasets.

Sensor name	GF-2	Quickbird	Worldview-2
Spectral ranges of selected bands (μm)	Pan: 0.45 to 0.90	Pan: 0.45 to 0.90	Pan: 0.45 to 0.80
	Blue: 0.45 to 0.52	Blue: 0.45 to 0.52	Blue: 0.45 to 0.51
	Green: 0.52 to 0.59	Green: 0.52 to 0.60	Green: 0.51 to 0.58
	Red: 0.63 to 0.69	Red: 0.63 to 0.69	Red: 0.63 to 0.69
	NIR: 0.77 to 0.89	NIR: 0.76 to 0.89	Red edge: 0.71 to 0.75
Spatial resolution (m)	MS: 3.2 m	MS: 2.4 m	MS: 2.0 m
	Pan: 0.8 m	Pan: 0.6 m	Pan: 0.5 m
Radiometric resolution (bits)	11 bits	11 bits	11 bits
Image size (line $\times$ sample)	MS: 1250 $\times$ 1250	MS: 250 $\times$ 375	MS: 250 $\times$ 250
	Pan: 5000 $\times$ 5000	Pan: 1000 $\times$ 1500	Pan: 1000 $\times$ 1000
Number of datasets	Four pairs	One pair	One pair
Location	Guangzhou, China	Kunming, China	Hangzhou, China
Dataset name	GF2-1, GF2-2, GF2-3, and GF2-4	QB-KM	WV-HZ
Main land classes	Water body, trees, dense buildings, naked land	Trees, water body, and square	Sports ground

SAM calculates the spectral distort in terms of the change angle of spectral vectors, while ERGAS is the normalized version of the root mean square error between the fused image and the MS image. Both SAM and ERGAS values indicate the spectral fidelity of a method. The best value for both SAM and ERGAS is 0. Conversely, the best value for both $Q2^n$ and QNR is 1. QNR consists of two factors, that is, a spectral distort index and a spatial distort index. The highest value of QNR (=1) is obtained when the spectral and spatial distortions are both zero. $Q2^n$ takes three change factors into account, i.e., the correlation coefficient, the mean bias, and the contrast, between images. Both spatial and spectral distortions are considered by this metric.

Among four considered metrics, the calculation of SAM, ERGAS, and $Q2^n$ request an ideal reference image (the ground truth) with both high spectral and spatial resolutions. When the ideal reference is absent, an alternative way is to degrade the original Pan and MS images resolutions by a certain ratio, pansharpen the degraded images, and use the original MS images as a reference image.^{5,32} One common argument on this process is that the performance of a given method on a synthetic dataset can not necessarily be extended to a higher resolution case, especially when the modulation transfer function of the instruments is not available.^{2,32} A recent discussion on the issue can be found in Ref. 33. However, the issue is also beyond the scope of this research.

In this context, in the case of the absence of the ground truth, i.e., all datasets except for the Pléiades dataset, original (upsampled) MS, and Pan images are utilized for calculating the spectral and spatial related factors in SAM, ERGAS, and $Q2^n$, respectively. From this point of view, a metric score, which indicates both a spectral closeness to the original MS images and a spatial closeness to the original Pan image, is preferable.

Furthermore, visual inspection is also provided. Output images are identically stretched and representative subsets are selected for the clear visualization. As reference-dependent and reference-independent, quantitative and visual evaluation methods are employed, the effectiveness of the algorithms can be properly verified.

3.3 Results and Discussion

The experiments mainly consist of two parts. The first part is to verify the specialty of the proposed method by comparing global and local adaptive models. The second part is to evaluate the performance by comparing state-of-the-art methods.

3.3.1 Influence of parameters

The proposed pansharpening method has two parameters, including the window size $s1$ for Pan simulation and the window size $s2$ for detail injection. In this section, based on the Pléiades dataset, the influence of the two parameters on the performance of the proposed method is analyzed.

We totally examined 1640 parameter combinations of $s1$ and $s2$, according to $[3, 5, \dots, 83] \times [3, 5, \dots, 81]$, respectively. As the operation time is independent to the processing windows, which takes about 15 s for each parameter combination. The pansharpening qualities, in terms of QNR, $Q2^n$, ERGAS, and SAM against the different parameter combinations are calculated.

As shown in Fig. 3, the first 10% best combinations, i.e., the best 164 combinations, are labeled with blue balls and the rest are labeled with red stars. As can be observed, except the term SAM, a large Pan-simulation window $s1$, e.g., 83, is preferable according to the remaining three metrics; meanwhile, a large processing window $s2$, e.g., 81, is preferable to $Q2^n$, SAM, and ERGAS. When $s1$ or $s2$ is larger than a median number, e.g., 40, the quality scores of the four metrics change slowly and continuously.

Figures 4(d)–4(f) illustrate the fused images that were obtained according to the parameters corresponding to each best metric value. The visual gaps among these fused images are small. The closeness between the fused images and the ground truth [Fig. 4(c)] is also obvious. We also found that, this status lasts until $s1$ and $s2$ are as large as around half the size of the image, i.e., 512. Thereafter, QNR and $Q2^n$ scores decrease gradually, SAM and ERGAS scores increase gradually. Similar phenomena are also observed from the experiments that we conducted on the remaining datasets, where the original MS and Pan were the reference images. These may be attributed to the same spatial resolution ratio (i.e., 4) between all Pan and MS image pairs.

From the previous observations, we can conclude that the proposed method is not sensitive to the changes of the two size parameters. When $s1$ is the range [21, 83] and $s2$ is in the range [7, 81], all parameters can obtain acceptable pansharpening results. This property of the

Fig. 3 Selection of the optimal window sizes $s1$ for the Pan simulation and $s2$ for the CS according to different metrics: (a) SAM, (b) QNR, (c) $Q2^n$, and (d) ERGAS. The first 10% best combinations are labeled with blue, filled circles, and the rest is labeled with red stars.

Fig. 4 Subset of the Pléiades dataset and fused results with different parameter settings, shown with the same contrast stretch. (a) Low resolution MS with the band combination 4-3-2; (b) Panchromatic image; (c) ground truth with the band combination 4-3-2; and (d)–(f) the fused results with the proposed method with $[s1, s2] = [83, 7]$, $[83, 81]$, and $[21, 81]$, respectively.

algorithm should be attributed to the average of weights employed in both Pan simulation and components substitution models. Therefore, both values for $s1$ and $s2$ are fixed as $8 \times r + 1$ in the following experiments, where r is the spatial resolution ratio.

3.3.2 Comparison of global- and local-based CS methods

A general CS process consists of the Pan simulation and detail injection models.^{1,34} By considering different combinations of the simulation and injection models, the following six algorithms can be derived. The characteristics of methods are available in Table 2 and described as follows:

- GS: The first method is the general GS method,⁶ which utilizes the average of MS bands $\frac{1}{d} \sum_{j=1}^d \mathbf{M}_{i,j}$ to simulate a Pan band and the injection weights is globally calculated by the covariance of each MS band and the simulated Pan over the variance of the simulated Pan image $\text{Cov}(\mathbf{M}, \hat{\mathbf{P}}) / \text{Var}(\hat{\mathbf{P}})$.
- GSA: While the second method, i.e., the adaptive Gram–Schmidt (GSA),²⁰ simulates the Pan band by means of a linear regression algorithm and the same injection model as GS is employed.

Table 2 A list characterizing six methods derived from general CS framework.¹ The notations are defined in Sec. 2.1.

Name	Pan-simulation model		Detail injection model	
	Type	Method	Type	Method
GS	Global	$\hat{\mathbf{P}}_i = \frac{1}{d} \sum_{j=1}^d \mathbf{M}_{ij}$	Global	$\mathbf{g} = \text{Cov}(\mathbf{M}, \hat{\mathbf{P}}) / \text{Var}(\hat{\mathbf{P}})$ $\tilde{\mathbf{M}}_i = \mathbf{g}(\mathbf{P}_i - \hat{\mathbf{P}}) + \mathbf{M}_i$
GSA	Global	$\hat{\mathbf{P}}_i = \beta^T \mathbf{M}_i + b$	Global	$\mathbf{g} = \text{Cov}(\mathbf{M}, \hat{\mathbf{P}}) / \text{Var}(\hat{\mathbf{P}})$ $\tilde{\mathbf{M}}_i = \mathbf{g}(\mathbf{P}_i - \hat{\mathbf{P}}_i) + \mathbf{M}_i$
GSA_CA	Global	$\hat{\mathbf{P}}_i = \beta^T \mathbf{M}_i + b$	Local	$\mathbf{g}_i = \text{Cov}_i(\mathbf{M}, \hat{\mathbf{P}}) / \text{Var}_i(\hat{\mathbf{P}})$ $\tilde{\mathbf{M}}_i = \mathbf{g}_i(\mathbf{P}_i - \hat{\mathbf{P}}_i) + \mathbf{M}_i$
GSA_CA2	Global	$\hat{\mathbf{P}}_i = \beta^T \mathbf{M}_i + b$	Local	$\mathbf{g}_k = \text{Cov}_k(\mathbf{M}, \hat{\mathbf{P}}) / \text{Var}_k(\hat{\mathbf{P}})$ $\bar{\mathbf{g}}_i = \sum_{k \in w_i} \mathbf{g}_k / w_i $ $\tilde{\mathbf{M}}_i = \bar{\mathbf{g}}_i(\mathbf{P}_i - \hat{\mathbf{P}}_i) + \mathbf{M}_i$
LCS_P	Local	$\bar{\beta}_i = \sum_{k \in w_i} \beta_k / w_i $ $\bar{\mathbf{P}}_i = \bar{\beta}_i^T \mathbf{M}_i + \bar{b}_i$	Global	$\mathbf{g} = \text{Cov}(\mathbf{M}, \hat{\mathbf{P}}) / \text{Var}(\hat{\mathbf{P}})$ $\tilde{\mathbf{M}}_i = \mathbf{g}(\mathbf{P}_i - \hat{\mathbf{P}}_i) + \mathbf{M}_i$
LCS	Local	$\bar{\beta}_i = \sum_{k \in w_i} \beta_k / w_i $ $\bar{\mathbf{P}}_i = \bar{\beta}_i^T \mathbf{M}_i + \bar{b}_i$	Local	$\mathbf{g}_k = \text{Cov}_k(\mathbf{M}, \hat{\mathbf{P}}) / \text{Var}_k(\hat{\mathbf{P}})$ $\bar{\mathbf{g}}_i = \sum_{k \in w_i} \mathbf{g}_k / w_i $ $\tilde{\mathbf{M}}_i = \bar{\mathbf{g}}_i(\mathbf{P}_i - \bar{\mathbf{P}}_i) + \mathbf{M}_i$

- GSA_CA: The third method is called context-adaptive GSA (GSA_CA),¹ where the global injection model in GSA is superseded by a local sliding-window-based strategy, and each $\mathbf{g}_i$ is locally calculated from the square window centered at i according to $\text{Cov}_i(\mathbf{M}, \hat{\mathbf{P}}) / \text{Var}_i(\hat{\mathbf{P}})$. The window size is set as 32×32 .
- GSA_CA2: The fourth method is named GSA-CA2. $\mathbf{g}_i$ in GSA_CA is replaced by $\bar{\mathbf{g}}_i$ in GSA_CA2, where $\bar{\mathbf{g}}_i$ is defined as $\sum_{k \in w_i} \mathbf{g}_k / |w_i|$, the average of all potential values associated with the pixel i .
- LCS_P and LCS: The fifth method replaces the Pan-simulation model of GSA with the proposed local-based Pan simulation (see Sec. 2.2), and the same injection model as GS is employed. This method is called the partially local-based CS method (LCS_P). The sixth method is the proposed totally local-based CS method (LCS).

As summarized in Table 2, all six algorithms can be divided into two groups, where GS and GSA belong to the global-based group and the rest belong to (partially) the local-based group. Figure 5 shows quantitative scores of the pansharpened images generated by the six methods.

Generally speaking, the local-based methods are superior to the global-based methods across all datasets and all metrics. This observation confirms the effectiveness of local-based processing strategies. In the first group, GSA is superior to GS in most cases, which indicates the necessity of simulating the Pan band by means of the linear regression algorithm, rather than a simple average of MS bands. In the second group, the synergism of the two local-based models, i.e., LCS, can create better results than any individual application of the two models. In most cases, LCS_P and LCS are superior to GSA_CA and GSA_CA2 in terms of the four metrics.

As GSA, GSA_CA, and GSA_CA2 share the same Pan-simulation model, by comparing the performance of them, the strengths and limitations of the local-based injection model can be revealed. As shown in Fig. 6, the employment of the local-based injection strategy, i.e., GSA_CA (e) and GSA_CA2 (f), leads to overall improvement on the spectral quality over the global method GSA.

Fig. 5 Quality scores of the pansharpened images generated by six model combinations (GS, GSA, GSA_CA, GSA_CA2, LCS_P, and LCS), based on six datasets (GF2-1, GF2-2, GF2-3, GF2-4, QB-KM, and WV-HZ). (a) QNR, (b) $Q2^n$, (c) ERGAS, and (d) SAM. The ideal values of the four metrics are 1, 1, 0, and 0, respectively.

Fig. 6 Subset of the GF2-1 dataset, fused results, and local details according to different combinations of the Pan-simulation and CS models, shown with the band combination 4-3-2. (a) MS, (b) Pan, (c) GS, (d) GSA, (e) GSA_CA, (f) GSA_CA2, (g) LCS_P, (h) LCS, (i) local patches of (a)–(h) for a detailed comparison. The meaning of abbreviations is described in details in Sec. 3.3.2.

The difference of GSA_CA2 over GSA_CA in terms of the four metrics seems to be not significant, the GSA_CA is even slightly better than GSA_CA in the terms of QNR. But the details shown in Fig. 6(i) can illustrate the differences among them. Although the GSA_CA result has achieved overall good quality, it still has severe spectral distortion

along the narrow boundaries of the roof, the narrow boundary area in the GSA_CA result is severely contaminated by the spectra of the surrounding tree pixels. This observation indicates the limitation of the local-based processing strategies. If the land cover type of pixels in the window is dominant by one class, whereas the pixel located at the center of the window belongs to another, the injection weights that are estimated by the whole window, can only represent the relationship between MS and Pan bands of the dominant class, rather than the relationship of the center pixel. In contrast, GSA_CA2 can alleviate this spectral distortion to some extent due to the average strategy embedded in the method. As the injection weights associated with a given pixel are calculated from multiple windows that cover the pixel.

It is also obvious in Fig. 6 that the global injection model can totally avoid this problem. The reason is that, regardless of the spatial position, all pixels are treated with the same injection weights. If only two pixels are similar in spectra, the corresponding fused pixels are also similar. However, because the injection weights that have been globally estimated cannot accurately fit the relationship among Pan and MS bands in a specified local area, the spectral fidelity between the original pixel and the fused counterpart cannot be assured, and the whole image suffers from over-all spectral distortion.

As GS, GSA, and LCS_P share the same global injection model, by comparing the performance of them, the property of the local-based Pan-simulation model can also be revealed. According to Figs. 5 and 6, LCS_P always provides better results than GS and GSA. This observation indicates the local-based Pan-simulation method can guarantee the local closeness of structures in the substituted and substitution components, and thus, benefit the final output.

In summary, the local-based Pan-simulation model can assure the local closeness of structures in the substituted and substitution components. Meanwhile, the local-based injection model can more accurately estimate the relationship between the MS and Pan than the global estimation model, and thus the spatial details can be properly added back to the original MS bands.

Fig. 7 Local patches from GF2-2 dataset. (a) Fused results and (b) corresponding edges extracted by the canny operator. From the left to right and from up to down, the patches are responding to MS, Pan, UNB,⁸ SRM,⁹ PRA,³⁵ NND,¹⁵ GSA_CA,¹ BDSD,¹⁰ and LCS, respectively. The spatial structures of the fused results can be appreciated in the edge maps, and BDSD presents a block effect.

However, even in a small window, the center pixel still may be not originated from the land class that is dominating the whole window. Therefore, the general local-based injection model, e.g., that embedded in GSA_CA,¹ may suffer from spectral distortion in the area of small structures. This problem can be partly addressed through the local average strategy that is utilized after the general local-based injection model. One potential solution to this problem is to further divide the local patch into different groups, and establish models for each group, respectively.

3.3.3 Comparison with state-of-the-art methods

In this section, based on six datasets listed in Table 1, we compared the proposed local-based CS method (LCS) with several representative methods, including UNB-Pansharpening (UNB),⁸ SRM,⁹ the partial replacement adaptive CS (PRA),³⁵ NND,¹⁵ GSA_CA,¹ and BDSD.¹⁰ All considered methods either involve the simulation of a low-resolution Pan band or include a CS step, and are known for their good performance in spectral fidelity. For the detailed descriptions on algorithms, refer to related papers.

Among them, UNB, SRM, and NND have been embedded in the commercial software, PCI, ERDAS, and ENVI, respectively. PRA and BDSD are implemented in Matlab and shared through the website available in Ref. 27. For BDSD, in order that the size is an aliquot part of the row and column dimensions of Pan, the default value 128×128 was used for the Pléiades dataset, and the window size was set as 100×100 for the remaining six datasets. The default parameters and settings recommended by the related papers or software are accepted for the remaining methods.

Visual inspections across various datasets and landscapes were first conducted. Some local patches composed of various land cover types are illustrated in Figs. 7–10. By extracting edges

Fig. 8 Comparison of fused results over the vegetation area in the GF2-2 dataset, displayed with true colors and the same contrast stretch. (a) MS, (b) Pan, (c) UNB,⁸ (d) SRM,⁹ (e) PRA,³⁵ (f) NND,¹⁵ (g) GSA_CA,¹ (h) BDSD,¹⁰ (i) LCS, and (j) local patches of (a)–(i) for a detailed comparison.

Fig. 9 Comparison of fused results over the trees and buildings in the GF2-1 dataset, shown with NIR, R, and G bands, and the same contrast stretch. (a) MS, (b) Pan, (c) UNB,⁸ (d) SRM,⁹ (e) PRA,³⁵ (f) NND,¹⁵ (g) GSA_CA,¹ (h) BDSD,¹⁰ and (i) LCS.

from the fused outputs, the methods' performance with respect to the spatial enhancement was assessed in Fig. 7. In detail, each fused result is first converted into a one-band gray image, and the canny detector³⁶ is applied to get edges. As can be seen, the edges extracted from all methods are quite consistent to the edges extracted from the original PAN image; therefore, all methods can properly integrate the spatial details of the Pan band. Whereas BDSD presents a block effect, where abrupt spectral transitions and spatial artifacts can be seen around block borders. This effect also can be observed from the forest details [Fig. 8(j)] and the white building area [Fig. 9(h)]. The cause of this effect is the block-by-block processing strategy utilized by BDSD, which ignores the spatial relationship among adjacent blocks. In contrast, the proposed LCS method is conducted in a sliding-window and local-average fashion, which always exhibits gradual changes.

All pansharpening methods can enhance spatial details effectively, but their performances on preserving spectral property are quite different. As shown in Fig. 8, with the true color viewing mode, UNB (c), SRM (d), PRA (e), and NND (f) have obvious spectral distortions on the vegetation regions; whereas GSA_CA (g), BDSD (h), and LCS (i) presents a similar spectral property as the original MS image. The spectral difference among GSA_CA, BDSD, and the

Fig. 10 Comparison of fused results over the dense buildings in the GF2-3 dataset, shown with NIR, R, and G bands, and the same contrast stretch. (a) MS, (b) Pan, (c) UNB,⁸ (d) SRM,⁹ (e) PRA,³⁵ (f) NND,¹⁵ (g) GSA_CA,¹ (h) BDSD,¹⁰ and (i) LCS.

proposed LCS [Figs. 8(g)–8(i)] is not remarkable in this view mode. In contrast, with the NIR-red-green viewing mode (Fig. 9), PRA (e), BDSD (h), and LCS (i) have slight color distortions in vegetation regions, and GSA_CA (g) has severe spectral distortion to the original MS image (a). The outputs of the counterpart methods over the dense building are shown in Fig. 10. The spectral fidelity of PRA (e), GSA_CA (g), BDSD (h), and LCS (i) is higher than the remaining methods.

Based on the above observations, we can conclude that the proposed LCS presents a similar spectral property as the original MS images across all kinds of land classes and viewing modes. PRA and BDSD provide comparable results as LCS.

In order to further evaluate the effectiveness of the proposed method, the quantitative comparison to the state-of-the-art methods is also reported in Fig. 11. The proposed method can

Fig. 11 Quality scores of the images generated by the seven methods (i.e., UNB,⁸ SRM,⁹ PRA,³⁵ NND,¹⁵ GSA_CA,¹ BDSD,¹⁰ and LCS), based on six datasets (GF2-1, GF2-2, GF2-3, GF2-4, QB-KM, and WV-HZ). (a) QNR, (b) $Q2''$, (c) ERGAS, and (d) SAM. The ideal values of the four metrics are 1, 1, 0, and 0, respectively.

achieve the best or second-best results over most of the cases; BDSD and PRA have slightly poor performance. These observations are quite consistent with the visual inspection. However, a conflicting conclusion can be drawn from the SAM metric for the NND method. NND has significantly better performance than other methods in terms of SAM over all datasets, whereas the performance judged by visual inspection is not unsatisfactory. As can be seen in Figs. 8(f)–10(f), NND tends to obtain fused results with serious spectral distortion.

Moreover, each group of metric values associated with a specific dataset and a fusion method is labeled as high-, medium-, or low-performance by a K -means clustering process. Thereafter, the overall performance of each pansharpening method over all six datasets was collected and is reported in Table 3. It is clear that PRA, BDSD, and the proposed LCS can be divided into the high-performance group, as they have achieved at least three high over all six datasets; GS, UNB, and SRM belong to the middle performance group; whereas NND should be grouped into the low performance.

We have also tested the computational time using a Windows 10 laptop with an Intel I7 2.7-GHz CPU using 16-GB memory. In this computing architecture, the processing time of the three high-performance methods (i.e., BDSD, PRA, and LCS methods) were compared and times in second are reported in Table 4. The proposed method takes almost five times as long to execute PRA, and more than seven times as long to execute BDSD. It is quite natural, as the method involves intensive localized operations. The first time-consuming part, which takes about 75% of the whole execution time, is calculating the inverse of the local covariance matrix in Eq. (3).

Table 3 The overall performance of each pansharpening algorithm on six datasets, where H, M, and L denote high, medium, and low performances, respectively.

PRA	LCS	BDSD	GSA_CA	UNB	SRM	NND
4H,2M	4H,2M	3H,3M	2H,4M	1H,4M,1L	1H,2M,3L	1M,5L

Table 4 The time cost (in second) of pansharpening methods over different datasets.

Image name	Image size	BDS	PRA	LCS
GF2-1	Pan: 5000 × 5000	11.35	49.13	230.35
	MS: 1250 × 1250			
QB-KM	Pan: 1000 × 1500	1.75	2.74	12.54
	MS: 250 × 375			
WV-HZ	Pan: 1000 × 1000	0.43	1.70	8.32
	MS: 250 × 250			

This calculation should repeat N times (N is the number of pixels of the whole image). The second time-consuming part of the method, which takes about 15% of the time, is the local average operations in Eqs. (5) and (8).

4 Conclusions and Perspectives

This paper proposes a context adaptive pansharpening method, which consists of two local-based processing blocks: Pan simulation and CS. Owing to its capability to accurately model the relationship between MS and Pan bands and to correctly extract the spatial details coming from the Pan image, the proposed method can more efficiently enhance spatial quality and presents better spectral quality than counterpart methods. In particular, the proposed simulation and injection strategy should be given priority in case of requiring high spectral fidelity.

However, the proposed algorithm is quite time-consuming as it involves intensive local operations. Fortunately, due to basic operations being highly parallel, the implementation of the proposed algorithm using parallel processing techniques is our on-going work, which may significantly decrease the processing time. Future works will also explore context adaptive models, which take both spectral and spatial homogeneity into account.

Acknowledgments

The authors would like to express sincere gratitude to Dr. Luciano Alparone and Gemine Vivone for sharing their pansharpening toolbox. The authors would like to thank Dr. Chen Zheng for the valuable discussions and suggestions. The authors also would like to thank the editor and anonymous reviewers for their insightful suggestions, which significantly improved this paper. This research was jointly supported by the National Natural Science Foundation of China under grant nos. 41571372, 41771375, and 41301470. The authors declare no conflict of interest.

References

1. B. Aiazzi et al., "A comparison between global and context-adaptive pansharpening of multispectral images," *IEEE Geosci. Remote Sens. Lett.* **6**(2), 302–306 (2009).
2. L. Alparone et al., "Comparison of pansharpening algorithms: outcome of the 2006 GRS-S data-fusion contest," *IEEE Trans. Geosci. Remote Sens.* **45**(10), 3012–3021 (2007).
3. X. Kang, S. Li, and J. A. Benediktsson, "Pansharpening with matting model," *IEEE Trans. Geosci. Remote Sens.* **52**(8), 5088–5099 (2014).
4. Q. Xu et al., "Pansharpening using regression of classified MS and Pan images to reduce color distortion," *IEEE Geosci. Remote Sens. Lett.* **12**(1), 28–32 (2015).
5. G. Vivone et al., "A critical comparison among pansharpening algorithms," *IEEE Trans. Geosci. Remote Sens.* **53**(5), 2565–2586 (2015).
6. C. A. Laben and B. V. Brower, "Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening," U.S. Patent No. 6,011,875 (2000).

7. A. R. Gillespie, A. B. Kahle, and R. E. Walker, "Color enhancement of highly correlated images. II-channel ratio and 'chromaticity' transformation techniques," *Remote Sens. Environ.* **22**(3), 343–365 (1987).
8. Y. Zhang, "System and method for image fusion," U.S. Patent No. 7,340,099 (2008).
9. Y. Zhang and R. K. Mishra, "A review and comparison of commercially available pan-sharpening techniques for high resolution satellite image fusion," in *IEEE Int. Geoscience and Remote Sensing Symp. (IGARSS '12)*, Munich, pp. 182–185 (2012).
10. A. Garzelli, F. Nencini, and L. Capobianco, "Optimal MMSE pan sharpening of very high resolution multispectral images," *IEEE Trans. Geosci. Remote Sens.* **46**(1), 228–236 (2008).
11. G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science* **313**(5786), 504–507 (2006).
12. W. Huang et al., "A new pan-sharpening method with deep neural networks," *IEEE Geosci. Remote Sens. Lett.* **12**(5), 1037–1041 (2015).
13. G. Masi et al., "Pansharpening by convolutional neural networks," *Remote Sens.* **8**(7), 594 (2016).
14. C. Dong et al., "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.* **38**(2), 295–307 (2016).
15. W. Sun, B. Chen, and D. W. Messinger, "Nearest-neighbor diffusion-based pan-sharpening algorithm for spectral images," *Opt. Eng.* **53**(1), 013107 (2014).
16. H. Zhang and B. Huang, "A new look at image fusion methods from a Bayesian perspective," *Remote Sens.* **7**(6), 6828–6861 (2015).
17. B. Xie, H. Zhang, and B. Huang, "Revealing implicit assumptions of the component substitution pansharpening methods," *Remote Sens.* **9**(5), 443 (2017).
18. R. Haydn et al., "Application of the IHS color transform to the processing of multisensor data and image enhancement," in *Proc. of the Int. Symp. on Remote Sensing of Environment, First Thematic Conf.: Remote Sensing of Arid and Semi-Arid Lands*, Cairo, Egypt (1982).
19. S. Rahmani et al., "An adaptive IHS pan-sharpening method," *IEEE Geosci. Remote Sens. Lett.* **7**(4), 746–750 (2010).
20. B. Aiazzi, S. Baronti, and M. Selva, "Improving component substitution pansharpening through multivariate regression of MS+Pan data," *IEEE Trans. Geosci. Remote Sens.* **45**(10), 3230–3239 (2007).
21. Y. Zhang, "Understanding image fusion," *Photogramm. Eng. Remote Sens.* **70**(6), 657–661 (2004).
22. Q. Xu et al., "High-fidelity component substitution pansharpening by the fitting of substitution data," *IEEE Trans. Geosci. Remote Sens.* **52**(52), 7380–7392 (2014).
23. C. Thomas et al., "Synthesis of multispectral images to high spatial resolution: a critical review of fusion methods based on remote sensing physics," *IEEE Trans. Geosci. Remote Sens.* **46**(5), 1301–1312 (2008).
24. H. Wang et al., "A robust image fusion method based on local spectral and spatial correlation," *IEEE Geosci. Remote Sens. Lett.* **11**(2), 454–458 (2014).
25. K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. Pattern Anal. Mach. Intell.* **35**(6), 1397–1409 (2013).
26. T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning Data Mining, Inference and Prediction*, Springer, New York (2009).
27. L. Alparone and G. Vivone, "Remote sensing image pansharpening toolbox," <http://openremotesensing.net/knowledgebase/a-critical-comparison-among-pansharpening-algorithms/> (July 2017).
28. R. H. Yuhas, A. F. H. Goetz, and J. W. Boardman, "Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm," in *JPL, Summaries of the Third Annual JPL Airborne Geoscience Workshop* (1992).
29. T. Ranchin and L. Wald, "Fusion of high spatial and spectral resolution images: the ARSIS concept and its implementation," *Photogramm. Eng. Remote Sens.* **66**, 49–61 (2000).
30. L. Alparone et al., "Multispectral and panchromatic data fusion assessment without reference," *Photogramm. Eng. Remote Sens.* **74**(74), 193–200 (2008).
31. A. Garzelli and F. Nencini, "Hypercomplex quality assessment of multi/hyperspectral images," *IEEE Geosci. Remote Sens. Lett.* **6**(4), 662–665 (2009).

32. F. Laporterie-Déjean et al., "Thematic and statistical evaluations of five panchromatic/multi-spectral fusion methods on simulated PLEIADES-HR images," *Inf. Fusion* **6**(3), 193–212 (2005).
33. F. Palsson et al., "Quantitative quality evaluation of pansharpened imagery: consistency versus synthesis," *IEEE Trans. Geosci. Remote Sens.* **54**(3), 1247–1259 (2016).
34. B. Aiazzi et al., "Context-driven fusion of high spatial and spectral resolution images based on oversampled multiresolution analysis," *IEEE Trans. Geosci. Remote Sens.* **40**(10), 2300–2312 (2002).
35. J. Choi, K. Yu, and Y. Kim, "A new adaptive component-substitution-based satellite image fusion by using partial replacement," *IEEE Trans. Geosci. Remote Sens.* **49**(1), 295–309 (2011).
36. J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. PAMI* **8**(6), 679–698 (1986).

Qinling Dai received her MS degree in the School of Printing and Packaging from Xi'an University of Technology in 2007. She is currently pursuing her PhD in graphic communication engineering at Wuhan University, Wuhan, China. She is engaged in digital image processing and understanding.

Bin Luo received his MSc degree from ENS Cachan, France, in 2003, and his PhD from ENST Paris, France, in 2007. He worked as a postdoctoral researcher in Grenoble Images Speech Signals and Automatics Laboratory, Grenoble, France, from 2008 to 2010. Currently, he is a professor in the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing (LIESMARS) at Wuhan University, Wuhan, China. His research interests include hyperspectral data analysis and high-resolution remote sensing image processing.

Leiguang Wang received his PhD in LIESMARS from Wuhan University in 2009. Since 2012, he has been an associate professor with Southwest Forestry University, Kunming, China. He is the author of more than 10 articles. His research interests include remote sensing image segmentation and pattern recognition.

Zhigang Tu started his MPhil PhD in image processing at the School of Electronic Information, Wuhan University, China, 2008. He received his PhD in communication and information systems from Wuhan University, 2013. In 2015, he received his PhD in computer science from Utrecht University, the Netherlands. From 2015 to 2016, he did one year of postdoctoral research at Arizona State University, USA. He is currently a postdoctoral research fellow at the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. His research interests include motion estimation, object segmentation, object tracking, action recognition, and anomaly detection.