



Article

A Pan-Sharpening Method with Beta-Divergence Non-Negative Matrix Factorization in Non-Subsampled Shear Transform Domain

Yuetao Pan ¹, Danfeng Liu ^{1,*}, Ligu Wang ¹, Jón Atli Benediktsson ² and Shishuai Xing ¹

¹ College of Information and Communication Engineering, Dalian Minzu University, Dalian 116600, China; panyuetao@dlnu.edu.cn (Y.P.); wangliguo@hrbeu.edu.cn (L.W.); xingshishuai@dlnu.edu.cn (S.X.)

² Faculty of Electrical and Computer Engineering, University of Iceland, 107 Reykjavik, Iceland; benedikt@hi.is

* Correspondence: liudanfeng@dlnu.edu.cn

Abstract: In order to combine the spectral information of the multispectral (MS) image and the spatial information of the panchromatic (PAN) image, a pan-sharpening method based on β -divergence Non-negative Matrix Factorization (NMF) in the Non-Subsampled Shearlet Transform (NSST) domain is proposed. Firstly, we improve the traditional contrast calculation method to build the weighted local contrast measure (WLCM) method. Each band of the MS image is fused by a WLCM-based adaptive weighted averaging rule to obtain the intensity component I . Secondly, an image matting model is introduced to retain the spectral information of the MS image. I is used as the initial α channel to estimate the foreground color F and the background color B . Depending on the NSST, the PAN image and I are decomposed into one low-frequency component and several high-frequency components, respectively. Fusion rules are designed corresponding to the characteristics of the low-frequency and high-frequency components. A β -divergence NMF method based on the Alternating Direction Method of Multipliers (ADMM) is used to fuse the low frequency components. A WLCM-based rule is used to fuse the high-frequency components. The fused components are inverted by NSST inverse transformation, and the obtained image is used as the final α channel. Finally, the final fused image is reconstructed according to the foreground color F , background color B , and the final α channel. The experimental results demonstrate that the proposed method achieves superior performance in both subjective visual effects and objective evaluation, and effectively preserves spectral information while improving spatial resolution.

Keywords: non-subsampled shearlet transform (NSST); weighted local contrast measure (WLCM); image matting model; alternating direction method of multipliers (ADMM)



Citation: Pan, Y.; Liu, D.; Wang, L.; Benediktsson, J.A.; Xing, S. A Pan-Sharpening Method with Beta-Divergence Non-Negative Matrix Factorization in Non-Subsampled Shear Transform Domain. *Remote Sens.* **2022**, *14*, 2921. <https://doi.org/10.3390/rs14122921>

Academic Editor: Edoardo Pasolli

Received: 18 May 2022

Accepted: 16 June 2022

Published: 18 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Multi-source remote sensing satellites can provide numerous remote sensing images with different spatial, spectral, and temporal resolutions. The panchromatic (PAN) image has a high spatial resolution and can reflect the overall spatial structure and detail the features of the remote sensing image. The multispectral (MS) image contains rich spectral information which can be used in various fields, such as the identification, classification, and interpretation of ground objects. By fusing the MS image with rich spectral information and the PAN image with high spatial resolution, a fused image with high spatial and spectral resolution can be obtained. This process is also called pan-sharpening. The pan-sharpening method can be used to obtain a fused image that contains more complete and richer information than a single type of remote sensing image. It is widely used in land use planning, vegetation cover analysis, earth resources surveying, and other fields, as shown in Figure 1.

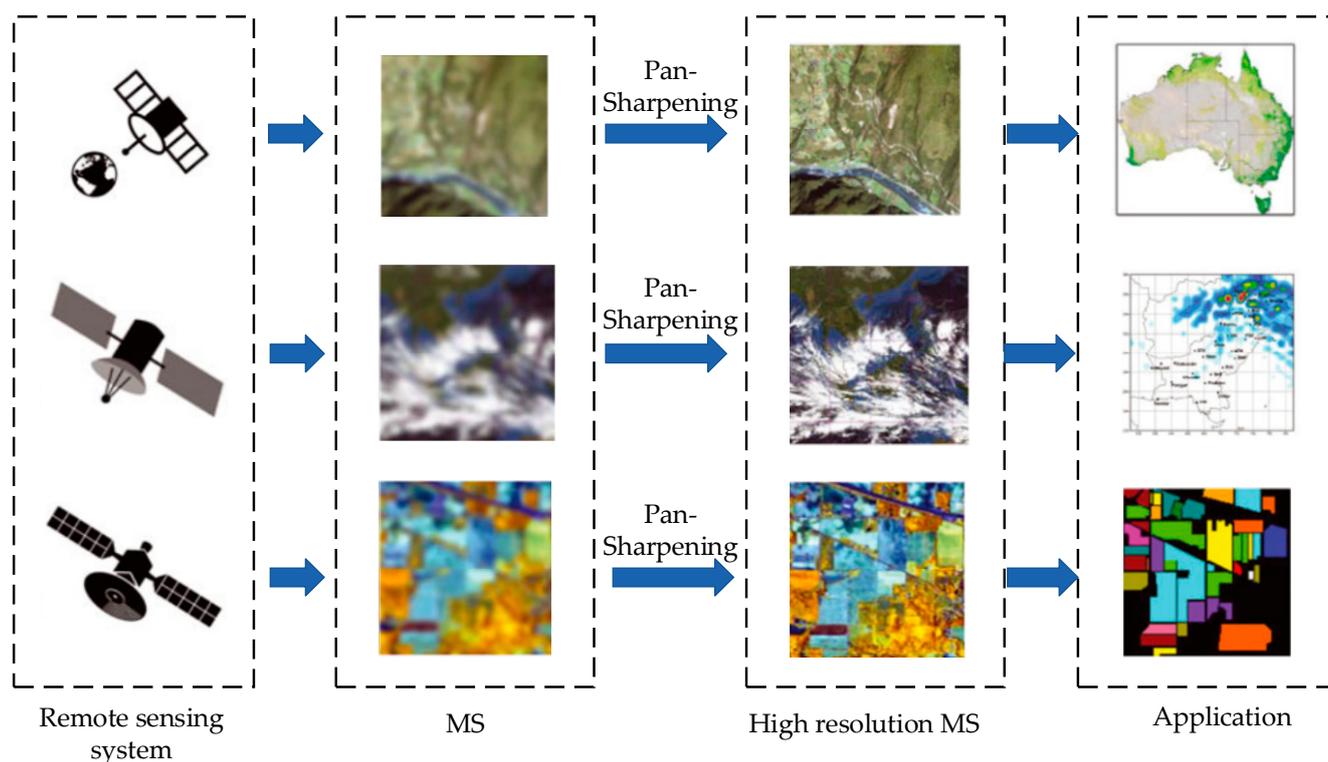


Figure 1. Application of the pan-sharpening method.

In the past, many pan-sharpening methods have been proposed to fuse the MS and PAN images. The component substitution-based (CS) methods can transform the MS image into a new projection space. In the new projection space, the MS image is decomposed into spectral and spatial components and then the spatial component is replaced with the PAN image. Finally, a fused image can be obtained by inverse transform. The CS-based methods mainly include the intensity–hue–saturation (IHS) transform method [1,2], principal component analysis-based (PCA) method [3], Gram–Schmidt-based (GS) method, adaptive GS (GSA) method [4], band-dependent spatial-detail with physical constrains (BDSD-PC) method [5], partial replacement-based adaptive component substitution (PRACS) method [6], etc. The CS-based methods operate more efficiently, and the spatial details of the fused image are clearer. They are robust to alignment errors and blending mistakes.

The multi-resolution analysis-based (MRA) methods inject the spatial details obtained from the PAN image by multi-resolution decomposition into the MS image, such as the wavelet transform-based method [7] and the additive wavelet luminance proportional (AWLP) method [8], etc. Compared with the CS-based methods, the MRA-based methods can maintain the spectral information of the MS image, but it will cause spatial structure distortion in the fused image.

Moreover, Fu et al. [9] proposed a variational local gradient constraints-based (VLGC) pan-sharpening method that can make full use of the spatial information contained in the PAN image. Wu et al. [10] proposed a new multi-objective decision-based (MOD) pan-sharpening method. This method models the parameters from a multi-objective perspective while maximizing the quality of all the pixels in the fused image. Khan et al. [11] proposed a pan-sharpening method that combines Brovey and Laplacian filter. The Laplacian edge sharpening plays an important role in enhancing edge contrast and improving image visibility. Li et al. [12] proposed a pan-sharpening method based on a guided filter method. This method first decomposes the MS and PAN images into high-frequency and low-frequency components. Then, a guided filter is used to enhance the spectral information.

While the existing pan-sharpening methods perform well in many aspects, there are still some areas in need of improvement. For example, the CS-based and MRA-based

methods can improve the spatial resolution of the fused images but cause spectral distortion in the fused images. Moreover, the deep learning-based methods often require a large number of training datasets, but specialized remote sensing datasets are scarce. Different satellites have different data types, so it is difficult to train different satellite datasets at the same time. In addition, it takes substantial time to train the network model.

To solve the spatial and spectral distortion problems in the remote sensing image fusion process, we propose a pan-sharpening method based on an image matting model, a Non-Subsampled Shearlet Transform (NSST), and an Alternating Direction Method of Multipliers-based (ADMM) β -divergence Non-negative Matrix Factorization (NMF). The proposed method combines the advantages of the CS-based and MRA-based methods, and it mainly comprises the following three processes:

Firstly, inspired by the superior spectral preservation ability of an image matting model, this model is introduced into the process of pan-sharpening. However, in the process of remote sensing imaging, due to different signal-to-noise ratio, the characteristics of the MS and PAN images are not exactly the same. If the PAN image is directly used as the α channel, spectral distortion will occur in the fused image. Thus, we improve the traditional local contrast measurement method and establish a weighted local contrast measure (WLCM) method. According to the WLCM method, each band of the MS image is fused to obtain the intensity component I as the initial α channel.

Then, the NSST decomposition is performed on the MS and PAN images separately, and a low-frequency component and several high-frequency components are obtained. Based on the NSST decomposition, two different fusion rules are designed according to the different characteristics of the high-frequency and low-frequency components. The high-frequency components contain rich detailed information of the source image, such as edges and textures. Thus, a WLCM-based rule is adopted to fuse the high-frequency coefficients. The low-frequency component is the approximation of the original image and describes the basic structure of the original image. Thus, an ADMM-based β -divergence NMF method is used to fuse the low frequency components.

Finally, the image fused by the PAN image and I is used as the final α channel. According to an image matting model, a fused image with high spatial and spectral resolution can be reconstructed based on the foreground color F , background color B , and the final α channel.

The main contributions of the method proposed in this paper are as follows:

- (1) An image matting model is introduced in the fusion process, which can effectively maintain the spectral resolution of the MS image.
- (2) A NSST is introduced in the multi-resolution analysis process, which has the advantages of multi-scale, multi-direction, and translation invariance. In addition, the NSST can overcome the pseudo-Gibbs effect when reconstructing images and can capture more feature information of the source image.
- (3) The low-frequency components are fused according to an ADMM-based β -divergence NMF method. Moreover, the ADMM-based β -divergence NMF method has a faster convergence speed and better solution results.
- (4) The traditional local contrast measure method is improved and a WLCM method is proposed in this paper. Initially, the local contrast measure value is calculated using the median of the neighborhood. Then, the mean of the difference between the local pixel values and the middle pixel value is introduced to weight the local contrast measure value. The WLCM method can enhance the faint spatial details and suppress the irrelevant backgrounds, which in turn improves the detection rate of detailed information and ultimately enhances the fusion effect.

The rest of this paper is organized as follows. Section 2 introduces the principles of NSST decomposition, image matting model, ADMM-based β -divergence NMF, and WLCM. Section 3 describes the detailed steps and principles of the proposed method. Section 4 conducts experiments and comparative analysis. Finally, Section 5 consists of the conclusion and some future plans.

2. Materials and Methods

2.1. NSST Decomposition

A Shear wave is a special case of continuous wavelet [13]. In two dimensions, it is defined as follows:

$$Q_{LJ}(\psi) = \left\{ \psi_{m,n,k}(x) = |\det L|_{m/2} \psi(J^n L^m x - k) : m, n \in M, k \in M^2 \right\} \quad (1)$$

where ψ is a function set and $\psi \in D^2(T^2)$. $D^2(\cdot)$ represents the two-dimensional energy finite function space; T denotes a real number; M denotes an integer; L represents the anisotropic matrix of multi-scale partition; J represents the shear matrix that is used for directional analysis; m , n , and k are scale, direction, and shift parameters, respectively. If $\forall f \in D^2(T^2)$, satisfying $\sum_{m,n,k} |\langle f, \psi_{m,n,k} \rangle| = \|f\|^2$, then the elements of $Q_{LJ}(\psi)$ are called synthetic wavelets.

When $a > 0$ and $J \in R$, L and J are 2-order invertible matrices as follows:

$$L = \begin{bmatrix} l & 0 \\ 0 & \sqrt{l} \end{bmatrix}, J = \begin{bmatrix} 1 & j \\ 0 & 1 \end{bmatrix} \quad (2)$$

Suppose that $l = 4$ and $j = 1$, the continuous wavelet is the Shearlet. Shearlet is a special case of synthetic wavelets, when L is an expansion matrix with anisotropy and J is a shear matrix. L and J are as follows.

$$L = \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix}, J = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \quad (3)$$

The discretization process of NSST is mainly divided into two steps: multiscale decomposition and direction localization. In the process of multi-scale decomposition, Non-Subsampled Laplace Pyramid transform (NSLP) [14] is adopted. Thus, it has translational invariance. If we perform m -scale decomposition, we can get $m + 1$ components with the same size as the original image, including m high-frequency components and one low-frequency component. The multi-directional decomposition is realized by the improved Shearlet filters (SF). This maps the standard SF in the Shearlet from pseudo-polarized coordinates to Cartesian coordinates, and the whole process is achieved directly by two-dimensional convolution to avoid down-sampling operations and to make it translation invariant. These filters are formed by avoiding secondary sampling to satisfy shift invariance. Thus, NSST has the advantages of structural simplicity, multi-scale, multi-directional, and translational invariance. Figure 2 shows a three-level NSST decomposition model.

2.2. Image Matting Model

According to an image matting model [15], an image can be divided into foreground and background color. This means that the color of the i -th pixel is the linear combination of the corresponding foreground color F_i and background color B_i . The details are as follows:

$$Z_i = \alpha_i F_i + (1 - \alpha_i) B_i \quad (4)$$

where F_i is the foreground color of the i -th pixel, B_i is the background color of the i -th pixel, and α is the opacity of F . When determining the input image Z and the α channel, the foreground color F and background color B can be estimated by solving the following function:

$$\min_i \sum_k (\alpha_i F_i^k + (1 - \alpha_i) B_i^k)^2 + |\alpha_{ix}| ((F_{ix}^k)^2 + (B_{ix}^k)^2) + |\alpha_{iy}| ((F_{iy}^k)^2 + (B_{iy}^k)^2) \quad (5)$$

where i denotes the i -th channel of the input image Z , and F_{ix}^k and F_{iy}^k are the horizontal and vertical derivatives of the foreground color F^k , respectively. B_{ix}^k and B_{iy}^k are the horizontal

and vertical derivatives of the background color B^k , respectively. α_{ix} and α_{iy} are the horizontal and vertical derivatives of the α channel, respectively. For more details, please refer to the literature [15].

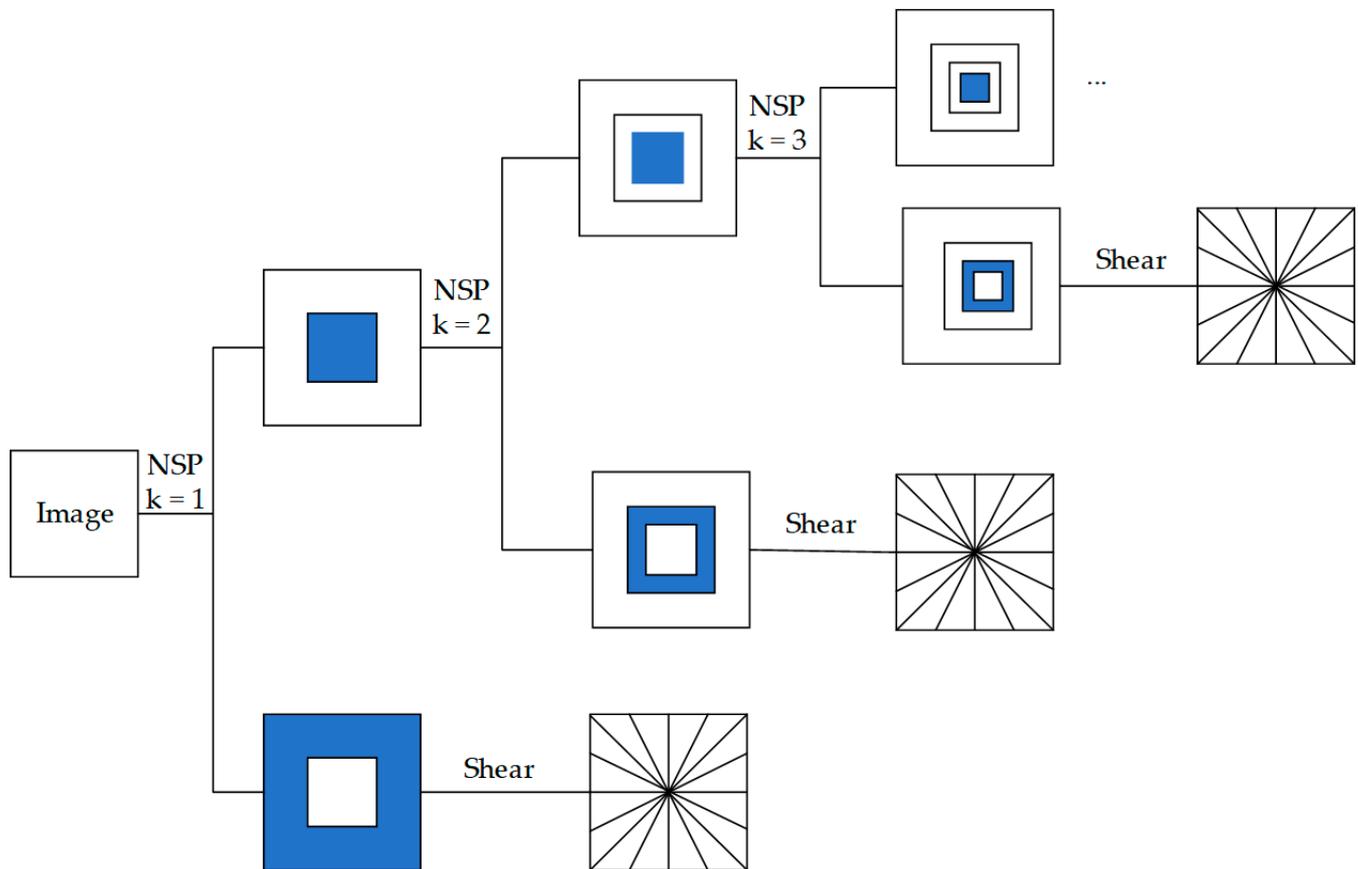


Figure 2. Three-level multi-scale and multi-directional decomposition of NSST.

2.3. β -Divergence Non-Negative Matrix Factorization Based on Alternating Direction Method of Multiplier

Non-negative Matrix Factorization (NMF) is a common method for learning interpretable features from non-negative data. The NMF method has a superior ability for extracting local features. When the NMF method is used for the remote sensing image fusion process, it can integrate the dominant regions of different remote sensing images. Thus, the NMF method can strengthen the regional features and obtain superior fusion results. A β -divergence NMF method in the literature [16] is derived based on the Alternating Direction Method of Multipliers (ADMM) framework.

The ADMM method is a simple and effective method to solve separable convex programming problems, especially for large-scale problems. It can be regarded as a new method developed based on the Lagrange augmentation method. In addition, it can equally separate the objective function, which takes advantage of the separability of the objective function. The original problem is decomposed into several sub-problems for which the local solutions are easy to find. Then, the global solution of the original problem is obtained by alternating analysis. The ADMM-based method has faster convergence speed and accurate sparsity. It is easy to implement and only one adjustment parameter, λ , is needed. In this paper, the value of λ is automatically chosen according to the literature [17]. In addition, the ADMM-based method can reach a given level of accuracy much faster, which is several orders of magnitude faster than the multiplication update rule. Thus, it is a small price for tuning λ . The multiplication update rule is particularly susceptible to falling into

local optimal state. However, the ADMM-based β -divergence NMF method has faster convergence and better solution results.

The general form of the NMF problem is as follows:

$$\begin{aligned} & \text{minimize } D_\beta(\mathbf{E}|\mathbf{UV}) \\ & \text{subject to } \mathbf{U} \geq 0, \mathbf{V} \geq 0 \end{aligned} \quad (6)$$

where $D_\beta(\mathbf{E}|\mathbf{UV})$ denotes the β -divergence measure between \mathbf{E} and its reconstruction, i.e., \mathbf{UV} .

The divergence between two matrices is defined as the sum of element divergence, which is shown below.

$$D_\beta(\mathbf{E}|\hat{\mathbf{E}}) = \sum d_\beta(\mathbf{E}|\hat{\mathbf{E}}) \quad (7)$$

The expression of the β -divergence NMF method is as follows:

$$d_\beta(m|n) = \begin{cases} \frac{m^\beta}{\beta(\beta-1)} + \frac{n^\beta}{\beta} - \frac{mn^{\beta-1}}{\beta-1}, \beta \in \mathbb{R} \setminus \{0, 1\} \\ \frac{1}{2}(m-n)^2, \beta = 2 \\ m \log \frac{m}{n} - m + n, \beta = 1 \\ \frac{m}{n} - \log \frac{m}{n} - 1, \beta = 0 \end{cases} \quad (8)$$

The non-negative constraints of \mathbf{U} and \mathbf{V} make the optimization problems on \mathbf{U} and \mathbf{V} more complex. The new variables \mathbf{U}_+ and \mathbf{V}_+ are introduced to apply non-negative constraints, and the constraints are $\mathbf{U} = \mathbf{U}_+$ and $\mathbf{V} = \mathbf{V}_+$.

In general, Formula (7) can be rewritten as follows:

$$\begin{aligned} & \text{minimize } D_\beta(\mathbf{E}|\mathbf{X}) \\ & \text{subject to } \mathbf{X} = \mathbf{UV} \\ & \mathbf{U} = \mathbf{U}_+, \mathbf{V} = \mathbf{V}_+ \\ & \mathbf{U}_+ \geq 0, \mathbf{V}_+ \geq 0 \end{aligned} \quad (9)$$

The above expression represents an augmented Lagrangian function consisting of eight variables, i.e., five primitive variables and three dual variables. For ADMM, this can be optimized in three parts: \mathbf{U} , \mathbf{V} , and $(\mathbf{X}, \mathbf{U}_+, \mathbf{V}_+)$. Since the optimization goal is split into \mathbf{X} , \mathbf{U}_+ , and \mathbf{V}_+ , optimizing them separately is equivalent to optimizing them together:

$$\begin{aligned} G_\lambda(\mathbf{X}, \mathbf{U}, \mathbf{V}, \mathbf{U}_+, \mathbf{V}_+, \alpha_X, \alpha_U, \alpha_V) = & \\ & D_\beta(\mathbf{E}|\mathbf{X}) + \langle \alpha_X, \mathbf{X} - \mathbf{UV} \rangle + \frac{\lambda}{2} \|\mathbf{X} - \mathbf{UV}\|_F^2 \\ & + \langle \alpha_U, \mathbf{U} - \mathbf{U}_+ \rangle + \frac{\lambda}{2} \|\mathbf{U} - \mathbf{U}_+\|_F^2 \\ & + \langle \alpha_V, \mathbf{V} - \mathbf{V}_+ \rangle + \frac{\lambda}{2} \|\mathbf{V} - \mathbf{V}_+\|_F^2 \end{aligned} \quad (10)$$

Alternating update optimization G_λ is performed for each of the five original variables. Then, a gradient ascent is performed for each of the three dual variables. The detailed process of the ADMM-based β -divergence NMF [16] are as follows (Algorithm 1):

Algorithm 1 The ADMM-based β -divergence NMF

Inputs	E
Initialize	$\mathbf{X}, \mathbf{U}, \mathbf{V}, \mathbf{U}_+, \mathbf{V}_+, \alpha_X, \alpha_U, \alpha_V$
Repeat	$\mathbf{U}^T \leftarrow (\mathbf{V}\mathbf{V}^T + \mathbf{I}) \setminus (\mathbf{V}\mathbf{X}^T + \mathbf{U}_+^T + \frac{1}{\lambda}(\mathbf{V}\alpha_X^T - \alpha_U^T))$ $\mathbf{V} \leftarrow (\mathbf{U}^T\mathbf{U} + \mathbf{I}) \setminus (\mathbf{U}^T\mathbf{X} + \mathbf{V}_+ + \frac{1}{\lambda}(\mathbf{U}^T\alpha_X - \alpha_V))$ $\mathbf{X} \leftarrow \underset{\mathbf{X} \geq 0}{\operatorname{argmin}} D_\beta(E \mathbf{X}) + \langle \alpha_X, \mathbf{X} \rangle + \frac{\lambda}{2} \ \mathbf{X} - \mathbf{U}\mathbf{V}\ _F^2$ $\mathbf{U}_+ \leftarrow \max(\mathbf{U} + \frac{1}{\lambda}\alpha_U, 0)$ $\mathbf{V}_+ \leftarrow \max(\mathbf{V} + \frac{1}{\lambda}\alpha_V, 0)$ $\alpha_X \leftarrow \alpha_X + \lambda(\mathbf{X} - \mathbf{U}\mathbf{V})$ $\alpha_U \leftarrow \alpha_U + \lambda(\mathbf{U} - \mathbf{U}_+)$ $\alpha_V \leftarrow \alpha_V + \lambda(\mathbf{V} - \mathbf{V}_+)$
Until	Convergence Return $\mathbf{U}_+, \mathbf{V}_+$

In the process of updating, it is difficult to update \mathbf{X} , and the updating method varies with the value of β . For more details, please refer to the literature [16].

2.4. Weighted Local Contrast Measure

The high-frequency components of different scales and directions can be obtained by NSST decomposition. It not only provides multi-scale and multi-directional information in the original image, but also contains abundant spatial information, such as textures and details. The components with more distinctly detailed features at the same scale and direction have a larger pixel value as an absolute value. In the fusion process of high-frequency components, if the component with the largest absolute value of pixel values is simply selected as the final fusion result, the correlation between adjacent pixels of the original image will be ignored. In addition, it will confer a lot of noise to the fused image.

The spatial details, such as edges and textures, have larger local contrast measure values, which are the targets of image fusion. The traditional local contrast measure (LCM) [18] method used the central pixel value and the maximum intensity of the eight adjacent pixels to calculate the LCM value of an image. This method is susceptible to highlight noise. In addition, the false alarm pixels will be introduced during operation, which will increase the false alarm rate. The central pixel and its eight adjacent pixels are shown in Figure 3.

In the proposed method, the median pixel value of the eight neighborhoods is used to calculate the local contrast measure values. This method can avoid misjudging the high luminance noise as detailed information, and reduce the false alarm caused by the high luminance isolated noise. Thus, the local contrast measure value C_n between the center pixel and eight adjacent pixels is defined as follows:

$$C_n = \frac{P_0 \times P_0}{P_{med}} \quad (11)$$

where P_0 is the pixel value of the central pixel in the local area, and P_{med} is the median of the pixel values in the eight pixels adjacent to the central pixel, which can be calculated by the following formula:

$$P_{med} = \text{median}(P_i), i = 1, 2, \dots, 8 \quad (12)$$

We can draw a conclusion from Formula (11): if the central area A_0 is the detail target, i.e., $\frac{P_0}{P_{med}} > 1$, so $C_n > P_0$. Then, the detail target will be enhanced. If the central area A_0 is the background, there may be details around or all the background, i.e., $\frac{P_0}{P_{med}} \leq 1$, so $C_n \leq P_0$. Then, the background will be suppressed.

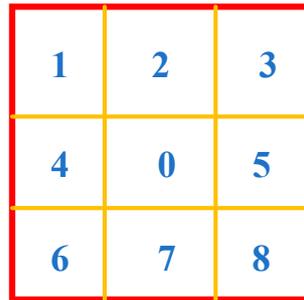


Figure 3. The central pixel and its eight adjacent pixels.

Firstly, the local contrast measure value is calculated by using the median pixels of the eight neighborhoods to avoid misclassifying noise as spatial detail. On this basis, the mean value of the pixel value difference between the central pixel and adjacent neighborhoods is introduced to weight the local contrast measure value, then the weighted local contrast measure (WLCM) value can be calculated. The WLCM method can enhance the weak spatial details and suppress the background, which greatly improves the significance of spatial details and increases the detection rate of spatial detail information. The smaller the mean value of the pixel value differences between the central area and the eight neighborhoods, the less likely the central area is spatial detail. On the contrary, the greater the mean value of the pixel value difference between the central area and the eight neighborhoods, the greater the possibility that the central area is spatial detail. M_n is the average of the pixel value differences between the central area and the eight neighborhoods. Then, the local contrast measure value is weighted by M_n . The calculation details of M_n are as follows:

$$M_n = \frac{1}{8} \times \sum_{i=1}^8 |P_0 - P_i| \quad (13)$$

Finally, the calculation formula of the WLCM method can be obtained. The details are as follows:

$$WLCM_n = C_n \times M_n \quad (14)$$

3. The Steps and Principles

3.1. The Overall Image Fusion Steps

Through NSST decomposition, one low-frequency component and several high-frequency components can be obtained. The low-frequency component is an approximate version of the original image, which contains the main information of the original image. In the fusion process of the low-frequency component, if only the low-frequency component of I is used as the fusion result, the spectral information of the MS image can be well maintained, but the spatial resolution of the fused image will be reduced. If only the low-frequency component of the PAN image is used as the fusion result, the spatial information of the PAN image can be well maintained, but the spectral resolution of the fused image will be affected.

In this paper, we design two different fusion rules according to the different characteristics of low-frequency and high-frequency components. The low-frequency component contains the main information and represents the approximate feature of the original image. Thus, an ADMM-based β -divergence NMF method is used to fuse the low-frequency coefficients. The high-frequency components represent the edge and texture information of the original image. Thus, a WLCM-based rule is used to fuse the high frequency coefficients.

Figure 4 shows the flow chart of the proposed method. The detailed fusion process is as follows:

- (1) Adaptive Weighted Average Calculates the MS Intensity Component

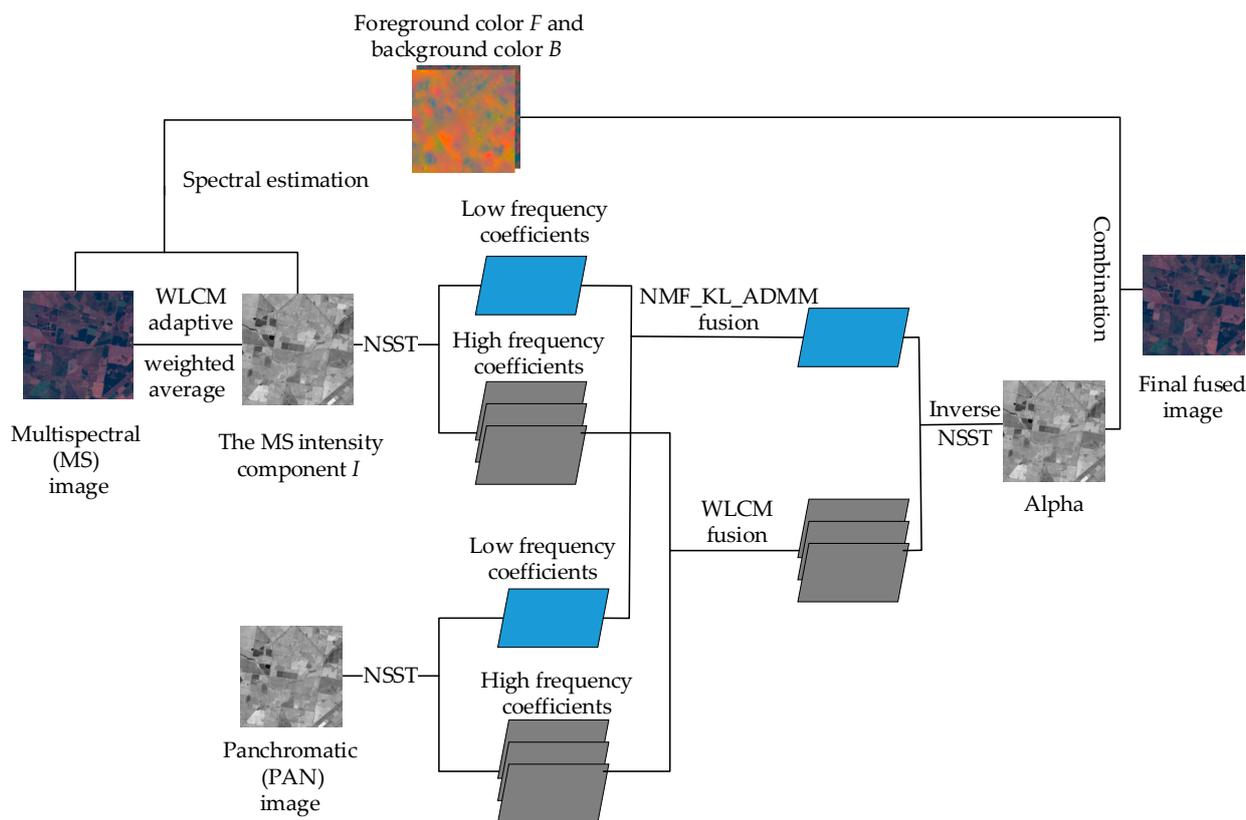


Figure 4. The flow chart of the proposed method.

If a simple average fusion rule is adopted, the details of the original image will be lost. The accurate selection of the weighting coefficient determines the quality of the fused image. In order to generate the MS intensity component I , an adaptive weighted average method based on the WLCM method is used to fuse each band of the MS image. The WLCM value can be used as an index to evaluate the detailed information in the spatial domain. The pixels with larger WLCM values are considered to be more weighted information, such as edges or textures, which are given more weight in the fusion process. Thus, an adaptive weighted average coefficient ω_i is designed according to the WLCM value. The details are as follows:

$$I(x, y) = \sum_{i=1}^n \frac{1}{\omega_i(x, y)} MS_i(x, y) \tag{15}$$

$$\omega_i(x, y) = \frac{WLCM_i(x, y)}{\sum_{i=1}^n WLCM_i} \tag{16}$$

where n is the band number of the MS image, $WLCM_i$ denotes the WLCM value of the i -th band of the MS image at the position (x, y) , $MS_i(x, y)$ denotes the pixel value of the i -th

band of the MS image at the position (x, y) , $\omega_i(x, y)$ denotes the weighting factor of the i -th band of the MS image at the position (x, y) , and $I(x, y)$ denotes the pixel value of the intensity component I at the position (x, y) .

(2) Spectral Estimation

Taking I as the initial α channel, the foreground color F and background color B are calculated according to Formula (5). F and B contain abundant spectral information, but not spatial information. The main purpose of the steps discussed below is to obtain spatial detail information from the PAN image by the fusion process.

(3) NSST Decomposition

The intensity component I and the PAN image are separately decomposed by NSST, and the corresponding components of different scales and directions can be obtained. Specifically, a low-frequency component and a plurality of high-frequency components can be obtained. Subsequently, we design two different fusion strategies according to the characteristics of low-frequency and high-frequency components.

(4) High-Frequency Components Fusion

The high-frequency components at different scales and directions not only provide multi-scale and multi-directional information, but also contain abundant edge and textural detail information. The edges, textures, and other spatial details have high local contrast values, which is the target of image fusion. A WLCM-based rule is used to fuse the high-frequency components. The detailed fusion process of the high-frequency components is described in Section 3.2.

(5) Low-Frequency Components Fusion

The low-frequency component is an approximation of the original image, which describes only the basic structure of the image and does not include spatial details such as edges and textures. An ADMM-based β -divergence NMF method is used to fuse the low-frequency components. The detailed fusion process of the low-frequency components is described in Section 3.3.

(6) NSST Inverse Transformation

The fused components are inverted by NSST inverse transformation to obtain the fused image. Then, the fused image is used as the final α channel to participate in the final reconstruction.

(7) Image Reconstruction

According to Formula (4), the final fusion result is obtained by using α , F , and B for reconstruction.

3.2. High-Frequency Components Fusion Algorithm

The high-frequency components at different scales and directions not only provide multi-scale and multi-directional information, but also contain abundant spatial detail information such as edges and textures. At the same scale and direction, the absolute values of the components representing the detailed information are relatively large. In the fusion process of high-frequency components, if the component with the largest absolute value of pixel values is simply selected as the final fusion result, the correlation between the adjacent pixels of the original image will be ignored. In addition, it will confer a lot of noise to the fused image. According to the literature [19], when the difference between the WLCM values of the two components is less than 0.015, we can say that the difference between these two components is small. On the contrary, above this value, we can say that the difference between these two components is relatively large. Thus, this fusion rule with a threshold value of 0.015 can be used to select high quality high frequency

components and fuse them into the final result. Finally, A WLCM-based rule is used to fuse the high-frequency components, and the details are as follows:

$$H_{m,n}^F(i,j) = \begin{cases} w_I(i,j)H_{m,n}^I(i,j) + w_P(i,j)H_{m,n}^P(i,j), & |MICM_D(i,j)| \leq 0.015 \\ H_{m,n}^I(i,j), & |MICM_D(i,j)| > 0.015, MICM_I(i,j) > MICM_P(i,j) \\ H_{m,n}^P(i,j), & |MICM_D(i,j)| > 0.015, MICM_I(i,j) < MICM_P(i,j) \end{cases} \quad (17)$$

$$MICM_D = MICM_I(i,j) - MICM_P(i,j) \quad (18)$$

$$w_I(i,j) = \frac{MICM_I(i,j)}{MICM_I(i,j) + MICM_P(i,j)} \quad (19)$$

$$w_P(i,j) = \frac{MICM_P(i,j)}{MICM_I(i,j) + MICM_P(i,j)} \quad (20)$$

where m and n are the decomposition order and direction number, respectively. $H_{m,n}^F(i,j)$ represents the high-frequency coefficient value of the fused image at the position (i,j) , $H_{m,n}^I(i,j)$ represents the high-frequency coefficient value at the position (i,j) in I , $H_{m,n}^P(i,j)$ represents the high-frequency coefficient value at the position (i,j) in the PAN image, $MICM_I(i,j)$ represents the WLCM value at the position (i,j) in I , and $MICM_P(i,j)$ represents the WLCM value at the position (i,j) in the PAN image.

3.3. Low-Frequency Component Fusion Algorithm

NMF is used to decompose the non-negative matrix $\mathbf{X} \in \mathbf{R}_+^{M \times N}$ into two smaller non-negative matrices, $\mathbf{W} \in \mathbf{R}_+^{M \times k}$ and $\mathbf{H} \in \mathbf{R}_+^{k \times N}$ multiplying by each other. Then, we can obtain $\mathbf{X} = \mathbf{WH} + \varepsilon$, where ε is the background noise. In addition, k is much smaller than M and N , that is, $k < \min\{M, N\}$. The original image can usually be regarded as a real image imaged in different types of sensors, which is obtained by adding a certain amount of background noise. When the NMF method is applied to the pan-sharpening process, the NMF method can maintain the overall features of the images involved in the fusion. Thus, the NMF method can obtain superior spatial details from the PAN image, while obtaining more spectral information from the MS image.

In the fusion process of the low-frequency components, we set $k = 1$. Firstly, an ADMM-based β -divergence NMF is used for iteration. The iterative solution is actually an optimization process. By iterating to minimize the reconstruction error between \mathbf{X} and \mathbf{WH} , the background noise can be effectively suppressed. After the iteration is completed, a unique feature base \mathbf{W} can be obtained. \mathbf{W} contains the overall features of the original images involved in the fusion and is regarded as an approximate reproduction of the source image. It can effectively converge and suppress the background noise. Finally, the fused image can be obtained by resetting the feature base \mathbf{W} and restoring it to the size of the source image.

The low-frequency component L_A of I and the low frequency component L_B of the PAN image are fused by an ADMM-based β -divergence NMF method. The detailed implementation steps are as follows:

- (1) The low-frequency components L_A and L_B are sorted into column vectors according to the priority of the rows. Then, the column vectors \mathbf{X}_A and \mathbf{X}_B are obtained. If the sizes of L_A and L_B are both $M \times N$, the sizes of \mathbf{X}_A and \mathbf{X}_B are $MN \times 1$. The details are as follows:

$$\mathbf{X}_A = \begin{bmatrix} xa_1 \\ xa_2 \\ xa_3 \\ \dots \\ xa_{MN} \end{bmatrix}, \mathbf{X}_B = \begin{bmatrix} xb_1 \\ xb_2 \\ xb_3 \\ \dots \\ xb_{MN} \end{bmatrix} \quad (21)$$

- (2) According to the column vectors X_A and X_B , the following original matrix X is constructed, and its size is $MN \times 2$.

$$X = [X_A, X_B] = \begin{bmatrix} xa_1 & xb_1 \\ xa_2 & xb_2 \\ xa_3 & xb_3 \\ \dots & \dots \\ xa_{MN} & xb_{MN} \end{bmatrix} \quad (22)$$

- (3) We set $k = 1$. NMF is the factorization with error, which means $X \approx WH$. In order to obtain an approximate factorization and minimize the reconstruction error between X and WH , a cost function must be defined. The cost function can measure the approximation effect of the solution. In the proposed method, we choose Kullback–Leibler (KL) divergence as the cost function. The maximum number of iterations is set to 2000. The initial iteration values W_0 and H_0 are randomly generated with sizes $M \times k$ and $k \times N$, respectively. The details are as follows:

$$W_0 = rand(M, k), H_0 = rand(k, N) \quad (23)$$

- (4) After setting the relevant parameters, the original matrix X is decomposed using an ADMM-based β -divergence NMF method. The detailed iterative process can be found in Section 2.3. When the iteration is finished, the basis matrix W and the weight coefficient matrix H can be obtained. W contains the overall features of the low-frequency components L_A and L_B , which can be regarded as the approximate reproduction of the original image.
- (5) We reset W to a matrix S of $M \times N$. Finally, S is the fusion result of the low-frequency components.

4. Experiments and Discussion

4.1. Experimental Images

This paper used a dataset of 38 image pairs containing six bands, which were taken by LANDSAT 7 ETM+ and can be easily obtained [20]. The spatial resolution of the MS and PAN images are 30 m and 15 m, respectively. The pixel sizes of the MS and PAN images are 200×200 and 400×400 , respectively.

Since there is no high-resolution MS image as a reference image in the dataset, we first up-sampled the original MS image to obtain an MS image with a pixel size of 400×400 . Then, the MS images with the pixel size of 400×400 and the PAN image were down-sampled to obtain the MS image and the PAN image with pixel size of 200×200 as the experimental image. Finally, the original MS image was used as the reference image and compared with the fused image of each method. Figure 4 shows four image pairs of the MS and PAN image. In Figure 5: (a) and (b) are the first image pair; (c) and (d) are the second image pair; (e) and (f) are the third image pair; (g) and (h) are the fourth image pair. The four image pairs of the MS and PAN image were subsequently used for experimental analysis.

4.2. Selected Comparison Method

In order to verify the validity and reliability of the method proposed in this paper, the proposed method is compared with ten existing representative fusion methods.

These ten fusion methods are the: Brovey Transform-based (BT) method [11], Gram–Schmidt Adaptive-based (GSA) method [4], Guided Filter-based (GF) method [12], Intensity–Hue–Saturation-based (IHS) method [2], Multi-objective Decision-based (MOD) method [10], Principal Component Analysis-based (PCA) method [3], Partial Replacement Adaptive Component Substitution-based (PRACS) method [6], Variational Local Gradient Constraints-based (VLGC) method [9], Band Dependent Spatial-detail with Physical Constrains-based (BDSD-PC) method [5], and Wavelet Transform-based (WT) method [7].

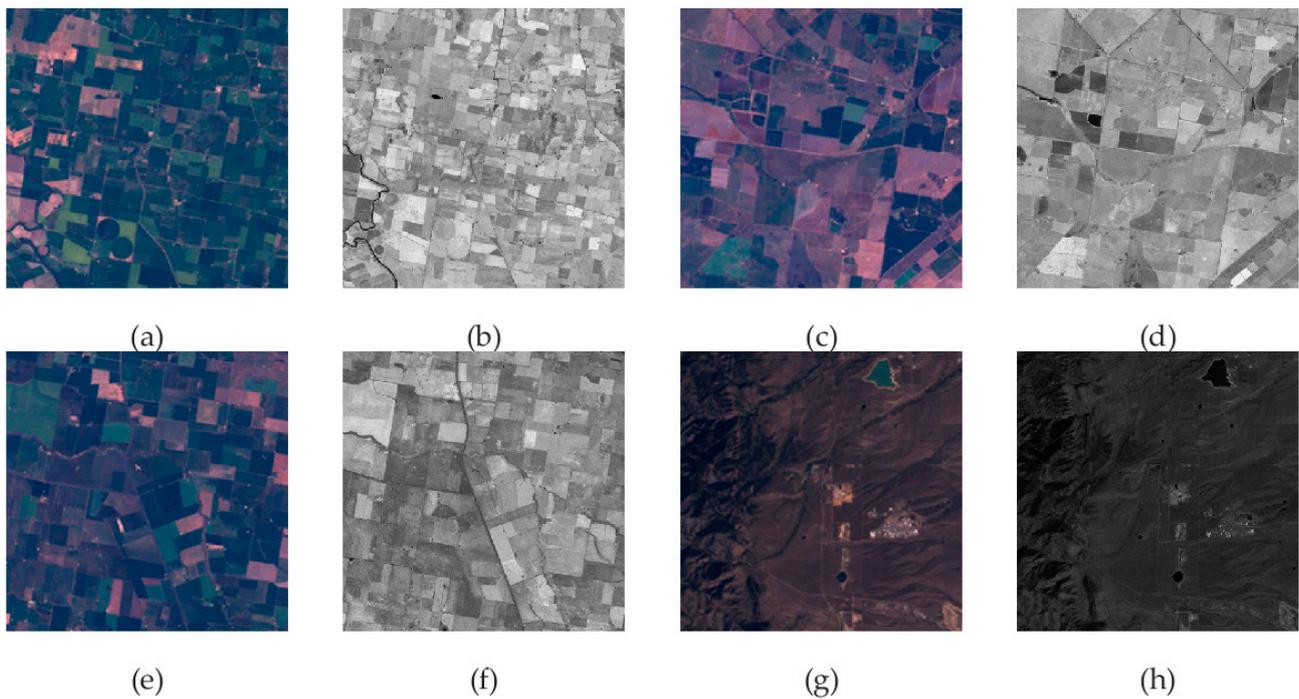


Figure 5. Four image pairs of the MS and PAN image.

4.3. Objective Evaluation Indices

It is difficult to accurately compare the quality of the fused images by each method based on subjective evaluation alone. To quantitatively evaluate the fusion quality, this paper adopts five well-known objective evaluation indices, which are the Correlation Coefficient (CC), Erreur Relative Global Adimensionnelle de Synthse (ERGAS), Relative Average Spectral Error (RASE), Spectral Information Divergence (SID), and No Reference Quality Evaluation (QNR), as detailed below. The performance of an image fusion method is evaluated scientifically and objectively through quantitative evaluation methods and indices. It is not influenced by human visual characteristics and psychological states.

- (1) The Correlation Coefficient (CC) [21] calculates the correlation between the reference image and a pan-sharpening result. Its ideal value is 1. It is defined as follows:

$$CC = \frac{\sum_{m=1}^M (R_m - \bar{R})(P_m - \bar{P})}{\sqrt{\sum_{m=1}^M (R_m - \bar{R})^2 \sum_{m=1}^M (P_m - \bar{P})^2}} \quad (24)$$

where m is the m -th pixel, M is the total number of pixels, R is the reference MS image, P is the pan-sharpening image, and \bar{R} and \bar{P} are the average values of R and P , respectively. Please refer to the literature [21] for more details.

- (2) Erreur Relative Global Adimensionnelle de Synthse (ERGAS) [22] can measure the fusion quality of a pan-sharpening method, which is defined as follows:

$$ERGAS = 100 \frac{H}{L} \sqrt{\frac{1}{C} \sum_{c=1}^C \left(\frac{(RMSE^c)^2}{\bar{R}^c} \right)} \quad (25)$$

where L and H are the spatial resolutions of the MS and PAN images, respectively. \bar{R}^c is the mean value of the c -th band in the reference MS image. $RMSE^c$ is the root mean square error (RMSE) [23] value between the c -th band of the reference MS image and the c -th band

of the pan-sharpening image. Its ideal value is 0. Please refer to the literature [22] for more details.

- (3) Relative Average Spectral Error (*RASE*) [24] reflects the average performance of a pan-sharpening method on spectral errors. Smaller values of *RASE* denote less spectral distortion. Its ideal value is 0. It is defined as follows:

$$RASE = \frac{100}{M} \sqrt{\frac{1}{N} \sum_{i=1}^N RMSE(R_i)^2} \quad (26)$$

where M is the mean radiance of the N -band original spectral images R , and $RMSE(R_i)$ is the *RMSE* value for each spectral band in R . Please refer to the literature [24] for more details.

- (4) Spectral Information Divergence (*SID*) [25] evaluates the difference between spectra. Its ideal value is 0. Please refer to the literature [25] for more details.
- (5) No Reference Quality Evaluation (*QNR*) [26] can evaluate the quality of a pan-sharpening image without a reference image, which consists of three parts: a spectral distortion index D_λ , a spatial distortion index D_s , and a global *QNR* value. The detailed definition is provided in Formula (27). For the global *QNR*, the higher the value, the better the fusion effect. Its ideal value is 1. Please refer to the literature [26] for more details.

$$QNR = (1 - D_\lambda)(1 - D_s) \quad (27)$$

- (6) D_λ is a sub-metric of *QNR*, which can measure the spectral distortion of a pan-sharpening image. The smaller the value, the better the fusion effect. Its ideal value is 0. It is defined as follows:

$$D_\lambda = \sqrt{\frac{1}{C(C-1)} \sum_{c=1}^C \sum_{d=1, d \neq c}^C |Q(L^c, L^d) - Q(P^c, P^d)|} \quad (28)$$

where Q is the universal image quality index (*UIQI*) [27] value; L is the low-resolution MS image; c and d are the c -th and d -th bands of the MS image; C is the total bands number of the MS image; and X and Y are low resolution and high-resolution PAN images, respectively. Please refer to the literature [26] for more details.

- (7) D_s is a sub-metric of *QNR*, which can measure the spatial distortion of the fused image. The smaller the value, the better the fusion effect. Its ideal value is 0. It is defined as follows:

$$D_s = \sqrt{\frac{1}{C} \sum_{c=1}^C |Q(L^c, X) - Q(P^c, Y)|} \quad (29)$$

where Q is the *UIQI* value; L is the low-resolution MS image; c is the c -th band of the MS image; C is the total number of bands of the MS image; and X and Y are low resolution and high-resolution PAN images, respectively. Please refer to the literature [26] for more details.

4.4. Implementation Details

In this section, we provide some implementation details of the proposed method, the experimental environment, and the development platform. The details are shown in the following Table 1.

Table 1. Implementation details.

Projects	Implementation Details
The number of band-pass directional sub-bands in each layer of NSST	32, 32, 16, 16
The level of NSST directional decomposition	4
Experimental environment	Windows 10 System PC Intel (R) Core (TM) i7-8700 CPU 3.20 GHz 16 GB Memory
Development platform	MATLAB R2018a

4.5. Experimental Results and Analysis

To verify the effectiveness and reliability of the method proposed in this paper, we used four real remote sensing image pairs of the MS and PAN image for the experimental verification. In addition, the results were analyzed based on subjective visual effects and objective quantitative evaluations, and the experimental results are presented and discussed after setting the parameters.

In Figures 6–9, (a)–(k) show the fusion results of four different image pairs obtained by the BT, GSA, GF, IHS, MOD, PCA, PRACS, VLGC, BDSD-PC, and WT methods, as well as the proposed method, respectively. In addition, the reference MS image is also included in the figure, i.e., (l).

In order to show the details more intuitively, the fusion results are locally enlarged. The locally enlarged details are placed in the lower right corner of the fused image. Tables 2–5 show the results of the objective quality evaluation. The objective quality assessment indices include spectral and spatial quality evaluations, which are CC, ERGAS, SID, RASE and QNR. For all quality evaluations, the best results are shown in bold red, the second best results are shown in bold green, and the third best results are shown in bold blue.

In Figure 6, the BT and IHS methods both suffer from spectral distortion, especially in the local magnified image; the dark green part becomes blue, and the pink part becomes brick red. The GSA and PCA methods also suffer from severe spectral distortion; in the local magnified image, the dark green part becomes pink, and the pink part becomes green. The GF method has less spectral distortion, but the spatial details are blurred in the local magnification image and the spatial distortion is more severe. The WT method has less spectral distortion, but artifacts appear in the local magnification image and the spatial details are blurred. The MOD, PRACS, VLGC, and BDSD-PC methods have better spectral preservation characteristics, but their spatial details are less clear than the proposed method in the local magnification image. The proposed method in this paper maintains the spectral information with clear spatial details. Thus, it is demonstrated that the proposed method improves the spatial detail information while maintaining spectral characteristics.

Table 2. Objective evaluation of the experimental results on the first image pair.

	CC (1)	ERGAS (0)	SID (0)	RASE (0)	QNR (1)
BT	0.320	6.684	0.060	28.735	0.384
GSA	0.113	7.504	0.065	26.255	0.151
GF	0.857	7.612	0.011	20.597	0.868
IHS	0.364	6.130	0.029	26.130	0.397
MOD	0.944	1.605	0.007	5.148	0.904
PCA	0.173	6.241	0.049	26.641	0.258
PRACS	0.944	1.607	0.008	5.175	0.916
VLGC	0.945	1.602	0.009	5.156	0.902
BDSD-PC	0.942	1.613	0.007	5.160	0.884
WT	0.734	3.054	0.021	11.238	0.573
Proposed	0.948	1.486	0.005	4.923	0.925

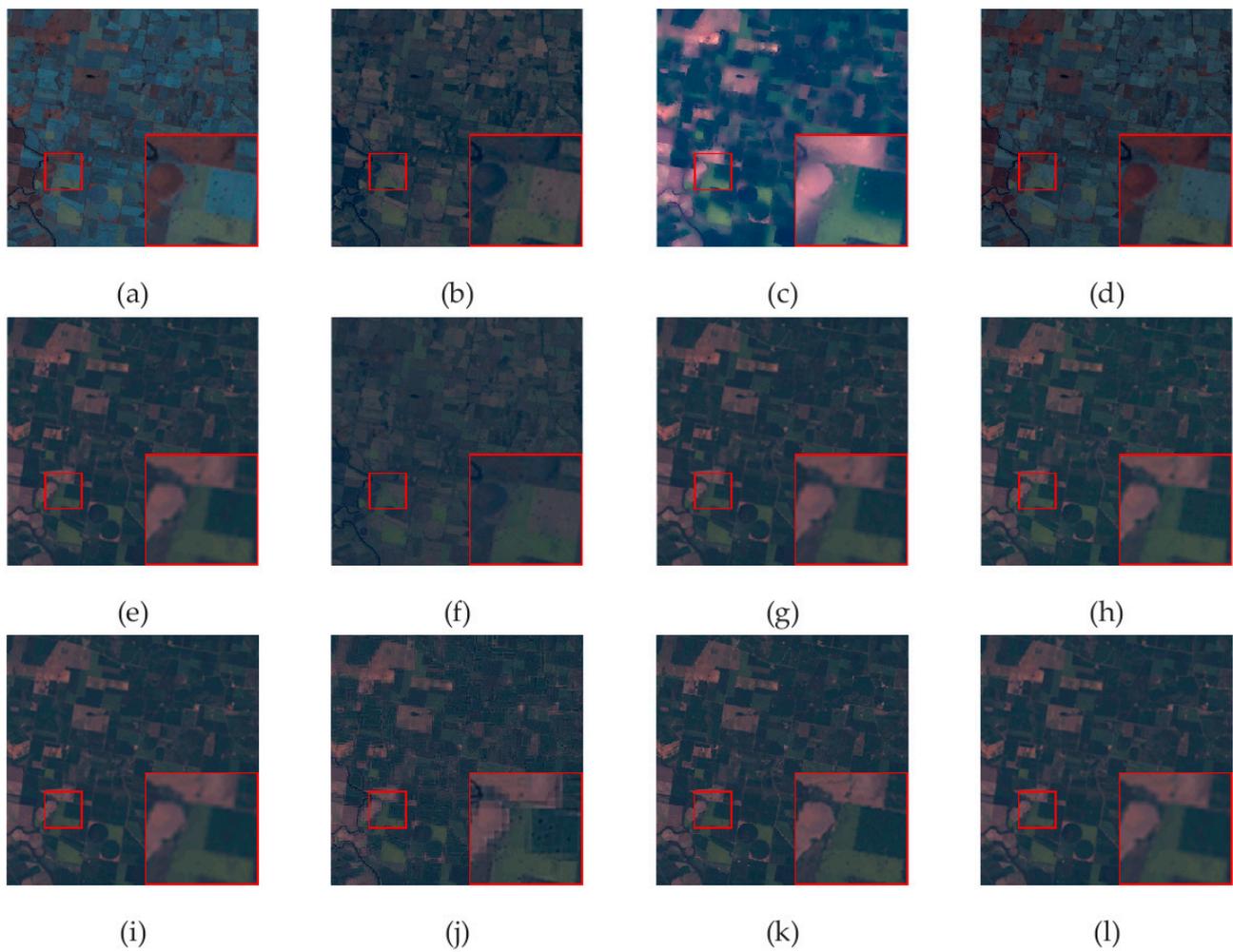


Figure 6. The fusion results of ten different methods on the first image pair: (a) BT; (b) GSA; (c) GF; (d) IHS; (e) MOD; (f) PCA; (g) PRACS; (h) VLGC; (i) BDS-D-PC; (j) WT; (k) the proposed method; (l) the reference MS image.

Table 3. Objective evaluation of the experimental results on the second image pair.

	CC (1)	ERGAS (0)	SID (0)	RASE (0)	QNR (1)
BT	0.317	5.800	0.018	22.405	0.347
GSA	0.052	5.377	0.036	20.560	0.110
GF	0.876	4.145	0.009	24.917	0.802
IHS	0.325	5.717	0.011	18.624	0.386
MOD	0.944	1.618	0.006	5.320	0.904
PCA	0.128	6.241	0.031	18.624	0.199
PRACS	0.945	1.607	0.006	5.515	0.909
VLGC	0.942	1.875	0.008	5.314	0.832
BDS-D-PC	0.946	1.613	0.007	5.460	0.918
WT	0.820	1.900	0.007	7.376	0.565
Proposed	0.949	1.324	0.004	3.817	0.930

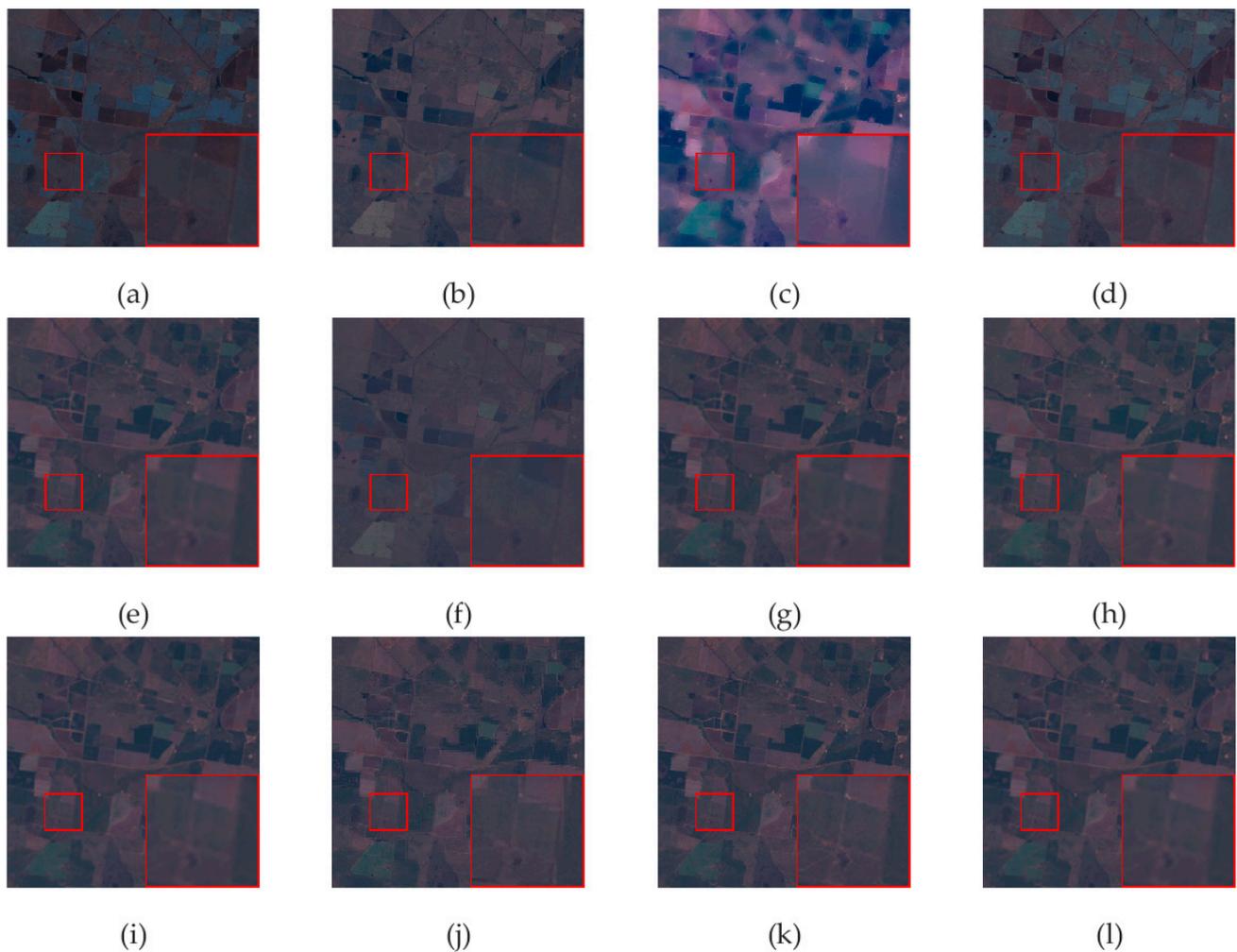


Figure 7. The fusion results of ten different methods on the second image pair: (a) BT; (b) GSA; (c) GF; (d) IHS; (e) MOD; (f) PCA; (g) PRACS; (h) VLGC; (i) BDS-PC; (j) WT; (k) the proposed method; (l) the reference MS image.

Table 4. Objective evaluation of the experimental results on the third image pair.

	CC (1)	ERGAS (0)	SID (0)	RASE (0)	QNR (1)
BT	0.386	4.090	0.019	14.609	0.397
GSA	0.109	4.945	0.034	17.795	0.133
GF	0.862	7.612	0.014	20.597	0.832
IHS	0.336	4.146	0.017	14.140	0.369
MOD	0.943	1.604	0.011	5.148	0.874
PCA	0.135	4.246	0.026	15.258	0.223
PRACS	0.941	1.607	0.011	3.875	0.849
VLGC	0.934	1.613	0.008	5.156	0.870
BDS-PC	0.942	1.602	0.011	5.160	0.869
WT	0.794	1.986	0.009	7.369	0.572
Proposed	0.946	1.218	0.006	2.952	0.882

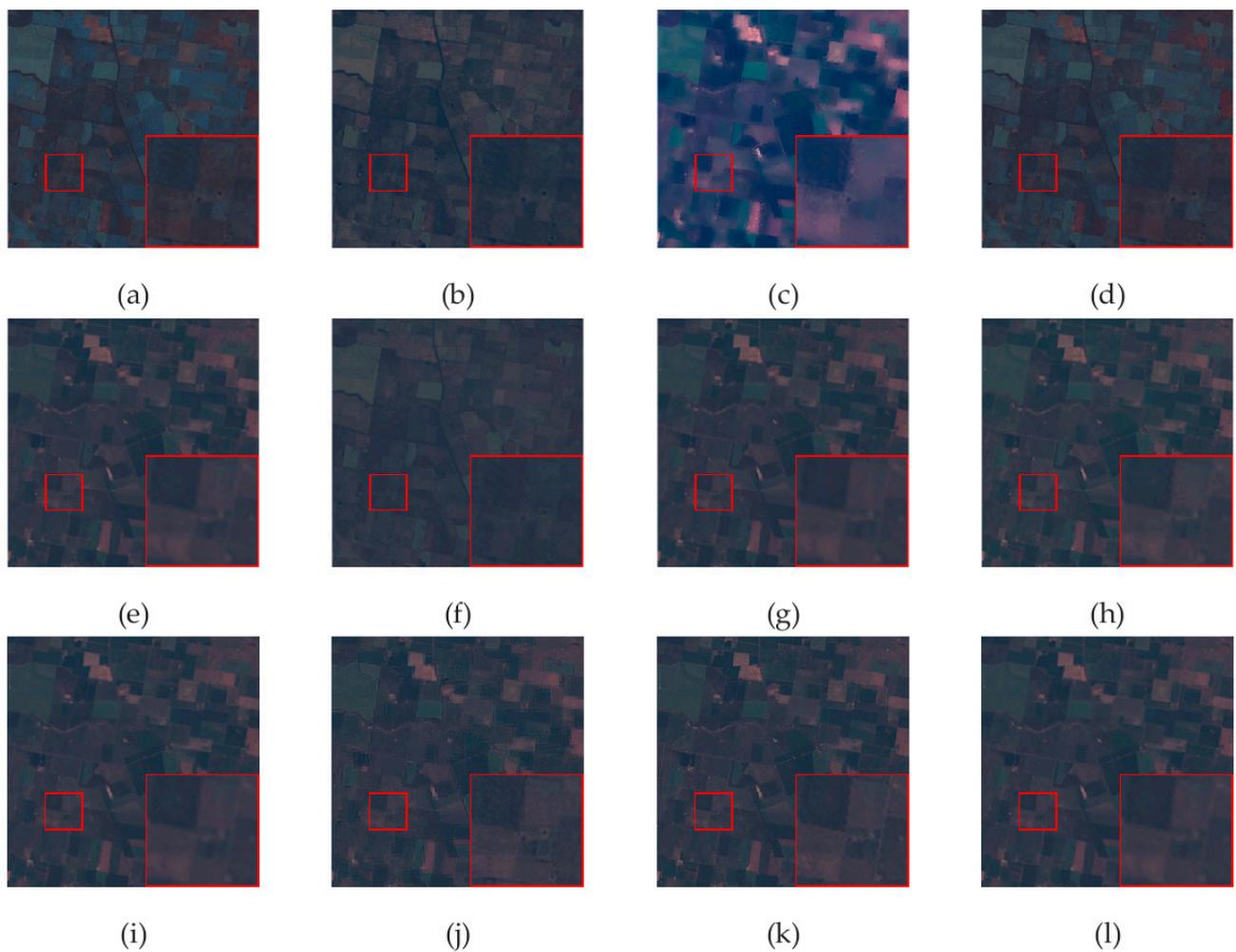


Figure 8. The fusion results of ten different methods on the third image pair: (a) BT; (b) GSA; (c) GF; (d) IHS; (e) MOD; (f) PCA; (g) PRACS; (h) VLGC; (i) BDSD-PC; (j) WT; (k) the proposed method; (l) the reference MS image.

Table 5. Objective evaluation of the experimental results on the fourth image pair.

	CC (1)	ERGAS (0)	SID (0)	RASE (0)	QNR (1)
BT	0.810	8.302	0.010	21.254	0.617
GSA	0.818	3.260	0.005	12.917	0.610
GF	0.890	8.412	0.006	20.597	0.804
IHS	0.814	3.464	0.009	12.584	0.606
MOD	0.943	1.612	0.003	5.153	0.839
PCA	0.827	3.537	0.005	14.013	0.621
PRACS	0.929	2.056	0.003	8.143	0.743
VLGC	0.945	1.602	0.004	5.156	0.839
BDSD-PC	0.950	1.593	0.003	5.140	0.845
WT	0.905	2.286	0.005	9.098	0.824
Proposed	0.953	1.370	0.002	2.103	0.982

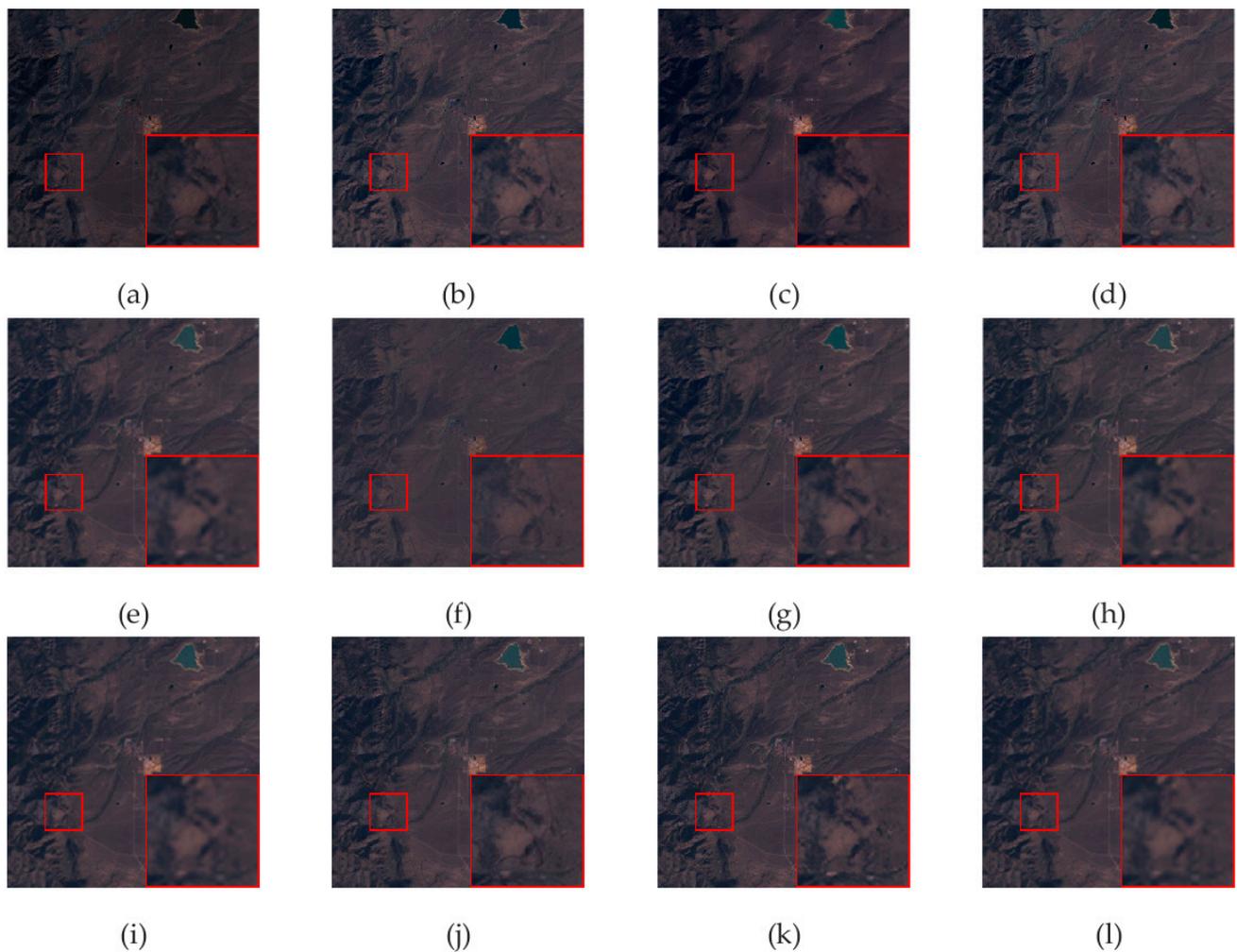


Figure 9. The fusion results of ten different methods on the fourth image pair: (a) BT; (b) GSA; (c) GF; (d) IHS; (e) MOD; (f) PCA; (g) PRACS; (h) VLGC; (i) BDSD-PC; (j) WT; (k) the proposed method; (l) the reference MS image.

As shown in Table 2, the proposed method has the best performance on CC, ERGAS, SID, RASE, and QNR indices, among the five indices, which is superior compared with ten existing pan-sharpening methods. In addition, the QNR value of the proposed method is 0.925, which is close to the optimal value. Compared with some existing pan-sharpening methods, the VLGC method performs second best in the CC and ERGAS indices; the MOD method performs second best in the SID and RASE indices; the BDSD-PC method performs second best in the SID index; and the PRACS method performs second best in the QNR index. Although these existing methods perform better in some indices, they still fall short of the method proposed in this paper.

As shown in Figure 7, the WT method shows some artifacts but with less spectral distortion, and the GF method has blurred spatial details. The GSA method has clearer spatial details but severe spectral distortion, as seen in the local magnification image, where pink turns into green. The BT and IHS methods have clearer spatial details but severe spectral distortion, as seen in the local magnification images, where pink turns into brick. The MOD, PRACS, VLGC, and BDSD-PC methods have better spectral preservation characteristics, but the clarity of their spatial details is less clear than the proposed method. The proposed method in this paper maintains the spectral information with clear spatial details. Thus, it has been demonstrated that the proposed method improves the spatial detail information while maintaining spectral characteristics.

As shown in Table 3, the proposed method has the best performance on CC, ERGAS, SID, RASE, and QNR indices, among the five evaluation indices, which is superior compared with ten existing pan-sharpening methods. The QNR value of the proposed method is 0.930, which is close to the optimal value. Compared with some existing pan-sharpening methods, the PRACS method performs second best in the ERGAS and SID indices; the MOD method performs second best in the SID index; the BDS-PC method performs second best in the CC and QNR indices; and the VLGC method performs second best in the RASE index. Although these existing methods perform well in some indices, they still fall short of the method proposed in this paper.

As shown in Figure 8, the GF method loses a lot of spatial details and the texture details are not clear. The WT method has less spectral distortion, but in the local magnification image, artifacts appear and the spatial details are blurred. The BT method and IHS methods have clearer spatial details, but the spectral distortion is severe. As seen in the local magnification image, the green turns into dark blue and the light green turns into brick red. The GSA and PCA methods also suffer from severe spectral distortion; in the upper left part, the dark green part turns into light green. The MOD, PRACS, VLGC, and BDS-PC methods have better spectral preservation characteristics, but the texture features are less clear than the proposed method in the local magnification image. The method proposed in this paper maintains the spectral information with clear spatial details. Thus, it has been demonstrated that the proposed method improves the spatial detail information while maintaining spectral characteristics.

As shown in Table 4, the proposed method has the best performance on the CC, ERGAS, SID, RASE, and QNR indices, among the five indices, which is superior compared with ten existing pan-sharpening methods. Compared with some existing pan-sharpening methods, the PRACS method performs second best in the RASE index; the MOD method performs second best in the CC and QNR indices; the BDS-PC method performs second best in the ERGAS index; the PRACS method performs second best in the RASE index; and the VLGC method performs second best in the SID index. Although these existing methods perform well in some indices, they still fall short of the method proposed in this paper.

As shown in Figure 9, the spatial details of the fused images obtained by the BT, GSA, IHS, and PCA methods are relatively clear, and the spatial detail information of the PAN image is completely preserved. However, these methods show spectral distortion in the overall region, which is more pronounced in the locally enlarged region. However, the fused images obtained by the BT, GSA, IHS, and PCA methods show spectral distortion, which is more pronounced in locally amplified regions. In the GF method, both the spatial details and the spectra features of the fused images are severely distorted. Other methods can keep the spectral features of the MS images entirely, but the spatial details of the local amplification part are blurred. The proposed method maintains the spectral information with clear spatial details. Thus, it has been demonstrated that the proposed method improves the spatial detail information while maintaining spectral characteristics.

As shown in Table 5, the proposed method has the best performance on the CC, ERGAS, SID, RASE, and QNR indices, among the six evaluation indicators, which is superior compared with ten existing pan-sharpening methods. In addition, the QNR value of the proposed method is 0.982, which is close to the optimal value of 1. Compared with some existing pan-sharpening methods, the BDS-PC method performs second best in the CC, ERGAS, SID, RASE, and QNR indices. The MOD and PRACS methods perform second best in the SID index. Although these existing methods perform well in some indices, they still fall short of the method proposed in this paper.

In summary, the proposed method can achieve superior results in both visual effects and objective evaluation compared with some existing pan-sharpening methods. It can obtain spatial details from the PAN image while preserving more spectral information from the MS image.

5. Conclusions

To solve the existing problems in the field of pan-sharpening, including spatial distortion and spectral distortion, this paper proposed a superior pan-sharpening method by applying an image matting model, an ADMM-based β -divergence NMF, and NSST. The proposed method makes full use of the multi-resolution analysis and multi-direction characteristics of NSST, and uses different fusion rules to realize the fusion of the MS and PAN images in different frequency domains. For the low-frequency components, an ADMM-based β -divergence NMF method was used for fusion, which can effectively suppress the background noise and maintain the spectral characteristics of the MS image. For the high frequency components, a WLCM-based rule was adopted for fusion, which can make the spatial detail information, such as edges and textures, more prominent in the fusion results.

Compared with some existing pan-sharpening methods, the proposed method can obtain more spatial detail from the PAN image while preserving more spectral information from the MS image. Thus, the proposed method is an effective pan-sharpening method. It could be widely used in land use planning, vegetation cover analysis, earth resources surveys and other fields.

Different applications may have different requirements for remote sensing image features. For example, some applications require clearer spatial detail, while others may have higher requirements for spectral fidelity. These factors need to be considered in the fusion process to improve the application results. In our future work, we will be driven by specific application requirements to develop more effective fusion strategies that will further improve the spatial and spectral resolution of the fused images. In addition, we would like to extend our methods to other multi-sensor fusion fields.

Author Contributions: Conceptualization, Y.P.; methodology, Y.P.; software, S.X.; validation, S.X.; formal analysis, Y.P.; investigation, S.X.; resources, Y.P.; data curation, S.X.; writing—original draft preparation, Y.P.; writing—review and editing, D.L. and J.A.B.; visualization, Y.P.; supervision, D.L. and J.A.B.; project administration, D.L.; funding acquisition, L.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Foundation of China, grant number 62071084.

Data Availability Statement: Not applicable.

Acknowledgments: This work was supported by the National Natural Science Foundation of China, grant number 62071084. Thanks to Danfeng Liu, Jón Atli Benediktsson, and Ligu Wang for their guidance and revision suggestions during the paper-writing process. We are grateful to the editors and reviewers for their valuable suggestions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Masoudi, R.; Kabiri, P. New intensity-hue-saturation pan-sharpening method based on texture analysis and genetic algorithm-adaptation. *J. Appl. Remote Sens.* **2014**, *8*, 083640. [[CrossRef](#)]
2. Jelének, J.; Kopačková, V.; Koucká, L.; Mišurec, J. Testing a modified PCA-based sharpening approach for image fusion. *Remote Sens.* **2016**, *8*, 794. [[CrossRef](#)]
3. Liu, C.; Qi, X.; Zhang, W.; Huang, X. Research of improved Gram-Schmidt image fusion algorithm based on IHS transform. *Eng. Surv. Mapp.* **2018**, *27*, 9–14.
4. Aiazzi, B.; Baronti, S.; Selva, M. Improving component substitution pansharpening through multivariate regression of MS + Pan data. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 3230–3239. [[CrossRef](#)]
5. Vivone, G. Robust band-dependent spatial-detail approaches for panchromatic sharpening. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 6421–6433. [[CrossRef](#)]
6. Choi, J.; Yu, K.; Kim, Y. A new adaptive component-substitution based satellite image fusion by using partial replacement. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 295–309. [[CrossRef](#)]
7. Cheng, J.; Liu, H.; Liu, T.; Wang, F.; Li, H. Remote sensing image fusion via wavelet transform and sparse representation. *ISPRS J. Photogramm. Remote Sens.* **2015**, *104*, 158–173. [[CrossRef](#)]

8. Otazu, X.; Gonzalez-Audicana, M.; Fors, O.; Nunez, J. Introduction of sensor spectral response into image fusion methods. Application to wavelet-based methods. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 2376–2385. [[CrossRef](#)]
9. Fu, X.; Lin, Z.; Huang, Y.; Ding, X. A variational pan-sharpening with local gradient constraints. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019.
10. Wu, L.; Yin, Y.; Jiang, X.; Cheng, T. Pan-sharpening based on multi-objective decision for multi-band remote sensing images. *Pattern Recognit.* **2021**, *118*, 108022. [[CrossRef](#)]
11. Khan, S.S.; Ran, Q.; Khan, M.; Ji, Z. Pan-sharpening framework based on laplacian sharpening with Brovey. In Proceedings of the 2019 IEEE International Conference on Signal, Information and Data Processing (ICSIDP), Chongqing, China, 11–13 December 2019.
12. Li, Q.; Yang, X.; Wu, W.; Liu, K.; Jeon, G. Pansharpening multispectral remote-sensing images with guided filter for monitoring impact of human behavior on environment. *Concurr. Comput. Pract. Exp.* **2021**, *32*, e5074. [[CrossRef](#)]
13. Yin, M.; Liu, W.; Zhao, X.; Yin, Y.; Guo, Y. A novel image fusion algorithm based on nonsubsampling shearlet transform. *Optik* **2014**, *125*, 2274–2282. [[CrossRef](#)]
14. Ullah, H.; Ullah, B.; Wu, L.; Abdalla, F.Y.; Ren, G.; Zhao, Y. Multi-modality medical images fusion based on local-features fuzzy sets and novel sum-modified-Laplacian in non-subsampling shearlet transform domain. *Biomed. Signal Process. Control* **2020**, *57*, 101724. [[CrossRef](#)]
15. Levin, A.; Lischinski, D.; Weiss, Y. A closed-form solution to natural image matting. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *30*, 228–242. [[CrossRef](#)] [[PubMed](#)]
16. Sun, D.L.; Fevotte, C. Alternating direction method of multipliers for non-negative matrix factorization with the beta-divergence. In Proceedings of the 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Florence, Italy, 4–9 May 2014; pp. 6201–6205.
17. Ghadimi, E.; Teixeira, A.; Shames, I.; Johansson, M. On the optimal step-size selection for the alternating direction method of multipliers. *IFAC Proc.* **2012**, *45*, 139–144. [[CrossRef](#)]
18. Chen, C.L.P.; Li, H.; Wei, Y.; Xia, T.; Tang, Y.Y. A local contrast method for small infrared target detection. *IEEE Trans. Geosci. Remote Sens.* **2013**, *52*, 574–581. [[CrossRef](#)]
19. Jin, X.; Jiang, Q.; Yao, S.; Zhou, D.; Nie, R.; Lee, S.-J.; He, K. Infrared and visual image fusion method based on discrete cosine transform and local spatial frequency in discrete stationary wavelet transform domain. *Infrared Phys. Technol.* **2018**, *88*, 1–12. [[CrossRef](#)]
20. US Gov. Available online: <https://earthexplorer.usgs.gov/> (accessed on 24 October 2019).
21. Alparone, L.; Wald, L.; Chanussot, J.; Thomas, C.; Gamba, P.; Bruce, L. Comparison of pansharpening algorithms: Outcome of the 2006 GRS-S data-fusion contest. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 3012–3021. [[CrossRef](#)]
22. Zhang, L.; Zhang, L.; Tao, D.; Huang, X. On combining multiple features for hyperspectral remote sensing image classification. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 879–893. [[CrossRef](#)]
23. Yang, Y.; Tong, S.; Huang, S.; Lin, P. Multifocus image fusion based on NSCT and focused area detection. *IEEE Sens. J.* **2015**, *15*, 2824–2838.
24. Ranchin, T.; Wald, L. Fusion of high spatial and spectral resolution images: The ARSIS concept and its implementation. *Photogramm. Eng. Remote Sens.* **2000**, *66*, 49–61.
25. Chang, C.-I. Spectral information divergence for hyperspectral image analysis. In Proceedings of the IEEE 1999 International Geoscience and Remote Sensing Symposium. IGARSS'99 (Cat. No.99CH36293), Hamburg, Germany, 28 June–2 July 1999; Volume 1, pp. 509–511.
26. Alparone, L.; Aiazzi, B.; Baronti, S.; Garzelli, A.; Nencini, F.; Selva, M. Multispectral and panchromatic data fusion assessment without reference. *Photogramm. Eng. Remote Sens.* **2008**, *74*, 193–200. [[CrossRef](#)]
27. Wang, Z.; Bovik, A.C. A universal image quality index. *IEEE Signal Process. Lett.* **2002**, *9*, 81–84. [[CrossRef](#)]