Spectral preservation fusion for remote sensing images using focus measure operators based on fast discrete curvelet transform and hyperspherical color space

Bin Zhong
Chen Zhang
MingWei Liao
HaiSheng Cai
Spectral preservation fusion for remote sensing images using focus measure operators based on fast discrete curvelet transform and hyperspherical color space

Bin Zhong,a,b Chen Zhang,c MingWei Liao,b and HaiSheng Cai*a

aJiangxi Agricultural University, College of Forestry, Key Laboratory of Poyang Lake Watershed Agricultural Resources and Ecology of Jiangxi Province, Nanchang, China
bJiangxi Provincial Geomatics Center, Jiangxi Province Engineering Research Center of Surveying Mapping and GeoInformation, Nanchang, China
c915 Geological Brigade of Jiangxi Provincial Bureau of Geology and Mineral Resources, Nanchang, China

Abstract. How to preserve the spectral information when enhancing the spatial details is a key issue of remote sensing image fusion. The component substitution (CS)-based fusion methods can effectively enhance the spatial details while suffering spectral distortion, and multiresolution analysis (MRA)-based methods have advantages in preserving spectral information but are not satisfactory in terms of spatial details. This paper proposes a hybrid method to integrate the advantages of CS- and MRA-based approaches. The intensity image is first obtained from an original multispectral (MS) image by hyperspherical color space (HCS) transform; then, the intensity image and original panchromatic (PAN) image are decomposed by fast discrete curvelet transform (FDCT). The focus measure operators are introduced to fused low-frequency, middle-high frequency, and fine scale subband coefficients in curvelet domain with specific fusion strategies. The final fusion image is achieved by inverse FDCT and inverse HCS transform. From the aspects of subjective and objective quality assessments, the experimental results on various types of remote sensing images including IKONOS, QuickBird, and WorldView-2 indicate that compared with existing well-known algorithms and commercial softwares, the proposed method shows obvious advantages in preserving spectral information and maintaining the spatial details. © The Authors. Published by SPIE under a Creative Commons Attribution 3.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: 10.1117/1.JRS.12.035017]

Keywords: remote sensing image fusion; fast discrete curvelet transform; hyperspherical color space; focus measure operator; spectral preservation.

Paper 180302 received Apr. 10, 2018; accepted for publication Aug. 20, 2018; published online Sep. 13, 2018.

1 Introduction

With the continuous development of satellite remote sensing technology, remote sensing images have greatly improved in spatial resolution and spectral resolution. However, due to the limitation of technical conditions, the spatial resolution and spectral resolution of the image are conflicting for the same imaging system. Most commercial satellites such as SPOT, IKONOS, QUICKBIRD, and WorldView provide high-resolution panchromatic (PAN) images and low-resolution multispectral (MS) images. For example, IKONOS offers 1-m high-resolution PAN images and 4-m low-resolution MS images. The spatial resolution of QuickBird’s PAN and MS images is 0.61 and 2.44 m, respectively. The high-resolution PAN image reflects the spatial structure information and can fully express the detailed features of the ground objects. The low-resolution MS image has rich spectral information and is conducive to the identification and interpretation of the ground targets. More and more remote sensing applications need to...
combine the advantages of these two types of images to generate MS images with high-spatial resolution to meet the deeper application.\(^1-^3\)

Image fusion is an effective way to solve this problem, which is to spatially register the image data of the same area acquired by different types of sensors, and then use certain methods to combine the advantages or complementarity of each image to generate new images. In general, the spectral resolution of MS images is high, but the spatial resolution is relatively low; that is, the spatial detail expression ability is poor; the PAN image has high spatial resolution but the spectral resolution is low. Therefore, the fused image has high spatial detail performance while preserving the spectral characteristics of the MS image.

Remote sensing image fusion can be divided into three levels, i.e., pixel level, feature level, and decision level. Pixel level is the lowest level. The image is first preprocessed such as spatial registration, and then weighted and summed. The obtained value is the pixel value of the new image on the coordinate. It is mainly to increase the useful information content in the image with higher accuracy, but the amount of information that needs to be processed is too large. Feature level fusion is to fuse at the feature extraction stage. Feature extraction is performed on different images, and fusion is performed on the same type of features on each image. After such fusion, the required thematic image features can be extracted from the fused image with a high degree of confidence. The fused image can provide the feature information required for decision analysis to a large extent. The disadvantage is that its fusion accuracy is relatively poor. Decision fusion is of the highest level. Each type of sensor image is first classified, and the characteristic images in each category are determined. These classification decisions are then combined into a decision tree. Because a complete decision tree requires more than two classification results, for a target, at least two sensors are needed to detect and classify targets at the same time before classification. This level is very fault-tolerant and open, but with low accuracy.

At present, most fusion methods are concentrated at the pixel level. There are four main types:

(1) Component substitution (CS). The basic idea of this kind of method is to transform the original MS image into another feature space, then replace one of the components in this feature space with PAN image, and then transform back to the original space to obtain a fused image. The most typical method is the IHS and PCA-based fusion method. The IHS transform algorithm is simple and can obtain high spatial resolution, but it can only be applied to three bands. It is incapable of multiband images. Moreover, since the difference between the PAN image and the intensity image causes the change of saturation, it will cause serious spectral distortion.\(^4\) PCA transform can fuse any multiband image, but the fused image will lose the original physical characteristics, and the fusion effect depends on the correlation between the two replaced main components.\(^5\) In order to reduce these adverse influences, various new CS-based methods are proposed, such as Gram–Schmidt (GS) method,\(^6\) generalized IHS (GIHS),\(^7\) Tu’s HIS,\(^8\) hyperspherical color space (HCS),\(^9\) the Amélioration de la Résolution Spatiale par Injection de Structures (ARSIS) concept.\(^10\) In these methods, the PAN image plays an important role. Therefore, the fusion image has more or less spectral distortions.

(2) Multiresolution Analysis (MRA). The introduction of MRA based on wavelet transform (WT) has opened up a new way to reduce spectral distortion, because the WT can describe and extract spatial detail information and spectral information of images at different scales.\(^11-^14\) In recent years, more and more MRA methods have been proposed, such as the discrete wavelet transform (DWT),\(^13\) à trous wavelet transform (ATWT),\(^15\) nonsubsampled Contourlet transform,\(^16\) and nonsubsampled Shearlet transform,\(^17\) fast discrete curvelet transform (FDCT),\(^18\) smoothing filter-based intensity modulation,\(^19\) MTF-GLP with high-pass modulation,\(^20\) and MTF-generalized Laplacian pyramid (MTF-GLP).\(^21\) The original MS and PAN images are decomposed into several scales and the spatial and spectral information are expressed as coefficients in scales. Then, the fused image is obtained by integrating coefficients at every scale using different strategies and the inversing MRA. Compared with CS-based approaches, the MRA-based methods can improve the performance of spectral preservation, but may not show many advantages in maintaining spatial information.

(3) Hybrid fusion. In order to make full use of the advantages of the CS- and MRA-based approaches, many hybrid methods have been proposed to combine the two together. First, the spatial and spectral information of the original MS image are separated by CS, and then
the spatial details are extracted by the MRA analysis method from PAN image. Spatial details are then injected into the MS image to achieve a balance between spatial information and spectral information. Such approaches include a large number of research results, such as additive wavelet-based proportioning, curvelet and ICA, discrete contourlet transform and PCA, guided filter PCA (GFPCA), and so on.

(4) Model-based fusion. Model-based fusion methods could use more prior information and other useful data from original MS and PAN images, hence in some cases they may have better performances. A hierarchical Bayesian model is proposed in Ref. 26 by employing an appropriate prior distribution obtained from Hamiltonian Monte Carlo algorithm. Using the learned spectral dictionary and the known high spatial resolution data, a sparse representation-based fusion model is used to desire high quality MS images with high resolution. In order to solve the problem of uncertain result with high reconstruction error, a compressive sensing-based fusion model with fewer nonzero coefficients is proposed to reduce the processing time and achieve high quality fused images. Some other model-based fusion methods such as total variation-based approach, dictionary learning-based approach, and Markov random field model-based approach also could guarantee the fusion quality in many applications.

Belonging to the hybrid fusion type, this paper combines the information extraction capabilities of FDCT and HCS together with the focus measure operators, to improve the spatial resolution while preserve better spectral information.

The outline of the paper is as follows: Sec. 2 briefly describes the related work including FDCT, HCS and focus measure operators. The proposed image fusion method with the FDCT and HCS is discussed in Sec. 3. In Sec. 4, the experimental results on various remote sensing images are presented. Finally, Sec. 5 discusses the conclusion of the proposed method.

2 Related Work

2.1 Fast Discrete Curvelet Transform

The digital implementation of the first-generation curvelet is more complex and requires a series of steps such as subband decomposition, smooth partitioning, normalization, and Ridgelet analysis, and the decomposition of the curvelet pyramid also brings a huge amount of data redundancy. So Candes et al. proposed a quick and easy-to-understand curvelet transform algorithm, namely the second-generation curvelet.

The second-generation curvelet is completely different from the first-generation. The first generation curvelet’s construction idea is to approximately treat the curve in each block as a straight line through enough small blocks, and then use the local ridgelet to analyze its characteristics. The second-generation curvelet and ridgelet theory have no relation, and the implementation process also does not need to use ridgelet. The only difference between the two theories lies in the abstract mathematical meaning such as tight support and framework.

The discrete curvelet transform is

\[ c(i, j, k) = \int \tilde{f}(\omega) \bar{U}_j(S_0^{-T} \omega) e^{i(S_0^{-T}b \omega)} d\omega. \]  

(1)

Since the shear block \( S_0^{-T}(k_1 \times 2^{-j}, k_2 \times 2^{-j/2}) \) is not a standard rectangle, in order to use the fast Fourier algorithm, the above equation could be rewritten as

\[ c(i, j, k) = \int \tilde{f}(\omega) \bar{U}_j(S_0^{-T} \omega) e^{i(S_0^{-T}b \omega)} d\omega = \int \tilde{f}(S_0^{-T} \omega) U_j(\omega) e^{i(b \omega)} d\omega. \]  

(2)

Then, the FDCT could be implemented by the Wrap algorithm in local Fourier Transform. The wrap around the origin is the core idea of the curvelet based on the Wrap algorithm. Any region is mapped to the affine region of the origin through the periodic technique. This mapping is a one-to-one correspondence. Specific steps are as follows:

Step 1: Obtaining \( \tilde{f}[n_1, n_2] \), \(-n/2 \leq n_1, n_2 < n/2 \) through two-dimensional (2-D) FFT on \( f[t_1, t_2] \in L^2(R) \).
Step 2: Multiplying $\mathcal{F}[n_1, n_2]$ by parabolic window $\mathcal{U}_{j,l}[n_1, n_2]$ for each scale and angle parameter set $(j, l)$ to localized $\mathcal{F}$.

Step 3: Localizing $\mathcal{F}$ by warping around the origin then getting $\mathcal{F}_{j,l}[n_1, n_2] = W(\mathcal{U}_{j,l})[n_1, n_2]$.

Step 4: Implementing 2-D IFFT to localized $\mathcal{F}_{j,l}$ and finally obtaining the discrete curvelet coefficients $c^D(j, l, k)$.

2.2 Hyperspherical Color Space-Based Fusion

Converting from the original color space to HCS follows the standard conversion from $n$-dimensional Cartesian space to $n$-dimensional hypersphere space. For an image with $N$ bands, an intensity component and $N - 1$ angular components in hyperspheres are obtained. The equation for the positive transformation from the $N$-band color space to the HCS is as follows:

\[
\begin{align*}
I &= \sqrt{x_1^2 + x_2^2 + \ldots + x_n^2} \\
q_1 &= \tan^{-1}\left[\frac{\sqrt{x_1^2 + x_2^2 + \ldots + x_n^2}}{x_1}\right] \\
q_{n-2} &= \tan^{-1}\left[\frac{\sqrt{x_n^2}}{x_{n-1}}\right] \\
q_{n-1} &= \tan^{-1}\left[\frac{x_n}{x_{n-1}}\right]
\end{align*}
\]

where $x_i$ is the $i$'th band of the original $n$-bands image.

The inverse transform is:

\[
\begin{align*}
x_1 &= I \cos q_1 \\
x_2 &= I \sin q_1 \cos q_2 \\
x_{n-1} &= I \sin q_1 \sin q_2 \ldots \sin q_{n-2} \cos q_{n-1} \\
x_n &= I \sin q_1 \sin q_2 \ldots \sin q_{n-2} \sin q_{n-1}
\end{align*}
\]

HCS-based image fusion is similar to the IHS fusion method, which belongs to the CS fusion, and is divided into naïve mode and smart mode. The procedures of the naïve mode are:

1. Calculating the sum of square of MS and PAN images

\[
\begin{align*}
I^2 &= \sum_{i=1}^{N} x_i^2 \\
P^2 &= (\text{Pan})^2
\end{align*}
\]

2. Calculating the matched PAN

\[
\text{PM}^2 = \frac{\sigma_0}{\sigma_1} (P^2 - \mu_1 + \sigma_1) + \mu_0 - \sigma_0.
\]

where $\mu_1$, $\sigma_1$, $\mu_0$, and $\sigma_0$ denote the mean and variance of $P^2$ and $I^2$.

\[
I_{\text{adj}} = \sqrt{\text{PM}^2}.
\]

3. getting the fused image by substituting $I$ in HCS transform as $I_{\text{adj}}$, and applying the inverse HCS transform.

2.3 Focus Measure Operators

1. Sum-modified-Laplacian (SML)\textsuperscript{35–37}

Nayar and Nakagawa\textsuperscript{38} proposed a modified Laplacian (ML) by the fact that the opposite signs will appear in the second derivatives in the $x$- and $y$-directions of the Laplacian which tend to cancel each other:

\[
\nabla_{\text{ML}}^2 f(x, y) = |2f(x, y) - f(x \text{ step, } y) - f(x + \text{ step, } y)| + |2f(x, y) - f(x, y \text{ step}) - f(x, y + \text{ step})|.
\]
where step is a variable spacing. Step always equals 1. The SML is

$$\text{SML} = \sum_{i=x-N}^{i=x+N} \sum_{j=y-N}^{j=y+N} \nabla^2_{m_l} f(i, j),$$  \hspace{1cm} (9)

where $N$ is the window size used to compute the SML focus measure.

(2) Energy of gradient (EOG)

Denoting the horizontal and vertical gradients of an image are $f_x = f(x + 1, y) - f(x, y)$ and $f_y = f(x, y + 1) - f(x, y)$, then the EOG is

$$\text{EOG} = \sum_{x} \sum_{y} (f_x^2 + f_y^2).$$  \hspace{1cm} (10)

(3) Energy of Laplacian of image (EOL)

For analyzing the high spatial frequencies and the edge sharpness of an image, the EOL is proposed as

$$\text{EOL} = \sum_{x} \sum_{y} (f_{xx} + f_{yy})^2,$$  \hspace{1cm} (11)

where

$$f_{xx} + f_{yy} = 20f(x, y) - f(x - 1, y - 1) - 4f(x - 1, y) - f(x - 1, y + 1) - 4f(x, y - 1) - 4f(x, y + 1) - f(x + 1, y) - f(x + 1, y + 1).$$  \hspace{1cm} (12)

(4) Tenengrad (TEN)

TEN is proposed by Tenenbaum, which employs Sobel operator to describe the gradient energy

$$\text{Tenengrad} = \sum_{x=2}^{M-1} \sum_{y=2}^{N-1} (\nabla S(x, y))^2,$$  \hspace{1cm} (13)

where $\nabla S(x, y) = [\nabla S_x(x, y)^2 + \nabla S_y(x, y)^2]^{1/2}$, $T$ is the threshold, $\nabla S_x(x, y)$ and $\nabla S_y(x, y)$ are the horizontal and vertical Sobel gradients, respectively:

$$\nabla S_x(x, y) = \{f(x + 1, y - 1) - f(x - 1, y - 1) + 2f(x + 1, y) - 2f(x - 1, y) + f(x + 1, y + 1) - f(x - 1, y + 1)\},$$  \hspace{1cm} (14)

$$\nabla S_y(x, y) = \{f(x - 1, y + 1) - f(x - 1, y - 1) + 2f(x, y + 1) - 2f(x, y - 1) + f(x + 1, y + 1) - f(x + 1, y - 1)\}. $$  \hspace{1cm} (15)

(5) Spatial frequency (SF)$^{39}$

SF is a modification of EOG with the following expression:

$$\text{SF} = \sqrt{(\text{RF})^2 + (\text{CF})^2},$$  \hspace{1cm} (16)

where

$$\text{RF} = \sqrt{\frac{1}{M \times N} \sum_{x=1}^{M} \sum_{y=2}^{N} [f(x, y) - f(x, y - 1)]^2},$$  \hspace{1cm} (17)
\[ CF = \sqrt{\frac{1}{M \times N} \sum_{x=2}^{M} \sum_{y=1}^{N} [f(x, y) - f(x - 1, y)]^2}. \]  

(18)

3 Proposed Method

3.1 Overall Workflow

The fusion scheme based on FDCT and HCS transform proposed in this paper is shown in the figure below. It is mainly divided into the following steps: first, the original MS image is transformed by HCS to obtain the intensity component and multiple angular components. Then, the intensity component and original PAN image are subjected to FDCT to obtain low-frequency, midhigh frequency, and fine-scale subband coefficients. Different fusion strategies are adopted for each subband coefficient to obtain new low-frequency, midhigh frequency, and fine-scale subband coefficients. According to the characteristic that low-frequency energy concentrates in the intensity image, low-frequency subbands are first calculated by focus measure operators, and then weighted by variance. According to the characteristics of rich high-frequency information in the midhigh frequency and fine-scale subbands of the PAN image, based on the calculation of focus measure operators, the midhigh frequency subbands are fused using the correlation coefficient matching strategy, and the fine-scale subband directly adopts the largest absolute value fusion strategy. Then, the new low frequency, midhigh frequency, and fine-scale subband coefficients are inversely transformed by FDCT to obtain a new intensity component. Finally, the HCS inverse transformation is performed on the new intensity component and the angular components obtained from the HCS transformation of the original MS image in the first step, and the final fused image is obtained (Fig. 1).

![Fig. 1 Workflow of the proposed method.](https://www.spiedigitallibrary.org/journals/Journal-of-Applied-Remote-Sensing)
3.2 HCS and FDCT Transform

First, the original MS image is resampled to the same pixel size as the original PAN image, and an intensity component and multiple angular components could be obtained by applying HCS transform to the resampled MS image.

Then, the FDCT is used to the intensity component and original PAN image, respectively, and low-frequency, midhigh frequency, and fine-scale subbands coefficients for intensity component and PAN image are obtained.

Based on the above procedure, the coefficients fusion is performed using different strategies for low-frequency, midhigh frequency, and fine-scale subbands.

3.3 Fusion of Low Frequency

The low-frequency coefficient of the image reflects the energy distribution of the image. According to the imaging principle of the MS sensor and the panchromatic sensor, the MS image reflects the spectral characteristics of the target, and the PAN image reflects the spatial details of the target. The gray-scale distribution characteristics of the same target in both images are often very different. Therefore, the traditional average weighting strategy for low-frequency coefficients does not consider the physical characteristics of the sensor, which will increase the spectral distortion of the fused image and lose some useful information. Therefore, this paper first calculates the metric of the low-frequency coefficients by focus evaluation operator, and then uses the local variance weighted strategy to calculate the low-frequency coefficient of the fused image, as

\[
C_{\text{fus}}(x, y) = C_I^{l}(x, y) + \frac{\sigma(W_{\text{pan}})}{\sigma(W_{\text{pan}}) + \sigma(W_{I})} \times [C_{\text{pan}}^{l}(x, y) - \min(W_{\text{pan}}, W_{I})],
\]

where \(C_I^{l}, C_{\text{pan}}^{l},\) and \(C_{\text{fus}}^{l}\) represent the low-frequency subband coefficients of the HCS transformed intensity image, PAN image, and fused image, respectively. \(W\) represents the local window, \(\sigma\) represents the variance in the local window of the focus measure operator metric value, \(\overline{W}\) represents the mean value in the local window of the original low-frequency coefficient, and \(\min(W_{\text{pan}}, W_{I})\) are the minimum value of the local window mean value of the original low-frequency coefficient which represents the common low-frequency characteristic of the MS intensity image and the PAN image. \(C_{\text{pan}}^{l}(x, y) - \min(W_{\text{pan}}, W_{I})\) represents the characteristic low-frequency characteristic of the PAN image. \(\frac{\sigma(W_{\text{pan}})}{\sigma(W_{\text{pan}}) + \sigma(W_{I})}\) is a weighting coefficient and is used to inject the low-frequency characteristic specific to the PAN image to the MS intensity image. The variance \(\sigma\) is a measure of the difference in energy distribution. After the calculation of the focus measure operator, the variance can better reflect the energy distribution of the low-frequency coefficients. It can be seen from the equation that the low-frequency coefficients of the fused image consist of the low-frequency coefficients of the MS image and the unique low-frequency coefficients of the PAN image, so the entire information of the MS image can be completely preserved and the unique information of the PAN image can be fully utilized. This strategy can achieve maximum preservation of the low-frequency information of original images.

3.4 Fusion of Middle-High Frequency

The middle-high frequency coefficients of the image reflect the distribution of the details of the image. The PAN image has rich high-frequency detail information. Relatively speaking, the detailed features of MS intensity images are not obvious. Therefore, this paper first calculates the metrics of the focus measure operator of a middle-high frequency subband, and then uses local correlation coefficients to calculate the matching degree of the metrics, and accordingly chooses the appropriate middle-high frequency coefficient. The local correlation coefficient is defined as
Zhong et al.: Spectral preservation fusion for remote sensing images using focus measure operators...

\[
CC_{A,B} = -\frac{1}{M \times N} \times \frac{\sum_{i=1}^{M} \sum_{j=1}^{N} [A(i,j) - \mu_A][B(i,j) - \mu_B]}{\sqrt{\left\{\sum_{i=1}^{M} \sum_{j=1}^{N} [A(i,j) - \mu_A]^2\right\}\left\{\sum_{i=1}^{M} \sum_{j=1}^{N} [B(i,j) - \mu_B]^2\right\}},
\]

where \(A\) and \(B\) represent two vectors, \(\mu_A\) and \(\mu_B\) denote the mean of \(A\) and \(B\), respectively. The value of \(CC\) is the matching degree of the middle-high frequency coefficients between the MS intensity image and the PAN image in a certain neighborhood. When selecting the appropriate middle-high frequency coefficients, a threshold \(T\) is first set, which is defined as the average of the correlation coefficients of all subbands, i.e.,

\[
T = \frac{1}{n} \sum_{i=1}^{n} CC[C^{mh}_I(i), C^{mh}_{pan}(i)],
\]

where \(n\) is the number of middle-high frequency subbands, \(C^{mh}_I(i)\) and \(C^{mh}_{pan}(i)\) are the \(i\)’th sub-band of the MS intensity image and the PAN image, respectively. \(CC[C^{mh}_I(i), C^{mh}_{pan}(i)]\) represents the correlation coefficient of the \(i\)’th subband. The \(CC\) value of a particular subband is compared with the threshold \(T\). \(CC(i) < T\) indicates that the matching degree between the two subbands is very low, hence it is necessary to retain both information at the same time. At this time, the middle-high frequency coefficient of the fused image is defined as

\[
C^{mh}_{fus}(x,y) = C^{mh}_I(x,y) + C^{mh}_{pan}(x,y).
\]

When \(CC(i) \geq T\), the matching degree is high. In order to reduce the redundancy of information, only high-middle frequency coefficients with relatively large absolute values are preserved. At this time, the middle-high frequency coefficients of the fused image are defined as

\[
C^{mh}_{fus}(x,y) = \begin{cases} 
C^{mh}_{pan}(x,y), & \text{if } |C^{mh}_{pan}(x,y)| > |C^{mh}_I(x,y)| \\
C^{mh}_I(x,y), & \text{otherwise}
\end{cases}
\]

\[3.5\] Fusion of Fine Scale Frequency

The fine-scale subband coefficients reflect the highest level of detail in the image. Since the spatial resolution of an MS image is much lower than that of a PAN image, the energy of its subband coefficients is much lower than that of a PAN image. Of course, it is not ruled out that in some subbands, the energy is higher than that of a PAN image. Therefore, for the fine-scale subband coefficients, this paper only needs to use the approach with the largest absolute value to keep the details as much as possible. Its fusion strategy is

\[
C^{fine}_{fus}(x,y) = \begin{cases} 
C^{fine}_{pan}(x,y), & \text{if } |C^{fine}_{pan}(x,y)| > |C^{fine}_I(x,y)| \\
C^{fine}_I(x,y), & \text{otherwise}
\end{cases}
\]

\[3.6\] Fusion Result

Through the above subband fusion strategies, the fused subbands of the low-frequency subband, the middle-high frequency subband, and the fine-scale subband are obtained. All of the fused subbands were subjected to FDCT inverse transformation to obtain a fused MS intensity image. The HCS inverse transformation is performed on each angular component obtained by HCS transformation of original MS image and the newly generated intensity image, and the final fused image can be obtained.
4 Experiments and Discussion

4.1 Experimental Setup

Three types of remote sensing images, i.e., IKONOS, QuickBird, and WorldView-2, are used to test the performance of the proposed method and the results are compared with existing well-known fusion algorithms and algorithms from commercial softwares. The basic information of these three types of images are listed as following (Table 1):

The size of images used in this paper is all 2048 pixels × 2048 pixels and the MS images are all resampled to the same pixel size with PAN images using bicubic interpolation algorithm. Some subsets images with size of 512 pixels × 512 pixels are shown in Fig. 2.

The GIHS algorithm is selected since it is the generalized representative CS algorithm. To compare the performance between with/without focus measure operators, the same framework named “FDCT + HCS” (“C + HCS” for short) is used without applying focus measure operators before low-, midhigh frequency, and fine scale fusion procedure. The GS algorithm from ENVI software, HCS algorithm from ERDAS IMAGINE software, and UNB algorithm from PCI GEOMATICA software are used to compare the fusion performance since they are three famous and mainstream commercial softwares in remote sensing field and can provide fair fusion results to be compared. Meanwhile, for a compressive comparison, six methods with outstanding

<table>
<thead>
<tr>
<th></th>
<th>PAN spatial resolution (m)</th>
<th>MS spatial resolution (m)</th>
<th>Number of bands</th>
</tr>
</thead>
<tbody>
<tr>
<td>IKONOS</td>
<td>1.0</td>
<td>4.0</td>
<td>4</td>
</tr>
<tr>
<td>QuickBird</td>
<td>0.61</td>
<td>2.44</td>
<td>4</td>
</tr>
<tr>
<td>WorldView-2</td>
<td>0.5</td>
<td>2.0</td>
<td>8</td>
</tr>
</tbody>
</table>

Fig. 2 Remote sensing image datasets used in this paper. (a) and (d) MS and PAN images of IKONOS, (b) and (e) MS and PAN from QuickBird images, (c) and (f) MS and PAN images of WorldView-2.
Fig. 3 Fusion results on IKONOS dataset. (a) Original MS image, (b) original PAN image, (c) GiHS, (d) GS, (e) C + HCS, (f) HCS, (g) UNB, (h) ATWT, (i) BDSD, (j) GSA, (k) PRACS, (l) MTF_GLP, (m) MTF_GLP_HPM, (n) FDCT + HCS + SML, (o) FDCT + HCS + EOG, (p) FDCT + HCS + EOL, (q) FDCT + HCS + TEN, and (r) FDCT + HCS + SF.
performances in Ref. 22 are also used, i.e., ATWT, band-dependent spatial detail (BDSD) algorithm, adaptive GS (GSA), partial replacement adaptive CS (PRACS), MTF-generalized LP (MTF_GLP), and MTF-GLP with HPM (MTF_GLP_HPM).

For the proposed fusion framework, five focus measure operators are used hence five fusion methods are obtained, i.e., FDCT + HCS + SML, FDCT + HCS + EOG, FDCT + HCS + EOL, FDCT + HCS + TEN, and FDCT + HCS + SF. In the experiment section, we use SML, EOG, EOL, TEN, and SF to represent the above five fusion methods. Therefore, 16 fusion methods (GIHS, GS, C + HCS, HCS, UNB, ATWT, BDSD, GSA, PRACS, MTF_GLP, MTF_GLP_HPM, FDCT + HCS + SML, FDCT + HCS + EOG, FDCT + HCS + EOL, FDCT + HCS + TEN, and FDCT + HCS + SF) are compared in this paper.

4.2 Fusion Quality Assessment Indices

Quality assessment indices are used to objectively compare the performance of all employed fusion algorithms. There are many types of indices which can be divided into three categories, i.e., spatial quality indices, spectral quality indices, and overall quality indices. According to Wald’s protocol, two modes with and without reference images are used in this paper. The first mode treats the original MS image as reference image. The original MS and PAN images

![Subset fusion results on IKONOS dataset. (a) Original MS image, (b) original PAN image, (c) GIHS, (d) GS, (e) C + HCS, (f) HCS, (g) UNB, (h) ATWT, (i) BDSD, (j) GSA, (k) PRACS, (l) MTF_GLP, (m) MTF_GLP_HPM, (n) FDCT + HCS + SML, (o) FDCT + HCS + EOG, (p) FDCT + HCS + EOL, (q) FDCT + HCS + TEN, and (r) FDCT + HCS + SF.](image-url)
are first downscaled using a bicubic resampling algorithm, for example, the original MS image with 4-m spatial resolution and PAN image with 1-m spatial resolution are downscaled to the MS image with 16-m spatial resolution and PAN image with 4-m spatial resolution. Then, the downscaled MS and PAN images are fused using some fusion algorithm, and finally the fused result is compared with the original MS image. The basic assumption and principle of this mode is that the fusion algorithm should keep consistency for image fusion procedure. For the second mode, the original MS and PAN images are directly fused in their original spatial resolution and the quality assessment could be conducted without references.

The indices for first mode include four spectral indices, i.e., Erreur Relative Globale Adimensionnelle de Synthese (ERGAS, or relative global dimensional synthesis error),\textsuperscript{40,41} spectral angle mapper (SAM),\textsuperscript{42} spectral correlation coefficient (CC),\textsuperscript{43} Universal Image Quality Index (UIQI),\textsuperscript{44} two spatial indices, i.e., spatial frequency index (SFI) and peak signal-to-noise ratio (PSNR), two overall indices, i.e., quality AB/F (QABF)\textsuperscript{45} and structure similarity index (SSIM).\textsuperscript{46}

<table>
<thead>
<tr>
<th>Method</th>
<th>ERGAS</th>
<th>SAM</th>
<th>CC</th>
<th>UIQI</th>
<th>SFI</th>
<th>PSNR</th>
<th>QABF</th>
<th>SSIM</th>
<th>QNR</th>
<th>GD</th>
</tr>
</thead>
<tbody>
<tr>
<td>GIHS</td>
<td>6.524</td>
<td>3.753</td>
<td>0.540</td>
<td>0.705</td>
<td>8.769</td>
<td>22.773</td>
<td>0.492</td>
<td>1.355</td>
<td>0.732</td>
<td>0.796</td>
</tr>
<tr>
<td>GS</td>
<td>22.524</td>
<td>40.965</td>
<td>0.480</td>
<td>0.368</td>
<td>22.669</td>
<td>12.434</td>
<td>0.097</td>
<td>0.634</td>
<td>0.779</td>
<td>0.678</td>
</tr>
<tr>
<td>C + HCS</td>
<td>2.671</td>
<td>2.009</td>
<td>0.938</td>
<td>0.947</td>
<td>9.587</td>
<td>30.528</td>
<td>0.459</td>
<td>1.426</td>
<td>0.910</td>
<td>0.925</td>
</tr>
<tr>
<td>HCS</td>
<td>2.452</td>
<td>1.882</td>
<td>0.948</td>
<td>0.946</td>
<td>10.314</td>
<td>31.209</td>
<td>0.423</td>
<td>1.435</td>
<td>0.918</td>
<td>0.943</td>
</tr>
<tr>
<td>UNB</td>
<td>2.155</td>
<td>1.863</td>
<td>0.951</td>
<td>0.956</td>
<td>10.830</td>
<td>32.347</td>
<td>0.442</td>
<td>1.452</td>
<td>0.921</td>
<td>0.940</td>
</tr>
<tr>
<td>ATWT</td>
<td>2.671</td>
<td>1.989</td>
<td>0.938</td>
<td>0.947</td>
<td>9.587</td>
<td>30.528</td>
<td>0.459</td>
<td>1.426</td>
<td>0.901</td>
<td>0.925</td>
</tr>
<tr>
<td>BDSG</td>
<td>1.848</td>
<td>1.749</td>
<td>0.958</td>
<td>0.963</td>
<td>8.657</td>
<td>34.354</td>
<td>0.438</td>
<td>1.485</td>
<td>0.915</td>
<td>0.945</td>
</tr>
<tr>
<td>GSA</td>
<td>1.753</td>
<td>1.828</td>
<td>0.961</td>
<td>0.964</td>
<td>6.954</td>
<td>30.186</td>
<td>0.452</td>
<td>1.467</td>
<td>0.917</td>
<td>0.941</td>
</tr>
<tr>
<td>PRACS</td>
<td>4.308</td>
<td>2.197</td>
<td>0.813</td>
<td>0.847</td>
<td>8.004</td>
<td>26.306</td>
<td>0.458</td>
<td>1.476</td>
<td>0.893</td>
<td>0.921</td>
</tr>
<tr>
<td>MTF_GLPA</td>
<td>1.866</td>
<td>2.023</td>
<td>0.944</td>
<td>0.955</td>
<td>7.183</td>
<td>33.015</td>
<td>0.450</td>
<td>1.461</td>
<td>0.915</td>
<td>0.937</td>
</tr>
<tr>
<td>MTF_GLPH</td>
<td>2.236</td>
<td>2.102</td>
<td>0.953</td>
<td>0.957</td>
<td>10.230</td>
<td>32.070</td>
<td>0.434</td>
<td>1.444</td>
<td>0.908</td>
<td>0.932</td>
</tr>
<tr>
<td>SML</td>
<td>2.108</td>
<td>1.780</td>
<td>0.948</td>
<td>0.952</td>
<td>6.508</td>
<td>32.516</td>
<td>0.483</td>
<td>1.474</td>
<td>0.921</td>
<td>0.948</td>
</tr>
<tr>
<td>EOG</td>
<td>1.745</td>
<td>1.783</td>
<td>0.966</td>
<td>0.969</td>
<td>7.323</td>
<td>34.159</td>
<td>0.483</td>
<td>1.489</td>
<td>0.928</td>
<td>0.949</td>
</tr>
<tr>
<td>EOL</td>
<td>1.745</td>
<td>1.782</td>
<td>0.966</td>
<td>0.969</td>
<td>7.280</td>
<td>34.161</td>
<td>0.482</td>
<td>1.490</td>
<td>0.927</td>
<td>0.948</td>
</tr>
<tr>
<td>TEN</td>
<td>1.806</td>
<td>1.784</td>
<td>0.964</td>
<td>0.967</td>
<td>7.434</td>
<td>33.863</td>
<td>0.487</td>
<td>1.495</td>
<td>0.923</td>
<td>0.949</td>
</tr>
<tr>
<td>SF</td>
<td>1.808</td>
<td>1.782</td>
<td>0.963</td>
<td>0.966</td>
<td>7.157</td>
<td>33.853</td>
<td>0.482</td>
<td>1.485</td>
<td>0.930</td>
<td>0.949</td>
</tr>
</tbody>
</table>

Note: The value in bold denotes the best value of the method.

The GS result shows much higher ERGAS and SAM, much lower CC and UIQI indicating that it has the worst spectral information, which is also obvious as seen in Figs. 3 and 4. The spectral quality of GIHS result is worse than the other methods since its spectral quality indices are worse. The C + HCS and HCS show similar quality indices values indicating that they show the similar fusion performance. BDSG has the highest PSNR while PRACS shows worse performance than MTF-based methods due to the over sharpened details. There are a little differences between five methods in the proposed framework, and they have better spectral quality than the other existing fusion methods. Benefiting from the better preservation of spectral information, the overall qualities of five methods in the proposed framework are also better than the other methods, which can be seen from higher values in SSIM, QNR, and GD.

are first downscaled using a bicubic resampling algorithm, for example, the original MS image with 4-m spatial resolution and PAN image with 1-m spatial resolution are downscaled to the MS image with 16-m spatial resolution and PAN image with 4-m spatial resolution. Then, the downscaled MS and PAN images are fused using some fusion algorithm, and finally the fused result is compared with the original MS image. The basic assumption and principle of this mode is that the fusion algorithm should keep consistency for image fusion procedure. For the second mode, the original MS and PAN images are directly fused in their original spatial resolution and the quality assessment could be conducted without references.

The indices for first mode include four spectral indices, i.e., Erreur Relative Globale Adimensionnelle de Synthese (ERGAS, or relative global dimensional synthesis error),\textsuperscript{40,41} spectral angle mapper (SAM),\textsuperscript{42} spectral correlation coefficient (CC),\textsuperscript{43} Universal Image Quality Index (UIQI),\textsuperscript{44} two spatial indices, i.e., spatial frequency index (SFI) and peak signal-to-noise ratio (PSNR), two overall indices, i.e., quality AB/F (QABF)\textsuperscript{45} and structure similarity index (SSIM).\textsuperscript{46}

The spatial quality indices describe the capacity that the fusion algorithm maintain the spatial information of the PAN image, hence the greater values of SFI and PSNR indicate that the fused image have high spatial quality. The spectral quality measures the performance that the fusion algorithm transmits spectral information of the original MS image to the fused result, therefore, the lower ERGAS, SAM, higher CC, and UIQI show that more spectral information are kept. However, there is no one algorithm that could get higher spatial quality and higher spectral...
Fig. 5 Fusion results on QuickBird dataset. (a) Original MS image, (b) original PAN image, (c) GIHS, (d) GS, (e) C + HCS, (f) HCS, (g) UNB, (h) ATWT, (i) BDS, (j) GSA, (k) PRACS, (l) MTF_GLP, (m) MTF_GLP_HPM, (n) FDCT + HCS + SML, (o) FDCT + HCS + EOG, (p) FDCT + HCS + EOL, (q) FDCT + HCS + TEN, and (r) FDCT + HCS + SF.
quality. Hence, the overall quality is used to strike a balance between spatial and spectral quality, and the higher the QABF and SSIM are, the better quality the fused image shows.

The second mode includes indices named quality with no reference (QNR)\textsuperscript{47} and Gaussian scale space-based index (GD).\textsuperscript{48} The higher QNR and GD values indicate that the better fused image is achieved.

4.3 Experiment on IKONOS Dataset

The original MS/PAN and 10 fused results are shown in Fig. 3. The original size of all images are 2048 pixels \( \times \) 2048 pixels and here only 512 pixels \( \times \) 512 pixels subset images are shown. Also the 128 pixels \( \times \) 64 pixels subset images labeled in the red rectangle in all original and fused images are enlarged in Fig. 4. The original MS image and fused images have four bands, and the true color images with red, green, and blue bands are shown in Figs. 3 and 4.

From Figs. 3(d) and 4(d) we can obviously see that the spectral information of GS-based result changes much compared with original MS image, where the green vegetation areas change to yellow white. Figures 3(c) and 4(c) tend to whiter than the original MS image in light green area. The other results show similar spectral information as the original MS image, except for a little sharper in the UNB, BDSD, PRACS, and MTF_GLP_HPM results in Figs. 3(g), 3(i), 4(g), and 4(i).

Fig. 6 Subset fusion results on QuickBird dataset. (a) Original MS image, (b) original PAN image, (c) GIHS, (d) GS, (e) C + HCS, (f) HCS, (g) UNB, (h) ATWT, (i) BDSD, (j) GSA, (k) PRACS, (l) MTF_GLP, (m) MTF_GLP_HPM, (n) FDCT + HCS + SML, (o) FDCT + HCS + EOG, (p) FDCT + HCS + EOL, (q) FDCT + HCS + TEN, and (r) FDCT + HCS + SF.
The objective quality assessment results using 10 indices are shown in Table 2, where “SML” represents “FDCT + HCS + SML,” and the other “EOG, EOL, TEN, and SF” are similar as “SML.”

### 4.4 Experiment on QuickBird Dataset

Figure 5 shows the fusion results on QuickBird dataset, where the size of original images is 2048 pixels × 2048 pixels and only 512 pixels × 512 pixels subset images are shown. Figure 6 is the enlarged areas along the road in Fig. 5 in red rectangles. There are four bands in the original MS and fused images, and the red, green, and blue bands are shown as true color image.

Similar to Figs. 3 and 4, the GIHS results in 5(c) and 6(c) show lighter green and the GS results in 5(d) and 6(d) show yellow green in grass and tree areas. Some areas are blurred in the C + HCS results especially along the road which can be seen obviously in Fig. 6(e). The color of UNB and the five methods of the proposed framework are similar to each other, except for some sharper edges in UNB results. The ATWT and MTF_GLP_HPM results are a little over sharpened. The objective quality assessment results are shown in Table 3.

It can be seen from Table 3 and Fig. 7 that the spectral qualities of GIHS, GS, HCS, ATWT, BDSD, and MTF_GLP_HPM are worse than the other methods since they have higher ERGAS and SAM, and lower CC and UIQI. For spatial quality, the GS, HCS, and UNB results are better than the other methods because they have sharper edges. From the overall quality aspect, the five methods in the proposed framework as well as MTF_GLP show better performance than the other existing methods, since they have better spectral quality and relatively better spatial quality which can be seen from spectral and spatial quality indices.

### 4.5 Experiments on WorldView-2 Datasets

The WorldView-2 images with size of 2048 pixels × 2048 pixels are tested using 10 fusion methods, and the results are shown in Figs. 7 and 8. Similar to the previous experiments on
Fig. 7 Fusion results on QuickBird dataset. (a) Original MS image, (b) original PAN image, (c) GIHS, (d) GS, (e) C + HCS, (f) HCS, (g) UNB, (h) ATWT, (i) BDS, (j) GSA, (k) PRACS, (l) MTF_GLP, (m) MTF_GLP_HPM, (n) FDCT + HCS + SML, (o) FDCT + HCS + EOG, (p) FDCT + HCS + EOL, (q) FDCT + HCS + TEN, and (r) FDCT + HCS + SF.
IKONOS and QuickBird dataset, only 512 pixels × 512 pixels subset and 128 pixels × 64 pixels enlarged subset with true color are shown.

The GS results in Figs. 7(d) and 8(d) are obviously overenhanced and the color distortion is much more than the other methods. Except the GS result, the other methods show similar spectral information and spatial details. In addition, the UNB, ATWT, and MTF_GLP_HPM results are a little sharper in edge areas. The objective quality assessment results are shown in Fig. 9.

The quality assessment results are consistent with the subjective observations that the GS result shows the worst overall quality which can also be seen from highest ERGAS and SAM values, and the other methods have similar spectral quality. The five methods in the proposed framework still have the best spectral performance resulting in the best overall qualities.

From the above experiments on three typical remote sensing image datasets with various spatial resolutions and different bands, we can see that the proposed framework could effectively preserve spectral information while maintaining spatial details resulting in the outstanding overall qualities. The introduction of focus measure operators could effectively improve the fusion performances compared with the FDCT + HCS framework, because the focus measure operators are helpful in aggregating energies in every frequency subbands of curvelet domain to enhance the spatial details, and finally achieving better overall performances together with the spectral preservation capacity of FDCT + HCS framework.
Fig. 9 The objective quality assessment results of WorldView-2 dataset. (a) ERGAS, (b) SAM, (c) CC, (d) UIQI, (e) SFI, (f) PSNR, (g) QABF, (h) SSIM, (i) QNR, and (j) GD.
5 Conclusion

This paper proposed a hybrid spectral preservation image fusion method combining with CS- and MRA-based approaches to try to achieve a balance between spatial details and spectral information. The HCS transform is used to separate the spatial and spectral information into an intensity component and multiple angular components. In order to inject the spatial details of PAN image into the intensity component, the FDCT is employed to decompose the intensity and PAN images into low-frequency, middle-high frequency, and fine scale subbands. By introducing the focus measure operators to concentrate energies of subband coefficients, each subband is fused to generate a new intensity component according to different strategies, i.e., the low-frequency subbands use the local variance weighted strategy, the middle-high subbands use the local correlation coefficients matching strategy, and the largest absolute value strategy is used for fine-scale subbands. The final fusion results are obtained by performing inverse HCS transform using the newly generated intensity component and original angular components. Various remote sensing images including IKONOS, QuickBird, and WorldView-2 are used to test the performance of the proposed method and existing well known and commercial softwares. According to Wald’s protocol, some widely used objective quality assessment indices including ERGAS, UIQI, QNR, and so on are used to compare the performance through the aspects of spatial quality, spectral quality, and overall quality. The experimental results indicate that the proposed method can better preserve the spectral information while enhancing the spatial details. The experimental results also show that the focus measure operators are helpful to improve the fusion performance. The future work will focus on developing more accurate strategies to further enhance the fusion capacities.

Acknowledgments

This work was jointly supported by the National Natural Science Foundation of China (31660140, 31560150), the National Key Technology R&D Program of China (2015BAH50F01), the Special Fund by Surveying & Mapping and Geoinformation Research in the Public Interest (201512026), the Science and Technology Program of Jiangxi Province (2017BBE50062,2017BBG70094, 2017BBE50073), and the Natural Science Foundation of Jiangxi Province (20161BAB204180).

References


Biographies for the authors are not available.