A Practical Pan-Sharpening Method with Wavelet Transform and Sparse Representation

Yu Liu[1, 3]   Zengfu Wang[1, 2, 3]

[1]Department of Automation, University of Science and Technology of China, Hefei, China
[2]Institute of Intelligent Machine, Chinese Academy of Sciences, Hefei, China
[3]National Engineering Laboratory for Speech and Language Information Processing,
University of Science and Technology of China, Hefei, China
liuyu1@mail.ustc.edu.cn

Abstract—**Pan-sharpening is an important remote sensing image pre-processing technique, which aims at obtaining a high-resolution multispectral (HRM) image by integrating the spectral information of a low-resolution multispectral (LRM) image and the spatial details of a high-resolution panchromatic (HRP) image. This paper proposes a new pan-sharpening method with sparse representation (SR) under the framework of wavelet transform. First, the wavelet transform is applied to the HRP image and the intensity component of LRM image. Then, the low-frequency components are fused based on SR to extract the spatial details in the HRP image as much as possible, and the dictionary is simply learned from high-quality nature images. Moreover, a novel strategy is also proposed to preserve the spectral information in the LRM image. On the other hand, the "numerous-but-sparse" high-frequency components are merged based on the local wavelet energy, which makes the algorithm more efficient than traditional SR-based methods. Finally, the fused result is obtained by performing inverse wavelet transform and inverse IHS transform. Experiments on WorldView-2 images demonstrate that the proposed method gives more spatial details and less spectral distortion compared with some conventional methods in terms of both visual quality and objective measurements.**

Keywords—Remote sensing, sparse representation, wavelet transform, image fusion, pan-sharpening

I. INTRODUCTION

Optical remote sensing images with both high spectral and high spatial resolution are always required in many remote sensing applications such as land-cover classification, map updating, reconnaissance, etc. However, due to the technological limitations of bandwidth and onboard storage, most remote sensing satellites such as QuickBird, IKONOS, and WorldView-2 often provide image data with spectral and spatial information separately. While the low spatial resolution multispectral (MS) images provide the spectral information, the panchromatic (PAN) images with high spatial resolution allows for accurate earth observation. The pan-sharpening technique, which aims at producing a high-resolution multispectral (HRM) image by fusing a low-resolution multispectral (LRM) image and a high-resolution panchromatic (HRP) image, has become an effective pre-processing technique in many remote sensing applications [1].

Various pan-sharpening approaches have been proposed in the past two decades. In general, the existing methods can be categorized into three classes: component-substitution-based methods, ARSIS-based methods, and model-based methods. The most popular component-substitution-based methods are intensity-hue-saturation (IHS) [2], principal component analysis (PCA) [3], and Gram-Schmidt (GS) [4]. As an example, the IHS-based method contains three main steps. First, the LRM image is transformed into the IHS color space. Then, the intensity component is replaced by the HRP image. Finally, the HRM image is obtained by the inverse transform. The fused images obtained through this category method always suffer from severe spectral distortion, although they can achieve high spatial resolution.

In the past decade, the methods based on the concept of *amélioration de la résolution spatiale par injection de structures* (ARSIS) have become popular [5], [6]. The ARSIS-based methods assume that the spatial details can be extracted from the HRP image and injected into the LRM image to integrate an HRM image. The most representative framework of ARSIS-based methods is wavelet transform (WT). Under it, the HRP image and each band of the LRM image are decomposed into low- and high-frequency components through WT. Then, the high-frequency component extracted from the HRP image is merged into the LRM bands. The fused image is obtained by performing the inverse WT in the final. Since the low-frequency components of LRM bands almost remain unchanged, this method preserves spectral information well. However, some spatial distortions may occur, such as blurring. In general, this category method offers a reliable framework for further exploration, and the key is to extract sufficient spatial details from the HRP image with preserving the spectral information in LRM image simultaneously.

More recently, the model-based methods have become a popular issue in pan-sharpening field [7], [8]. The LRM image and HRP image are viewed as the degraded versions of the HRM image, and various image restoration-based methods are used to reconstruct the HRP image by solving some optimization problems. Li and Yang [8] proposed a novel remote sensing imaging model and used compressed sensing technique to solve the regularization problem. After that, some improved versions based on sparse representation (SR) have been proposed [9], [10]. Similar to natural image super-resolution task, these methods can always get the state-of-the-art results with the effectiveness of SR. However, this category method has two main shortcomings. The first one is that it is difficult to obtain a universal dictionary to meet all the needs of

remote sensing images acquired from various satellites. The other is that these methods are often very time-consuming without parallel computation.

In this paper, we propose a practical pan-sharpening method, which belongs to the ARSIS-based methods, under the framework of wavelet transform. In it, the low-frequency components are merged via SR with the dictionary simply learned from natural images, which makes the method more practical. In this way, the high-frequency information in the HRP image can be extracted as much as possible. Moreover, a novel strategy is proposed to preserve spectral information to the most extent. On the other hand, the "numerous-but-sparse" high-frequency components are fused with a traditional rule based on wavelet energy, which makes the method more efficient than the general SR-based methods that directly process the original images in spatial domain.

II. SPARSE REPRESENTATION

Sparse representation (SR) addresses the essential sparisty of signals according to the characteristic of human visual system. It is a very powerful image modeling technique which has been successfully used in many image processing applications, such as image denoising [11] and super-resolution [12]. The basic idea behind SR is to assume that a natural signal can be well approximated by a sparse linear combination of atoms with respect to a dictionary, i.e., $x \approx D\alpha$, where $x \in R^n$ is the signal, $\alpha \in R^m$ is the sparse coefficients vector, and $D \in R^{n \times m}$ $(n < m)$ is an over-complete dictionary, which contains m prototype signals be referred to atoms. The goal of SR is to calculate the sparsest α for an input signal x with a given dictionary D through the following optimization problem:

$$\alpha = \arg\min_{\alpha} \| \alpha \|_0 \quad subject\ to\ \|x - D\alpha\|_2 < \varepsilon, \quad (1)$$

where $\varepsilon > 0$ is the error tolerance and the sparsity of vector α is often measured by its $l_0 - norm$, which counts the number of nonzero entries. Some pursuit algorithms have been proposed to solve this NP-hard problem [13], such as basis pursuit (BP) and orthogonal matching pursuit (OMP).

The dictionary plays a very important role in SR. Generally, there are two main approaches to obtain a dictionary. The first is using the analytical models, such as DCT and WT. However, most of these dictionaries are restricted to signals of a certain type, and cannot be used for an arbitrary family of signals. The second approach is applying the learning technique to obtain the dictionary from a large number of training examples. This type of dictionaries has a better representative ability and always outperforms the pre-constructed ones.

III. PROPOSED FUSION METHOD

A. Dictionary Learning

In the model-based methods, such as [8], the dictionary is often learned from a training set consists of many HRM images. However, the HRM images are not easy to get for they are just the target of pan-sharpening. Another limitation is the source images obtained from different types of remote sensors require different dictionaries to prevent the spectral information from being distorted. To make the proposed method more practical, the dictionary is trained from high-quality natural images and only used for extracting the spatial details. Fig. 1 shows ten example images of the training set.

Suppose M image patches of size $\sqrt{n} \times \sqrt{n}$ pixels are sampled and ordered lexicographically as column vectors in R^n space. Thus, the training database $\{y_i\}_{i=1}^M$ is constructed with $y_i \in R^n$. The dictionary learning model can be presented as the following optimization problem:

$$\min_{D, \{\alpha_i\}_{i=1}^M} \sum_{i=1}^M \|\alpha_i\|_0 \quad subject\ to\ \|y_i - D\alpha_i\|_2 \le \varepsilon,\ i = 1......M, \quad (2)$$

where $D \in R^{n \times m}$ is the unknown dictionary and $\{\alpha_i\}_{i=1}^M$ is the sparse coefficients corresponding to $\{y_i\}_{i=1}^M$. In this paper, we use the K-SVD algorithm [14] to solve (2). It should be noted that the mean value of each patch is subtracted to zero before training, so that the mean value of each atom in the learned dictionary is also zero, which indicates that the dictionary does not contain any spectral information. In addition, each atom of the dictionary is normalized to unit length in Euclid space.

Fig. 1 Ten example natural images of the training set

B. Detailed Fusion Scheme

In the ARSIS-based pan-sharpening methods, the two most important issues are extracting the high-frequency spatial information in the HRP image and injecting it into the LRM image without missing spectral information. Here we first apply wavelet transform to separate the high-frequency component from the HRP image. Furthermore, to extract the spatial details as much as possible, an SR-based fusion framework [15] is applied to merge the low-frequency components with carefully preserving the spectral information. The detailed fusion scheme is shown as the following steps:

Step 1. Resample and IHS transform

Resample the LRM image to the same resolution with the HRP image and apply IHS transform to the resampled version to get its intensity component denoted as *INT*.

Step 2. Wavelet transform (WT)

Apply WT to HRP and *INT* to obtain their low-frequency components *LP* and *LI*, and high-frequency components *HP* and *HI*, respectively.

Step 3. Fusion of high-frequency components

The high-frequency components are fused based on the wavelet energy which is defined as follows:

$$E(w) = \sum_{m=1}^{M}\sum_{n=1}^{N} H(m,n)^2 / (M \times N), \qquad (3)$$

where w is a local window with size $M \times N$ and $H(m,n)$ is the high-frequency wavelet coefficient at pixel (m,n). Let $w_P(i,j)$ and $w_I(i,j)$ denote the local windows centered at pixel (i,j) in *HP* and *HI*, respectively. The high-frequency fused result *HF* is obtained by applying the following rule:

$$HF(i,j) = \begin{cases} HP(i,j), & E(w_P(i,j)) > E(w_I(i,j)) \\ HI(i,j), & otherwise \end{cases}. \qquad (4)$$

Step 4. Fusion of low-frequency components

Construct an empty image *LF* with the same size of *LP*. Apply the "sliding window" technique to divide *LP* and *LI* into $\sqrt{n} \times \sqrt{n}$ patches from left-top to right-bottom with a step length of one pixel. Suppose the size of *LF* is $H \times W$, then there are $(H - \sqrt{n} + 1) \times (W - \sqrt{n} + 1)$ pairs of corresponding patches from *LP* and *LI*. For each pair $\{p_{LP}, p_{LI}\}$, transform them into column vectors $\{v_{LP}, v_{LI}\}$ via lexicographic ordering and normalize the mean value of v_{LP} and v_{LI} to zero through:

$$v_{NLP} = v_{LP} - mean(v_{LP}) \cdot 1, \qquad (5)$$

$$v_{NLI} = v_{LI} - mean(v_{LI}) \cdot 1, \qquad (6)$$

where 1 is an $n \times 1$ vector with all entries equaling to one. Decompose v_{NLP} and v_{NLI} into their sparse representations with the sparse model (1) using OMP. Let α_{LP} and α_{LI} denote the sparse coefficient vectors of v_{NLP} and v_{NLI}, respectively. Then, Merge α_{LP} and α_{LI} with the following rule:

$$\alpha_{LF} = \begin{cases} \alpha_{LP}, & \|a_{LP}\|_1 \times E(p_{HP}) > \|a_{LI}\|_1 \times E(p_{HI}) \\ \alpha_{LI}, & otherwise \end{cases}, \qquad (7)$$

where $\|\bullet\|_1$ is the $l_1 - norm$ and $\{p_{HP}, p_{HI}\}$ is the corresponding pair of high-frequency patches with its position in $\{HP, HI\}$ same as the position of $\{p_{LP}, p_{LI}\}$ in $\{LP, LI\}$. This fusion rule simultaneously takes the sparse coefficients of low-frequency component as well as the high-frequency component into consideration, which makes the algorithm more robust. Then, the fused patch is:

$$LF(p) = D\alpha_{LF} + mean(v_{LI}). \qquad (8)$$

In (8), the first term represents the spatial details injected from the HRP image to the LRM image, while the second term is always set as the mean value of the patch from the intensity component of LRM to preserve the spectral information in LRM. It should be noted that this strategy is different from the traditional SR-based image fusion method [15]. Otherwise, the

spectral distortion cannot be avoided. Finally, for each pixel in *LF*, the value is divided by the accumulation times at its position to obtain the low-frequency fused result.

Step 5. Inverse WT and inverse IHS transform

Apply inverse wavelet transform to *HF* and *LF* to get the fused intensity component, and the final HRM image can be obtained by performing inverse IHS transform.

IV. EXPERIMENTAL RESULTS

A. Experiment Setups

In this paper, we use the WorldView-2 data to evaluate the effectiveness of the proposed method. WorldView-2 is a new high-resolution satellite which provides 8-band multispectral images at 1.84-m resolution and panchromatic images at 0.46-m resolution. The data used in this paper is downloaded from the website http://www.datatang.com/data/43234, which is originated from Digital Globe and arranged by Beijing key laboratory of digital media, Beihang university. The dataset contains 23 pairs of LRM and HRP images covering the area of urban, seaside, bridge, etc. In real pan-sharpening task, the reference image that used for assessment does not exist, and a popular method to get a reference image is degrading the original LRM and HRP images before experiments. However, as the spatial resolution of the LRM images in this dataset is only 128×128 pixels, it is not advisable to use the above protocol. Fortunately, the dataset provides some high-quality fused examples of HRM images almost without spectral distortion. Thus, it is reasonable to use them as the reference images. Without loss of generality, we consider a pan-sharpening case only with three spectral bands: Red (R), Green (G) and Blue (B), as the LRM images in this dataset only contain these three bands. The proposed method can be easily extended to processing four or even more spectral bands pan-sharpening problem. Fig. 2 shows two sets of images which are captured over urban area and seaside area.
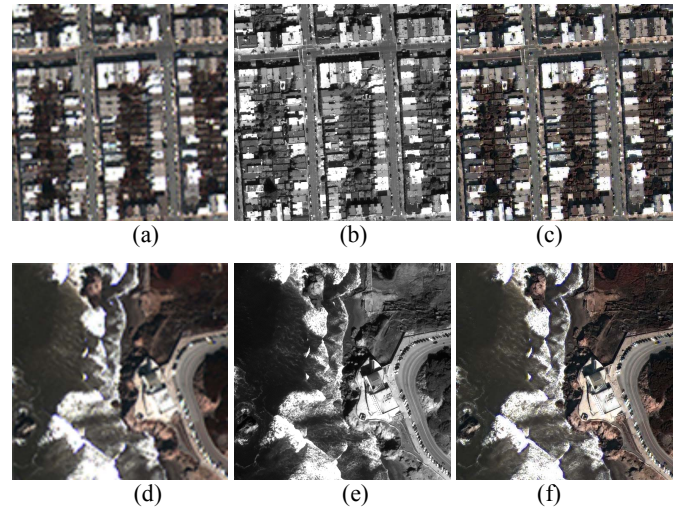

Fig. 2 The source and reference images. (a) Resampled LRM image (urban). (b) Original HRP image (urban). (c) The reference image (urban). (d) Resampled LRM image (seaside). (b) Original HRP image (seaside). (c) The reference image (seaside).

The proposed method (HS-WV-SR) is compared with five popular pan-sharpening methods: the PCA method (PCA), the IHS method (IHS) [2], the adaptive IHS method (AIHS) [16], the wavelet-based method (IHS-WV) [5], and the SR-based method (IHS-SR). The IHS-WV method performs a two-level wavelet transform on the intensity component of LRM image. The IHS-SR method, which is similar to the proposed method except for it applies SR directly in the spatial domain, is used to verify the effectiveness of wavelet transform in the IHS-WV-SR method. We choose "db4" wavelet in both the IHS-WV method and the IHS-WV-SR method, and the transform level in IHS-WV-SR is set to one. Empirically, the dictionary has 256 atoms and the patch size is set as 8×8 pixels. The parameter ε in the two SR-based methods is set to 0.01 when the source images are not corrupted by noise.

In addition, several objective assessment indexes are applied to evaluate the fusion results. With the reference image, the universal image quality index (UIQI) [17], the spectral angle mapper (SAM) [18], and the ERGAS [19] are used as the assessment indexes. Moreover, the "quality with no reference" (QNR) [20] that consists of the spectral distortion index D_λ and the spatial distortion index D_s is also used. The optimal value of UIQI and QNR are one. The optimal values of SAM, ERGAS, D_λ, and D_s are zero.

B. Pan-sharpening Results

Fig. 3 and Fig. 4 show the pan-sharpening results of the two sets of source images in Fig. 2. By visually comparing the six results, we can see that the results of PCA and IHS methods suffer severe spectral distortion and the situation becomes a little better in AIHS result. The IHS-WV, IHS-SR, and IHS-WV-SR methods preserve spectral information well. In terms of the spatial details, the pan-sharpened image with IHS-WV method is the mostly blurred while the other five results are fine. By comparing Fig. 3(e) (Fig. 4(e)) with Fig. 3(f) (Fig. 4(f)) carefully, we can find that the spatial information of the IHS-WV-SR is more accurate than that of the IHS-SR method, especially in the strong edge areas, and the main reason may be that the reconstruction results of SR in image domain is not very stable in high-frequency areas, such as the edges. So when the SR is applied in low-frequency domain, the results will be better.

The objective quality indexes of the results in Fig. 3 and Fig. 4 are reported in Table I and Table II respectively, in which the best results for each index are labeled in bold. We can see that the result is mainly in agreement with the visual perception. The proposed method provides the best UIUQ, SAM, D_s, and QNR, while the IHS-SR method gives the best ERGAS and D_λ. It should be noted that the IHS-SR method preserves the spectral information a little better than the proposed approach as it directly obtains the most original spectral information from the LRM image without any truncation error. However, the aforementioned point that the spatial details of the proposed method is more better is also confirmed by the indexes D_s and UIUQ. Furthermore, the

IHS-WV-SR method has a significant superiority than the IHS-SR method in terms of the algorithm efficiency.

(a)   (b)
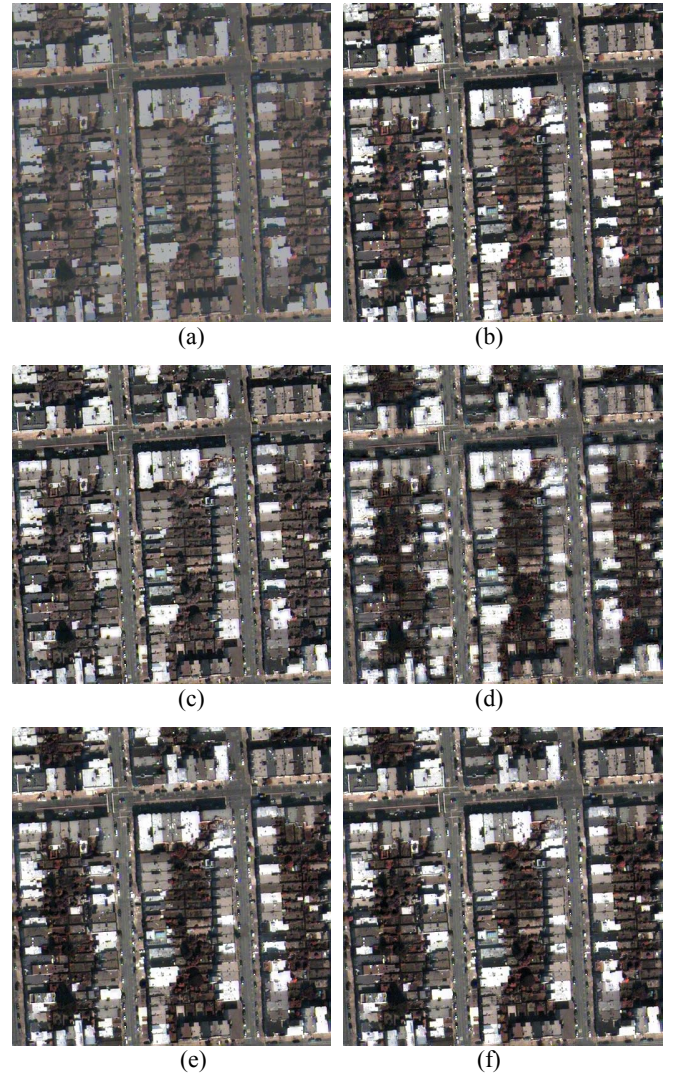(c)   (d)
(e)   (f)

Fig. 3 Fusion results of the *urban* images by different methods. (a) PCA method. (b) IHS method. (c) AIHS method. (d) IHS-WV method. (e) IHS-SR method. (f) The proposed IHS-WV-SR method.

TABLE I

QUANTITATIVE INDEXES OF THE FUSION RESULTS IN FIG. 3

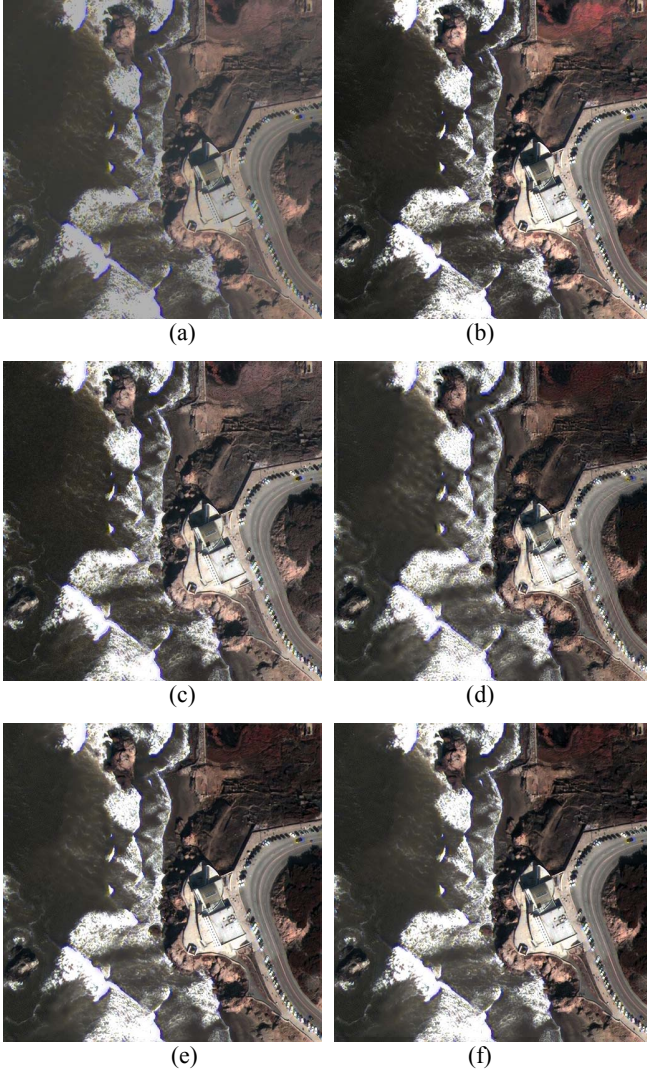| Indexes	PCA	IHS	AIHS	IHS-WV	IHS-SR	Proposed
SAM	2.1613	1.9119	1.6432	1.6261	1.6045	**1.5931**
ERGAS	8.5942	3.8482	3.5088	3.5763	**2.0687**	2.5676
UIQI-R	0.7790	0.9103	0.9255	0.8744	0.9335	**0.9417**
UIQI-G	0.7682	0.9186	0.9172	0.8617	0.9385	**0.9432**
UIQI-B	0.7515	0.9124	0.9072	0.8488	0.9314	**0.9366**
D_λ	0.1003	0.0302	0.0269	0.0109	**0.0096**	0.0203
D_s	0.1155	0.0505	0.0315	0.1279	0.0459	**0.0159**
QNR	0.7958	0.9208	0.9424	0.8626	0.9449	**0.9641**

Fig. 4 Fusion results of the *seaside* images by different methods. (a) PCA method. (b) IHS method. (c) AIHS method. (d) IHS-WV method. (e) IHS-SR method. (f) The proposed IHS-WV-SR method.

TABLE II

QUANTITATIVE INDEXES OF THE FUSION RESULTS IN FIG. 4

Indexes	PCA	IHS	AIHS	IHS-WV	IHS-SR	Proposed
SAM	1.7385	1.5844	1.2147	1.1958	1.1819	**1.1751**
ERGAS	9.5522	5.7130	5.0383	2.8087	**1.5821**	1.9883
UIQI-R	0.7734	0.7837	0.8600	0.8571	0.8845	**0.9051**
UIQI-G	0.7639	0.7735	0.8683	0.8471	0.8808	**0.9088**
UIQI-B	0.7468	0.7625	0.8669	0.8387	0.8694	**0.9062**
D_λ	0.0593	0.0560	0.0516	0.0307	**0.0248**	0.0379
D_s	0.0784	0.0550	0.0148	0.0927	0.0434	**0.0146**
QNR	0.8669	0.8921	0.9344	0.8794	0.9329	**0.9481**

V. CONCLUSION

This paper proposes a new pan-sharpening method with sparse representation under the framework of wavelet transform. In our method, the low-frequency components of wavelet are fused based on sparse representation with the dictionary learned from natural images, which makes the method more practical. Moreover, a novel strategy is presented to preserve the spectral information to the most extent. On the other hand, the "numerous-but-sparse" high-frequency components are fused with the traditional wavelet-energy-based rule, which makes the method more efficient than the general SR-based fusion methods. The proposed method is compared with several conventional pan-sharpening methods, and the experimental results demonstrate that the proposed method can obtain satisfactory high spatial multispectral images with preserving the spectral information simultaneously.

REFERENCES

[1] C. Thomas, T. Ranchin, L. Wald, and J. Chanussot, "Synthesis of multispectral images to high spatial resolution: A critical review of fusion methods based on remote sensing physics," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1301–1312, May 2008.

[2] T.-M. Tu, P. S. Huang, C.-L. Hung, and C.-P. Chang, "A fast intensity-hue-saturation fusion technique with spectral adjustment for IKONOS imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 1, no. 4, pp. 309–312, Oct. 2004.

[3] V. P. Shah and N. H. Younan, "An efficient pan-sharpening method via a combined adaptive PCA approach and contourlets," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1323–1335, May 2008.

[4] C. A. Laben and B. V. Brower, "Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening," *U.S. Patent 6011875*, Jan. 4, 2000.

[5] P. S. Pradhan, R. L. King, N. H. Younan, and D. W. Holcomb, "Estimation of the number of decomposition levels for a wavelet-based multi-resolution multisensor image fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 12, pp. 3674–3686, Dec. 2006.

[6] X. Otazu, M. González-Audícana, O. Fors, and J. Núñez, "Introduction of sensor spectral response into image fusion methods. Application to wavelet-based methods," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 10, pp. 2376–2385, Oct. 2005.

[7] M. Joshi and A. Jalobeanu, "MAP estimation for multiresolution fusion in remotely sensed images using an IGMRF prior model," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 3, pp. 1245–1255, Mar. 2010.

[8] S. Li and B. Yang, "A new pan-sharpening method using a compressed sensing technique," I*EEE Trans. Geosci. Remote Sens.*, vol. 49, no. 2, pp. 738–746, Feb. 2011.

[9] J. Cheng, H. Zhang, H. Shen, and L. Zhang, "A practical compressed sensing-based pan-sharpening method," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 4, pp. 629–633, Jul. 2012.

[10] X. Zhu and R. Bamler, "A sparse image fusion algorithm with application to pan-sharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2827–2836, May. 2013.

[11] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.

[12] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, Nov. 2010.

[13] A. Tropp and J. Wright, "Computational methods for sparse solution of linear inverse problems," *Proc. IEEE*, vol. 98, no. 6, pp. 948–958, Jun. 2010.

[14] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[15] B. Yang and S. Li, "Multifocus image fusion and restoration with sparse representation," *IEEE Trans. Instrum. Meas.*, vol. 59, no. 4, pp.884–892, Apr. 2010.

[16] S. Rahmani, M. Strait, D. Merkurjev, M. Moeller, and T. Wittman, "An adaptive IHS pan-sharpening method," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 4, pp. 746–750, Oct. 2010.

[17] Z. Wang and A. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Mar. 2002.

[18] L. Alparone, L. Wald, J. Chanussot, C. Thomas, P. Gamba, and L. M. Bruce, "Comparison of pansharpening algorithms: Outcome of the 2006 GRS-S data-fusion contest," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3012–3021, Oct. 2007.

[19] M. M. Khan, L. Alparone, and J. Chanussot, "Pansharpening quality assessment using the modulation transfer functions of instruments," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 11, pp. 3880–3891, Nov. 2009.

[20] L. Alparone, B. Aiazzi, S. Baronti, A. Garzelli, F. Nencini, and M. Selva, "Multispectral and panchromatic data fusion assessment with-out reference," *Photogramm. Eng. Remote Sens.*, vol. 74, no. 2, pp. 193–200, Feb. 2008.