

# Two-Step Sparse Coding for the Pan-Sharpening of Remote Sensing Images

Cheng Jiang, Hongyan Zhang, *Member, IEEE*, Huanfeng Shen, *Member, IEEE*, and Liangpei Zhang, *Senior Member, IEEE*

**Abstract**—Remote sensing image pan-sharpening is an important way of enhancing the spatial resolution of a multispectral (MS) image by fusing it with a registered panchromatic (PAN) image. The traditional pan-sharpening methods often suffer from color distortion and are still far from being able to synthesize a real high-resolution MS image, as could be directly acquired by a better sensor. Inspired by the rapid development of sparse representation theory, we propose a two-step sparse coding method with patch normalization (PN-TSSC) for image pan-sharpening. Traditional one-step sparse coding has difficulty in choosing dictionary atoms when the structural information is weak or lost. By exploiting the local similarity between the MS and PAN images, the proposed sparse coding method deals with the dictionary atoms in two steps, which has been found to be an effective way of overcoming this problem. The experimental results with IKONOS, QuickBird, and WorldView-2 data suggest that the proposed method can effectively improve the spatial resolution of a MS image, with little color distortion. The pan-sharpened high-resolution MS image outperforms those images fused by other traditional and state-of-the-art methods, both quantitatively and perceptually.

**Index Terms**—Image fusion, pan-sharpening, remote sensing image, two-step sparse coding.

## I. INTRODUCTION

HIGH spatial resolution multispectral (MS) images are favored by a lot of remote sensing applications, and a number of high-resolution remote sensing systems have been launched and widely used, such as IKONOS, QuickBird, GeoEye-1, WorldView-1, and WorldView-2. Due to the physical limitations, it is difficult to acquire a high-resolution MS image directly. One commonly used strategy is to use two sensors to separately acquire a high-resolution panchromatic (HRP) image and a low-resolution multispectral (LRM) image,

and then utilize a pan-sharpening technique to produce a high-resolution multispectral (HRM) image by fusing the two images.

For the pan-sharpening task, the key problem is to find and utilize the relationship between the LRM image and the HRP image. A lot of pan-sharpening methods have been developed to date, and they can be categorized into four types: transformation and substitution based, arithmetic based, ARSIS [1] concept based (from its French name Amélioration de la Résolution Spatiale par Injection de Structures), and restoration based. The transformation and substitution based methods first transform the resampled LRM image to a space where one of its components is strongly correlated with the low-resolution panchromatic (LRP) image degraded from the HRP image; the HRP image is then stretched to replace the component; finally, an inverse transform is performed. The transformation and substitution based methods have been widely used, e.g., intensity-hue-saturation (IHS or LHS) [2], principal component analysis (PCA) [3], and the Gram-Schmidt (GS) process [4]. As to the arithmetic based methods, the operations of multiplication, division, addition, and subtraction are combined in different ways to achieve a better fusion performance [5]. Among them, the Brovey transform [5] is the most well known. For the ARSIS concept based methods, these approaches assume that the missing spatial information in the LRM image can be obtained from the high frequencies of the HRP image, so the relationships between the high frequencies in the HRP image and the LRM image are searched and exploited [1]. The typical ARSIS concept based methods are high-pass filtering (HPF) [3] and the wavelet based methods [6]. The above three types of methods can be viewed as traditional methods [7], which tend to cause color distortions due to the wavelength extension of the new satellite PAN images [7]. Inspired by the rapid development of the single-image super-resolution technique, restoration based methods have recently become popular. For the restoration based methods, the LRM image and the HRP image are both viewed as observations of the HRM image via the image degradation model, and a prior of the target HRM image should be assumed [8]–[10]. In addition to the four main types of methods, hybrid methods have also been developed [11]–[13] and have shown improved performance.

Recently, the compressed sensing (CS) technique has been applied to pan-sharpening [14]. Since the fused image patch is adaptively estimated with different sets of dictionary atoms, the CS based method achieves impressive results. However, the applicability is limited as the dictionary construction method needs HRM training images, which may not be available. To

Manuscript received April 23, 2013; revised July 16, 2013 and August 31, 2013; accepted September 20, 2013. Date of publication October 11, 2013; date of current version June 13, 2014. This work was supported in part by the National Basic Research Program of China (973 Program) under Grant 2011CB707105, by the 863 program under Grant 2013AA12A301, by the National Natural Science Foundation of China under Grants 61201342 and 40930532, and by Program for Changjiang Scholars and Innovative Research Team in University (IRT1278), the Fundamental Research Funds for the Central Universities. (*Corresponding author: H. Zhang.*)

C. Jiang, H. Zhang, and L. Zhang are with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, P.R. China (e-mail: zhanghongyan@whu.edu.cn).

H. Shen is with the School of Resource and Environmental Science, Wuhan University, P.R. China.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/JSTARS.2013.2283236

deal with this problem, Jiang *et al.* [15] constructed a joint dictionary from upsampled LRM and HRP training images, which are easily acquired from the available remote sensing systems. The shortcoming of the method in [15] is that it still needs a large collection of training images, and the trained dictionary is also relatively large, which takes much computing time for sparse solutions. Following the well-known coupled sparse coding technique for natural image super-resolution [16], the latest work [17] brings the coupled sparse coding technique to image pan-sharpening. The method in [17] shows good potential by only using the HRP image and its degraded version, which has the same resolution as the LRM image. The coupled image pairs are directly used, without training, to construct two coupled dictionaries, which consist of a high-resolution dictionary and a low-resolution dictionary. Since we assume that the sparse coefficients of the target HRM patch over the high-resolution (HR) dictionary are the same as the sparse coefficients of the LRM patch over the low-resolution dictionary, the sparse coefficients of every LRM patch should be estimated as accurately as possible. However, at the sparse coding stage, the method in [17] relies strongly on the structural information of the LRM patch, which may be lost or very weak, due to its low resolution, so the coding of the LRM patch may be contaminated by unsuitable low-resolution (LR) dictionary atoms, for which the corresponding HR dictionary atoms cannot effectively represent the ideal HRM patch. From the experimental results in Section IV, it is found that the method in [17] preserves the spectral properties well, but it shows poor spatial results in the regions with small objects or fine details.

In view of this, we propose a novel sparse coding based remote sensing image pan-sharpening method. The main contribution of this paper is twofold. First, a two-step sparse coding is used to obtain the sparse coefficients, instead of the traditional one-step sparse coding. For the two-step sparse coding, we fully utilize the strong correlation between the LRM patch and the corresponding LRP patch, and thus overcome the problem of structural information loss in some LRM patches. This greatly improves the spatial details but may introduce a little color distortion in some cases. To handle this issue, the patch normalization strategy, which aims to improve the stability of the sparse coding, is exploited and used to prevent the color distortion that may happen when using the two-step sparse coding on the image patch directly, and further improves the fusion result. Moreover, since the proposed method deals with each patch independently, it can be fast with a multi-thread implementation. For simplicity, the proposed method also processes each MS band separately, as in [17], and thus less memory and computational burden are required.

## II. PAN-SHARPENING WITH SPARSE REPRESENTATION

### A. Sparse Representation

Inspired by the research into the receptive fields, sparse representation has been proposed as a powerful statistical image modeling technique [18]. Since the emergence of CS theory [19], [20], the theory and application of sparse representation, which is strongly correlated with CS, has been widely studied.

In fact, natural images tend to be sparse in the redundant image domain [21]. The redundant image domain is often called as the dictionary, where the image can be represented as a linear combination of only a small number of the dictionary atoms. For practical applications, the images are processed patch by patch, and it has been shown that a trained dictionary is better than the fixed redundant bases such as discrete cosine transform (DCT) [22]. The sparse representation problem is often represented as:

$$\min_{\alpha} \|\mathbf{x} - \mathbf{D}\alpha\|_2^2 + \lambda \|\alpha\|_0 \quad (1)$$

where  $\|\cdot\|_2$  indicates the  $\ell_2$ -norm, and  $\|\cdot\|_0$  denotes the  $\ell_0$ -norm.  $\mathbf{x}$  is a column vector representing a signal or a lexicographically ordered image patch.  $\mathbf{D}$  is a matrix representing the dictionary, with each column called as an atom.  $\alpha$  is the vector of the sparse coefficients, with most of its coefficients being close to or zero.  $\lambda$  is the regularization parameter.

Since (1) is well known as an NP-hard problem, greedy algorithms such as orthogonal matching pursuit (OMP) are often used to tackle this problem. An alternative approach is to relax (1) to an  $\ell_1$ -norm convex optimization problem [19]:

$$\min_{\alpha} \|\mathbf{x} - \mathbf{D}\alpha\|_2^2 + \lambda \|\alpha\|_1 \quad (2)$$

where  $\|\cdot\|_1$  indicates the  $\ell_1$ -norm.

Many  $\ell_1$ -norm optimization methods have been proposed to solve this convex problem [23], and the least angle regression stagewise (LARS) algorithm [24] has been widely used due to its good performance and stability. In this paper, the LARS algorithm package developed by Julien Mairal *et al.* [25], optimized with multi-thread implementation, is adopted for our second sparse coding step.

### B. Sparse Representation With Coupled Dictionaries

Sparse representation was first applied to single natural image super-resolution by Yang *et al.* [16], and the main idea of the method is to assume an occurrence relation in that the upsampled LR and HR image patch pairs share the same sparse coefficients with respect to their own dictionaries. Supposing that coupled dictionaries are available (Section II-C describes how to construct the coupled dictionaries for pan-sharpening), then for a certain upsampled LR image patch, the sparse coefficients of the LR image patch over the LR dictionary are first solved by using (2). Then, with this assumption, the sparse coefficients of the LR image patch are taken as the sparse coefficients of the HR image patch over the HR dictionary. Finally, the target HR image patch is obtained by multiplying the HR dictionary with the sparse coefficients. Recently, Zhu and Bamler [17] brought this idea to the application of pan-sharpening, with each band of the LRM image being processed one by one. Due to its effectiveness, this idea is also adopted in our proposed method. Note that in the work of Yang *et al.* [16], the first- and second-order gradients of the upsampled LR image patch are concatenated into one vector to represent the LR image patch, and so are the LR training samples when training the dictionaries. However, in our work, the images are used directly, which is more suitable.

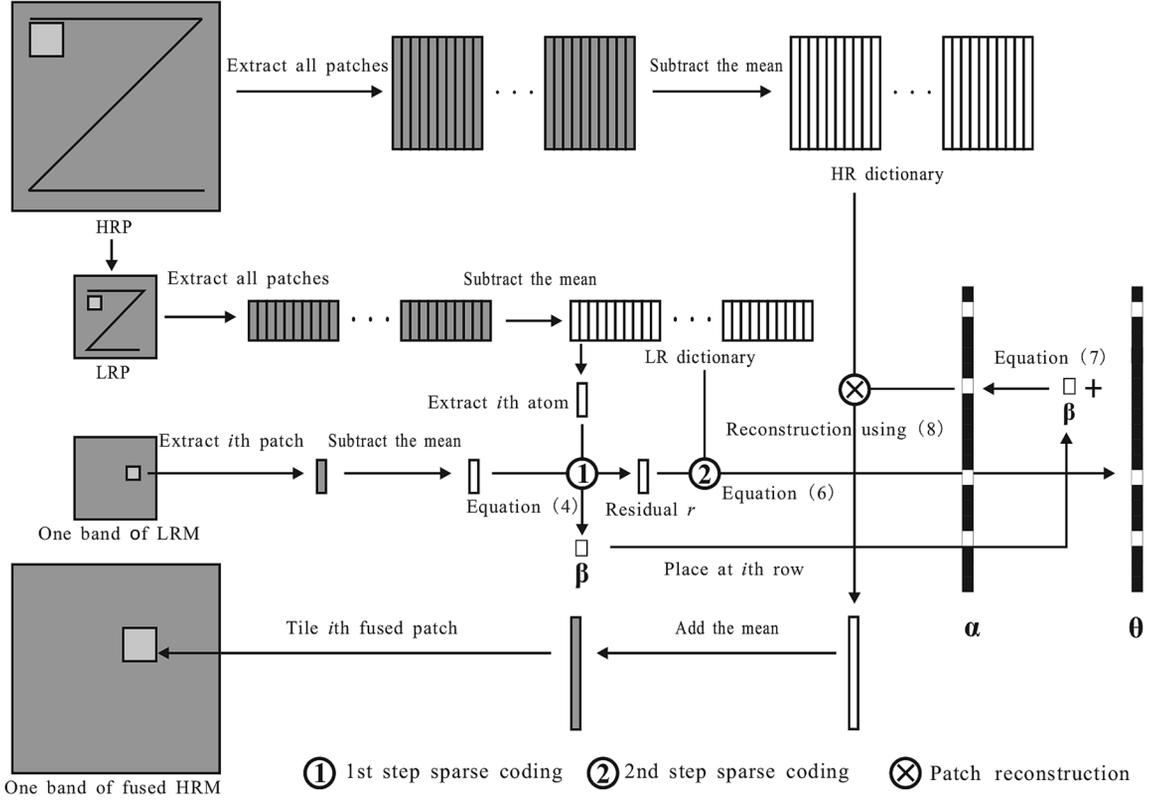


Fig. 1. The flowchart of the proposed method.

### C. Construction of the Coupled Dictionaries for Pan-Sharpener

For pan-sharpening, a corresponding LRM image  $\mathbf{X}_l$  and HRP image  $\mathbf{Y}_h$  are available, and the goal is to produce the HRM image  $\mathbf{X}_h$  by fusing  $\mathbf{X}_l$  and  $\mathbf{Y}_h$ . Unlike the traditional sparse representation based methods, which need external image collections to train the dictionaries, the coupled dictionaries are constructed directly from the panchromatic (PAN) image and its degraded version [17]. The HRP image  $\mathbf{Y}_h$  is first blurred and downsampled by a factor  $F_{DS}$  to get the LRP image  $\mathbf{Y}_l$ , which has the same resolution and the same image size as the LRM image  $\mathbf{X}_l$ . We extract all the patches from the LRP image  $\mathbf{Y}_l$  with a patch size  $[p_1 p_2]$  and a step size  $[s_1 s_2]$ , then rearrange every patch to a vector. Finally, we build a matrix  $\mathbf{D}_l$  with these vectors, which is viewed as the LR dictionary. Following the same method, the HR dictionary  $\mathbf{D}_h$  is obtained from the HRP image  $\mathbf{Y}_h$  with a patch size  $F_{DS} \times [p_1 p_2]$  and a step size  $F_{DS} \times [s_1 s_2]$ . In this way, the numbers of the dictionary atoms of the LR dictionary  $\mathbf{D}_l$  and the HR dictionary  $\mathbf{D}_h$  are the same.

### III. PAN-SHARPENING WITH TWO-STEP SPARSE CODING

The overall flowchart of the proposed pan-sharpening method is shown in Fig. 1. For the proposed pan-sharpening method, the coupled dictionaries are constructed as described in Section II-C, and the next task is to find the sparse representations of the LR patches over the LR dictionary, and reconstruct the fused patches with the HR dictionary, as mentioned in Section II-B. In the latest work [17], at the sparse coding

stage, every patch of the LRM image is coded with one step by directly using (2) to obtain the sparse coefficients over the same LR dictionary. In our method, a two-step sparse coding approach is utilized that fully considers the structural similarity between the LRP patch and the LRM patch. In addition, the patch normalization technique is used to improve the stability of the sparse coding.

#### A. Patch Structural Similarity Property

For the  $j$ th band of the LRM image  $\mathbf{X}_{l_j}$ , all the patches are extracted with a patch size  $[p_1 p_2]$  and a step size  $[s_1 s_2]$ . For the  $i$ th patch  $\mathbf{x}_{l_j}^i$  (we use upper-case to represent an image and lower-case to denote a patch from it), it is strongly correlated with the corresponding patch  $\mathbf{y}_l^i$  from the LRP band, as the patches share the same scene and are often acquired by different sensors with an overlapping spectral domain. As shown in Fig. 2, despite the differences among the spectral responses, each patch from the LRM bands tends to share a similar structural pattern to the patch at the same location on the LRP band. We name this important discovery the “patch structural similarity property.” In addition, we call the  $i$ th LR dictionary atom (also denoted by  $\mathbf{y}_l^i$ ), which comes from the column stacked patch  $\mathbf{y}_l^i$ , the “adjoint-atom” of the LRM patch  $\mathbf{x}_{l_j}^i$ .

#### B. Two-Step Sparse Coding

For the traditional sparse coding, when the LR image patch to be coded is weak in its structural information, it will be easily contaminated by unsuitable dictionary atoms, which may have a negative effect on the performance of the image pan-sharpening. Therefore, an intuitive way is to use a two-step sparse

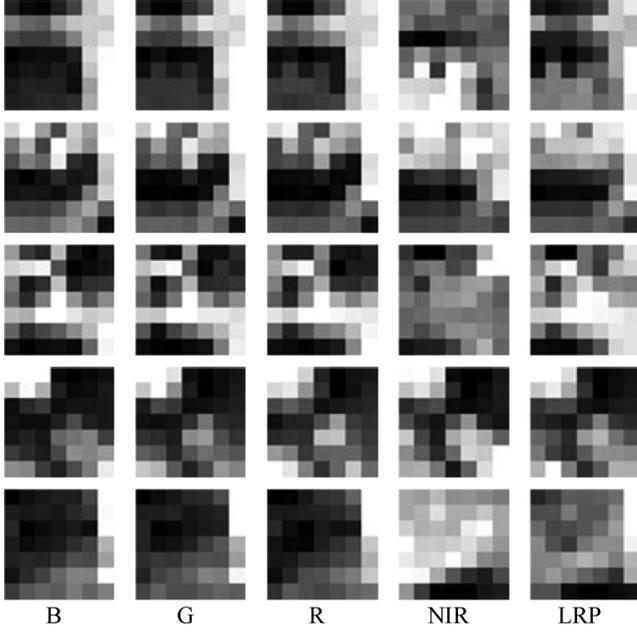


Fig. 2. Similar structural patterns among patches at the same location from the LRM and LRP images. Each row of the  $7 \times 7$  patches is randomly selected at the same spatial location from the IKONOS image. B, G, R, and NIR represent the blue, green, red, and near-infrared bands, respectively. Each image is magnified 10 times.

coding to distinguish the more reliable atoms from the less reliable ones. That is, at the first sparse coding step, the reliable atoms are used to sparsely represent the vector, and at the second sparse coding step, the rest of the atoms are used to sparsely represent the residual produced by the first sparse coding step. How to find the more reliable atoms remains a problem, which may differ between different applications. However, motivated by the patch structural similarity property of the LRP and LRM images, the adjoint-atom of the patch to be coded is sure to be a reliable atom. The proposed two-step sparse coding method is therefore simplified by using only one reliable atom at the first sparse coding step. In this way, the adjoint-atom  $\mathbf{y}_l^i$  is treated differently when we find the sparse coefficients of  $\mathbf{x}_{l,j}^i$  over the LR dictionary  $\mathbf{D}_l$ . We first use this reliable atom  $\mathbf{y}_l^i$  to construct a special one-atom dictionary (also denoted by  $\mathbf{y}_l^i$ ), and then use (3) to sparsely represent the LRM patch  $\mathbf{x}_{l,j}^i$  at the first sparse coding step:

$$\min_{\beta_j^i} \|\mathbf{x}_{l,j}^i - \mathbf{y}_l^i \beta_j^i\|_2^2 + \lambda_1 \|\beta_j^i\|_0 \quad (3)$$

where  $\beta_j^i$  is the coefficient and  $\lambda_1$  is the regularization parameter. Since the special dictionary  $\mathbf{y}_l^i$  consists of only one atom,  $\|\beta_j^i\|_0$  is a constant, and the sparse term can be omitted. We then get:

$$\min_{\beta_j^i} \|\mathbf{x}_{l,j}^i - \mathbf{y}_l^i \beta_j^i\|_2^2 \quad (4)$$

It is easy to solve (4) by letting its derivative be zero. Clearly,  $\mathbf{x}_{l,j}^i$  cannot be well represented by only one atom, so the second step is to represent the first step's residual  $\mathbf{r}_j^i$  ( $\mathbf{r}_j^i = \mathbf{x}_{l,j}^i - \mathbf{y}_l^i \beta_j^i$ )

over the general LR dictionary  $\mathbf{D}_l$ , which takes all the patches from the corresponding LRP image as atoms:

$$\min_{\theta_j^i} \|\mathbf{r}_j^i - \mathbf{D}_l \theta_j^i\|_2^2 + \lambda_2 \|\theta_j^i\|_0 \quad (5)$$

where  $\theta_j^i$  is the vector of sparse coefficients by the second sparse coding step, and  $\lambda_2$  is the regularization parameter. Equation (5) is a traditional sparse coding problem. Since (5) cannot be solved exactly with many dictionary atoms, we relax it to (6) and use the LARS algorithm optimized with multi-thread implementation (which can be downloaded from Julien Mairal's personal page [26]) to speed up the proposed method.

$$\min_{\theta_j^i} \|\mathbf{r}_j^i - \mathbf{D}_l \theta_j^i\|_2^2 + \lambda_2 \|\theta_j^i\|_1 \quad (6)$$

With the two-step sparse coding, the final sparse vector is acquired by:

$$\boldsymbol{\alpha}_j^i = \theta_j^i + \boldsymbol{\delta} \times \beta_j^i \quad (7)$$

where  $\boldsymbol{\delta}$  is a vector with the same dimension as  $\theta_j^i$ , with only the  $i$ th element being one and the others being zero. Finally, as mentioned in Section II, the fused patch  $\mathbf{x}_{h,j}^i$  is reconstructed by:

$$\mathbf{x}_{h,j}^i = \mathbf{D}_h \times \boldsymbol{\alpha}_j^i \quad (8)$$

By using (4), (6), (7), and (8), all the patches from each band of the LRM image are processed independently to acquire the corresponding fused HRM patches, and the overlapping areas of the fused patches are averaged.

Note that in the first sparse coding step, the one-atom dictionary varies with the current image patch being processed. However, at the second sparse coding step, the dictionary is the same for all the LRM patches. One may wonder why the adjoint-atom  $\mathbf{y}_l^i$  is still in the common dictionary at the second sparse coding step, since it has already been coded in the first sparse coding step. The reason for this is that in the first sparse coding stage, the other dictionary atoms are not considered, so there may be too much priority given to the atom  $\mathbf{y}_l^i$ , especially for the cases where the structural similarity between the LRM band patch and the LRP band patch is not very strong, e.g., the NIR patch with the LRP patch in the first row in Fig. 2, and the blue, green, and red band patch with the LRP patch in the fifth row in Fig. 2. Therefore, at the second sparse coding step, the adjoint-atom is still considered, to adaptively remedy this problem.

With the first sparse coding step, the proposed method puts more emphasis on the adjoint-atom than a single-step sparse coding based method. Thus, the two-step sparse coding adaptively utilizes the patch structural similarity property in a simple way, rather than only relying on the structural information of the LRM image patch to infer the correct sparse representation. The experiments in Section IV show that this consideration improves the performance greatly over the traditional one-step sparse coding method.

### C. Patch Normalization

Patch normalization is widely used in regression problems [27] to subtract a constant before the main procedures. It has

also been used in sparse representation based natural image processing to improve the numerical stability of sparse coding, and has resulted in a better visual quality of the results [28]. We adopt this strategy in the proposed method, and find that it is applicable to our two-step sparse coding based remote sensing image pan-sharpening process.

Patch normalization consists of two steps in the proposed method. First, for each atom of the common LR dictionary and HR dictionary, the mean intensity value of the vector is subtracted from all its elements. Second, for the LRM patch  $\mathbf{x}_{l,j}^i$ , the mean intensity value is also subtracted from all its elements before the first sparse coding step in (4), and is added back to all the elements of the fused patch  $\mathbf{x}_{h,j}^i$  after the patch reconstruction in (8). With this simple processing strategy, the spectral quality of the proposed two-step sparse coding is improved greatly, and the resulting fused image looks even smoother.

#### IV. EXPERIMENTS AND DISCUSSIONS

In Wald's [29] view, a synthetic image should be as similar as possible to the image that the corresponding sensor would observe at the highest spatial resolution. This idea is adopted here to assess the different pan-sharpening methods. In this section, we use IKONOS, QuickBird, and WorldView-2 data to test the performance of the proposed method. To quantitatively assess the quality of the results, the original data are low-pass filtered and degraded (for convenience, we use MATLAB's "imresize" function, with the "antialiasing" parameter being "true", for degrading images in this paper). The experiments are performed on the degraded data, and the original MS image is used as the reference. All the data used in the following experiments are Geo format images, and the MS image and PAN image are well corresponded with sub-pixel accuracy. All the images are 11-bit without dynamic range adjustment, and the final results are displayed by ENVI 4.7.

We compare the proposed method with three popular traditional methods: generalized IHS (GIHS) [2], GS [4] (implemented with ENVI 4.7, and we select the average of the LRM as the LRP), AWLP [30], and the state-of-the-art adjustable model based method (denoted by AM) [10]. The method used in the related study [17], which uses the traditional sparse coding for pan-sharpening, is also included (denoted by SC). To more comprehensively examine the performance of the proposed approach, we give two sets of results for the proposed method: the results using only two-step sparse coding (denoted by TSSC), and the results using both two-step sparse coding and patch normalization (denoted by PN-TSSC). For a fair comparison, the SC method is implemented on every patch independently, with an averaging operation for the overlapping areas, and the same multi-thread LARS is employed as the  $\ell_1$ -norm optimization tool.

We evaluate the results both visually and quantitatively. To quantitatively evaluate the experiments, five widely used quality indices are used: the correlation coefficient (CC), the structural similarity metric (SSIM), the spectral angle mapper (SAM), the erreur relative globale adimensionnelle de synthèse (ERGAS), and the Q4 index. Based on Wald's view, these indices call upon distances that measure the discrepancy between a fused image

(monomodal or multimodal) and its reference [31]. The details of these indices can be found in Appendix A.

##### A. Experiments With IKONOS Data

The IKONOS system simultaneously offers a 4 m-resolution MS image with four bands and a single-band 1 m-resolution PAN image. The original MS and PAN images are low-pass filtered and downsampled to acquire the 16 m-resolution and 4 m-resolution test images, respectively, and the original 4 m-resolution MS image is used as the reference. The data set used in this experiment is shown in Fig. 3. For the AM method, the Huber parameter  $\mu$  is set as 2, the adjustable parameter  $\lambda_{tradeoff}$  between the enhancement of the spatial information and the preservation of the spectral information is set as 10, and the termination threshold  $d$  is set as  $10^{-7}$ . We choose a patch size and step size of [7 7] and [3 3], respectively, for the SC, TSSC, and PN-TSSC methods. For the SC method,  $\lambda$  in (2) is set as  $2^{13}$ . For the TSSC and PN-TSSC methods,  $\lambda$  in (6) is set as  $2^{11}$  and  $2^9$ , respectively. We set the parameters of the proposed method by taking both performance and computation time into consideration, and the influence of these parameters is analyzed in Appendix B.

The experimental results of the seven pan-sharpening methods are shown in Fig. 4(b)–(h), respectively. To facilitate a spatial comparison, detailed regions are illustrated in the top-left corner of the images. From the middle-right vegetation area, it can be clearly observed that the GIHS result suffers from severe color distortion when compared with the original MS image, and many details of the PAN image which do not belong to the original MS image are contained in the fused image. The AWLP result has good color, but some small details are lost, and the result does not look very sharp. For the GS method, the result seems good, perceptually, with high contrast. However, compared with the original MS image, some areas are over-sharpened, and the color in the middle-right vegetation area is brighter. The result of the AM method is similar to the GS result, with better color. The SC result shows a good spectral quality, but contains blurring effects and outlier pixels in some regions with small objects or fine details. The main reason for this is that the SC method only relies on the structural information of the LR patch to infer the structural information of the HR patch. As shown in Fig. 3(a), the structural information of some areas is almost lost, so it is difficult to infer the correct sparse representation of the LRM image patch via one-step sparse coding. By using the two-step sparse coding, which can adaptively utilize the structural similarity property between the LRM patch and its corresponding LRP patch, the TSSC result overcomes the blurring effects observed in the SC result and shows very good spatial detail information, but suffers from some color distortion when compared with the result of the SC method. By simultaneously incorporating both the two-step sparse coding approach and the patch normalization strategy, the proposed method achieves the best overall performance among all the methods. The color of the PN-TSSC method is very similar to that of the SC method and the AWLP method, and it is much better than the colors of the other methods, while the spatial details of the PN-TSSC method are much better than

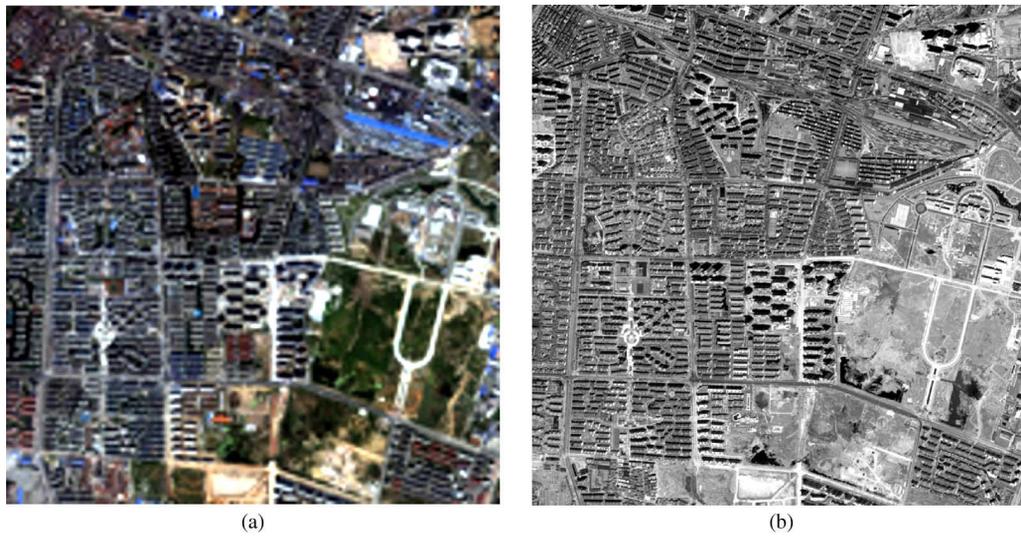


Fig. 3. Input  $600 \times 600$  IKONOS images: (a) upsampled 16 m MS image; (b) 4 m PAN image.

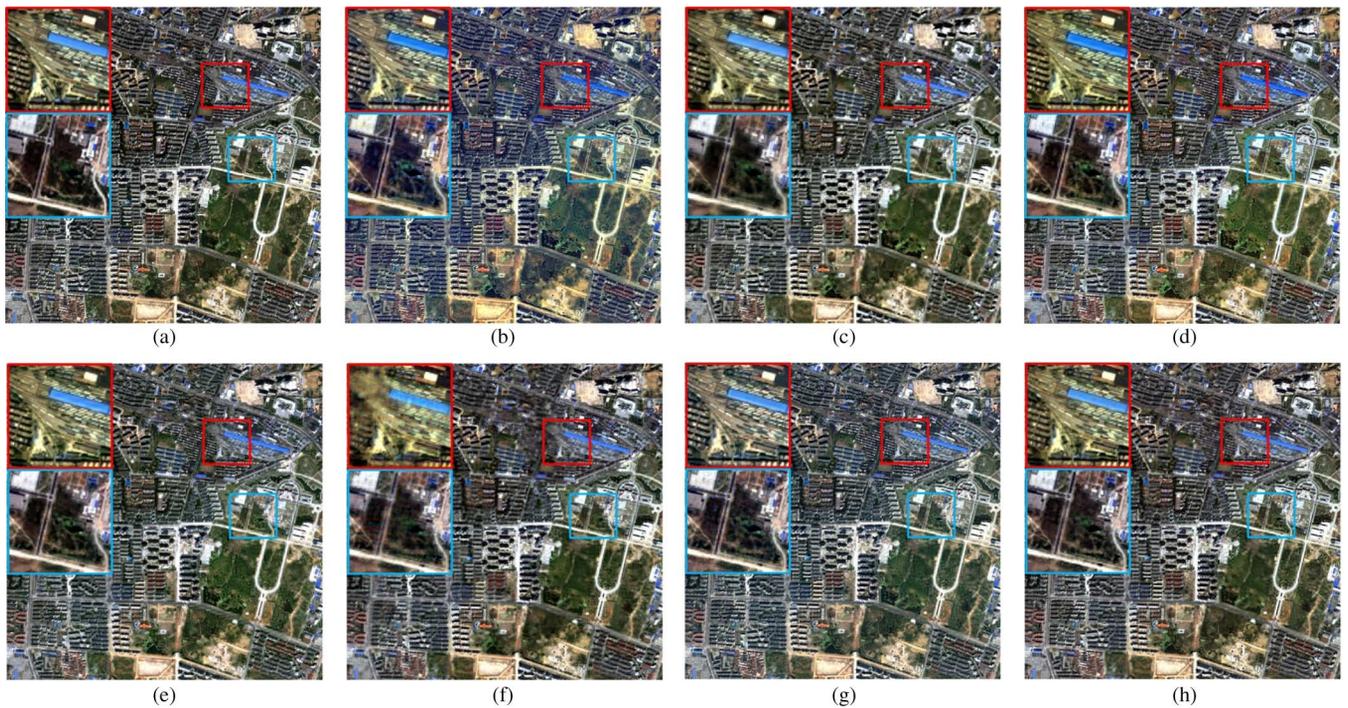


Fig. 4. Pan-sharpening results with the degraded IKONOS image. (a) Original; (b) GIHS; (c) AWLP; (d) GS; (e) AM; (f) SC; (g) TSSC; (h) PN-TSSC.

those of the SC method, and are comparable with or even better than those of the other methods.

The quantitative assessment results are shown in Table I, in which the best results for each quality index are marked in bold, and the second best are underlined. B, G, R, and NIR represent the results of the blue, green, red, and near-infrared bands, respectively, and Avg. is the average result of the B, G, R, and NIR results. In our experience, a difference of 0.01 for CC, SSIM, and Q4, and a difference of 0.1 for SAM and ERGAS, indicate significant differences which can be noticed by visual inspection. Bearing this in mind, it can be seen that the quantitative assessment results are consistent with the visual evaluations. The

AWLP method injects the high frequencies of the PAN image into every MS band, in proportion to their original values, in such a way that the spectral angle is not changed before and after fusion for the MS image. Due to the blurring effect, SC fails to surpass the traditional methods. Comparing the results of SC, TSSC, and PN-TSSC, the results improve successively from SC to PN-TSSC, which indicates the effectiveness of the two-step sparse coding approach and the patch normalization strategy for the sparse coding based pan-sharpening methods from a quantitative perspective. As a state-of-the-art method, AM is distinctly better than the traditional methods, and PN-TSSC is comparable with or even better than AM in most of the quality measures.

TABLE I  
THE QUANTITATIVE ASSESSMENT RESULTS OF THE SIMULATED EXPERIMENTS WITH THE IKONOS DATA SHOWN IN FIG. 4

	GIHS	AWLP	GS	AM	SC	TSSC	PN-TSSC
CC	B	0.8519	0.9214	<u>0.9321</u>	<b>0.9352</b>	0.8943	0.9165
	G	0.9500	0.9389	0.9482	<u>0.9521</u>	0.9280	<b>0.9537</b>
	R	0.9455	0.9366	0.9431	0.9478	0.9252	<u>0.9509</u>
	NIR	0.9182	0.9320	0.9206	<u>0.9435</u>	0.9152	<b>0.9436</b>
	Avg.	0.9164	0.9322	0.9360	<b>0.9446</b>	0.9157	<u>0.9444</u>
SSIM	B	0.8712	0.9010	0.8882	<b>0.9059</b>	0.8459	0.8807
	G	0.8668	0.8862	0.8707	<u>0.8959</u>	0.8328	0.8929
	R	0.8187	0.8577	0.8382	0.8591	0.7992	<u>0.8675</u>
	NIR	0.7770	<u>0.8424</u>	0.7960	<b>0.8460</b>	0.7530	0.8419
	Avg.	0.8334	0.8718	0.8483	<u>0.8767</u>	0.8077	0.8707
SAM	4.5112	<b>3.8755</b>	4.1639	<u>3.9134</u>	4.4013	4.1389	3.9255
ERGAS	3.5162	2.7649	3.0780	<u>2.4790</u>	2.9857	2.5020	<b>2.4557</b>
Q4	0.7745	0.8487	0.7948	0.8532	0.7602	<u>0.8541</u>	<b>0.8615</b>

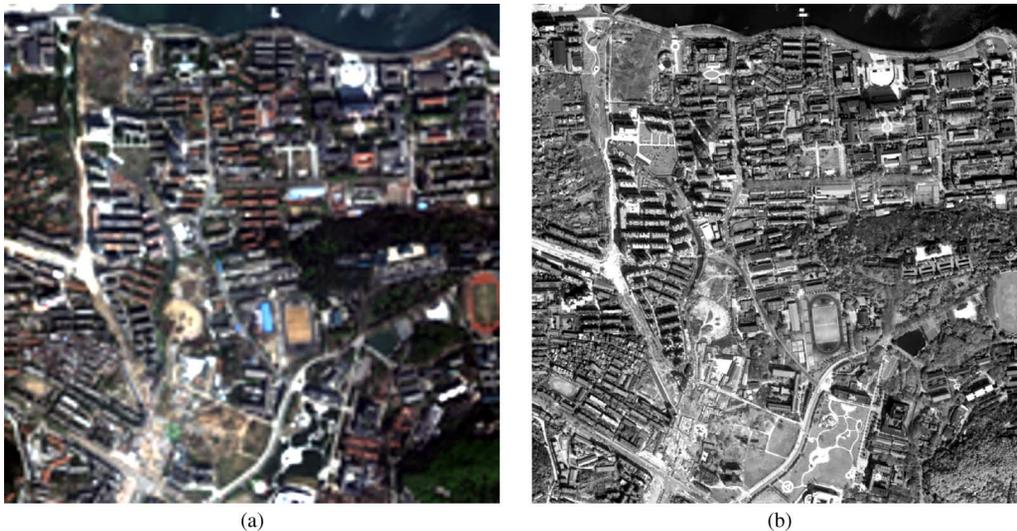


Fig. 5. Input  $600 \times 600$  QuickBird images: (a) upsampled 11.2 m MS image; (b) 2.8 m PAN image.

### B. Experiments With QuickBird Data

To test the proposed method with a different sensor, the four-band 2.8 m-resolution MS image and the 0.7 m-resolution PAN image of the QuickBird data are low-pass filtered and resampled to 11.2 m-resolution and 2.8 m-resolution test images, respectively, and the original 2.8 m-resolution MS image is used as the reference, as shown in Fig. 5. For the AM method, the Huber parameter  $\mu$  is set as 3, the adjustable parameter  $\lambda_{tradeoff}$  is set as 120, and the termination threshold  $d$  is set as  $10^{-7}$ . The patch size and step size are set as [7 7] and [3 3], respectively, for SC, TSSC, and PN-TSSC. For the SC method,  $\lambda$  in (2) is set as  $2^{11}$ . For the TSSC and PN-TSSC methods,  $\lambda$  in (6) is set as  $2^7$  for both methods.

The results of the seven different methods are presented in Fig. 6(b)–(h), respectively. To facilitate a spatial comparison, magnified regions are displayed in the top-left corner of the images. Compared with the traditional methods, the SC method shows a good color appearance, but blurring effects and some outlier pixels are obvious. With the two-step sparse coding, the TSSC method shows good spatial details, but the result is accompanied by slight color distortion. By simultaneously exploiting the two-step sparse coding method and the patch normalization strategy, the proposed PN-TSSC method succeeds in achieving the best overall quality among all the

methods. It should also be noted that, despite the proposed method showing superiority in both spatial and spectral qualities, an exception in the tiny region inside the yellow rectangle in Fig. 6 can be observed. Here, there is still a blurring effect for the small building surrounded by trees. The reason for this may be as follows: on the one hand, the small building covers an area of about  $17 \times 4$  pixels, which is small compared with the reconstructed  $28 \times 28$  image patch, and the sparse coding step is only concerned with the overall reconstruction accuracy of the whole patch. On the other hand, the occasion of a small building surrounded by trees is a relatively rare occurrence in the whole image, which also means that dictionary atoms similar to this kind of occasion will not be abundant. Therefore, to preserve the spectral quality, tiny details are sacrificed with the PN-TSSC method, and further research into adopting a multi-resolution dictionary may be able to solve this imperfection.

The quantitative assessment results are shown in Table II, in which the best results for each quality index are marked in bold, and the second best are underlined. Again, it can be seen that the quantitative assessment results are consistent with the visual evaluations, and are similar to those in Table I, which indicates that the proposed PN-TSSC method is stable with different kinds of sensors.

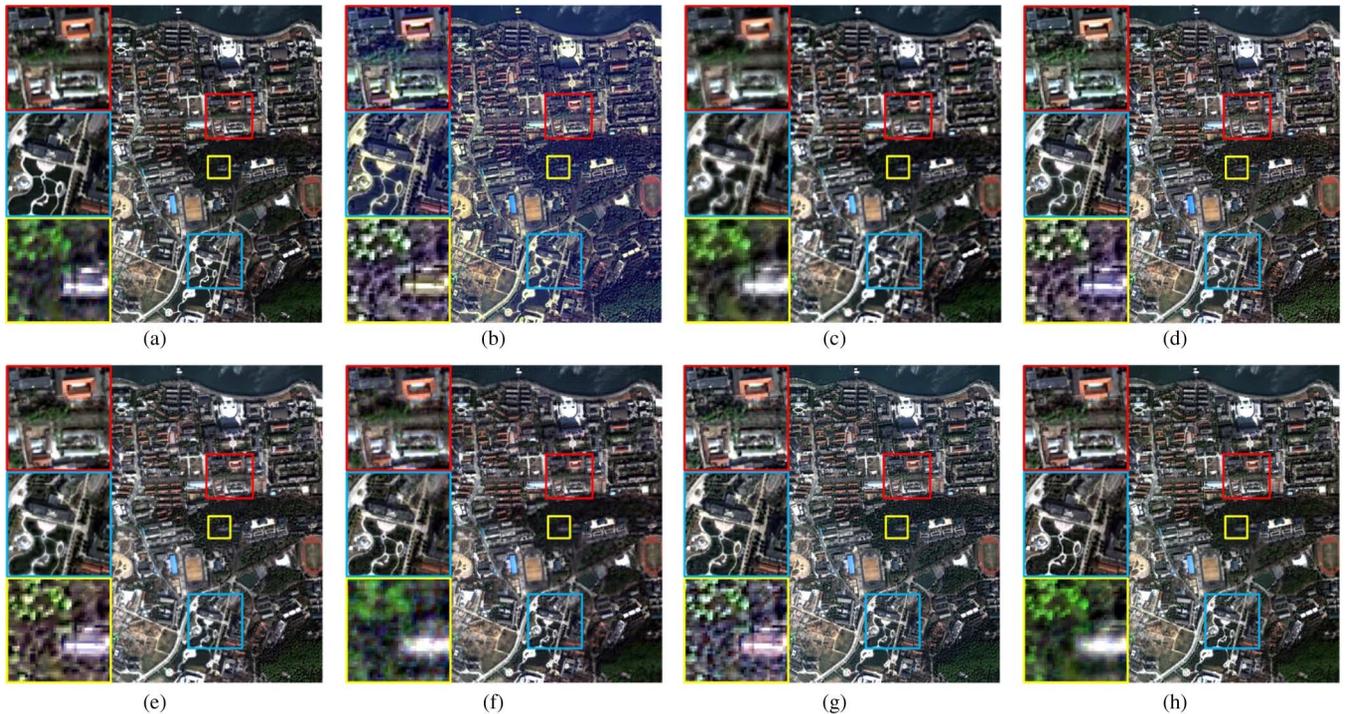


Fig. 6. Pan-sharpening results with the degraded QuickBird image. (a) Original; (b) GIHS; (c) AWLP; (d) GS; (e) AM; (f) SC; (g) TSSC; (h) PN-TSSC.

TABLE II  
THE QUANTITATIVE ASSESSMENT RESULTS OF THE SIMULATED EXPERIMENTS WITH THE QUICKBIRD DATA SHOWN IN FIG. 6

		GIHS	AWLP	GS	AM	SC	TSSC	PN-TSSC
CC	B	0.7731	0.9575	0.9425	<b>0.9644</b>	0.9522	0.9334	<u>0.9582</u>
	G	0.9545	0.9607	0.9508	<b>0.9663</b>	0.9629	0.9530	<u>0.9649</u>
	R	0.9575	0.9634	0.9501	<b>0.9690</b>	0.9657	<u>0.9659</u>	0.9647
	NIR	0.9607	0.9660	0.9570	0.9579	0.9666	<b>0.9712</b>	<u>0.9699</u>
	Avg.	0.9115	0.9619	0.9501	<b>0.9644</b>	0.9618	0.9558	<b>0.9644</b>
SSIM	B	0.9400	<u>0.9847</u>	0.9810	<b>0.9863</b>	0.9810	0.9698	0.9830
	G	0.9563	<u>0.9595</u>	0.9548	<b>0.9614</b>	0.9564	0.9407	0.9591
	R	0.9524	<u>0.9578</u>	0.9492	<u>0.9603</u>	0.9563	0.9547	0.9544
	NIR	0.9283	0.9402	0.9378	0.9254	0.9311	<b>0.9430</b>	<u>0.9420</u>
	Avg.	0.9442	<b>0.9606</b>	0.9557	<u>0.9584</u>	0.9562	0.9521	<u>0.9596</u>
SAM	2.3043	<b>1.7258</b>	1.9437	1.9212	1.9400	1.8634	<u>1.7447</u>	
ERGAS	2.0384	1.5799	1.8805	1.6230	<u>1.5498</u>	1.5939	<b>1.5350</b>	
Q4	0.8132	0.8612	0.8433	<u>0.8670</u>	0.8271	0.8488	<b>0.8695</b>	

TABLE III  
THE SPECTRAL RANGES OF THE BANDS OF WORLDVIEW-2

Band	Coastal	Blue	Green	Yellow	Red	Red edge	NIR-1	NIR-2	PAN
Range (nm)	400–450	450–510	510–580	585–625	630–690	705–745	770–895	860–1040	450–800

### C. Experiments With WorldView-2 Data

To further test the proposed method with a different sensor type and more bands, we utilize a 2 m-resolution MS image and the corresponding 0.5 m-resolution PAN image, as acquired by the WorldView-2 earth observation satellite. The spectral ranges of all the bands of the WorldView-2 data are given in Table III. The original data are degraded to obtain the test images, as shown in Fig. 7. The results of all the pan-sharpening methods are then obtained by fusing the 8 m-resolution MS image and the

2 m-resolution PAN image, and the original 2 m-resolution MS image is used as the reference. For the AM method, the Huber parameter  $\mu$  is set as 0.5, the adjustable parameter  $\lambda_{tradeoff}$  is set as 30, and the termination threshold  $d$  is set as  $10^{-7}$ . The patch size and the step size are set as [7 7] and [3 3], respectively, for SC, TSSC, and PN-TSSC. For the SC method,  $\lambda$  in (2) is set as  $2^{12}$ . For the TSSC and PN-TSSC methods,  $\lambda$  in (6) is set as  $2^{11}$  and  $2^{10}$ , respectively.

The results of the different methods are presented in Fig. 8, and zoomed regions are displayed in the top-left corner of the

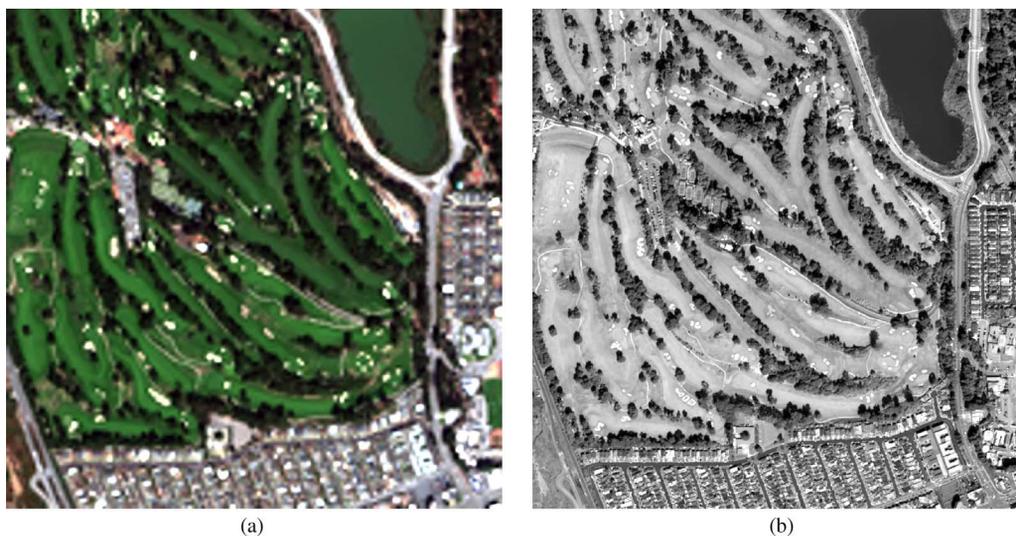


Fig. 7. Input  $600 \times 600$  WorldView-2 images: (a) upsampled 8 m MS image; (b) 2 m PAN image.

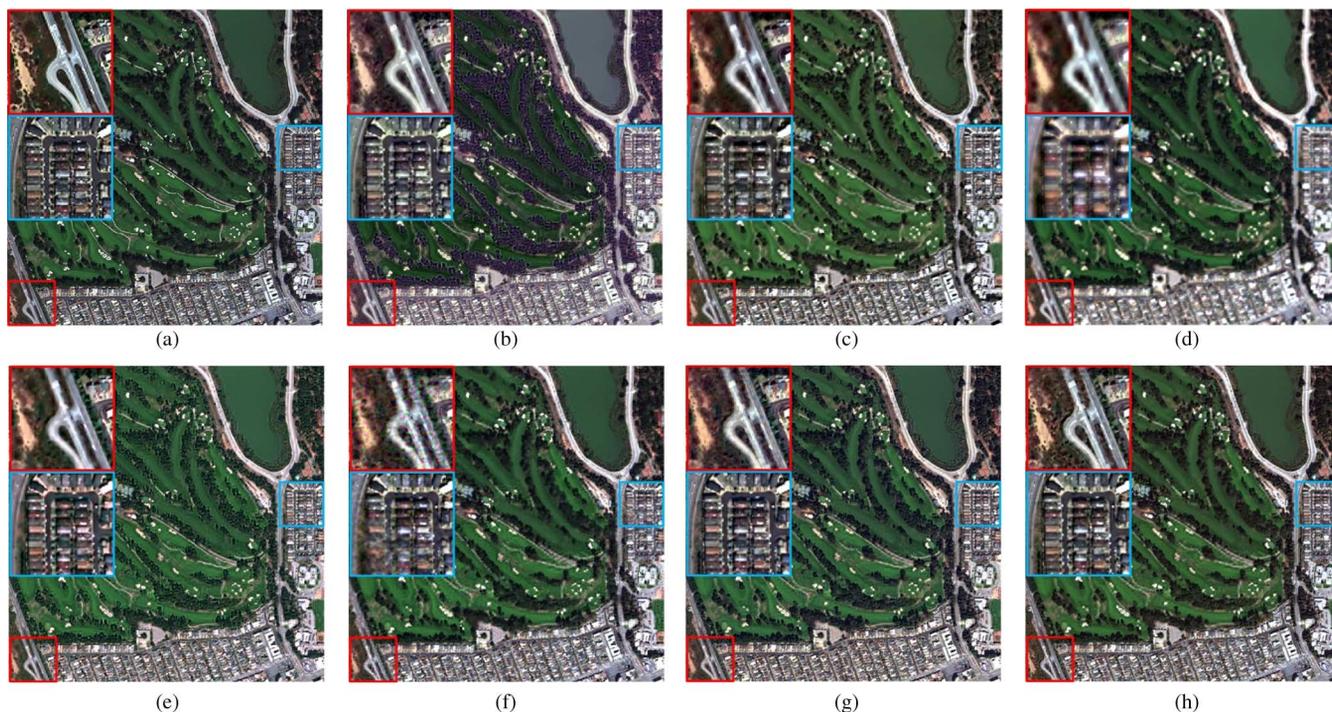


Fig. 8. Pan-sharpening results with the degraded WorldView-2 image. (a) Original; (b) GIHS; (c) AWLP; (d) GS; (e) AM; (f) SC; (g) TSSC; (h) PN-TSSC.

images to facilitate a spatial comparison. For the GS method, the blurring effect is obvious. This is because GS selects the average of all the LRM bands as the LRP, and the spectral ranges of the coastal band, the NIR-1 band, and the NIR-2 band are both out (almost out for NIR-1) of the range of the panchromatic band. This also affects the AM method, which assumes that the panchromatic band is a linear combination of all the multispectral bands with the same resolution. Since the PN-TSSC method can process every band and every local patch adaptively by choosing atoms from the dictionary, it is relatively robust to this kind of problem. From the results, it can be easily concluded that the proposed PN-TSSC method ranks first among all the methods, with a much better spatial quality.

The quantitative assessment results are shown in Table IV, in which the best results for each quality index are marked in bold, and the second best are underlined. C, B, G, Y, R, R-E, NIR-1, and NIR-2 represent the results of the coastal, blue, green, yellow, red, red edge, near-infrared-1, and near-infrared-2 bands, respectively, and Avg. is the average result of the C, B, G, Y, R, R-E, NIR-1, and NIR-2 results. As with the visual evaluations, the proposed PN-TSSC method performs the best among all the methods, in terms of the quantitative evaluations.

#### D. Cost of Time

The GS method is implemented in ENVI 4.7, AM is implemented in Microsoft Visual C++ 6.0, and the other methods are

TABLE IV  
THE QUANTITATIVE ASSESSMENT RESULTS OF THE SIMULATED EXPERIMENTS WITH THE WORLDVIEW-2 DATA SHOWN IN FIG. 8

	GIHS	AWLP	GS	AM	SC	TSSC	PN-TSSC	
CC	C	0.8692	0.9118	0.8375	<u>0.9264</u>	0.8897	0.9201	<b>0.9282</b>
	B	0.8917	0.9117	0.8310	0.9298	0.8991	<b>0.9357</b>	<u>0.9306</u>
	G	0.9009	0.9243	0.8668	0.9469	0.9317	<b>0.9520</b>	<u>0.9499</u>
	Y	0.9054	0.9226	0.8535	<u>0.9446</u>	0.9235	0.9414	<b>0.9452</b>
	R	0.9074	0.9174	0.8397	0.9195	0.9133	<u>0.9282</u>	<b>0.9363</b>
	R-E	0.9529	0.9599	0.9224	0.9688	0.9613	<u>0.9701</u>	<b>0.9718</b>
	NIR-1	0.9449	0.9491	0.8903	0.9458	0.9416	<u>0.9517</u>	<b>0.9542</b>
	NIR-2	0.9475	0.9509	0.8890	0.9510	0.9453	<u>0.9536</u>	<b>0.9556</b>
	Avg.	0.9150	0.9309	0.8663	0.9416	0.9257	<u>0.9441</u>	<b>0.9465</b>
SSIM	C	0.8765	0.9604	0.9321	<u>0.9616</u>	0.9437	0.9548	<b>0.9636</b>
	B	0.8960	0.9420	0.8964	0.9443	0.9215	<b>0.9476</b>	<u>0.9467</u>
	G	0.9282	0.9224	0.8709	0.9225	0.9071	<b>0.9398</b>	<u>0.9370</u>
	Y	0.8967	0.9028	0.8279	0.9077	0.8734	<u>0.9088</u>	<b>0.9134</b>
	R	0.8698	<u>0.9158</u>	0.8453	0.9058	0.8940	0.9138	<b>0.9220</b>
	R-E	0.8395	0.8781	0.8772	<u>0.8942</u>	0.8496	0.8912	<b>0.8988</b>
	NIR-1	0.6838	<u>0.7504</u>	0.7082	0.6993	0.6632	0.7442	<b>0.7594</b>
	NIR-2	0.7061	<u>0.7641</u>	0.7116	0.7364	0.6834	0.7583	<b>0.7714</b>
	Avg.	0.8371	0.8795	0.8337	0.8715	0.8420	<u>0.8823</u>	<b>0.8890</b>
SAM	7.2493	<u>5.5357</u>	7.6779	6.1675	6.4520	5.9033	<b>5.4392</b>	
ERGAS	5.2027	4.1451	5.8494	3.9634	4.2765	<u>3.8418</u>	<b>3.7186</b>	
Q8	0.6932	0.7481	0.6999	0.7308	0.6767	<u>0.7543</u>	<b>0.7728</b>	

TABLE V  
TIME COSTS FOR THE DIFFERENT METHODS WITH  $600 \times 600$  DATA

	GIHS	AWLP	GS	AM	SC	PN-TSSC
IKONOS	0.1	0.7	3	22	31	42
QuickBird	0.1	0.7	3	91	31	39
WorldView-2	0.2	0.9	4	161	62	62

implemented in MATLAB 2012a. The personal computer we use is a Dell T1500. The central processing unit (CPU) is an Intel Core i3 540 @ 3.07 GHz with dual-cores and four threads, the RAM is 3 Gb with 3.06 GHz, and the operating system is 32-bit Windows XP. The time costs for the experiments with the IKONOS, QuickBird, and WorldView-2 data are shown in Table V. From Table V, it can be seen that the time cost of the proposed PN-TSSC method is comparable with the AM method; however, it should be noted that the proposed method could be easily implemented with parallel processing. It is also reasonable to believe that with the rapid development in computer hardware and computation techniques, the time cost of the proposed method will soon no longer be an issue. For the proposed method, the whole of the PAN image is used to construct the dictionary. Therefore, when the PAN image is very large, i.e.,  $1000 \times 1000$ , or even larger, the dictionary will also be large, which will increase the computation time for each patch. A simple solution is to partition the whole image into several parts with overlaps, and process each part with the proposed method separately. In addition to this, a pre-trained dictionary pair for a certain kind of remote sensing system could be a better choice, and this will be investigated in our future work.

## V. CONCLUSIONS

This paper presents a novel sparse coding based pan-sharpening method for remote sensing images. Spatial resolution

enhancement and spectral information preservation are the two key issues in the remote sensing image pan-sharpening task. On the one hand, to improve the spatial resolution of the existing one-step sparse coding based method, a two-step sparse coding method is utilized on the basis of the patch structural similarity between the low-resolution panchromatic image patch and the low-resolution multispectral image patch. On the other hand, a patch normalization strategy is used to improve the stability of the sparse coding and preserve the spectral information. The experimental results with IKONOS, QuickBird, and WorldView-2 data show that the performance of the proposed method is competitive in both the spatial and spectral qualities, and it outperforms both traditional and state-of-the-art methods when considering the overall performance. Moreover, since a whole image can be processed in patches independently with multi-thread calculation, the proposed method can also be fast. With further work, better results may be obtained by considering all the multispectral bands together. A better dictionary should also be considered to speed up the method when the processed image is very large.

## APPENDIX A

The five indices used in this paper are defined as follows.

1) *The Correlation Coefficient (CC)*: The CC shows the similarity in small-size structures between the reference image band

and the fused image band [29], [31]. As a measure of correlation, it should be as close as possible to 1, and it is defined as:

$$CC_j = \frac{\sigma_{\mathbf{X}_j, \mathbf{Y}_j}}{\sigma_{\mathbf{X}_j} \sigma_{\mathbf{Y}_j}} \quad (9)$$

where  $\mathbf{X}_j$  and  $\mathbf{Y}_j$  are the  $j$ th bands of the reference image and the fused image, respectively.  $\sigma_{\mathbf{X}_j, \mathbf{Y}_j}$  is the covariance between  $\mathbf{X}_j$  and  $\mathbf{Y}_j$ .  $\sigma_{\mathbf{X}_j}$  and  $\sigma_{\mathbf{Y}_j}$  represent the standard deviation of  $\mathbf{X}_j$  and  $\mathbf{Y}_j$ , respectively.

2) *The Structural Similarity Metric (SSIM)*: Based on the fact that the human visual system has evolved to extract structural information from images, the SSIM [32] was developed to capture the loss of image structure. It has been widely used and is considered to be better than the root mean square error (RMSE) since RMSE only measures the radiometric distortion of the fused image band from the reference image band. The value of the SSIM should be as close as possible to 1, and the SSIM metric between the  $j$ th band of the reference and that of the fused image is defined as:

$$SSIM(\mathbf{X}_j, \mathbf{Y}_j) = [l(\mathbf{X}_j, \mathbf{Y}_j)]^\alpha [c(\mathbf{X}_j, \mathbf{Y}_j)]^\beta [s(\mathbf{X}_j, \mathbf{Y}_j)]^\gamma \quad (10)$$

where  $l(\mathbf{X}_j, \mathbf{Y}_j)$ ,  $c(\mathbf{X}_j, \mathbf{Y}_j)$ , and  $s(\mathbf{X}_j, \mathbf{Y}_j)$  are the luminance, contrast, and structural comparison measures.  $\alpha$ ,  $\beta$ , and  $\gamma$  are parameters used to define the relative importance of the three components, given as follows:

$$l(\mathbf{X}_j, \mathbf{Y}_j) = \frac{2\mu_{\mathbf{X}_j}\mu_{\mathbf{Y}_j} + C_1}{\mu_{\mathbf{X}_j}^2 + \mu_{\mathbf{Y}_j}^2 + C_1} \quad (11)$$

$$c(\mathbf{X}_j, \mathbf{Y}_j) = \frac{2\sigma_{\mathbf{X}_j}\sigma_{\mathbf{Y}_j} + C_2}{\sigma_{\mathbf{X}_j}^2 + \sigma_{\mathbf{Y}_j}^2 + C_2} \quad (12)$$

$$s(\mathbf{X}_j, \mathbf{Y}_j) = \frac{\sigma_{\mathbf{X}_j\mathbf{Y}_j} + C_3}{\sigma_{\mathbf{X}_j}\sigma_{\mathbf{Y}_j} + C_3} \quad (13)$$

where  $\mu_{\mathbf{X}_j}$  and  $\mu_{\mathbf{Y}_j}$  are the mean values of  $\mathbf{X}_j$  and  $\mathbf{Y}_j$ , respectively.  $C_1$ ,  $C_2$ , and  $C_3$  are small constants given as:  $C_1 = (K_1L)^2$ ,  $C_2 = (K_2L)^2$ , and  $C_3 = C_2/2$ , respectively.  $L$  is the dynamic range of the pixel values. In this paper, we set  $\alpha = \beta = \gamma = 1$ ,  $L = 2047$ ,  $K_1 = 0.01$ , and  $K_2 = 0.03$ , as suggested in [32].

3) *The Spectral Angle Mapper (SAM)*: The SAM measures the absolute angle between the spectral vectors of the reference and fused images, which reflects the spectral distortion introduced by the fusion process. The ideal value is 0, and it is defined as:

$$SAM = \frac{1}{M_1M_2} \sum_{i=1}^{M_1M_2} \arccos \frac{\sum_{j=1}^b (\mathbf{X}_j^i \mathbf{Y}_j^i)}{\sqrt{\sum_{j=1}^b (\mathbf{X}_j^i)^2} \sqrt{\sum_{j=1}^b (\mathbf{Y}_j^i)^2}} \quad (14)$$

where  $\mathbf{X}_j^i$  and  $\mathbf{Y}_j^i$  are the  $i$ th pixels of the  $j$ th bands of the reference and fused images, respectively.  $M_1$  and  $M_2$  are the horizontal and vertical sizes of the image.

4) *The Erreur Relative Globale Adimensionnelle De Synthèse (ERGAS)*: ERGAS is widely used and it gives a global depiction of the radiometric difference between the reference

and fused images. ERGAS fulfills three requirements: the independence of units, the independence of the number of spectral bands, and the independence of the ratio of the scales [33]. The ideal value for ERGAS is 0, and it is defined as:

$$ERGAS = 100 \frac{h}{l} \sqrt{\frac{1}{b} \sum_{j=1}^b \frac{RMSE_j^2}{\mu_{\mathbf{X}_j}^2}} \quad (15)$$

where  $h$  and  $l$  are the spatial resolutions of the PAN image and the MS image, respectively. For the IKONOS, QuickBird, and WorldView-2 data,  $l$  is four times as large as  $h$ , and  $b$  is the number of the bands. The RMSE is defined as:

$$RMSE_j = \sqrt{\frac{\|\mathbf{X}_j - \mathbf{Y}_j\|^2}{(M_1M_2)}}$$

5) *The Q4 Index*: Based on the theory of quaternions, the Q4 index [34] is an extension of the Q index [35], which is a product of CC, mean bias, and contrast variation. The Q4 index is suitable for testing both the spectral and spatial qualities of fused images [14]. The ideal value for the Q4 index is 1, and it is defined as:

$$Q4 = E(Q4_{D \times D}) \quad (16)$$

where  $Q4_{D \times D}$  is defined as:

$$Q4_{D \times D} = \frac{|\sigma_{\mathbf{z}_1\mathbf{z}_2}|}{\sigma_{\mathbf{z}_1} \cdot \sigma_{\mathbf{z}_2}} \cdot \frac{2\sigma_{\mathbf{z}_1} \cdot \sigma_{\mathbf{z}_2}}{\sigma_{\mathbf{z}_1}^2 + \sigma_{\mathbf{z}_2}^2} \cdot \frac{2|\bar{\mathbf{z}}_1| \cdot |\bar{\mathbf{z}}_2|}{|\bar{\mathbf{z}}_1|^2 + |\bar{\mathbf{z}}_2|^2} \quad (17)$$

where  $\mathbf{z}_1$  and  $\mathbf{z}_2$  are the quaternions, defined as:  $\mathbf{z}_1 = \mathbf{X}_1^i + i\mathbf{X}_2^i + j\mathbf{X}_3^i + k\mathbf{X}_4^i$ , and  $\mathbf{z}_2 = \mathbf{Y}_1^i + i\mathbf{Y}_2^i + j\mathbf{Y}_3^i + k\mathbf{Y}_4^i$ .  $|\cdot|$  denotes the modulus of the quaternion.  $\sigma_{\mathbf{z}_1\mathbf{z}_2}$  is the hypercomplex covariance between  $\mathbf{z}_1$  and  $\mathbf{z}_2$ .  $\sigma_{\mathbf{z}_1}$  and  $\sigma_{\mathbf{z}_2}$  are the square roots of the variances of  $\mathbf{z}_1$  and  $\mathbf{z}_2$ , respectively.  $\bar{\mathbf{z}}_1$  and  $\bar{\mathbf{z}}_2$  are the expected values of  $\mathbf{z}_1$  and  $\mathbf{z}_2$ , respectively. The first term of (17) is the modulus of the hypercomplex CC between  $\mathbf{z}_1$  and  $\mathbf{z}_2$ , and the second and third terms measure contrast changes and the mean bias on all bands, respectively.

The conjugates of  $\mathbf{z}_1$  and  $\mathbf{z}_2$  are defined as  $\mathbf{z}_1^* = \mathbf{X}_1^i - i\mathbf{X}_2^i - j\mathbf{X}_3^i - k\mathbf{X}_4^i$  and  $\mathbf{z}_2^* = \mathbf{Y}_1^i - i\mathbf{Y}_2^i - j\mathbf{Y}_3^i - k\mathbf{Y}_4^i$ , respectively. As quaternions,  $\mathbf{z}_1$  and  $\mathbf{z}_2$  have the following rules:

$$|\mathbf{z}_1| = \sqrt{\mathbf{z}_1 \cdot \mathbf{z}_1^*} = \sqrt{(\mathbf{X}_1^i)^2 + (\mathbf{X}_2^i)^2 + (\mathbf{X}_3^i)^2 + (\mathbf{X}_4^i)^2} \quad (18)$$

$$|\mathbf{z}_2| = \sqrt{\mathbf{z}_2 \cdot \mathbf{z}_2^*} = \sqrt{(\mathbf{Y}_1^i)^2 + (\mathbf{Y}_2^i)^2 + (\mathbf{Y}_3^i)^2 + (\mathbf{Y}_4^i)^2} \quad (19)$$

$$\sigma_{\mathbf{z}_1\mathbf{z}_2} = E[(\mathbf{z}_1 - \bar{\mathbf{z}}_1) \cdot (\mathbf{z}_2 - \bar{\mathbf{z}}_2)^*] = E(\mathbf{z}_1 \cdot \mathbf{z}_2^*) - \bar{\mathbf{z}}_1 \cdot \bar{\mathbf{z}}_2^* \quad (20)$$

$$\sigma_{\mathbf{z}_1}^2 = E[|\mathbf{z}_1 - \bar{\mathbf{z}}_1|^2] = E[|\mathbf{z}_1|^2] - |\bar{\mathbf{z}}_1|^2 \quad (21)$$

$$\sigma_{\mathbf{z}_2}^2 = E[|\mathbf{z}_2 - \bar{\mathbf{z}}_2|^2] = E[|\mathbf{z}_2|^2] - |\bar{\mathbf{z}}_2|^2 \quad (22)$$

By using (18)–(22), (17) can be rewritten as:

$$Q4_{D \times D} = \frac{4|E(\mathbf{z}_1 \cdot \mathbf{z}_2^*) - \bar{\mathbf{z}}_1 \cdot \bar{\mathbf{z}}_2^*|}{E(\mathbf{z}_1 \cdot \mathbf{z}_1^*) - \bar{\mathbf{z}}_1 \cdot \bar{\mathbf{z}}_1^* + E(\mathbf{z}_2 \cdot \mathbf{z}_2^*) - \bar{\mathbf{z}}_2 \cdot \bar{\mathbf{z}}_2^*} \cdot \frac{\sqrt{|\bar{\mathbf{z}}_1| \cdot \bar{\mathbf{z}}_1^*} \sqrt{|\bar{\mathbf{z}}_2| \cdot \bar{\mathbf{z}}_2^*}}{|\bar{\mathbf{z}}_1| \cdot \bar{\mathbf{z}}_1^* + |\bar{\mathbf{z}}_2| \cdot \bar{\mathbf{z}}_2^*} \quad (23)$$

where  $E(\cdot)$  is obtained by averaging the pixel quaternions within a  $D \times D$  block. We choose  $D = 8$  in this paper, and

TABLE VI  
INFLUENCE OF  $\lambda$  ON THE IKONOS DATA

$\lambda$	16384	2048	1024	512	256	128	1	1/16384
ERGAS	2.5646	2.4578	<b>2.4556</b>	2.4557	2.4562	2.4567	2.4572	2.4572
Q4	0.8455	...	0.8605	0.8612	0.8615	0.8615	<b>0.8616</b>	...
Time (s)	<b>11</b>	30	37	42	45	48	50	54

TABLE VII  
INFLUENCE OF  $\lambda$  ON THE QUICKBIRD DATA

$\lambda$	16384	512	256	128	64	32	1	1/16384
ERGAS	1.7916	1.5379	1.5352	<b>1.5350</b>	1.5353	1.5356	1.5360	1.5360
Q4	0.8311	...	0.8679	0.8690	0.8695	0.8698	<b>0.8700</b>	...
Time (s)	<b>5</b>	28	35	39	41	46	47	124

TABLE VIII  
INFLUENCE OF  $\lambda$  ON THE WORDVIEW-2 DATA

$\lambda$	16384	4096	2048	1024	512	256	1	1/16384
ERGAS	3.8740	3.7392	3.7227	<b>3.7186</b>	3.7191	3.7201	3.7225	3.7225
Q8	0.7596	...	0.7709	0.7723	0.7728	0.7730	<b>0.7731</b>	...
Time (s)	<b>19</b>	39	51	62	72	78	90	97

extend the Q4 index to Q8 for the 8-band WorldView-2 data by redefining  $\mathbf{z}_1$  and  $\mathbf{z}_2$  as:

$$\mathbf{z}_1 = \mathbf{X}_1^i + i\mathbf{X}_2^i + j\mathbf{X}_3^i + k\mathbf{X}_4^i + a\mathbf{X}_5^i + b\mathbf{X}_6^i + c\mathbf{X}_7^i + d\mathbf{X}_8^i \quad (24)$$

$$\mathbf{z}_2 = \mathbf{Y}_1^i + i\mathbf{Y}_2^i + j\mathbf{Y}_3^i + k\mathbf{Y}_4^i + a\mathbf{Y}_5^i + b\mathbf{Y}_6^i + c\mathbf{Y}_7^i + d\mathbf{Y}_8^i \quad (25)$$

Note that  $\mathbf{z}_1$  and  $\mathbf{z}_2$  are still hypercomplex numbers with only their number of dimensions changing from 4 to 8, and we can assume that we can still use (16) and (23) to calculate Q8 with dimensional extension when calculating  $\mathbf{z}_1 \cdot \mathbf{z}_1^*$ ,  $\mathbf{z}_2 \cdot \mathbf{z}_2^*$ ,  $\mathbf{z}_1 \cdot \mathbf{z}_2^*$ ,  $\bar{\mathbf{z}}_1 \cdot \bar{\mathbf{z}}_1^*$ ,  $\bar{\mathbf{z}}_2 \cdot \bar{\mathbf{z}}_2^*$ , and  $\bar{\mathbf{z}}_1 \cdot \bar{\mathbf{z}}_2^*$ , e.g., (18) is replaced with (26):

$$|\mathbf{z}_1| = \sqrt{\mathbf{z}_1 \cdot \mathbf{z}_1^*} = \sqrt{\begin{matrix} (\mathbf{X}_1^i)^2 + (\mathbf{X}_2^i)^2 + (\mathbf{X}_3^i)^2 \\ + (\mathbf{X}_4^i)^2 + (\mathbf{X}_5^i)^2 + (\mathbf{X}_6^i)^2 \\ + (\mathbf{X}_7^i)^2 + (\mathbf{X}_8^i)^2 \end{matrix}} \quad (26)$$

## APPENDIX B

There are three parameters for the proposed method, i.e., the patch size, the step size, and the regularization parameter  $\lambda$ . For simplicity, the ERGAS and Q4 indices are used to evaluate the performance here. How these three parameters affect the performance and computation time of the proposed method is discussed as follows. Similar results were also observed for SC and TSSC, but have been omitted to save space.

We first set the patch size and step size as [7 7] and [3 3], respectively. The influence of  $\lambda$  on the IKONOS, QuickBird, and WorldView-2 data is shown in Table VI, Table VII, and Table VIII, respectively. By setting  $\lambda$  from 16384 to 1/16384, more dictionary atoms are allowed to be used to represent a certain patch; thus, the performance is first improved, then converges because the patch has already been well represented with enough dictionary atoms. For the computation time, it increases continuously with a smaller  $\lambda$ . Similar results are also observed

TABLE IX  
INFLUENCE OF STEP SIZE ON THE IKONOS DATA  
WITH PATCH SIZE [7 7] AND  $\lambda = 1$

Step size	[1 1]	[2 2]	[3 3]	[4 4]	[5 5]	[6 6]	[7 7]
ERGAS	<b>2.3904</b>	2.4185	2.4572	2.5326	2.6274	2.7401	2.8805
Q4	<b>0.8690</b>	0.8658	0.8616	0.8534	0.8438	0.8317	0.8169
Time (s)	5933	674	50	16	6	4	<b>2</b>

with other patch sizes and step sizes. In fact, we can simply set  $\lambda = 1$ , because for different data, the results converge with  $\lambda = 1$ , which has a very similar performance and consumes only a little more time than the value we chose in this paper.

To test the influence of the step size, we fix  $\lambda = 1$ , and test different step sizes with patch size [7 7] on the data sets used in this paper. The results are shown in Table IX, Table X, and Table XI, respectively. It can be seen that a smaller step size achieves an obviously better result; however, the computation time will be greatly increased. Similar results are also observed with the other patch sizes. We chose the step size as [3 3] in this paper because it is able to show the advantages of the proposed method over the other methods, and the computation time is tolerable with the basic PC used in this study. With much better computation power offered by a workstation or distributed computation technique, the step size could be set as [1 1] to achieve the best performance.

Similarly, to test the influence of the patch size, we fix  $\lambda = 1$  and set the step size as [3 3]. The different results obtained by the different patch sizes with the three data sets are displayed in Table XII, Table XIII, and Table XIV, respectively. It can be seen that a bigger patch size needs more computation time. The best patch size for the IKONOS data is [11 11], and the best patch size for the WorldView-2 data is [7 7]. For the QuickBird data, a patch size of [7 7] can achieve the best Q4 index, while a

TABLE X  
INFLUENCE OF STEP SIZE ON THE QUICKBIRD DATA  
WITH PATCH SIZE [7 7] AND  $\lambda = 1$

Step size	[1 1]	[2 2]	[3 3]	[4 4]	[5 5]	[6 6]	[7 7]
ERG	<b>1.5076</b>	1.5195	1.5360	1.5778	1.6231	1.6794	1.7528
AS	<b>0.8757</b>	0.8734	0.8700	0.8627	0.8544	0.8474	0.8337
Time (s)	5012	425	47	14	6	3	<b>2</b>

TABLE XI  
INFLUENCE OF STEP SIZE ON THE WORDVIEW-2 DATA  
WITH PATCH SIZE [7 7] AND  $\lambda = 1$

Step size	[1 1]	[2 2]	[3 3]	[4 4]	[5 5]	[6 6]	[7 7]
ERG	<b>3.6809</b>	3.6945	3.7225	3.8010	3.9229	4.0102	4.2212
AS	<b>0.7839</b>	0.7801	0.7731	0.7608	0.7442	0.7360	0.7099
Time (s)	7383	595	90	29	12	6	<b>4</b>

TABLE XII  
INFLUENCE OF THE PATCH SIZE ON THE IKONOS DATA  
WITH STEP SIZE [3 3] AND  $\lambda = 1$

Patch size	[5 5]	[7 7]	[9 9]	[11 11]	[13 13]
ERGAS	2.5500	2.4572	2.4342	<b>2.4169</b>	2.4177
Q4	0.8544	0.8616	0.8625	<b>0.8633</b>	0.8628
Time (s)	<b>20</b>	50	121	358	771

TABLE XIII  
INFLUENCE OF THE PATCH SIZE ON THE QUICKBIRD DATA  
WITH STEP SIZE [3 3] AND  $\lambda = 1$

Patch size	[5 5]	[7 7]	[9 9]	[11 11]	[13 13]	[15 15]
ERGAS	1.5838	1.5360	1.5266	1.5202	<b>1.5191</b>	1.5241
Q4	0.8652	<b>0.8700</b>	0.8689	0.8675	0.8659	0.8635
Time (s)	<b>18</b>	47	110	314	720	1300

TABLE XIV  
INFLUENCE OF THE PATCH SIZE ON THE WORLDVIEW-2 DATA  
WITH STEP SIZE [3 3] AND  $\lambda = 1$

Patch size	[5 5]	[7 7]	[9 9]	[11 11]	[13 13]
ERGAS	3.7466	<b>3.7225</b>	3.7397	3.7542	3.7696
Q8	0.7689	<b>0.7731</b>	0.7722	0.7695	0.7656
Time (s)	<b>37</b>	90	226	620	1497

patch size of [13 13] leads to the best ERGAS index. Therefore, unlike the step size, the best patch size varies with different data. Moreover, in our experiments, we also found that the best patch size varied from [7 7] to [11 11] with different data acquired by the same sensor. Since the computation time for [7 7] is much less than for [9 9] or [11 11], we chose [7 7] for all the data in this paper.

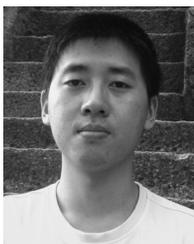
## ACKNOWLEDGMENT

The authors would like to thank Julien Mairal *et al.* for sharing the multi-thread implementation of the LARS algorithm. Thanks are also due to Ron Rubinstein and Michael Elad for sharing the fast MEX implementation of the patch operations. The authors would also like to thank DigitalGlobe for providing the WorldView-2 images used in this study, and the IEEE GRSS Data Fusion Technical Committee for organizing the 2006 [36] and 2012 Data Fusion Contest.

## REFERENCES

- [1] T. Ranchin, B. Aiuzzi, L. Alparone, S. Baronti, and L. Wald, "Image fusion—The ARSIS concept and some successful implementation schemes," *ISPRS J. Photogramm. Remote Sens.*, vol. 58, no. 1, pp. 4–18, Jun. 2003.
- [2] T.-M. Tu, S.-C. Su, H.-C. Shyu, and P. S. Huang, "A new look at IHS-like image fusion methods," *Inf. Fusion*, vol. 2, no. 3, pp. 177–186, Sep. 2001.
- [3] P. S. Chavez, S. C. Sides, and J. A. Anderson, "Comparison of three different methods to merge multiresolution and multispectral data: Landsat TM and SPOT panchromatic," *Photogramm. Eng. Remote Sens.*, vol. 57, no. 3, pp. 295–303, 1991.
- [4] C. A. Laben and B. V. Brower, "Process for Enhancing the Spatial Resolution of Multispectral Imagery Using Pan-Sharpener," U.S. Patent 6,011,875, Jan. 4, 2000, Tech. Rep., Eastman Kodak Company.
- [5] Y. Zhang, "Problems in the fusion of commercial high-resolution satellite images as well as Landsat-7 images and initial solutions," in *Proc. Int. Archives Photogramm. Remote Sens. Spatial Inf. Sci.*, 2002, vol. 34, no. 4, pp. 587–592, Part 4.
- [6] K. Amolins, Y. Zhang, and P. Dare, "Wavelet based image fusion techniques—An introduction, review and comparison," *ISPRS J. Photogramm. Remote Sens.*, vol. 62, no. 4, pp. 249–263, Sep. 2007.
- [7] Y. Zhang, "Understanding image fusion," *Photogramm. Eng. Remote Sens.*, vol. 70, no. 6, pp. 657–661, Jun. 2004.
- [8] M. V. Joshi, L. Bruzzone, and S. Chaudhuri, "A model-based approach to multiresolution fusion in remotely sensed images," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 9, pp. 2549–2562, Sep. 2006.
- [9] Z. Li and H. Leung, "Fusion of multispectral and panchromatic images using a restoration-based method," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 5, pp. 1482–1491, May 2009.
- [10] L. Zhang, H. Shen, W. Gong, and H. Zhang, "Adjustable model-based fusion method for multispectral and panchromatic images," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 42, no. 6, pp. 1693–1704, Dec. 2012.
- [11] Y. Zhang and G. Hong, "An IHS and wavelet integrated approach to improve pan-sharpening visual quality of natural colour IKONOS and QuickBird images," *Inf. Fusion*, vol. 6, no. 3, pp. 225–234, Sep. 2005.
- [12] V. P. Shah, N. H. Younan, and R. L. King, "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.
- [13] B. Luo, M. M. Khan, T. Bienvenu, J. Chanussot, and L. Zhang, "Decision-based fusion for pansharpening of remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 1, pp. 19–23, Jan. 2013.
- [14] S. Li and B. Yang, "A new pan-sharpening method using a compressed sensing technique," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 2, pp. 738–746, Feb. 2011.
- [15] C. Jiang, 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.
- [16] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, Nov. 2010.
- [17] 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.
- [18] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 6583, pp. 607–609, 1996.
- [19] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

- [20] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- [21] M. Elad, M. A. T. Figueiredo, and Y. Ma, "On the role of sparse and redundant representations in image processing," *Proc. IEEE*, vol. 98, no. 6, pp. 972–982, Jun. 2010.
- [22] 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.
- [23] J. A. Tropp and S. J. Wright, "Computational methods for sparse solution of linear inverse problems," *Proc. IEEE*, vol. 98, no. 6, pp. 948–958, Jun. 2010.
- [24] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *Ann. Statist.*, vol. 32, no. 2, pp. 407–499, 2004.
- [25] J. Mairal, F. Bach, J. Ponce, G. Sapiro, R. Jenatton, and G. Obozinski, Sparse Modeling Software [Online]. Available: <http://spams-devel.gforge.inria.fr/index.html> Feb. 16, 2013
- [26] J. Mairal, Julien Mairal's Homepage [Online]. Available: <http://lear.inrialpes.fr/people/mairal/> Feb. 16, 2013
- [27] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Stat. Soc. B*, vol. 58, no. 1, pp. 267–288, 1996.
- [28] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Non-local sparse models for image restoration," in *Proc. IEEE Int. Conf. Computer Vision*, Tokyo, Japan, 2009, pp. 2272–2279.
- [29] L. Wald, T. Ranchin, and M. Mangolini, "Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images," *Photogramm. Eng. Remote Sens.*, vol. 63, no. 6, pp. 691–699, 1997.
- [30] X. Otazu, M. Gonzalez-Audicana, O. Fors, and J. Nunez, "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.
- [31] C. Thomas and L. Wald, "Comparing distances for quality assessment of fused products," in *Proc. 26th EARSeL Annu. Symp. New Develop. Challenges Remote Sens.*, Z. Bochenek, Ed., Warsaw, Poland, May 29–31, 2006, pp. 101–111, Rotterdam, The Netherlands: Balkema, 2007.
- [32] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error measurement to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [33] L. Wald, "Quality of high resolution synthesized images: Is there a simple criterion?," in *Proc. Int. Conf. Fusion of Earth Data*, Nice, France, Jan. 2000, vol. 1, pp. 99–105.
- [34] L. Alparone, S. Baronti, A. Garzelli, and F. Nencini, "A global quality measurement of pan-sharpened multispectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 1, no. 4, pp. 313–317, Oct. 2004.
- [35] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Processing Lett.*, vol. 9, pp. 81–84, Mar. 2002.
- [36] L. Alparone, L. Wald, J. Chanussot, C. Thomas, P. Gamba, and L. Bruce, "Comparison of pansharpening algorithms: Outcome of the 2006 GRSS data-fusion contest," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3012–3021, Oct. 2007.



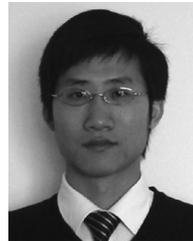
**Cheng Jiang** received the B.S. degree in remote sensing science and technology from Wuhan University, Wuhan, China, in 2010. He is currently pursuing the Ph.D. degree at the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University, Wuhan, China.

His research interests include image fusion, denoising and super-resolution in remote sensing images.



**Hongyan Zhang** (M'13) received the B.S. degree in geographic information system and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2005 and 2010, respectively.

Since 2010, he has been a lecturer with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University. His current research interests focus on image reconstruction and remote sensing image processing.



**Huanfeng Shen** (M'11) received the B.S. degree in surveying and mapping engineering and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2002 and 2007, respectively.

In July 2007, he joined the School of Resource and Environmental Science, Wuhan University, where he is currently a full professor. His research interests include image processing (for quality improvement), remote sensing application, data fusion and assimilation. Dr. Shen has published more than 60 research papers. He has been supported by several talent programs, including the New Century Excellent Talents by the Ministry of Education of China (2011) and the Hubei Science Fund for Distinguished Young Scholars (2011).



**Liangpei Zhang** (M'06–SM'08) received the B.S. degree in physics from Hunan Normal University, ChangSha, China, in 1982, the M.S. degree in optics from the Xi'an Institute of Optics and Precision Mechanics of Chinese Academy of Sciences, Xi'an, China, in 1988, and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 1998.

He is currently with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, as the head of the Remote Sensing Division. He is also a "Chang-Jiang Scholar" Chair Professor appointed by the Ministry of Education, China. He has more than 200 research papers and 5 patents. He is now Principal Scientist for the China State Key Basic Research Project (2011–2016) appointed by the Ministry of National Science and Technology of China to lead the remote sensing program in China. His research interests include hyperspectral remote sensing, high resolution remote sensing, image processing and artificial intelligence.

Dr. Zhang regularly serves as a Co-Chair of the series SPIE Conferences on Multispectral Image Processing and Pattern Recognition (MIPPR), Conference on Asia Remote Sensing, and many other conferences. He edits several conference proceedings, issues, and the Geoinformatics Symposiums. He also serves as an Associate Editor of *International Journal of Ambient Computing and Intelligence* (IJACI), *International Journal of Image and Graphics*, *International Journal of Digital Multimedia Broadcasting*, *Journal of Geo-spatial Information Science*, and the *Journal of Remote Sensing*. He is a Fellow of IEE, Executive Member (Board of Governor) of the China National Committee of International Geosphere-Biosphere Programme and Executive Member for the China Society of Image and Graphics.