Sparse Representation Based Pansharpening Using Trained Dictionary

Ming Cheng, Cheng Wang, and Jonathan Li, *Senior Member, IEEE*

***Abstract*—Sparse representation has been used to fuse high-resolution panchromatic (HRP) and low-resolution multispectral (LRM) images. However, the approach faces the difficulty that the dictionary is generated from the high-resolution multispectral (HRM) images, which are unknown. In this letter, a two-step method is proposed to train the dictionary from the HRP and LRM images. In the first step, coarse HRM images are obtained by additive wavelet fusion method. The initial dictionary is composed of randomly sampled patches from the coarse HRM images. In the second step, a linear constraint K-SVD method is designed to train the dictionary to improve its representation ability. Experimental results using QuickBird and IKONOS data indicate that the trained dictionary yields comparable fusion products with raw-patch-dictionary sampled from HRM images.**

***Index Terms*—Image fusion, K-SVD, multispectral image, panchromatic image, sparse representation.**

I. Introduction

SPACEBORNE imagery usually provides separated but complementary product types due to technological and physical constraints. Typically, optical satellite sensors provide high-resolution panchromatic (HRP) and low-resolution multispectral (LRM) images. In order to benefit from both spectral and spatial information, image fusion or pansharpening techniques have been developed to produce high-resolution multispectral (HRM) image by combining the HRP and LRM images.

So far, many pansharpening methods have been developed [1], [2]. Most of the methods take the HRP image as the source of high-resolution spatial information and induct the information into the LRM images. Initial efforts are based on component substitution (intensity-hue-saturation (IHS) [3], [4], principal component substitution (PCS) [5]). In IHS, the LRM images are resampled and transformed to IHS domain, and then the intensity component I is substituted with the HRP and transformed back to the spectral domain. For most advanced high resolution imagers such as IKONOS and QuickBird, the spectral responses of the multispectral channels are not perfectly overlapped with the bandwidth of PAN, and component substitution approaches may generate large spectral distortion [1]. Another family of methods is multiresolution analysis (MRA) which injects the zero-mean high-pass spatial details only. In some literatures it is called ARSIS. The spatial details can be extracted or described by Laplacian pyramids [6], wavelets [7], contourlets [8], or support value transform (SVT) [9]. The injection model can be injecting directly into each LRM band, or injecting through an IHS transform.

In recent years sparse representations (SR) have been widely used in image processing. The assumption that natural images admit a sparse decomposition over a redundant dictionary leads to many efficient algorithms [10]. In [11] this idea is applied to pansharpening (the authors use the term *compressed sensing* (CS) instead of SR). The dictionary is generated by randomly sampling raw patches from HRM images, and the sparse coefficients are calculated according to the relations between HRM, HRP, and LRM images. The fusion results are the products of the dictionary and the sparse coefficients.

One obstacle to the CS based pansharpening method is that the dictionary is generated from the HRM images, which are unknown. In fact they are just what we want to obtain. In [11] the authors use the decimated PAN and MS images to demonstrate their method, and the dictionary is generated from the original LRM images. However, no answer is given on how to construct the dictionary when fusing the original HRP and LRM images. One possible way to obtain the true HRM images is imaging from an airborne platform carrying the same sensor. But as far as we know, no such work is reported.

In this letter, a method is proposed to generate the dictionary from the HRP and LRM images. The method includes two steps. The first step is pansharpening using a simple and fast method to obtain coarse HRM images. The second step is dictionary training using patches sampled from the results of the first step. A dictionary training scheme is designed based on the well-known K-SVD method [12]. The training process incorporates information from the HRP image, which improves the ability of the dictionary to describe spatial details. Experimental results using QuickBird and IKONOS data indicate that the trained dictionary produces comparable fusion results with the method in [11].

II. CS Based Pansharpening

In this section we briefly describe the SR model and the CS based pansharpening method proposed in [11].

Manuscript received September 9, 2012; revised February 1, 2013; accepted March 29, 2013.

M. Cheng and C. Wang are with the Key Laboratory of Underwater Acoustic Communication and Marine Information Technology, Ministry of Education, and Department of Computer Science, Xiamen University, Xiamen 361005, China (e-mail: chm99@xmu.edu.cn; cwang@xmu.edu.cn).

J. Li is with the Department of Geography and Environmental Management, University of Waterloo, Waterloo, ON, Canada (e-mail: junli@uwaterloo.ca).

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/LGRS.2013.2256875

We consider image patches of size $\sqrt{n} \times \sqrt{n}$ pixels, ordered lexicographically as column vectors $\{\mathbf{x}_i, i = 1, 2, \cdots\}$, $\mathbf{x}_i \in \mathbf{R}^n$. SR theory supposes the existence of a matrix $\mathbf{D} \in \mathbf{R}^{n \times K}$, each column of which corresponds to a possible image. These possible images are referred to as atomic images, and the matrix $\mathbf{D}$ as a dictionary of the atomic images. An image patch $\mathbf{x}$ can be represented as $\mathbf{x} = \mathbf{D}\boldsymbol{\alpha}$. For overcomplete $\mathbf{D}(K \gg n)$, there are many possible $\boldsymbol{\alpha}$ satisfying $\mathbf{x} = \mathbf{D}\boldsymbol{\alpha}$. Our aim is to find the $\boldsymbol{\alpha}$ with the fewest nonzero elements. Thus, the $\boldsymbol{\alpha}$ is called the sparse coefficients of $\mathbf{x}$ with dictionary $\mathbf{D}$. Formally, this problem can be obtained by solving the following optimization problem:

$$\tilde{\boldsymbol{\alpha}} = \arg\ \min ||\boldsymbol{\alpha}||_0 \text{ subject to } ||\mathbf{D}\boldsymbol{\alpha} - \mathbf{x}||_2^2 = 0 \tag{1}$$

where $||\cdot||_0$ denotes the number of nonzero entries in a vector.

In practice, because of various restrictions, we cannot get $\mathbf{x}$ directly. Instead, only a small set of measurements $\mathbf{y}$ of $\mathbf{x}$ is observed. The observation $\mathbf{y}$ can be represented as

$$\mathbf{y} = \mathbf{M}\mathbf{x} + \mathbf{v} \tag{2}$$

where $\mathbf{M} \in \mathbf{R}^{m \times n}$ with $m < n$ is the measurement matrix and $\mathbf{v}$ is additive noise. The CS theory ensures that under sparsity regularization [13], we can recover $\mathbf{x}$ correctly from the observation $\mathbf{y}$ by solving the problem

$$\tilde{\boldsymbol{\alpha}} = \arg\min ||\boldsymbol{\alpha}||_0 \text{ subject to } ||\mathbf{y} - \mathbf{M}\mathbf{D}\boldsymbol{\alpha}||_2^2 \leq \varepsilon \tag{3}$$

where ε is the reconstruction error depending on noise level of the source image, and $\mathbf{x} = \mathbf{D}\tilde{\boldsymbol{\alpha}}$.

The above optimization is an NP-hard problem. If $\boldsymbol{\alpha}$ is sufficiently sparse, this problem can be replaced with minimizing the l_1-norm problem

$$\tilde{\boldsymbol{\alpha}} = \arg\min \|\boldsymbol{\alpha}\|_1 \text{ subject to } \|\mathbf{y} - \mathbf{M}\mathbf{D}\boldsymbol{\alpha}\|_2^2 \leq \varepsilon. \tag{4}$$

The proof from (3) to (4) can be found in [14]. The l_1-norm problem (4) can be solved by orthogonal matching pursuit (OMP) or basis pursuit (BP) algorithm effectively.

In the pansharpening problem, the HRP and LRM images are referred as the measurements $\mathbf{y}$. The matrix $\mathbf{M}$ is constructed by the model from HRM images to HRP and LRM images. We consider a pansharpening case with four spectral bands: B, G, R, and NIR, and the decimation factor from high to low spatial resolution is 4. Let

$$\mathbf{x} = \{x_{1,1}^B, \cdots, x_{1,8}^B, x_{2,1}^B, \cdots, x_{8,8}^B, x_{1,1}^G, \cdots, x_{8,8}^{NIR}\}^T$$

be the lexicographically ordered HRM image patch of size $8 \times 8 \times 4$ pixels. The measurements $\mathbf{y}$ is composed of two parts

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_{LRM} \\ \mathbf{y}_{PAN} \end{pmatrix}$$

where

$$\mathbf{y}_{LRM} = \{y_{1,1}^B, y_{1,2}^B, y_{2,1}^B, y_{2,2}^B, y_{1,1}^G, \cdots, y_{2,2}^{NIR}\}^T$$
$$\mathbf{y}_{PAN} = \{y_{1,1}^{PAN}, \cdots, y_{1,8}^{PAN}, y_{2,1}^{PAN}, \ldots, y_{8,8}^{PAN}\}^T$$

are the corresponding lexicographically ordered LRM and HRP image patches respectively.

Each entry in $\mathbf{y}_{LRM}$ can be seen as the average of a 4×4 image patch in the HRM images, and we can write

$$\mathbf{y}_{LRM} = \mathbf{M}_1\mathbf{x} + \mathbf{v}_1 \tag{5}$$

where $\mathbf{v}_1$ is the 16×1 zero-mean additive noise vector, and $\mathbf{M}_1$ is the decimation matrix of size 16×256 given by

$$\mathbf{M}_1 = \frac{1}{16}\mathbf{I}_8 \otimes [\mathbf{1}^T \otimes (\mathbf{I}_2 \otimes \mathbf{1}^T)]$$

where $\mathbf{I}_N$ is the $N \times N$ identity matrix, $\mathbf{1}$ is a 4×1 vector with all entries equal to one, and $\otimes$ is the Kronecker tensor product.

The HRP image is approximated by a linear combination of the four bands of the HRM images, and we can write

$$\mathbf{y}_{PAN} = \mathbf{M}_2\mathbf{x} + \mathbf{v}_2 \tag{6}$$

where $\mathbf{v}_2$ is the 64×1 zero-mean additive noise vector, and $\mathbf{M}_2 = \{w_1\mathbf{I}_{64}, w_2\mathbf{I}_{64}, w_3\mathbf{I}_{64}, w_4\mathbf{I}_{64}\}$ where w_p, $p = 1, 2, 3, 4$, is the weight of each band.

Finally, we get the relation between $\mathbf{y}$ and $\mathbf{x}$ given by

$$\mathbf{y} = \mathbf{M}\mathbf{x} + \mathbf{v}, \tag{7}$$

where $\mathbf{M} = \begin{pmatrix} \mathbf{M}_1 \\ \mathbf{M}_2 \end{pmatrix}$ and $\mathbf{v} = \begin{pmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \end{pmatrix}$.

The dictionary $\mathbf{D}$ is generated by randomly sampling raw patches from HRM images, then the optimization problem (4) is solved to get the sparse coefficients $\tilde{\boldsymbol{\alpha}}$. The fused HRM image patch is obtained by $\tilde{\mathbf{x}} = \mathbf{D}\tilde{\boldsymbol{\alpha}}$.

III. Dictionary Training Method

It is difficult to generate the dictionary from the HRM images directly since they are unknown. We design a two-step method to train the dictionary from the given HRP and LRM images. Firstly, coarse HRM images are obtained using proportional additive wavelet LHS (AWLP) method [7]. Then, an extended K-SVD method is designed to train the dictionary sampled from the coarse HRM images.

A. AWLP Method

Any pansharpening method can be used to generate the coarse HMR images. We select the multiresolution wavelet based method AWLP to fulfill the task because it can preserve the spectral signature while the operations are simple and fast. In this method, the HRP detail injected into each LRM band is proportional to the LRM band original radiance value. This method can be summarized as follows.

1) Upsample (resize) the LRM images to the size of the HRP image, the result of each band is HRM_i', $i = 1, 2, 3, 4$.
2) Obtaining the luminance band by $L = (1/4)\sum_{i=1}^{4} \text{HRM}_i'$.
3) Perform a histogram matching between the HRP band and the L band, modifying the former to obtain a new HRP band, denoted by HRP'.
4) Decompose HRP' into 2 wavelet planes, i.e., $\text{HRP}' = \omega_1 + \omega_2 + c$, c being the residual plane.
5) Add proportion of the wavelet planes to each HRM_i' to obtain the merged HRM band, i.e., HRM_i =

$(n_i/\sum_{i=1}^{4} n_i)(\omega_1+\omega_2)+\mathrm{HRM}'_i,\ i=1,2,3,4,$ where n_i is the original radiance value of each LRM band.

A shift-invariant discrete wavelet decomposition known as "*à trous*" algorithm is used to generate the wavelet planes. The algorithm performs successive convolutions with a mask, and the zero-mean wavelet planes are computed as the differences between the results of two consecutive convolutions.

B. Linear Constraint K-SVD Method

The initial dictionary $\mathbf{D}=\{\mathbf{d}_j\}_{j=1}^{K}$ is constructed by randomly sampling raw patches from the HRM images obtained by AWLP. Each atom $\mathbf{d}_j$ is a normalized lexicographically ordered image patch of size $8\times8\times4$ pixels. Though we can use this dictionary directly in the CS based method, we hope to improve its representation ability by training.

Let $\mathbf{y}_{HRM}$ be an arbitrary chosen HRM image patch and $\mathbf{y}_{PAN}$ be the corresponding HRP image patch

$$\mathbf{y}_{HRM}=\{y^B_{1,1},\dots y^B_{1,8},y^B_{2,1},\dots,y^B_{8,8},y^G_{1,1},\dots,y^{NIR}_{8,8}\}^T$$

$$\mathbf{y}_{PAN}=\{y^{PAN}_{1,1},\dots,y^{PAN}_{1,8},y^{PAN}_{2,1},\dots,y^{PAN}_{8,8}\}^T.$$

The SR of $\mathbf{y}_{HRM}$ and $\mathbf{y}_{PAN}$ are formulated by

$$\min\|\boldsymbol{\alpha_1}\|_0 \text{ subject to } \|\mathbf{y}_{HRM}-\mathbf{D}\boldsymbol{\alpha_1}\|_2^2\leqslant\varepsilon_1 \tag{8}$$

$$\min\|\boldsymbol{\alpha_2}\|_0 \text{ subject to } \|\mathbf{y}_{PAN}-\mathbf{M_2}\mathbf{D}\boldsymbol{\alpha_2}\|_2^2\leqslant\varepsilon_2. \tag{9}$$

Let

$$\mathbf{y}=\begin{pmatrix}\mathbf{y}_{HRM}\\ \mathbf{y}_{PAN}\end{pmatrix}$$

$$\mathbf{R}=\begin{pmatrix}\mathbf{I}_{256}\\ \mathbf{M}_2\end{pmatrix}=\begin{pmatrix}\mathbf{I}_{256}\\ w_1\mathbf{I}_{64}\ w_2\mathbf{I}_{64}\ w_3\mathbf{I}_{64}\ w_4\mathbf{I}_{64}\end{pmatrix}.$$

We redefine the SR problems of (8) and (9) in unified form with a linear constraint $\mathbf{R}$:

$$\min\|\boldsymbol{\alpha}\|_0 \text{ subject to } \|\mathbf{y}-\mathbf{RD}\boldsymbol{\alpha}\|_2^2\leqslant\varepsilon. \tag{10}$$

This is the basis for dictionary training. Let $\mathbf{Y}=\{\mathbf{y}_i\}_{i=1}^{N}$ be a collection of training samples, each column $\mathbf{y}_i$ being the concatenation of a HRM image patch and the corresponding HRP image patch, and $\mathbf{A}=\{\boldsymbol{\alpha}_i\}_{i=1}^{N}$, $\boldsymbol{\alpha}_i$ being the sparse coefficients of $\mathbf{y}_i$. Similar to the K-SVD method, we describe the dictionary training problem as follows:

$$\min_{\mathbf{D},\mathbf{A}}\{\|\mathbf{Y}-\mathbf{RDA}\|_F^2\}\text{ subject to }\forall i,\ \|\boldsymbol{\alpha}_i\|_0\leqslant T_0 \tag{11}$$

where $\|\cdot\|_F$ is the Frobenius norm of matrix, and T_0 is the predefined maximal l_0 norm of sparse coefficients.

Expression (11) is a linear constraint training problem [15], and could be minimized iteratively. Each iteration includes two stages. The first stage is sparse coding, in which we assume that $\mathbf{D}$ is fixed, and find the best coefficient matrix $\mathbf{A}$. The problem can be decoupled to N distinct problems of the form

$$\min\{\|\mathbf{y}_i-\mathbf{RD}\boldsymbol{\alpha}_i\|_2^2\}\text{ subject to }\|\boldsymbol{\alpha}_i\|_0\leqslant T_0, i=1,\cdots,N. \tag{12}$$

If T_0 is small enough, this problem can be adequately solved by any pursuit algorithms.

A second stage is performed to search for a better dictionary. This process updates one atom and the corresponding coefficients at a time. Allowing changes in the coefficients while updating the dictionary atoms accelerates convergence, since the subsequent atom updates will be based on more relevant coefficients. We put in question one atom $\mathbf{d}_k$ and the coefficients that correspond to it, the kth row in $\mathbf{A}$, denoted as $\boldsymbol{\alpha}_T^k$ (a row vector). The penalty term in (11) can be rewritten as

$$\begin{aligned}\|\mathbf{Y}-\mathbf{RDA}\|_F^2&=\Big\|\mathbf{Y}-\mathbf{R}\sum\nolimits_{j=1}^{K}\mathbf{d}_j\boldsymbol{\alpha}_T^j\Big\|_F^2\\&=\Big\|\Big(\mathbf{Y}-\mathbf{R}\sum\nolimits_{j\neq k}\mathbf{d}_j\boldsymbol{\alpha}_T^j\Big)-\mathbf{R}\mathbf{d}_k\boldsymbol{\alpha}_T^k\Big\|_F^2\\&=\big\|\mathbf{E}_k-\mathbf{R}\mathbf{d}_k\boldsymbol{\alpha}_T^k\big\|_F^2.\end{aligned} \tag{13}$$

Considering that $\boldsymbol{\alpha}_T^k$ is very likely to be sparse (since each column of $\mathbf{A}$ is sparse), we hope to remain the nonzero indices of $\boldsymbol{\alpha}_T^k$ unchanged during the update. We pick out the nonzero columns of $\boldsymbol{\alpha}_T^k$ to compose a new row vector $\boldsymbol{\alpha}_R^k$, and pick out the same columns of $\mathbf{E}_k$ to compose a new matrix $\mathbf{E}_k^R$. Then we replace the penalty term in (13) by $\big\|\mathbf{E}_k^R-\mathbf{R}\mathbf{d}_k\boldsymbol{\alpha}_R^k\big\|_F^2$.

Let the SVD factorization of $\mathbf{R}$ be

$$\mathbf{R}=\mathbf{U}_R\Sigma_R\mathbf{V}_R^T \tag{14}$$

where $\Sigma_R\in\mathbf{R}^{320\times256}$ includes just 256 nonzero entries on the main diagonal. Then,

$$\begin{aligned}||\mathbf{E}_k^R-\mathbf{R}\mathbf{d}_k\boldsymbol{\alpha}_R^k||_F^2&=\big\|\mathbf{E}_k^R-\mathbf{U}_R\Sigma_R\mathbf{V}_R^T\mathbf{d}_k\boldsymbol{\alpha}_R^k\big\|_F^2\\&=\big\|\mathbf{U}_R^T\mathbf{E}_k^R-\Sigma_R\mathbf{V}_R^T\mathbf{d}_k\boldsymbol{\alpha}_R^k\big\|_F^2.\end{aligned} \tag{15}$$

Let $\mathbf{F}=\mathbf{U}_R^T\mathbf{E}_k^R=\begin{pmatrix}\mathbf{F}_1\\ \mathbf{F}_2\end{pmatrix}$ such that $\mathbf{F}_1$ and $\mathbf{F}_2$ include 256 and 64 rows, respectively, and let Σ_R^σ be the top 256 rows of Σ_R, then

$$\begin{aligned}\big\|\mathbf{E}_k^R-\mathbf{R}\mathbf{d}_k\boldsymbol{\alpha}_R^k\big\|_F^2&=\left\|\begin{pmatrix}\mathbf{F}_1\\ \mathbf{F}_2\end{pmatrix}-\begin{pmatrix}\Sigma_R^\sigma\\ \mathbf{0}\end{pmatrix}\mathbf{V}_R^T\mathbf{d}_k\boldsymbol{\alpha}_R^k\right\|_F^2\\&=\big\|\mathbf{F}_1-\Sigma_R^\sigma\mathbf{V}_R^T\mathbf{d}_k\boldsymbol{\alpha}_R^k\big\|_F^2+\|\mathbf{F}_2\|_F^2.\end{aligned} \tag{16}$$

The second term is independent of $\mathbf{d}_k$ and $\boldsymbol{\alpha}_R^k$. Again we use the SVD factorization of matrix $\mathbf{F}_1=\mathbf{U}_F\Sigma_F\mathbf{V}_F^T$. The SVD finds the closest rank-one matrix (in Frobenius norm) that approximates $\mathbf{F}_1$, and this will effectively minimize the first term. Let $\Sigma_R^\sigma\mathbf{V}_R^T\mathbf{d}_k$ be the first column of $\mathbf{U}_F$, and $\boldsymbol{\alpha}_R^k$ be the first row of $\mathbf{V}_F^T$ multiplied by $\Sigma_F(1,1)$, we obtain the update formulas of $\mathbf{d}_k$ and $\boldsymbol{\alpha}_R^k$ given by

$$\mathbf{d}_k=\left(\Sigma_{\boldsymbol{R}}^\sigma\mathbf{V}_R^T\right)^{-1}\mathbf{U}_F^{\{1\}}=\mathbf{V}_R(\Sigma_{\boldsymbol{R}}^\sigma)^{-1}\mathbf{U}_F^{\{1\}} \tag{17}$$

$$\boldsymbol{\alpha}_R^k=\Sigma_F(1,1)\left\{\mathbf{V}_F^{\{1\}}\right\}^T. \tag{18}$$

The superscript $\{1\}$ represents the first column of a matrix. The inverse of Σ_R^σ is easy to obtain since it is a diagonal

matrix. The updated $\boldsymbol{\alpha}_R^k$ is used to replace the nonzero entries in $\boldsymbol{\alpha}_T^k$. A final normalization of the atom $\mathbf{d}_k$ in addition to a respective change in $\boldsymbol{\alpha}_T^k$ might also be needed.

An important question is whether the training process converges. When T_0 is small enough relative to K, the OMP or BP methods are known to perform very well to solve (12), and a good approximation to each $\mathbf{y}_i$ can be retrieved. In this case the sparse coding step decreases the total representation error $\|\mathbf{Y} - \mathbf{RDA}\|_F^2$. Moreover, when updating $\mathbf{d}_k$, an additional reduction or no change (when $\boldsymbol{\alpha}_T^k = \mathbf{0}$) in the representation error can be expected. Executing a series of such steps ensures a convergence to a local minimum. Practically, we saw in our experiments that a convergence can always be reached.

IV. Experimental Results

Experiments have been conducted to evaluate the performance of the trained dictionary using images of QuickBird and IKONOS. Four methods were selected for comparison: AWLP, generalized IHS with genetic algorithm (GIHS-GA) [16], Gram-Schmidt adaptive-context adaptive (GSA-CA) [17], and local correlation based gradient field optimization (LC-GFO) [18]. The original HRP and LRM images were fused by each method and the results were assessed by observations. Then, quantitative assessments were carried out using spatially degraded images and the original LRM images were used as the reference images. The following quality indicators were calculated: correlation coefficients (CC) of each band, root mean squared error (RMSE) of each band, spectral angle mapper (SAM), ERGAS, and $Q4_{8\times8}$.

The initial dictionary $\mathbf{D}$ included 2500 patches randomly sampled from the coarse HRM images. Another 10 000 HRM image patches together with the corresponding HRP image patches were randomly selected to construct the training samples $\mathbf{Y}$. The maximal l_0 norm T_0 in (11) was set to 60, and the maximum number of iterations was 80. The weight values in $\mathbf{M}_2$ were set to the same values in [11]. The optimization problem (12) was solved by OMP algorithm. In the quantitative evaluation, two types of dictionary were used for the CS based method. The first type was composed of randomly sampled raw patches from the original LRM images (denoted as CS1). The second type was the dictionary trained using the method proposed in this letter (denoted as CS2).

A. Fusion Results of QuickBird Data

QuickBird provides PAN image at 0.7 m resolution and MS images at 2.8 m resolution. Some datasets were downloaded from the website http://glcf.umiacs.umd.edu/data/quickbird/, other datasets were bought from DigitalGlobe, Inc. directly. Four datasets were used to train the dictionary, which cover buildings, roads, vegetation, and water.

The original images and the fusion results of four methods are shown in Fig. 1 (to save space, results of AWLP are not shown). The size of the HRP image is 400×400 pixels. All methods can fuse the data effectively. The edges of the buildings are all sharp enough. However, Fig. 1(f) can provide more spatial details (e.g., the cars in the lower left corner and the bushes in the center are more distinguishable). The spectral distortion of Fig. 1(e) is a little large. For the other three results, the spectral distortions are not obvious.

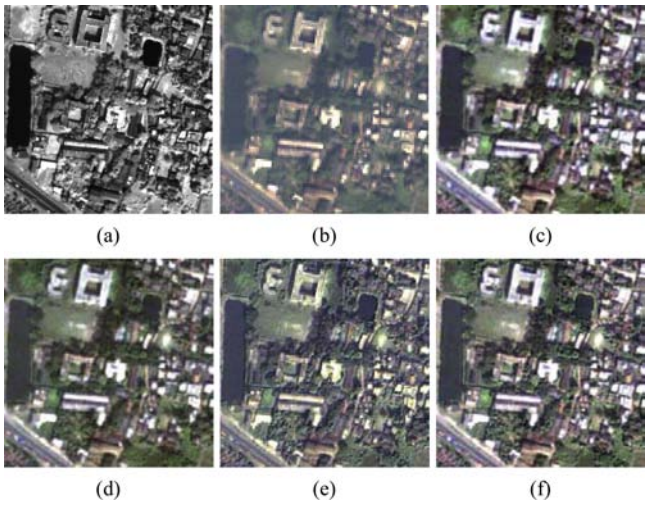

Fig. 1. Original QuickBird images and fusion results using different methods. (a) HRP image (0.7 m, 400×400). (b) LRM images (2.8 m). (c) GIHS-GA fused image. (d) GSA-CA fused image. (e) LC-GFO fused image. (f) CS fused image using trained dictionary. For all MS images, RGB bands are shown.

TABLE I
Quality Measurements of Fused Images for QuickBird Data

		AWLP	GIHS-GA	GSA-CA	LC-GFO	CS1	CS2
CC	B	0.981	**0.983**	0.914	0.885	0.976	0.981
	G	**0.971**	0.969	0.907	0.869	0.964	0.970
	R	0.931	0.931	0.895	0.920	0.931	**0.938**
	NIR	0.926	0.966	0.964	0.921	0.971	**0.972**
RMSE	B	7.711	7.536	**6.655**	10.066	8.821	7.854
	G	**14.118**	14.703	14.557	17.669	15.972	14.459
	R	16.899	16.840	**15.227**	19.058	17.552	16.544
	NIR	34.052	30.961	32.233	34.945	27.090	**26.605**
SAM		2.882	2.619	2.729	3.293	2.606	**2.437**
ERGAS		2.159	2.052	1.900	2.411	1.986	**1.896**
Q4		0.778	0.842	0.829	0.768	0.835	**0.850**

In the quantitative evaluation, the size of the decimated PAN image was 1600×1600 pixels. The quality indicators of the fused images of each method are listed in Table I. The best value of each indicator is labeled in bold. From the table we can see CS2 wins in two bands for CC and one band for RMSE, and gives the lowest ERGAS. CS2 also gives the best SAM and Q4.

B. Fusion Results of IKONOS Data

IKONOS provides PAN and MS images at 1.0 and 4.0 m resolutions, respectively. The datasets were downloaded from the website http://glcf.umiacs.umd.edu/data/ikonos/index. Two datasets were used to train the dictionary.

The original images and the fusion results of four methods are shown in Fig. 2. The size of the HRP image is 400×400 pixels. Fig. 2(e) is blurred and the spectral distortion is obvious. The other three results do not show great differences in spatial details. However, Fig. 2(f) seems to have the lowest spectral distortion.

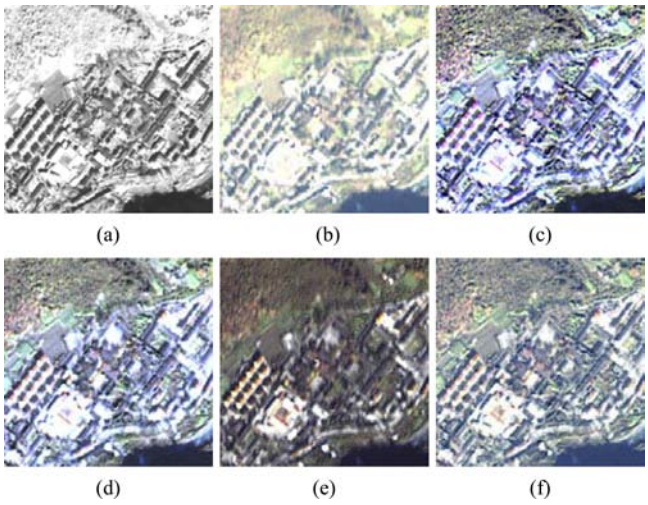

Fig. 2. Original IKONOS images and fusion results using different methods. (a) HRP image (1.0 m, 400×400). (b) LRM images (4.0 m). (c) GIHS-GA fused image. (d) GSA-CA fused image. (e) LC-GFO fused image. (f) CS fused image using trained dictionary. For all MS images, RGB bands are shown.

TABLE II
QUALITY MEASUREMENTS OF FUSED IMAGES FOR IKONOS DATA

		AWLP	GIHS-GA	GSA-CA	LC-GFO	CS1	CS2
CC	B	0.891	**0.935**	0.910	0.873	0.896	0.915
	G	0.867	0.864	0.919	0.848	0.910	**0.921**
	R	0.848	0.878	0.921	0.838	**0.924**	**0.924**
	NIR	0.842	0.904	0.917	0.849	0.922	**0.927**
RMSE	B	18.858	**9.367**	10.844	21.425	12.375	11.757
	G	25.657	25.377	23.475	31.293	**22.046**	23.397
	R	25.462	27.451	24.014	35.462	22.816	**22.785**
	NIR	55.665	57.974	54.524	56.136	54.512	**52.540**
SAM		4.528	5.155	4.495	4.633	**4.218**	4.251
ERGAS		3.981	4.160	3.825	4.532	3.748	**3.685**
Q4		0.646	0.686	0.681	0.641	0.697	**0.699**

In the quantitative evaluation, the size of the decimated PAN image was 1600×1600 pixels. The quality indicators are listed in Table II. CS2 wins in three bands for CC and two bands for RMSE, and also gives the best ERGAS and Q4. CS1 is better than CS2 in SAM but the difference is not significant.

To sum up, for all quality indicators CS2 is comparable with CS1, which suggests that in CS based pansharpening method, the trained dictionary can be used to replace the dictionary composed of raw patches sampled from HRM images. The trained dictionary is derived from the results of AWLP. Though the fusion quality of AWLP is moderate, a following training step incorporates spatial information from the HRP image and improves the representation ability of the dictionary.

V. CONCLUSION

CS technique has been used in pansharpening while it faces the difficulty that the dictionary is generated from the unknown HRM images. In this letter we proposed a dictionary training method to obtain the dictionary from HRP and LRM images. The method generates the initial dictionary from fused HRM images, and then improves the representation ability of the dictionary by training. The training process incorporates information from the HRP image, which improves the ability of the dictionary to describe spatial details. Experimental results using QuickBird and IKONOS data indicate that the trained dictionary can be used to replace the raw-patch-dictionary sampled from HRM images.

The CS approach takes more time than traditional methods. In future work, the time-consuming OMP can be speeded up by parallelization since it is done independently on each patch. Another research direction is to design multiscale dictionaries, which may be used to handle larger image patches. In addition, the linear relationship used in the letter is only an approximate model, and may not be applied to all sensors.

REFERENCES

[1] 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.

[2] 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.

[3] M. Choi, "A new intensity-hue-saturation fusion approach to image fusion with a tradeoff parameter," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 6, pp. 1672–1682, Jun. 2006.

[4] T. Tu, P. Huang, C. Hung, and C. 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.

[5] P. S. Chavez, S. C. Sides, and J. A. Anderson, "Comparison of 3 different methods to merge multiresolution and multispectral data-Landsat TM and SPOT panchromatic," *Photogramm. Eng. Remote Sens.*, vol. 57, no. 3, pp. 295–303, Mar. 1991.

[6] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, and M. Selva, "MTF-tailored multiscale fusion of high-resolution MS and pan imagery," *Photogramm. Eng. Remote Sens.*, vol. 72, no. 5, pp. 591–596, May 2006.

[7] 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.

[8] A. G. Mahyari and M. Yazdi, "Panchromatic and multispectral image fusion based on maximization of both spectral and spatial similarities," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 1976–1985, Jun. 2011.

[9] S. Zheng, W. Shi, H. Liu, and J. Tian, "Remote sensing image fusion using multiscale mapped LS-SVM," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1313–1322, May 2008.

[10] 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.

[11] S. T. 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.

[12] 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.

[13] E. J. Candès, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate information," *Commun. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1233, Aug. 2006.

[14] D. Donoho, "For most large underdetermined systems of linear equations the minimal l_1-norm solution is also the sparsest solution," *Commun. Pure Appl. Math.*, vol. 59, no. 6, pp. 797–829, Jun. 2006.

[15] M. Aharon, "Overcomplete dictionaries for sparse representation of signals," Ph.D. dissertation, Dept. Comput. Sci., Israel Inst. Technol., Haifa, 2006.

[16] A. Garzelli and F. Nencini, "Fusion of panchromatic and multispectral images by genetic algorithms," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Denver, CO, USA, Jul. 31–Aug. 4 2006, pp. 3810–3813.

[17] B. Aiazzi, S. Baronti, F. Lotti, and M. Selva, "A comparison between global and context-adaptive pansharpening of multispectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 2, pp. 302–306, Apr. 2009.

[18] Z. Zhou, S. Peng, B. Wang, Z. Hao, and S. Chen, "An optimized approach for pansharpening very high resolution multispectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 4, pp. 735–739, Jul. 2012.