A Model-Based Method for Pan-Sharpening of Multi-Spectral Images using Sparse Representation

Mohammad Khateri  
Image Processing and Information Analysis Lab.  
Faculty of Electrical and Computer Engineering  
Tarbiat Modares University  
Tehran, Iran  
m.khateri@modares.ac.ir

Hassan Ghassemian  
Image Processing and Information Analysis Lab.  
Faculty of Electrical and Computer Engineering  
Tarbiat Modares University  
Tehran, Iran  
ghasemi@modares.ac.ir

Fardin Mirzapour  
Faculty of Electrical Engineering  
Sadra Institute of Higher Education  
Tehran, Iran  
mirzapour@sadra.ac.ir

Abstract— Pan-sharpening (PS) is fusion of the low-resolution multi-spectral (LRM) image with the corresponding high resolution panchromatic (HRP) one, which aims to reach the high-resolution multi-spectral (HRM) image. Due to the importance of designing well-adapted dictionaries for the pan-sharpening problem as the fundamental challenge in the pan-sharpening methods utilizing sparse representation, we present a novel strategy to tackle this issue. Our method takes the image formation model into account in a patch-based strategy and exploits local spectral-spatial information. To make a fair trade-off between spectral and spatial information of each patch, some local parameters are considered and estimated adaptively. In addition, the energy ratio between LRM/ LRP patches is considered to locally reconstruct the same ratio for the corresponding HRM/ HRP patches. This model-based pan-sharpening strategy led to a closed-form solution, which results in precise HRM dictionary atoms. To obtain the unknown HRM image, after designing the HRM dictionary from input Pan and multi-spectral images, sparse representation over LRM and HRM dictionaries will be conducted in the sparse coding stage. The proposed method has been applied to two different datasets collected by WorldView-3 and GeoEye-1, and then compared with some popular and state-of-the-art methods, qualitatively and quantitatively.

Keywords—fusion, panchromatic, multi-spectral, pan-sharpening, sparse-representation, spectral, spatial, dictionary.

I. INTRODUCTION

Nowadays, remote sensing satellites provide inordinate information about the physical properties of the Earth. Need for high-resolution multi-spectral (HRM) images in different Earth observation applications are increasingly growing, whereas acquiring these images is of high cost due to technological constraints on the design of satellite sensors. Nevertheless, a widely used approach to obtain an HRM image is a cost-effective way is the fusion of the complementary images obtained by different sensors: low-resolution multi-spectral (LRM) image, and high resolution panchromatic (HRP) image. This approach is known as pan-sharpening (PS) in the literature [1-3]. The primary PS algorithms are based on intensity-hue-saturation (IHS) [4, 5], Gram-Schmidt (GS) [6], principal component analysis (PCA) [7]. These methods are based on substitution of HRP image for a component obtained by a transformation on up-sampled LRM image. These algorithms often provide high spatial resolution but fail to keep well spectral information of LRM image. Recently, many advanced versions of the mentioned methods have been presented to alleviate spectral distortion of pan-sharpened image: Shahdoosti and Ghassemian [8] proposed a filter-based method which uses spectral and spatial PCA, method in [9] presented an adaptive approach to create intensity component in PS procedure, authors in [10] introduced a self-learning approach based on self-similarity of structures in natural images in a multi-stage procedure, methods presented in [11, 12] utilized optimization techniques, and recently new approach based on cartoon and texture decomposition [13] has been proposed. Some methods use multi resolution analysis (MRA) which try to inject spatial information of Pan image into up-sampled MS image through the multi-resolution stack. Finding a criterion for controlling the level of injection is a bottleneck of these methods. Some of the well-known MRA tools used in the literature are wavelet [14], curvelet [15, 16], and contourlet [17], which are presented in different frameworks and priors.

In the last decade, several model-based PS approaches have been proposed in which the response of given pixels and data channels have been considered. It’s worth to note that disparate compositions of falling light may give the same response on the panchromatic channel, whereas may be different on the specific spectral channel; so regularization terms is necessary to restrict the solution space [18]. Aanæs, et al. [18] consider spectral consistency between low and high-resolution MS images and presented a pixel neighborhood regularization. Many papers attempt to address PS using this model in different frameworks; e.g., Li and Leung [19] used a constrained least square framework, Joshi and Jalobeanu [20] employed MAP framework with inhomogeneous Gaussian Markov random field (IGMRF) as prior. Recently, Upla, et al. [21] utilized self-learning and Gabor prior alongside IGMRF to produce HRM images from LRM and HRP images. Li and Yang [22] introduced a compressed-sensing (CS) based method employing HRM/LRM patch pairs extracted from external datasets. This method was significantly improved by Jiang, et al. [23] by utilizing a joint learning dictionary from Pan and MS training images. Similarly, Li, et al. [24] used sparse representation but HRM dictionary was constructed from source images, in contrast. An analogous framework exploited by Song, et al. [25] to fuse Thematic Mapper/Enhanced Thematic Mapper Plus (TM/ETM+) and SPOT5 images; detailed surveys on PS methods can be found in [2, 26]. Well-adapted dictionaries
play a decisive role in sparse-based PS methods. Therefore, in this paper, we present an efficient model-based PS method benefiting from sparse representation over well-adapted dictionaries. The proposed PS method takes remote sensing image formation model into account in an all-band patch-based strategy because: a) small patches can easily be formulated – whole image is intricate and non-stationary; b) all-band patches help to control spectral consistency during spatial improvement; c) it can benefit from sparse representation over dictionaries. Moreover, several parameters are used to make a fair trade-off between spectral and spatial information – all parameter will be estimated automatically for each patch. In addition, the energy ratio of LRM/LRP patches is considered to locally reconstruct the same ratio for the corresponding HRM/HRP patches. This model-based PS strategy brings about a closed-form solution which provides precise well-adapted atoms of HRM dictionary. By finding sparse representation of low resolution (LR) patches over LR dictionary and supposing the same representation of high resolution (HR) patches over HR dictionary, the desired HRM patches will be obtained. Finally, HRM image is obtained only by tilling the obtained HRM patches.

II. SPARSE REPRESENTATION

Sparse representation (SR) is one of the most important methodologies of the linear representation methods and has also been proven to be a powerful solution to a variety of image processing fields. SR stems from compressed sensing theory, that states if a signal is sparse (or compressive), the original signal can be recovered from a few measured values, which are much less than previous sampling theory (such as Shannon’s theorem) [27]. Suppose \( x \in \mathbb{R}^n \) is an over-complete dictionary, \( a \in \mathbb{R}^{k} \) is sparse coefficient vector, and each column of \( D \) known as an atom. The sparse coefficient obtained by following minimization problem:

\[
\hat{a} = \arg \min_a \|a\|_0 \quad \text{subject to} \quad \|x - Da\|_2^2 \leq \varepsilon
\]

where \( \|a\|_0 \) indicate the number of non-zero elements and \( \varepsilon \) is pre-defined \( \ell_2 \) norm error. Direct solution of (1) requires a combinatorial search of exponential size by considering all possible supports; so, it is computationally demanding. To tackle this problem, various alternative techniques, such as relaxation is used to approximate the solution of (1) by replacing \( \|a\|_0 \) by \( \|a\|_0 \) as follow:

\[
\hat{a} = \arg \min_a \|a\|_0 \quad \text{subject to} \quad \|x - Da\|_2^2 \leq \varepsilon
\]

For solving (2) many algorithms, such as basis pursuit (BP) have been presented [28]. In (2), \( D \) has important role, which in the literature a comprehensive studies have been carried out around the dictionaries, but last researches showed that trained dictionaries lead to better results than some fixed redundant bases such as undecimated wavelet and discrete cosine transformation [29]; it also has been successfully used in remote sensing image fusion [24, 30, 31].

III. PROPOSED METHOD

A. Image Formation

To reconstruct more reliable HRM image, considering physical characterizations of remote sensing satellite sensors could be beneficial. These characterizations clearly denote how the real scene is captured by the sensors. In this section, the image formation model will be presented for either of multi-spectral and panchromatic. To gain control over different areas of HRM image -spatially and spectrally, simultaneously- and to make use of sparse representation, hereafter, all equation will be written in the patch form.

1) Multi-spectral image formation

Let the LRM patches be arranged in the column vector and denoted by \( y = [y_1, \ldots, y_L]^T \in \mathbb{R}^{L \times k} \), which \( y_i \in \mathbb{R}^k \) represents an extracted patch from \( i \)-th LRM band and \( L \) is the number of MS bands. Likewise, \( X \in \mathbb{R}^{(2k + pn) \times L} \) is an extracted patch from HRP image (which \( \rho \) is resolution ratio between HRP and LRM image, in this letter paper \( \rho = 4 \)), and also point spread function (PSF) is a Gaussian kernel of size \((2k + 1) \times (2k + 1)\). Based on MS image formation procedure, LRM patch is spectrally consistent with the corresponding HRM patch, \( \hat{y} \in \mathbb{R}^{(2k + pn) \times L} \) as below:

\[
y = QY + n_i
\]

where \( Q = MB \in \mathbb{R}^{L \times (2k + pn)} \) is degradation matrix included decimating \( (M) \) and blurring \( (B) \) operators and \( n_i \in \mathbb{R}^{L \times (2k + pn)} \) is considered as zero-mean Gaussian noise. Owing to the filtering process in (3), marginal pixels are severely undergone the undesired effects; this issue is more critical when the PSF filter size and patch size are near to each other. Thanks to the importance of the before mentioned issue, addressing this subject is vital to create more precise HRM patches in PS problem. It will be elaborated in the following section. Assume PSF for \( i \)-th band be as below:

\[
H_i^{\rho} = [h_{n,i}] \quad \text{where} \quad m, n = 1, \ldots, (2k + 1)
\]

The modulation transfer function (MTF) of \( i \)-th band (blurring operator) denoted by \( B_i^{\rho} \in \mathbb{R}^{\rho^{n \times (2k + pn)}} \) and in our patch-based strategy is structured as below:

\[
B_i^{\rho} = \begin{bmatrix}
0, 0, 0, \ldots, 0, 0, 0
\end{bmatrix}_i, \quad i = 1, \ldots, L
\]

where, \( B_i^{\rho} \in \mathbb{R}^{\rho^{(n \times (2k + pn))}} \) are rows of \( H^{\rho} \), \( B_i^{\rho} \) is a vector of size \((\rho n - 1) \times 1\) with all entries equal to zero and \( H_i^{\rho} \) is a vector with zero elements, and also \( B_i^{\rho} \) is a row-cyclic matrix. By considering all spectral bands \( B \in \mathbb{R}^{\rho^{n \times (2k + pn)}} \) is formulated as below:
\[ \mathbf{B} = [\text{diag}(\mathbf{B}^1, \ldots, \mathbf{B}^L)] \]  

\[ \mathbf{B} \text{ is MTF filter of all multi-spectral bands and } M = \rho^{-2}(I_{2k+1}) \otimes (I_{2k+1}^T \otimes (I_{n \times n} \otimes I^T_{2k+1})) \in \mathbb{R}^{(2k+1) \times (2k+1)^2} \text{ is decimation operator, in which } I_{n \times n} \text{ is } n \times n \text{ identity matrix, } \otimes \text{ is Kronecker product and } I \text{ is a } \rho \times 1 \text{ vector that all entries equal to one.} \]

\[ \beta \lambda = - + + - + + \]  

2) Panchromatic image formation

Panchromatic image is wideband and approximately covers all wavelength of multi-spectral bands, therefore, HRP image can be written as a linear combination of HRM bands [24] as below:

\[ \mathbf{X} = \mathbf{W} \mathbf{Y} + c\mathbf{J} + \mathbf{n} \]

\[ \mathbf{W} = (\alpha_l I_{(2k+1)(2k+1)}), \ldots, \alpha_n I_{(2k+1)(2k+1)} \in \mathbb{R}^{(2k+1)^2 \times (2k+1)^2} \]  

and \( \alpha_l(l = 1, \ldots, L) \) denote the contribution of each band involved in (7). Some papers in the literature have utilized normalized response curves to calculate \( \alpha_l \) for whole image [19, 24]; because of relying on equation (7) for either of LRM/LRP and HRM/HRP image pairs, and also to control on different areas spectrally and spatially, in this paper \( \mathbf{W} \) is obtained locally for each patch as below:

\[ \omega = \sum_{i=0}^{\alpha_l} \omega_i \left( \frac{[\mathbf{Y}_l]}{\mathbf{X}_{\text{Down}}^i} \left| \mathbf{Y}_l \right| \right) \]

\[ \phi_0 \left( \frac{[\mathbf{Y}_l]}{\mathbf{X}_{\text{Down}}^i} \right) = \left( \frac{[\mathbf{Y}_l]}{\mathbf{X}_{\text{Down}}^i} \right) \]

Where, \( \omega_{i+d} \) is i-th element of \( \omega_0 \). By using \( \omega \), the energy ratio between LRM/LRP patches leads to locally reconstruct the same ratio for the corresponding HRM/HRP patches. In (8), \( \mathbf{X}_{\text{Down}} \) is LRP patch (down-sampled version of HRP patch), \( c \) is a constant which is taken from (9). \( \mathbf{J} \) is a \((2k+1)^2 \times 1 \) vector that all elements equal to one, and \( \mathbf{n} \) is a zero-mean Gaussian noise.

\[ c = \mathbf{X}_{\text{Down}} - \sum_{i=1}^{\alpha_l} \omega_i \mathbf{Y}_l / n^2 \]

B. Reconstruction of HRM patch

To achieve HRM patch \( \mathbf{Y} \) , (3) and (7) as integral parts involved in PS problem should be satisfied simultaneously, which imply on spectral and spatial consistency. In other words, the following problem should be optimized:

\[ \hat{\mathbf{Y}} = \min \left\{ \left\| \mathbf{Q} \hat{\mathbf{Y}} - \mathbf{y} \right\| + \beta \left\| \mathbf{W} \mathbf{Y} + c - \mathbf{X} \right\|^2 + \lambda \left\| \mathbf{Y} \right\|^2 \} \]

To mitigate the probable singularity issue for some patches, Tikhonov regularization term (\( \lambda \) ) is considered in optimization problem. The \( \beta \) parameter is defined to make trade-off between spectral and spatial terms. The equation (10) is a quadratic function in \( \mathbf{Y} \), upon differentiating and setting to zero, the global minimum reached in a closed-form expression as below:

\[ \hat{\mathbf{Y}} = (\mathbf{Q}^T \mathbf{Q} + \beta \mathbf{W}^T \mathbf{W} + \lambda \mathbf{I})^{-1} (\mathbf{Q}^T \mathbf{y} + \beta \mathbf{W}^T (\mathbf{X} - c)) \]

\[ \beta \text{ parameter in (10) is obtained by following iterative algorithm:} \]

\[ \text{For } i = 1: \text{Iter} \]

\[ \begin{align*}
\beta_i &= \beta_{i-1} \\
\text{End}
\end{align*} \]

(12)

\[ \mathbf{1} \]

(2) \[ \beta_{i+1} = \mathbf{Q} \hat{\mathbf{Y}}_{i-1}^2 - \mathbf{y}^2 + \mathbf{W} \hat{\mathbf{Y}}_{i-1}^2 + c - \mathbf{X} \]

Step 2 in the above iterative algorithm is the ratio of spectral reconstructed error to overall error, included spectral and spatial fidelity term’s errors. The \( \beta \) parameter has a decisive role in keeping spatial and spectral information; some papers [22, 24] did not consider it. In this letter paper we consider spectral-spatial trade-off parameter and proposed an iterative algorithm (12) to calculate it. Therefore, by considering (9) and (12) for a given patch, the equation (11) is utilized to reconstruct HRM patch \( \hat{\mathbf{Y}} \in \mathbb{R}^{(2k+1)^2} \). To achieve more reliable reconstructed pixels patch, the \( k \) bordering pixels are eliminated (these pixels portrayed by gray color in the Fig. 1). Accordingly, reliable HRM (RHRM) patch \( \hat{\mathbf{Y}} \in \mathbb{R}^{(2k+1)^2} \) is obtained.

C. Pan-sharpening from sparsity

Now, all HHRM patches arranged in \( \mathbf{D}_{ms}^{HR} \in \mathbb{R}^{L \times \rho^2 \times N} \), which \( N \) is the number of dictionary patches and \( \mathbf{D}_{ms}^{HR} \) known as the HRM dictionary. According to SR theory, every HRM patch has a sparse representation over a HR redundant dictionary. In the other word, \( \mathbf{Y}_d = \mathbf{D}_{ms}^{HR} \alpha_h \), where \( \alpha_h \) is the sparse coefficients over HRM dictionary and \( \mathbf{Y}_d \) is the desired HRM patch. In this section we aim to recover \( \mathbf{Y}_d \) from HRM dictionary, which it needs to find the sparse representation coefficient \( \alpha_h \) over \( \mathbf{D}_{ms}^{HR} \); to this end, firstly, sparse coefficients \( \alpha_h \) over LRM dictionary will be obtained by solving following optimization problem:

\[ \hat{\mathbf{a}}_i = \arg\min_{\mathbf{a}_i} \left\| \mathbf{a}_i \right\| \text{ subject to } \left\| \mathbf{X}_i \right\| + \left( \mathbf{D}_{ms}^{HR} \right)^{-1} \left( \mathbf{D}_{ms}^{HR} \right)^{-1} \left\| \mathbf{a}_i \right\| \leq \epsilon \]

(13)

Then, by assuming same sparsity pattern for \( \mathbf{D}_{ms}^{HR} \) and \( \mathbf{D}_{ms}^{LR} \) ( \( \alpha_h = \mathbf{a}_i \), hereinafter denoted by \( \mathbf{a} \) ), according to \( \mathbf{Y}_d = \mathbf{D}_{ms}^{LR} \mathbf{a} \), desired HRM patch will be obtained (this scheme drawn in Fig.1). Where, \( \mathbf{D}_{ms}^{LR} \) is HRP dictionary included HRP patches, \( \mathbf{X}_i \in \mathbb{R}^{\rho \times i} \) (central part of \( \mathbf{X} \), which shown by grid area on the HRP in Fig.1), and \( \mathbf{D}_{ms}^{LR} \) is LRM dictionary included LRM patches \( \mathbf{y} \in \mathbb{R}^{\rho \times i} \).
IV. EXPERIMENTAL RESULTS

Two datasets collected by WorldView-3 and GeoEye-1 utilized to evaluate the performance of the proposed fusion method. For WorldView-3 dataset, HRP and LRM images have 0.31-m and 1.24-m resolutions, respectively. For GeoEye-1 dataset, HRP and LRM images have 0.46-m and 1.84-m resolutions, respectively. Resolution ratio between HRP and LRM is four for either of datasets ($\rho = 4$). To assess fusion results Wald synthesis protocol used; to implement this protocol, firstly, the HRP and LRM images degraded, then the PS algorithm applied to these degraded images, and also the original LRM image considered as the reference image. The performance of the proposed fusion method compared with several methods, such as GS (Gram-Schmidt), MIHS (modified intensity-hue-saturation), SFIM (smoothing filter-based intensity modulation) and Li-Yin-Fang. LRM image was up-sampled by ‘bicubic’ interpolation for GS, MIHS and SFIM methods; in the Li-Yin-Fang method, LRM and HRP patches are extracted in size of $2 \times 2$ and $8 \times 8$. In the proposed method, the size of PSF $(2k+1)$ set to 5. As shown in the Fig. 1 patches extracted from LRM bands with the size of $2 \times 2$ and its corresponding HRP image with the size of $12 \times 12$ ($(\rho(2) + 2k) \times (\rho(2) + 2k)$) in a raster scan order. The gray margin shows the k pixels covered by PSF kernel. For Li-Yin-Fang and the proposed methods the overlap areas are 1 and 4 for LRM bands and HRP image; These patches used as dictionary atoms in (13), so they should not be too large. Additionally, equation (11), used to reconstruct HRM atoms, includes the inverse matrix, which is demanding for large patches. We set the size of patches manually to obtain the best results. Tikhonov regularization term ($\lambda$) and initial trade-off parameter ($\beta_0$) set to $10^{-5}$; to solve (13), BP algorithm implemented in SparseLab MATLAB toolbox, provided by Stanford university, used. The quality of the fused images appraised by the root mean square error (RMSE), relative global dimensional synthesis error (ERGAS), correlation coefficient (CC), universal image quality index (UIQI) and spectral angle mapper (SAM); aforementioned assessments elaborately discussed Vivone, et al. [26]. The WorldView-3 dataset as shown in Fig. 2 (c) is comprised of varied areas, such as building, vegetation, soil, water, etc. with diverse spectral information. The fusion results of WorldView-3 dataset depicted in Fig. 2(d-h). Moreover, to assist visual comparison, two areas of Fig. 2 (c-h) enlarged and displayed in Fig.2 (i-n). As shown in Fig. 2 (d) the GS method distorts spectral information of LRM image while retaining spectral information of HRP image well. It can easily be seen in Fig. 2 (j) that boast’s color highly darkened; additionally, vegetation and buildings are not spectrally consistent with the original image. MIH method severely distorts spectral information of source images, furthermore, some halo artifacts appear in edges of harbor areas (Fig. 2 (k)). SFIM method presents pleasant results in comparison to the aforementioned methods, but as Fig. 2 (l) represent, some halos occurred in harbor areas around the boats, and also spectrally lead to dark results. Li-Yin-Fang method somewhat reduces the halo artifacts happened in before mentioned methods but results in spectral distortion in vegetation and building areas (Fig. 2 (m)). In the proposed method (Fig 2 (n)) halo artifacts reduced; furthermore, it presented more consistent result between fused image and source images (see the vegetation and building areas in Fig 2(n)). The GeoEye-1 dataset includes building, road, vegetation, soil, etc. This dataset analogous to WorldView-3 dataset compared with several methods and shown in Fig. 3. To facilitate visual comparison one area restricted by the blue rectangle, magnified and shown in Fig. 3 (i-n). Fig. 3 (j) and Fig. 3 (k) illustrate that the GS and MIHS seriously distort
spectral information of source images. The SFIM led to dark results (white and red buildings, and also vegetation area Fig.3 (l)). The Li method presents more pleasant results in comparison with the aforementioned methods while some areas such as pink and red buildings shown in Fig. 3 (m) are The proposed method, further, it led to more consistent result with reference image (Fig. 3 (i)). Additionally, the superiority of the proposed method has been proved by the most well-known assessment metrics. Table 1 shows the quantitative results of WorldView-3 and GeoEye-1 datasets; the best results for each assessment index is labeled in bold. As can be seen from Table 1, all metrics including RMSE, ERGAS, CC, UIQI and SAM in the proposed method gained better results than its comparators in GeoEye-1 dataset, moreover, the proposed method yields better results in RMSE, ERGAS, CC and UIQI for WorldView-3 dataset (although the proposed method did not gain a better result in the SAM index, its result is near to the best one). The superiority in RMSE, ERGAS, CC and UIQI metrics clearly suggest that the proposed method provides more consistent fused products with the reference image.

V. C ONCLUSIONS

We proposed an efficient PS method based on image formation model which benefits from sparse Representation; panchromatic and multi-spectral image formation models were exploited to reach HRM image in a patch-based strategy. In the proposed procedure, some adaptable parameters were considered, which made good trade-off between local spatial and spectral information to create HRM
patches to be used as dictionary atoms. These atoms obtained by a closed-form formula, in which all parameter estimated automatically. Moreover, the energy ratio between LRM/LRP patches was considered to locally reconstruct the same ratio for the corresponding HRM/HRP patches, which in turn give rise to consistent output with the reference images. After designing the HRM dictionary from input Pan and MS images, sparse representation over LRMS and HRM dictionaries conducted in the sparse coding stage to obtain the unknown HRM image. The experimental results showed the superiority of the proposed method with respect to the comparators, visually and quantitatively.

REFERENCE