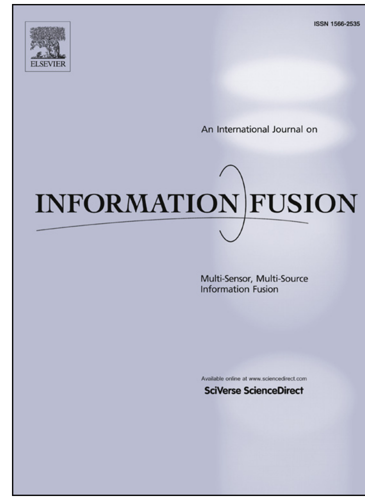
Accepted Manuscript



Spatio-Spectral Fusion of Satellite Images based on Dictionary-pair Learning

Huihui Song, Bo Huang, Kaihua Zhang, Hankui Zhang

PII: S1566-2535(13)00085-7
DOI: http://dx.doi.org/10.1016/j.inffus.2013.08.005
Reference: INFFUS 593

To appear in: *Information Fusion*

Received Date: 30 July 2012
Revised Date: 30 July 2013
Accepted Date: 19 August 2013

Please cite this article as: H. Song, B. Huang, K. Zhang, H. Zhang, Spatio-Spectral Fusion of Satellite Images based on Dictionary-pair Learning, *Information Fusion* (2013), doi: http://dx.doi.org/10.1016/j.inffus.2013.08.005

Spatio-Spectral Fusion of Satellite Images based on Dictionary-pair Learning

Huihui Song[a], Bo Huang[a*], Kaihua Zhang[b], Hankui Zhang[a]

[a] *Department of Geography and Resource Management, Chinese University of Hong Kong, Hong Kong, China*

[b] *Department of Computing, The Hong Kong Polytechnic University, Hong Kong, China*

[*]Corresponding author. Tel./fax: +852 39436536.

E-mail addresses: hhsongsherry@gmail.com (H. Song), bohuang@cuhk.edu.hk (B. Huang), zhkhua@gmail.com (K. Zhang), zhanghankui@gmail.com (H. Zhang)

Abstract- This paper proposes a novel spatial and spectral fusion method for satellite multispectral and hyperspectral (or high-spectral) images based on dictionary-pair learning. By combining the spectral information from sensors with low spatial resolution but high spectral resolution (LSHS) and the spatial information from sensors with high spatial resolution but low spectral resolution (HSLS), this method aims to generate fused data with both high spatial and spectral resolution. Based on the sparse non-negative matrix factorization technique, this method first extracts spectral bases of LSHS and HSLS images by making full use of the rich spectral information in LSHS data. The spectral bases of these two categories data then formulate a dictionary-pair due to their correspondence in representing each pixel spectra of LSHS data and HSLS data, respectively. Subsequently, the LSHS image is spatial unmixed by representing the HSLS image with respect to the corresponding learned dictionary to derive its representation coefficients. Combining the spectral bases of LSHS data and the representation coefficients of HSLS data, fused data are finally derived which are characterized by the spectral resolution of LSHS data and the spatial resolution of HSLS data. The experiments are carried out by comparing the proposed method with two representative methods on both simulation data and actual satellite images, including the fusion of Landsat/ETM+ and Aqua/MODIS data and the fusion of EO-1/Hyperion and SPOT5/HRG multispectral images. By visually comparing the fusion results and quantitatively evaluating them in term of several measurement indices, it can be concluded that the proposed method is effective in preserving both the spectral information and spatial details and performs better than the comparison approaches.

Keywords: spatio-spectral fusion; high spatial resolution; high spectral resolution; dictionary-pair learning; sparse non-negative matrix factorization

1. Introduction

A specific feature of remote sensing images, captured from different satellites and different sensors, is a tradeoff between spatial resolution and spectral resolution. This is caused, on the one hand, by the system tradeoffs related to data volume and signal-to-noise ratio (SNR) limitations and, on the other hand, by the specific requirements of different applications for a high spatial resolution or a high spectral resolution. For example, to fulfill the high spatial resolution requirement in many land-oriented applications, sensors with spatial resolution of half meter to tens of meters are designed, including, but not limited to, ETM+ (30m) on a platform of Landsat, the sensor on QuickBird (2.4m for multispectral bands), and the instruments on SPOT (2.5m~10m). Sensors like the MODerate resolution Imaging Spectroradiometer (MODIS) on-board Aqua or Terra, the MEdium Resolution Imaging Spectrometer (MERIS) on-board ENVISAT, and the Hyperion instrument on-board the EO-1 satellite, can provide remote sensing images with high spectral resolution but with low spatial resolution. Although these remote sensing instruments have many different properties, such as the revisit period, the swath width, the purpose of the launch (commercial or science), a specific user can obtain large amounts of data from different instruments on a given study area. This promotes the development of new algorithms to obtain remote sensing data with the best resolution available by merging their complementary information.

From the application point of view, remote sensing data with high spatial resolution are beneficial for the interpretation of satellite images and the extraction of spatial details of land cover, such as in land use/cover mapping and change detection [1]; whereas remote sensing data with high spectral resolution are capable of identifying those targets that cannot be easily distinguished by human eyes, such as in geological analysis and chemical contamination analysis [2]. To merge panchromatic (with a higher spatial resolution) and multispectral (with a lower spatial resolution) images from the same or separate sensors, pansharpening algorithms have been extensively studied in the past two decades [3, 4]. In this paper, we focus on the fusion of images from two categories of sensors: one category has low spatial resolution and high spectral resolution (hereinafter abbreviated as LSHS), such as Aqua (or Terra)/MODIS and EO-1/Hyperion; while the other category has high spatial resolution but low spectral resolution (hereinafter abbreviated as HSLS), the data from which are usually termed as multispectral images, such as Landsat/ETM+ and SPOT5/HRG. The low resolution or high resolution herein is in a relative sense. Through integrating the spectral information of LSHS data and the spatial information of HSLS data, we expect to extend the applications of available satellite images, thereby meeting various demands of data users.

To address this spatial and spectral fusion problem, one category of classic method is spatial-unmixing-based algorithms [5-9]. The processing steps of these methods are: (1) geometrical co-registration between the HSLS and LSHS images; (2) the multispectral classification on the high spatial resolution image (HSLS) to unmix the low spatial resolution image (LSHS); and (3) determination of the class spectra via regularized linear unmixing. These methods showed good performance at fusing Landsat images with MERIS [6, 7, 9] images or with ASTER [8] images for land applications. However, it should be noted that whether pure spectra exist resides in the sensor's spatial resolution of the given HSLS data. These methods also have high demanding on the geometric registration accuracy of the given data (e.g., less than 0.1-0.2 of the low resolution pixel size according to reference [5]). The authors in [10-12] proposed to fuse multispectral and hyperspectral images in a maximum a posteriori (MAP) estimation framework by establishing an observation model between the desired image and the known image. The method reported in [10] employed a spatially varying statistical model to help exploit the correlations between multispectral and hyperspectral images. The methods developed in [11, 12] made use of a stochastic mixing model of the underlying scene content to enhance the spatial resolution of hyperspectral image. The authors in [13] proposed to improve the spatial resolution of hyperspectral image by fusing with high resolution panchromatic band images based on super-resolution technique. Firstly, this method learns the spatial information from panchromatic images by sparse representation; then the high-resolution hyperspectral image is constructed based on the learned spatial structures with a spectral regularization.

Due to the low spatial resolution or the existence of homogenous mixtures in the hyperspectral images, unmixing techniques are needed to decompose the mixed pixels observed into a set of constituents and the corresponding fractional coefficients, which denote the proportion of each constituent [14]. The first step in this spectral unmixing task is to collect a suitable set of endmembers to model the spectra of measured pixels by weighting these spectral signatures. There are two categories of endmembers according to different extraction methods: image endmember derived directly from the images and library endmember derived from known target materials in field or laboratory spectra [15]. However, employing library endmembers is risky because it's difficult to ensure that these spectra are captured under the same physical conditions as the observed data. Whereas image endmembers can avoid this problem due to the collection at the same scale as the observed data, thereby being linked to the scene features more easily [14]. A number of endmember extraction algorithms for hyperspectral data were quantitatively compared in [15]. In this paper, we employ this similar endmember extraction strategy for both HSLS and LSHS data and term the extracted spectral bases as dictionaries. For the second step of abundances estimation in hyperspectral unmixing, a popular and effective method is sparse unmixing [16]. The basic principle of this method is based on the observation that only a small portion of endmembers participated in the formation of each mixed pixel. Therefore, with the sparsity-inducing regularizers, i.e., the constraint of a few non-zero components for the abundance vectors, the abundances can be estimated by calling for linear sparse regression techniques [16]. In this paper, we adopt this similar sparse regularization when solving the representation coefficients of HSLS and LSHS data with respect to their representation atoms (or dictionaries).

In this paper, we seek to extract a dictionary-pair for representing LSHS and HSLS data, respectively. Specifically, the representation atoms of LSHS and HSLS data are firstly extracted from the given images, respectively, and then to form the dictionary-pair. Each representation atom of the dictionary-pair herein is in correspondence. Accordingly, each pixel spectra of LSHS and HSLS data can be expressed as a linear combination of their corresponding dictionary atoms, which have the same functions as endmembers in hyperspectral unmixing. Based on this dictionary-pair, the proposed spatio-spectral fusion algorithm consists of two stages: in the first stage, the good spectral properties of the LSHS image are employed to extract the basis functions of spectra (representation atoms) and further to form the dictionary-pair by enforcing the same representation coefficients of the HSLS and the LSHS images with respect to their dictionaries; in the second stage, the good spatial properties of the HSLS image are utilized to derive the representation coefficients with respect to its dictionary. The representation coefficients herein have similar functions as abundances in hyperspectral unmixing but provide spatial location properties due to the high spatial resolution of the HSLS image. Finally, the desired high spatial and high spectral resolution image can be obtained by the multiplication of representation atoms for the LSHS image (i.e., the LSHS dictionary) and representation coefficients for the HSLS image.

The following section presents the theoretical basis of this paper. Section 3 describes the proposed method for the fusion of HSLS and LSHS data. Section 4 shows the experimental validation of the proposed algorithm through comparison with two representative algorithms on both simulated and actual satellite datasets. Finally, we conclude this paper with a discussion on the application of the proposed method and remarks about its inherent features.

2. Theoretical Basis

As introduced in Section 1, a dictionary-pair needs to be trained from the HSLS and LSHS data. Hence, the basic principles of dictionary-pair learning will first be introduced. Taking the non-negative properties of the bases spectra and fractional abundances of HSLS and LSHS data into account, we learn the required dictionary-pair by using sparse non-negative matrix factorization method, which will be presented in the second part of this section.

2.1 Dictionary-pair learning

Huge amounts of information captured by human eyes or satellite sensors are superfluous in a large part caused by the related signals in real world and the oversampling of sensors. Compared to the observed signals, the underlying generating signals actually have very small dimensions. Identifying the generating signals from the relevant information is, in fact, a dimension reduction process or finding the subspace where the data lie [17]. The combination of these generating signals or bases of subspace constitutes the dictionary, each component of which is a representation atom for the observed dataset. Given the observed data, building the dictionary to provide efficient representations for the given signals, is the dictionary learning process, which has greatly promoted the development of novel dictionary learning algorithms [17-19].

For representation convenience, we will use the superscript of symbols to discriminate variables and the subscript to denote the order of elements in this paper. Suppose for a given signal $x^1 \in R^{B\times 1}$, we have determined an overcomplete dictionary, which is denoted as $D^1 \in R^{B\times K}$ (K>B) and its atoms as d_k^1, k = 1, 2,…, K, where K is the dictionary's size. Usually, the dictionary atoms are set as unit norm functions in the signal representation space. Then, these dictionary atoms can be utilized to represent each signal x^1 in this subspace via a linear combination, i.e.,

$$x^1 = D^1\alpha^1 = \sum_{k=1}^{K} \alpha_k^1 d_k^1 \tag{1}$$

where α_k^1 is the weighting coefficient for atom d_k^1 and α^1 is termed as the representation coefficient for x^1 in this paper. Since the number of dictionary atoms needed to represent one signal is usually very small compared to the total number of atoms, the sparsity constraint for representation coefficients comes into play, which has been proved efficient as regularization [17-19]. To this end, the representation coefficient vector α^1 is restricted to hold few non-zero components, which is achieved via the following formula,

$$\min_{\alpha^1} \left\|\alpha^1\right\|_0, \quad s.t. \ \left\|x^1 - D^1\alpha^1\right\|_2^2 \le \varepsilon \tag{2}$$

where $\| \|_0$ denotes the number of non-zero elements and ε is an approximation error. To solve the sparse vector α^1 in (2), various sparse coding algorithms have been developed. One category of these algorithms is greedy algorithms such as the orthogonal matching pursuit (OMP) [20], which picks the most appropriate local atoms iteratively. Another category of sparse coding methods rely on convex relaxation. Among them, the representative one is the least absolute shrinkage and selection operator (LASSO) [21], which replaces the non-convex l_0 norm in (2) by a convex l_1 norm as follows:

$$\min_{\alpha^1} \left(\left\|x^1 - D^1\alpha^1\right\|_2^2 + \lambda\left\|\alpha^1\right\|_1 \right) \tag{3}$$

where λ is a regularization parameter. We recommend the interested readers on sparse coding algorithms to refer to the recently published review [22].

Suppose there exists another subspace expanded by atoms of $D^2 \in R^{b\times K}$ (b<K), which has a different dimension with D^1. If we want to establish correspondence between the signals in these two subspaces, a simple and intuitive approach is to build correspondence between their representation atoms and enforce the same representation coefficients for each signal pair.

Suppose the corresponding signal of x^1 is $x^2 \in R^{b\times 1}$ in the subspace expanded by atoms of D^2, then we can represent x^2 by atoms of dictionary D^2 with the same representation coefficient as x^1 in (1). Denote the atoms of D^2 as d_k^2, $k = 1, 2,\ldots, K$, then

$$x^2 = D^2\alpha^1 = \sum_{k=1}^{K}\alpha_k^1 d_k^2 \tag{4}$$

The two dictionaries D^1 and D^2 in the above case are defined as a dictionary-pair, the representation atoms of which are in correspondence and can generate signal pairs by restricting the same representation coefficients. The application of dictionary-pair learning has been proved effective in image super-resolution [23, 24].

The key task in dictionary learning approaches is to train an appropriate dictionary from the given training samples. There are three main categories of these learning algorithms: (1) probabilistic methods; (2) clustering-based methods; and (3) learning methods with specific structures. For a detailed description of these algorithms, we recommend the interested readers to see [17, 25]. For one given training set $X^1 = [x_1^1, x_2^1,\ldots, x_N^1]$ ($N>>K$), we can choose one dictionary learning method to best fit the given task. Under the application context of building spectral correspondence (i.e., the spectral atoms for one material captured by different sensors) between two signal spaces, the chosen dictionary learning method needs to be extended to two dictionaries (i.e., the dictionary-pair). Suppose the other training set is $X^2 = [x_1^2, x_2^2,\ldots, x_N^2]$, whose signals are in correspondence with those of dataset X^1. Two strategies have been proposed to generate the dictionary-pair from the training set pair X^1- X^2: in the first strategy, the two training sets are concatenated to $[X^1; X^2]$, then, two dictionaries $[D^1; D^2]$ with normalization are simultaneously learned from them by enforcing training sample pairs in X^1 and X^2 to have the same representation coefficients with respect to dictionaries D^1 and D^2, respectively [23]; in the second strategy, the training set with the higher amount of information, here we suppose it is X^1, is firstly trained to obtain its dictionary D^1 and then the dictionary D^2 for the training set X^2 is obtained by the multiplication of X^2 and Pseudo-Inverse of the derived representation coefficients from X^1 [24]. In this paper, we adopt an alternative update method by firstly updating the spectral bases of the LSHS image due to its affluent spectral information, and then updating the spectral bases of the HSLS image with the same representation coefficients derived from the LSHS image.

2.2 Sparse Non-negative Matrix Factorization (NMF)

As one matrix factorization technique, non-negative matrix factorization (NMF) seeks to determine a parts-based representation [26]. Originally, this method was applied to learn facial features and textual semantics. Given a non-negative data matrix $X \in R^{B\times N}$, NMF seeks to decompose it into the multiplication of non-negative matrix $D \in R^{B\times K}$ and $A \in R^{K\times N}$, i.e., $X \approx DA$, where each column of D is called representation basis, each column of A is an encoding for the corresponding data vector in X and K is the number of representation basis. The non-negative constraint, hereof, is to be compatible with the physical properties in practical applications (e.g., the quantities involved cannot be negative) [27]. To solve the matrices D and A, the most widely used optimization method is to minimize the approximation error between X and DA based on square error (Euclidean distance), i.e.,

$$E(D,A) = \|X - DA\|^2 = \sum_{i,j}\left(X_{ij} - (DA)_{ij}\right)^2 \tag{5}$$

The objective function (5) can be solved by a gradient algorithm or a multiplicative algorithm [27]. The standard multiplicative update rule in element-wise is as follows:

$$D_{ij} \leftarrow D_{ij}\frac{\left(XA^T\right)_{ij}}{\left(DAA^T\right)_{ij}}, \quad A_{ij} \leftarrow A_{ij}\frac{\left(D^T X\right)_{ij}}{\left(D^T DA\right)_{ij}} \tag{6}$$

One feature of NMF is that it usually generates a sparse representation for the given data. However, such sparseness cannot be controlled by NMF in a direct manner. Concerning this disadvantage of NMF, Hoyer [27] proposed to extend NMF to include the option of controlling sparseness explicitly by introducing a sparseness term via incorporating the L_1 norm and the L_2 norm:

$$sparseness(x) = \frac{\sqrt{B} - \left(\sum|x_i|\right)\Big/\sqrt{\sum(x_i)^2}}{\sqrt{B} - 1} \tag{7}$$

where B is the dimensionality of x. It should be noted that B hereof is strictly larger than 1 and thus this method would not work on panchromatic (single band) imagery. The measurement in (7) equals to one when x includes only one non-zero element, and takes the value of zero when all elements are equal. This sparseness constraint can be applied to both basis matrix D and coefficient matrix A depending on the specific applications in question. Thus, the objective function in (5) can be solved under the following optional constraints:

$$\begin{aligned} sparseness(d_i) &= S_D, \quad \forall i \\ sparseness(\alpha_i) &= S_A, \quad \forall i \end{aligned} \tag{8}$$

where d_i is the i-th column of D and α_i is the i-th column of A. S_D and S_A denote the desired sparseness of D and A, respectively. For the matrix with sparseness constraint, it can be solved by a projected gradient descent algorithm devised by Hoyer in [27].

3. Proposed Methodology

Based on the theoretical bases introduced in Section 2, this section presents the proposed method based on dictionary-pair learning and sparse NMF. Given two satellite datasets with LSHS and HSLS, respectively, the purpose of this paper is to combine their complementary information to obtain dataset with the spectral information of LSHS dataset and the spatial information of HSLS dataset (we abbreviate hereafter the desired dataset with high spatial resolution and high spectral resolution as HSHS). To achieve this, the two stages of the proposed method, i.e., the dictionary-pair learning stage from the given two datasets and the spatio-spectral fusion stage based on the projected gradient descent algorithm will be presented respectively. In the first stage, the spectral bases of LSHS and HSLS images are extracted by making use of the rich spectral information in the LSHS image; in the second stage, the desired HSHS image is predicted based on the spectral bases of the LSHS image and the spatial unmixing of the HSLS image. An overall graphical flow of the proposed methodology is shown in Fig. 1.

For the given LSHS and HSLS images from different remote sensors, we assume that they are captured on the same day and cover the same area of the earth surface. For convenience of description, the involved 3-dimensional data cube (i.e., the multi-spectral images) is converted to 2-dimensional matrices, each column of which denotes a pixel in all spectral bands. Suppose the input LSHS image is $X^{lh} \in R^{w \times h \times B}$, where w, h and B denote the image width, the image height and the number of bands, respectively, then it can be converted to $X^{lh} \in R^{B \times n}$, where $n = w \times h$ is the total number of pixels in the LSHS dataset. Similarly, the input HSLS data $X^{hl} \in R^{W \times H \times b}$ can be converted to $X^{hl} \in R^{b \times N}$, where b ($b$<$B$) is the number of spectral bands in HSLS dataset

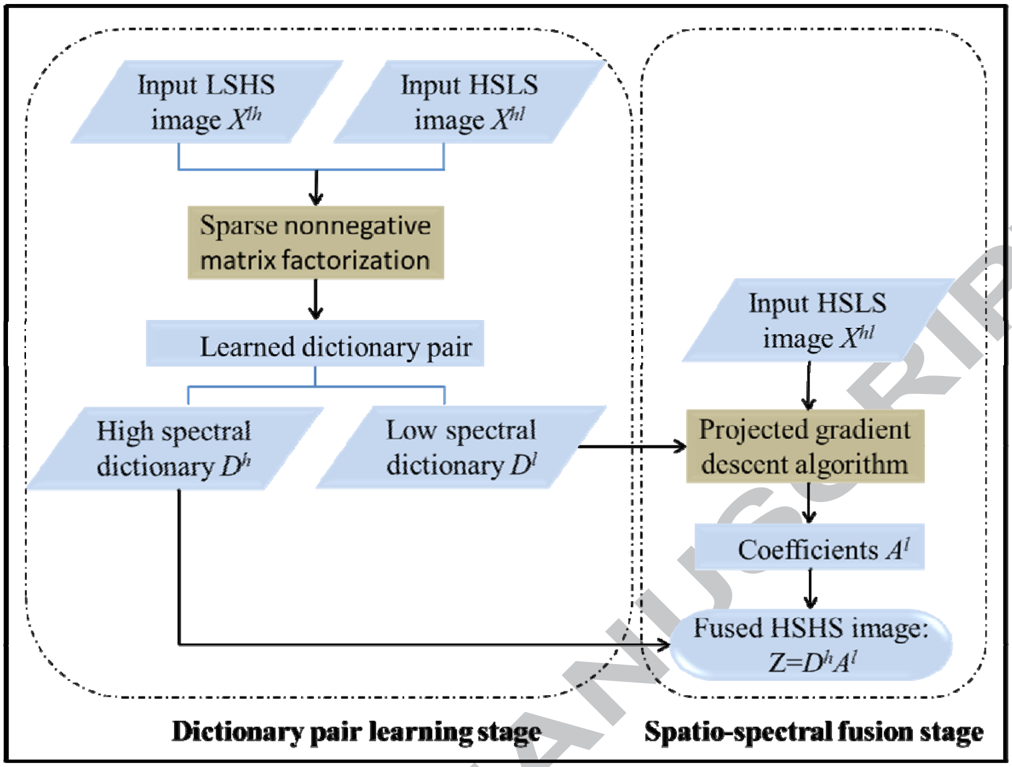


Fig.1. An overall graphical flow of the proposed methodology.

and $N = W \times H$ ($W>w$, $H>h$, $N>n$) is the total number of pixels, and the desired HSHS data can be denoted as $X^{hh} \in R^{B \times N}$. In order to build the spatial correspondence between pixels of LSHS and HSLS images we up-sample the LSHS image to have the same width and height size as HSLS using bicubic interpolation method. Accordingly, the resized LSHS image is denoted as $X^{lh} \in R^{B \times N}$.

3.1 Dictionary-pair learning

Given the input LSHS data $X^{lh} \in R^{B \times N}$ and HSLS data $X^{hl} \in R^{b \times N}$, we aim to determine two dictionaries to represent the spectral bases of X^{lh} and X^{hl}, respectively. We should note that the spectral bases hereof are not necessarily the pure signatures as defined in [28] in the sense that they are spectrally distinct signal sources, which may contain identified or unidentified image endmembers, anomalies and natural signatures [28]. This facilitates the consideration of land-cover variability induced by noise and nonhomogeneous substances. From the perspective view of signal generating, the mixed pixels in X^{lh} and X^{hl} are produced by the linear combination of a small number of spectral bases in their respective dictionaries. The spectral bases of LSHS and HSLS data are deemed as different due to the following two reasons: their different spectral resolutions and the potential radiance difference caused by different illumination, atmospheric or sensor conditions. However, we can build correspondence between these two categories of spectral bases based on the assumption that the LSHS and HSLS images are consistent in spatial location and capture time. In previous literature [23, 24, 29], dictionary-pair learning is utilized to correlate two subspaces representing spatial features. In this paper, we employ the dictionary-pair learning technique to establish the correspondence between two subspaces representing spectral properties.

Taking the high spectral resolution of the LSHS image (usually much higher than the HSLS image) into account, we first determine the spectral bases with high spectral resolution by decomposing the LSHS image into a spectral dictionary D^h and the corresponding representation coefficients. Because only a small portion of spectral bases participates in each mixed pixel [16], the sparseness constraint is imposed on the representation coefficients for each mixed pixel. On the other hand, we consider the fact that the involved physical quantities in this research, such as radiance and reflectance, are nonnegative. Therefore, the dictionary D^h is learned from the input LSHS data $X^{lh} \in R^{B\times N}$ by utilizing the sparse nonnegative matrix factorization method proposed in [27]. With an appropriate sparseness parameter for the representation coefficient, the objective function is as follows:

$$\begin{gathered}\left(D^h, A\right) = \underset{D^h, A}{\arg\min}\left\{\left\|X^{lh} - D^h A\right\|^2\right\} \\ s.t., \quad d_i > 0, \alpha_i > 0, sparseness\left(\alpha_i\right) = S_A, \quad \forall i\end{gathered} \tag{9}$$

where d_i is the i-th column of D_h, α_i is the i-th column of A and S_A is the sparseness parameter for representation coefficients. Suppose the number of spectral bases is K, then $D^h \in R^{B\times K}$ and $A \in R^{K\times N}$. K can be determined by the VD method proposed in [30], which was developed for identifying the number of spectrally distinct signal sources in hyperspectral data. According to the dictionary-pair learning method in [24], we can directly derive the spectral dictionary for the

HSLS image $X^{hl} \in R^{b \times N}$ via a Pseudo-Inverse expression (suppose that A has full row rank):

$$D^l = X^{hl} A^+ = X^{hl} A^T \left(A A^T \right)^{-1} \tag{10}$$

where A is the same representation coefficient matrix derived from (9) and $D^l \in R^{b \times K}$.

However, equation (10) cannot guarantee the non-negativity of D^l. We therefore adopt an alternate update mode to solve D^h, D^l, and A by minimizing the following formulation,

$$\begin{aligned} \left(D^h, D^l, A \right) &= \underset{D^h, D^l, A}{\arg\min} \left\{ \left\| X^{lh} - D^h A \right\|^2 + \left\| X^{hl} - D^l A \right\|^2 \right\} \\ s.t., \quad & d_i^h > 0, d_i^l > 0, \alpha_i > 0, sparseness\left(\alpha_i \right) = S_A, \quad \forall i \end{aligned} \tag{11}$$

where d_i^h is the i-th column of D^h and d_i^l is the i-th column of D^l. (11) can still be optimized by utilizing the sparse nonnegative matrix factorization method proposed in [27]. Although the sparse NMF method can constrain the scale of both dictionary atoms and representation coefficient vectors, to be consistent with the dictionary learning methods as in [17-19], we fix the L_2 norm of dictionary atoms to unity and make the scale of representation coefficients free, i.e., the norm of representation coefficients varies with the signals represented. The specific dictionary-pair training process based on the method in [27] is illustrated in Fig. 2. To assure that the spectral bases are extracted accurately, we first update the spectral bases of D^h in algorithm 1 by making full use of the rich spectral information in the LSHS image. For the updating process of representation coefficients, see reference [27] for more details. Due to the spatial correspondence of training samples in X^{lh} and X^{hl} and the enforcement of the same sparse representation coefficients for them, the spectral bases in D^h and D^l are expected to be in correspondence. That means if one spectral basis in D^h denotes the spectral property of one specific material in the LSHS sensor, then the corresponding spectral basis in D^l reflects the spectral property of the same material in the HSLS sensor. The representation accuracy of these spectral bases is determined by the number of training samples, the co-registration accuracy between the LSHS and HSLS images, and the setting for the relevant parameters including the dimensionality of the dictionary-pair and the sparseness degree for the representation coefficients.

3.2 Spatio-spectral fusion

Based on the learned dictionary-pair, we can derive the desired HSHS image by integrating the spatial information of the HSLS image into the LSHS image. Firstly, each pixel vector x^{hl} of the HSLS image X^{hl} can be sparsely represented as a linear combination of spectral bases in D^l. We solved these representation coefficients by utilizing the projected gradient descent algorithm proposed in [27] via optimizing the following function with non-negativity and sparseness constraints,

$$\begin{aligned} &\alpha^l = \arg\min_{\alpha} \left\| x^{hl} - D^l \alpha \right\|_2^2, \\ &s.t., \quad \alpha_i \geq 0, \quad sparseness(\alpha_i) = S_A, \quad \forall i \end{aligned} \tag{12}$$

where the sparseness parameter S_A was set as the same value in Eq. (11) to keep the sparse consistence in training and fusion stages. Combining the resolved coefficient vectors for all pixels, the representation coefficient matrix A^l for the HSLS image is derived from (12). With A^l and the high spectral resolution bases in D^h, the desired HSHS image X^{hh} can be obtained as follows:

$$X^{hh} = D^h A^l \tag{13}$$

Due to the low spatial resolution of X^{lh} and the high spatial resolution of X^{hl}, this fusion procedure can also be deemed as the spatial unmixing of X^{hl} for X^{lh}. On the other hand, because of the high spectral resolution of X^{lh} and the low spectral resolution of X^{hl}, this fusion procedure can be deemed as the spectral unmixing of X^{lh} for X^{hl}. In a unified point of view, the detailed spatial information of X^{hh} is endowed with the HSLS image and the spectral characteristic of X^{hh} is endowed with the LSHS image.

Algorithm 1. Training dictionary-pair

Input:

· LSHS dataset $X^{lh} \in R^{B\times N}$ and HSLS dataset $X^{hl} \in R^{b\times N}$;

· The dictionary dimensionality K and the sparseness degree S_A;

· The maximum iteration number J.

Initialize:

· $D^h \leftarrow \{d_1^h, d_2^h, ..., d_K^h\}$, $D^l \leftarrow \{d_1^l, d_2^l, ..., d_K^l\}$, where d_k^h and d_k^l are randomly sampled unity vectors from X^{lh} and X^{hl}, respectively.

· j=0; Calculate the initializing error of the objective function according to (11) as E_{old}.

While $j<J$:

1. Fix D^h and D^l, update each column of sparse coefficients A with the sparseness degree S_A; during updating process, calculate E_{new} and guarantee that $E_{\text{new}}< E_{\text{old}}$;
2. Fix D^l and A, update D^h by the standard NMF multiplicative algorithm as in (6); now the objective function is reduced to $\arg\min_{\{D^h\}}\{\| X^{lh} - D^h A \|_2^2\}, \quad s.t., d_i^h > 0$ during updating process, calculate E_{new} and guarantee that $E_{\text{new}}< E_{\text{old}}$;
3. Fix D^h and A, update D^l by the standard NMF multiplicative algorithm as in (6); now the objective function is reduced to $\arg\min_{\{D^l\}}\{\| X^{hl} - D^l A \|_2^2\}, \quad s.t., d_i^l > 0$ during updating process, calculate E_{new} and guarantee that $E_{\text{new}}< E_{\text{old}}$;

Output:

· High spectral resolution dictionary D^h and low spectral resolution dictionary D^l.

Fig. 2. Dictionary-pair learning process

4. Experimental results and Comparisons

In this section, we apply the proposed algorithm to both simulated data and actual satellite data and compare it with the spatial unmixing method proposed in [7] and the sparse representation with spectral regularization method proposed in [13]. For the actual satellite images, we take the Landsat-7/ETM+ reflectance as the HSLS image and Aqua/ MODIS reflectance as the LSHS image in their VIR (visual and infrared) spectrum to obtain the fused image, which is characterized by the spatial resolution of ETM+ image and the spectral properties of MODIS image. Then, we combine the hyperspectral EO-1/Hyperion image and the multispectral (MS) SPOT5/HRG image to obtain fused data with the spectral resolution of the Hyperion image and the spatial resolution of the SPOT5 MS image. For description convenience, we term the proposed dictionary-pair learning method as DPLM, the sparse representation with spectral regularization method in [13] as SRSR and the spatial unmixing method in [7] as SUM in this section. For SRSR, we learned the spatial structures from all HSLS image bands and used the same training parameters as set in [13]; for SUM, the HSLS image was classified into M classes by means of an ISODATA algorithm and the sliding window size was set appropriately according to the relevant class number. Since the function of class number for spectral signatures in SUM and SRSR is the same as the number of dictionary atoms in DPLM, their values are set to the same in all experiments by means of VD method [30].

4.1 Quality evaluation measurements for image fusion results

For the simulated data, the fusion results are visually compared with the ground truth images and are quantitatively evaluated in terms of Root Mean Square Error (RMSE), the erreur relative global adimensionnelle de synthèse (ERGAS) [31], the spectral angle mapper (SAM) [32] in degrees and the structural similarity (SSIM) index [33]. For SSIM index, the values closer to 1 indicate good preservation of the spatial details. For other measurements, lower values indicate better fusion performance in preserving radiometric or spectral properties.

Regarding the satellite images, because there are no ground truth images, we adopted the evaluation methods devised for assessing the fusion results of multispectral and panchromatic data without reference [34]. To evaluate the spectral distortion, the correlation coefficients (CC) between fused HSHS and original LSHS images are calculated in each band. Similarly, a spatial quality index is to calculate the spatial correlation coefficients (CC) between the spatial details of the fused bands and those of the HSLS image bands. Such details are extracted by means of a Laplacian filter and the calculation of spatial CC is executed on the corresponding bands of LSHS and HSLS images in spectrum. For both spatial CC and spectral CC, good fusion performance is indicated when their values are close to 1.

For satellite images, we also employed the spectral distortion index, the spatial distortion index and the jointly spectral and spatial quality index proposed in [34], which are denoted as I_λ, I_s and QNR, respectively. The spectral distortion index is defined by a statistical similarity measurement between any couple of LSHS bands calculated before and after fusion. Analogously, the spatial distortion index is defined by the same statistical similarity measurement between each LSHS band and the intensity band of HSLS calculated before and after fusion. The intensity band of the HSLS image hereof is calculated by averaging all the bands. I_λ (or I_s) is equal to zero when there are no spectral (or spatial) distortions during the image transformation from the spatial scale of LSHS data to that of the HSLS data. To combine the spatial and spectral assessments together, QNR is defined as follows:

$$QNR = (1 - I_{\lambda})(1 - I_{s}) \tag{14}$$

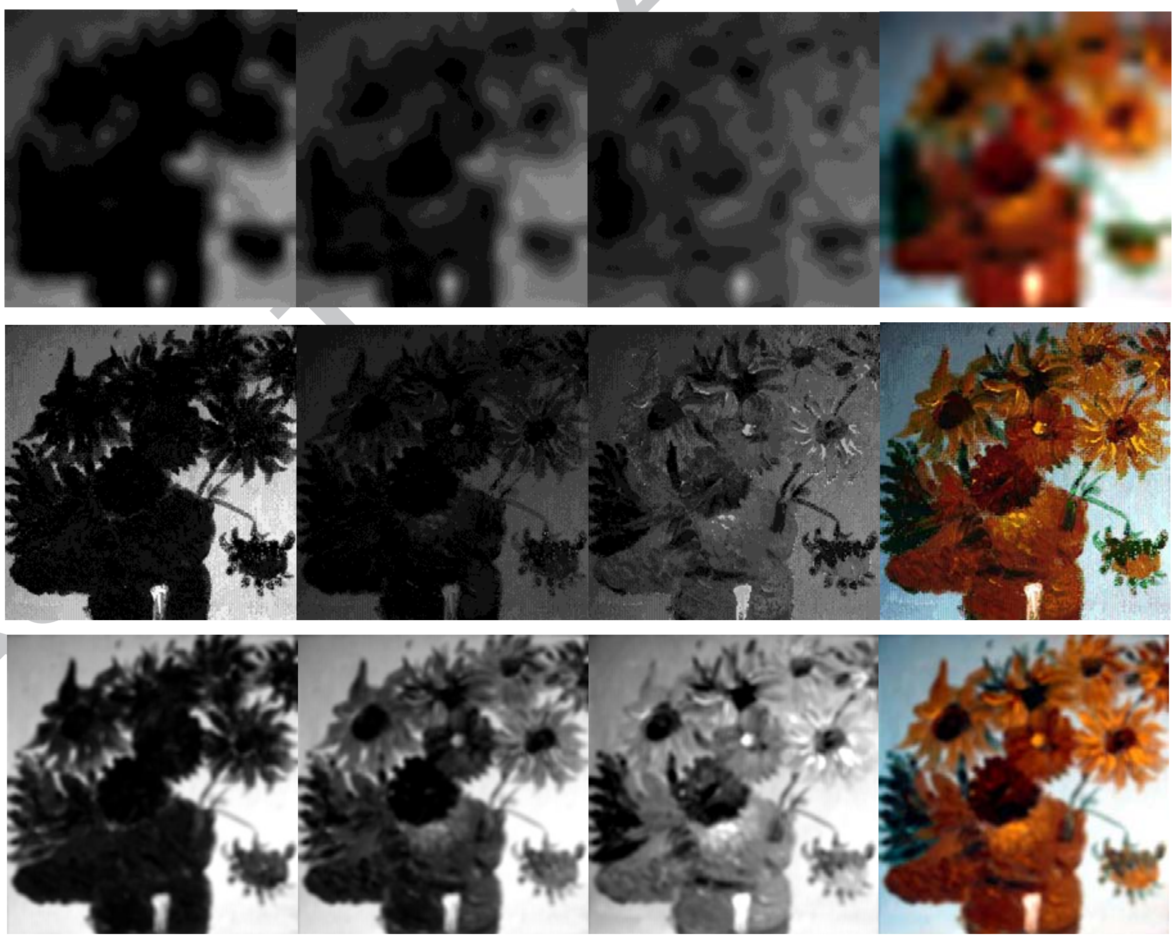
The highest value of *QNR* is one and is obtained when the spectral and spatial distortions are both zero.

4.2 Experiments with simulated data

We conducted the simulation experiment on a spectral image database described in [35], which provides a multispectral dataset from 400nm to 700nm at 10nm intervals (31 bands in total) and an RGB image (three bands in total). Each band of this dataset is stored in 16-bit PNG format with an image size of 480 by 352 pixels. The RGB image and the three multispectral bands 4, 15 and 31, which are most close to the blue, green and red bands of the RGB image in spectrum, are shown in Fig. 3. To simulate the LSHS data we down-sampled the multispectral data to 30 by 22 pixels with a scale factor of 16. Taking the original RGB image as the HSLS data, we aim to fuse them via DPLM, SRSR and SUM, respectively, to obtain the HSHS data with the spectral resolution of the LSHS image and the spatial resolution of the HSLS image.

Fig. 3. Illustration of simulation data. From the left, the images are RGB composite, multispectral bands 4, 15 and 31, respectively.

In the fusion process, the dictionary dimensionality of DPLM and the class number of SUM and SRSR were both set at 33, the sparseness degree of representation coefficients in DPLM was set at 0.85 and the window size of SUM was set at 19 after fine-tuning. The spatial and spectral fusion results of both methods are shown in Fig. 3. Due to the limitations of space, we chose the partial region of bands 7, 14 and 28 with an image size of 358×352 pixels to illustrate and compare the fusion results. In Fig. 4, from the top downwards, the rows demonstrate the input LSHS multispectral images (up-sampled by using bicubic interpolation), the SUM fusion results, the SRSR fusion results, the DPLM fusion results and the original multispectral images, respectively; from left to right, the columns show bands 7, 14, 28 and their RGB composite (28-14-7 bands as R-G-B), respectively. The comparison in Fig. 4 shows that the SUM results preserved the overall spectral color; however, there exists obvious spatial incoherence, especially in object edges, which is caused by the imprecise classification and insufficient LSHS pixels for unmixing. For SRSR fusion results, we can observe that some high-frequency spatial details and part of spectral color were lost, which are mainly caused by the large gap in spatial resolution between the input LSHS and RGB images. In contrast, the DPLM fusion result maintained good spatial coherence and overall spectral color, but there appears spectral distortion in certain objects (see the bright yellow regions in the RGB image), which is caused by insufficient training samples for this category of spectral signature in the LSHS image. To further analyze the fusion results, the absolute error images between the actual image and the fusion results of SUM, SRSR and DPLM in bands 7, 14, 28 and the average absolute error image for all bands are



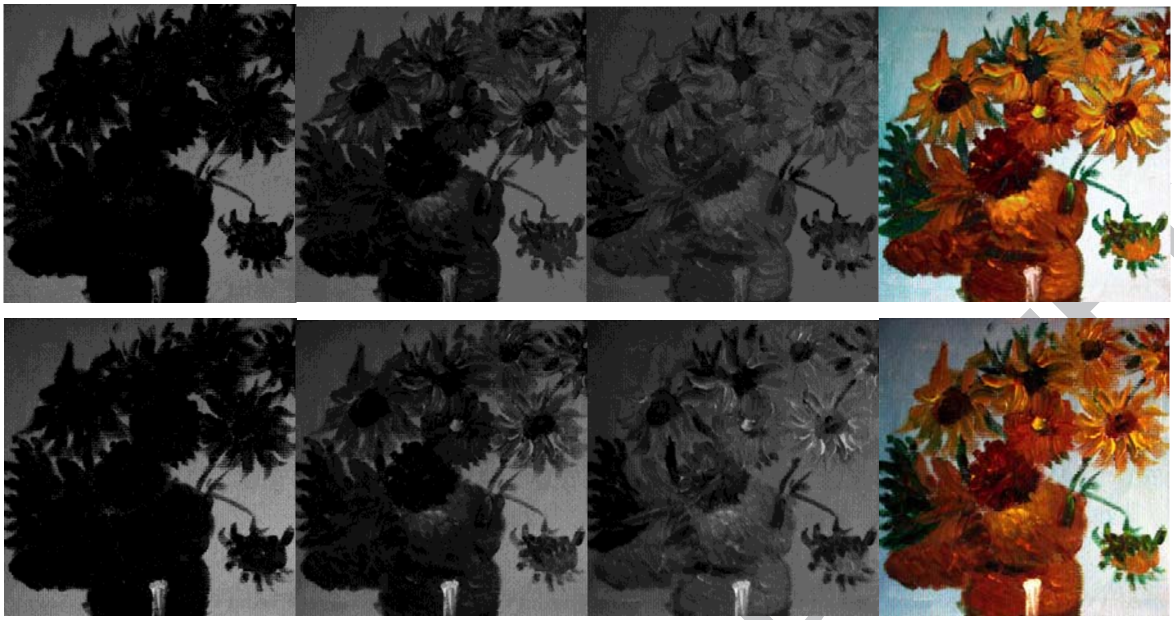
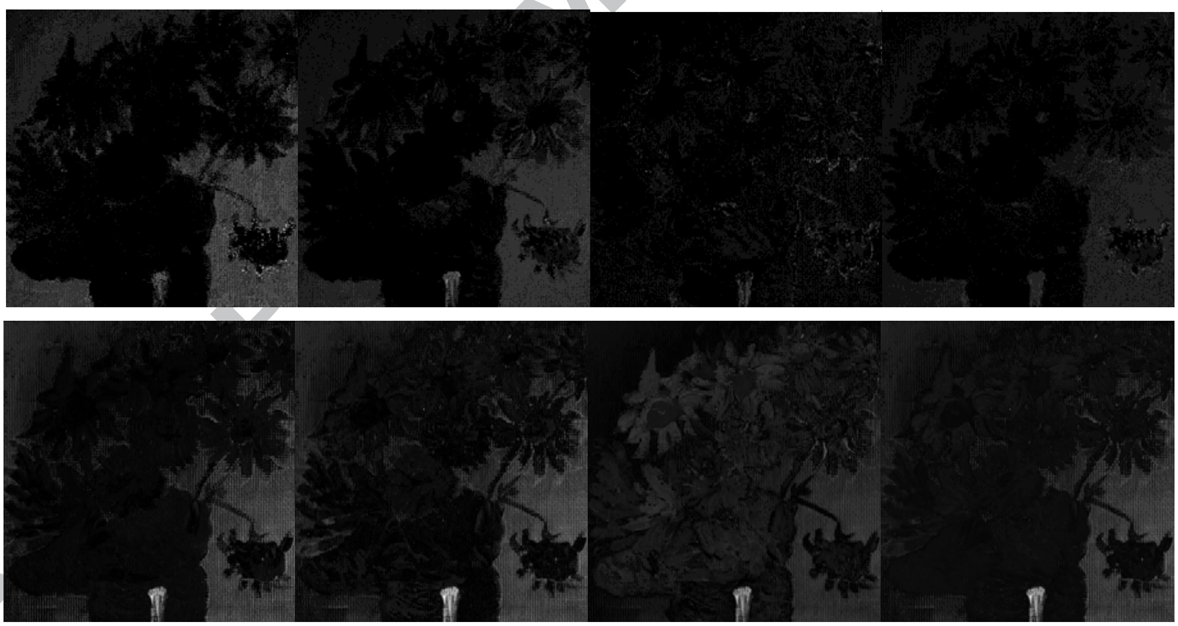


Fig. 4. Comparison of fusion results on simulated data. From the top downwards, the rows illustrate the input LSHS images, the fusion results of SUM, the fusion results of SRSR, the fusion results of DPLM and the original multispectral images, respectively. From left to right, the columns demonstrate bands 7, 14, 28 and their RGB composite, respectively.



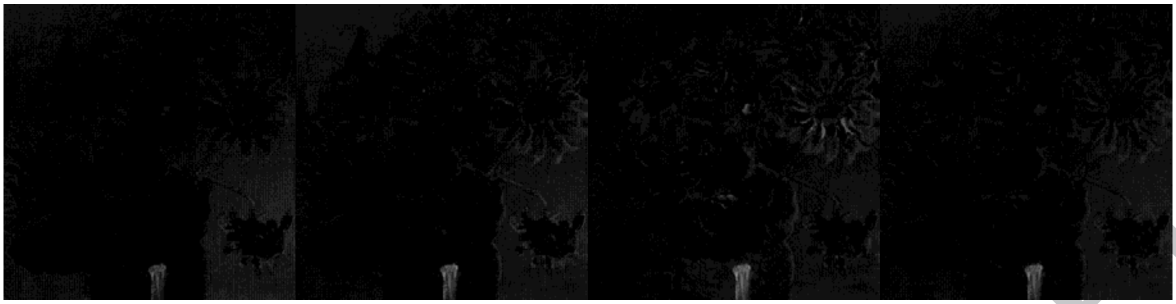


Fig. 5. The absolute error images between the actual image and the fusion results of SUM, SRSR and DPLM. From top downwards, the rows illustrate the absolute error images of SUM, SRSR and DPLM, respectively; from left to right, the columns show the error images of bands 7, 14, 28 and the average error image of all bands, respectively.

shown in Fig. 5 (from black to white the values denote 0 to 255). By comparing error images, we can observe that the prediction errors of SUM, SRSR and DPLM concentrate on object edges, specific band images and areas with less training samples, respectively.

To illustrate the fused spectral quality, we selected two regions of interest (ROI) with 3×3 pixels (Fig. 6(a)) and plotted the mean spectra of ROIs from the above fusion results as shown in Fig. 6(b). Comparing the spectra in Fig. 6(b), we can observe that the spectra from our algorithm are best fitted with the ground truth spectra. To quantitatively evaluate the fusion results, we calculated measurements between the actual image and the fusion results in terms of AAD, RMSE, ERGAS, SAM and SSIM, which are shown in Table 1. According to the comparisons in AAD and RMSE, we can conclude that our fusion results are the best at preserving radiometric property among three methods; the comparisons in ERGAS and SAM indicate that our fusion result preserved better spectral property than SUM and SRSR; based on the evaluations in SSIM, we can state that DPLM is better than the comparison approaches at preserving spatial details.

Since the sparseness constraint S_A is the main parameter for the proposed fusion algorithm, we analyze its impact on the fusion results. By setting S_A from 0.65 to 0.92 at 0.03 intervals, we got 10 fusion results from the proposed algorithm. The curve of RMSE of fusion results and S_A is shown in Fig. 6(c), from which we can see that when S_A locates at a small range (0.8~0.9 in this case), the fusion result is not sensitive to it (small RMSE) but the fusion error increases quickly when the parameter is out of this range. Besides, the processing time increases when S_A decreases because more projections are needed to deal with more representation bases.

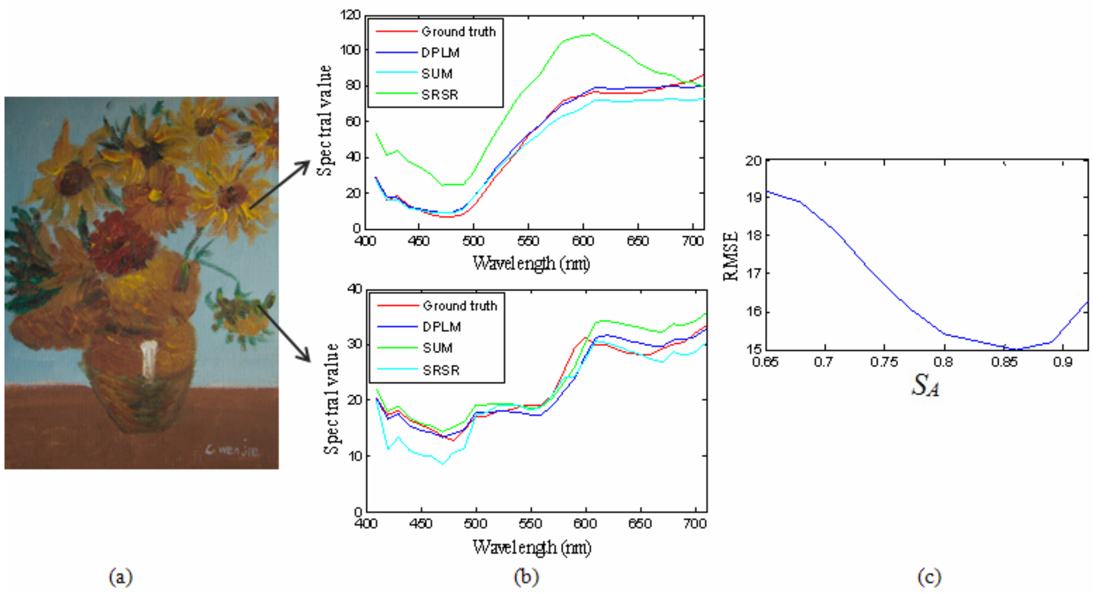


Fig. 6. (a) Selected two ROIs; (b) Spectra comparison from different fusion methods; (c) RMSE-S_A curve

Table 1. Quantitative evaluation comparisons of fusion results on simulation data. The numbers are shown in the range of 8-bit images. Those in **bold** are the best scores between three methods

Method	AAD	RMSE	ERGAS	SAM	SSIM
SUM	15.5875	21.8042	0.5513	15.4816	0.6874
SRSR	13.7706	19.0457	0.4527	10.7634	0.7302
DPLM	**10.4404**	**14.9414**	**0.3676**	**6.9237**	**0.7654**

4.3 Fusion of Landsat/ETM+ and Aqua/ MODIS

In this section, the spatial and spectral fusions of Landsat/ETM+ and Aqua/MODIS reflectance data are carried out by SUM, SRSR and DPLM, respectively. The ETM+ image was captured on November 20, 2001 and the MODIS image was acquired on November 17, 2001; we therefore considered they recorded the same scene of earth's surface due to their short capture interval. The ETM+ imager provides 7 bands in VIR region (30 m spatial resolution) and one band in the thermal infrared (TIR) region (60 m spatial resolution); and the MODIS sensor generates 36 bands (from 0.4 to 14.4 μm) with a spatial resolution of 250/500/1000 m. We focus on the fusion of the VIR region of ETM+ and MODIS in this paper, i.e., bands 1~5, 7, 8 of ETM+ and bands 1~19 of MODIS (1~7 bands with 500 m spatial resolution and 8~19 bands with 1000 m spatial resolution). Before fusion, the MODIS data was geometrically registered to the Landsat ETM+ data. The study area is located in Hong Kong, China, covering an area of 1.5 km by 1.5 km with a Landsat image size of 500×500 pixels and MODIS image size of 30×30 pixels (8~19 bands with 1000 m spatial resolution are resampled with a scale factor of 2). The partial region of the study area is shown in Fig. 7, in which the Landsat image size is 328×328 pixels and the MODIS image is up-sampled by using bicubic interpolation to the same spatial size of Landsat. From left to right the scenes are ETM+ composite image (NIR-red-green as the R-G-B), and the MODIS bands of 2, 7, and 17, respectively.

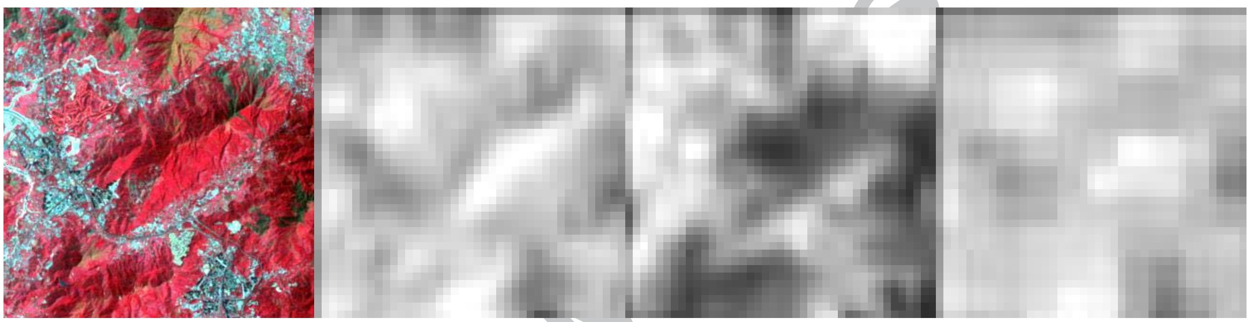


Fig. 7. Illustration of ETM+ and MODIS images. From left to right, the scenes are the ETM+ RGB composite image, MODIS bands 2, 7 and 17, respectively.

Taking the MODIS image as LSHS data and the ETM+ image as HSLS data, we expect to fuse them to obtain HSHS data with the spatial details of ETM+ and the spectral properties of MODIS by utilizing SUM, SRSR and DPLM, respectively. In the fusion process, the dictionary dimensionality of DPLM and the class number of SUM and SRSR were both set at 30, the sparseness degree of representation coefficients in DPLM was set at 0.8 and the window size of SUM was set at 25 via fine-tuning. The fusion results of SUM, SRSR and DPLM are demonstrated in Fig. 8 without enhancement. To save space, we chose a partial region of interest (with an image size of 240×240 pixels) and bands of 3, 7, 12 and 18 to illustrate the fusion results. From the top downwards, the rows show the input MODIS image, the fusion result of SUM, the fusion result of SRSR and the fusion result of DPLM, respectively; from left to right, the columns show bands of 3, 7, 12 and 18, respectively. From the comparisons in Fig. 8, we can observe that the SUM fusion result is lack of smoothing due to the classification step; the SRSR result is over-smoothing due to the super-resolution strategy; while the spatial information in our fusion result is best predicted due to the spectral bases extraction strategy.

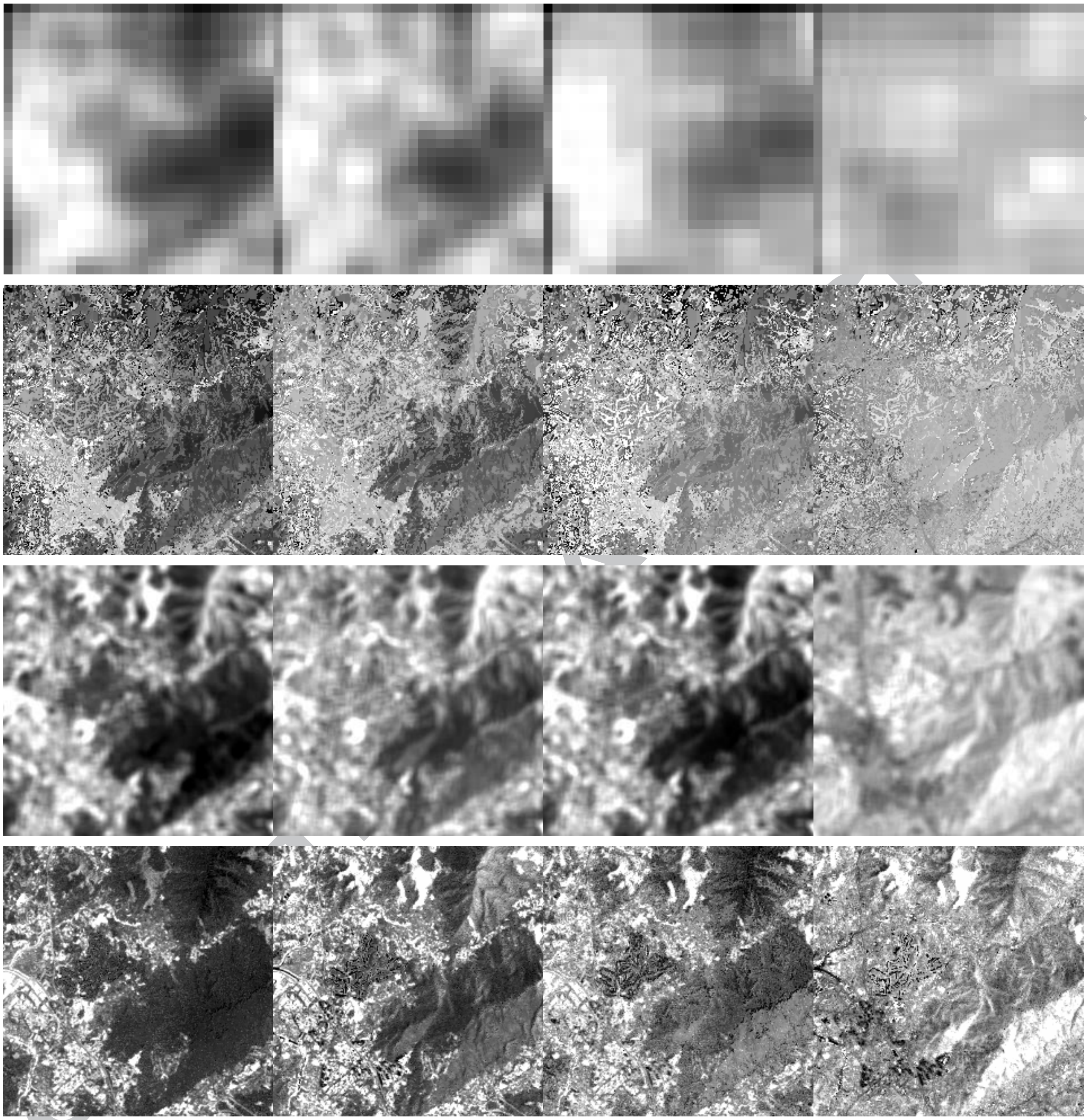


Fig. 8. Comparison of fusion results on ETM+ and MODIS data. From the top downwards, the rows illustrate the input MODIS image, the fusion result of SUM, the fusion result of SRSR and the fusion result of DPLM, respectively; from left to right, the columns show the scenes of bands 3, 7, 12 and 18, respectively.

To quantitatively evaluate the fusion results, we calculated the spectral CC, the spatial CC, the spectral distortion index I_λ, the spatial distortion index I_s and the joint spatial and spectral quality index *QNR*, respectively. When calculating the spectral CC, we down-sampled the fused results to the same spatial resolution of MODIS and calculated the CCs between the down-sampled fusion results and the actual MODIS image for all bands. For the spatial CCs, they were calculated between the corresponding bands of fused results and the original ETM+ image. The position of the ETM+ and MODIS bands in the spectrum is shown in Fig. 9. Since the fused HSHS data has the same spectral information as MODIS, their spectral position with ETM+ is the same as MODIS. The average spectral CC of 19 bands, the average spatial CC for the corresponding 6 bands in Fig. 9, I_λ, I_s and *QNR* are shown in Table 2. From the comparisons of spectral CC and I_λ, we can conclude that the performance of SRSR is the worst and SUM and DPLM perform very similarly at preserving spectral information; from the comparisons of spatial CC and I_s, we can conclude that DPLM is the best at preserving spatial information.

Table 2. Quantitative evaluation comparisons of fusion results on ETM+ and MODIS without reference. Those in **bold** are the best scores among the three methods

Method	Spectral CC	Spatial CC	I_λ	I_s	*QNR*
SUM	0.7838	0.4577	**0.0562**	0.1012	0.8482
SRSR	0.7047	0.6925	0.0752	0.0804	0.8504
DPLM	**0.7872**	**0.7523**	0.0642	**0.0717**	**0.8687**

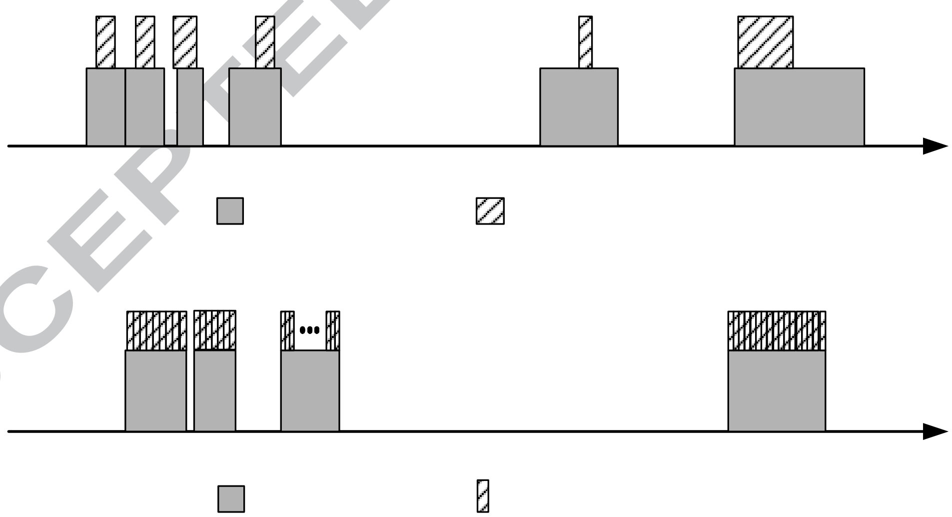


Fig. 9. The top row shows the position of ETM+ and MODIS bands in the spectrum; the bottom row shows the position of SPOT5 and Hyperion bands in the spectrum.

4.4 Fusion of EO-1/Hyperion and SPOT5/HRG

In this section, we applied the proposed method, SRSR and SUM to fuse hyperspectral and multispectral images: EO-1/Hyperion and SPOT5/HRG multispectral data. The Hyperion generates 220 bands in spectra range of 0.4~2.5 μm with a 30 m spatial resolution and the SPOT5/HRG offers a panchromatic image and a multispectral (MS) image (including green, red, NIR and SWIR bands from 0.5 to 1.75 μm) with spatial resolution of 5 m and 10 m, respectively. Taking the Hyperion image as the LSHS data and SPOT5 MS image as HSLS data, we expect to generate the desired HSHS data with the spectral information of Hyperion and the spatial information of SPOT5 MS by employing SUM, SRSR and DPLM, respectively. These two datasets were both captured on 28 November, 2008 and are located in Hong Kong, China, covering an area of 5km by 10 km with an image size of 336×171 pixels for Hyperion image and an image size of 1008×513 pixels for SPOT5 MS image. In the pre-processing step, we de-noised the Hyperion image by means of the method in [36] and kept 167 good-quality bands by removing those with strong noise; then the regions of interest in these two datasets are co-registered by choosing the ground control points in ENVI software. The partial region of interest is shown in Fig. 10, in which the image size of SPOT5 MS is 513×513 pixels and the Hyperion image is up-sampled by using bicubic interpolation to the same spatial size of SPOT5 MS with a scale factor of 3. From left to right the scenes are the SPOT5 composite image (NIR-red-green as an R-G-B), the Hyperion bands 20, 30 and 48 (which are closest to green, red and the NIR bands of SPOT5 in the spectrum), respectively.

Fig. 10. Illustration of SPOT5 MS and Hyperion images. From left to right, the scenes are the SPOT5 RGB composite image, Hyperion bands of 20, 30 and 48, respectively.

In the fusion process, the dictionary dimensionality of DPLM and the class number of SUM and SRSR were both set at 30, the sparseness degree for the representation coefficients in DPLM was set at 0.85 and the window size of SUM was set to 25 via fine tuning. The fusion results of

SUM, SRSR and DPLM are demonstrated in Fig. 11. To save space, we chose a partial region of interest with an image size of 290×290 pixels and bands of 8, 31, 56 and 120 to illustrate the fusion results. From the top downwards, the rows show the input Hyperion image, the fusion results of SUM, the fusion results of SRSR and the fusion results of DPLM, respectively; from left to right, the columns show bands of 8, 31, 56 and 120, respectively. From the comparisons in Fig. 11, we can clearly see that the spatial information is best preserved in our fusion result among three methods. Besides, we can observe that the existing noise in bands 8 and 120 of Hyperion image is removed in all fusion results.

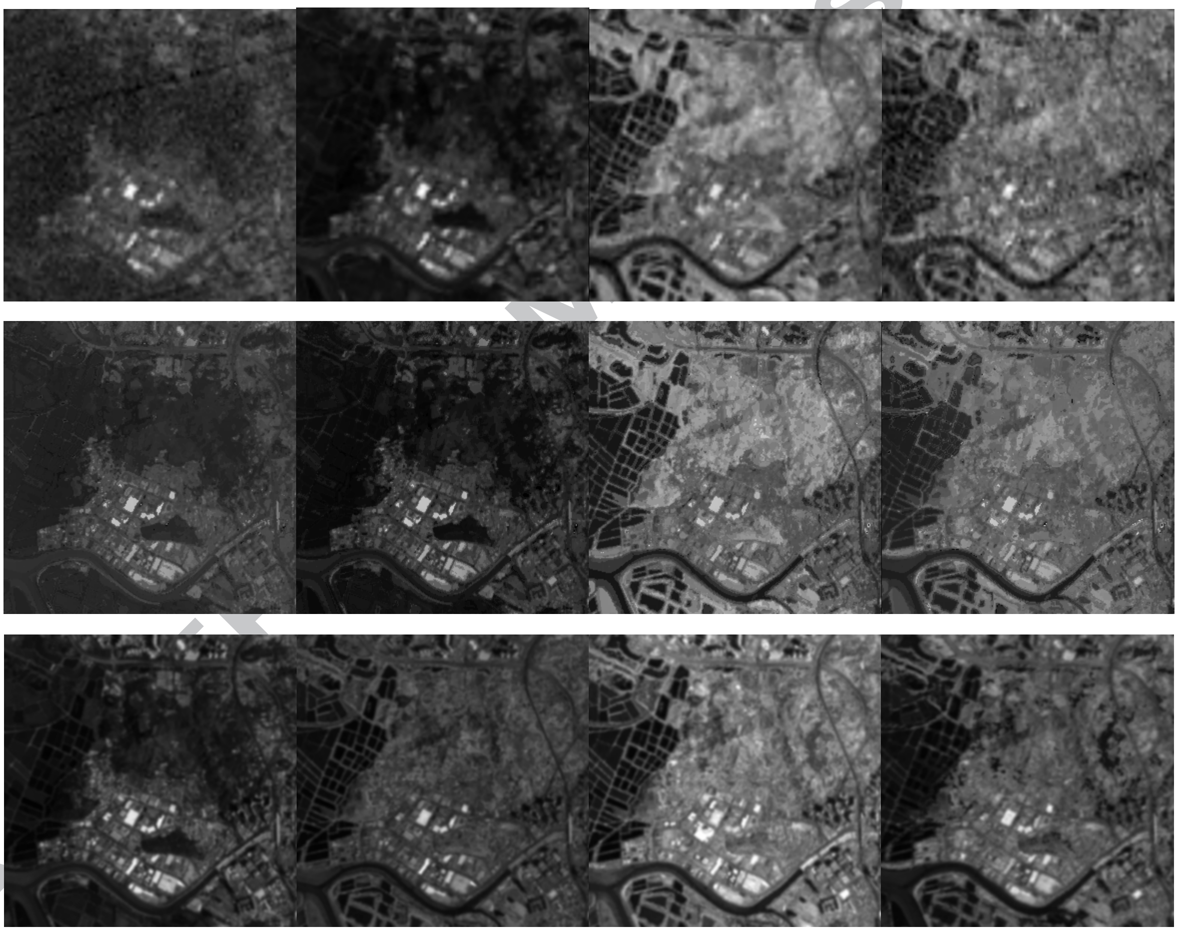


Fig. 11. Comparison of fusion results on Hyperion and SPOT5 data. From the top downwards, the rows illustrate the input Hyperion image, the fusion result of SUM, the fusion result of SRSR and the fusion result of DPLM, respectively; from left to right, the columns show the images of bands 8, 31, 56 and 120, respectively.

To quantitatively evaluate the fusion results, we calculated the spectral CC, the spatial CC, the spectral distortion index I_λ, the spatial distortion index I_s and the joint spatial and spectral quality index QNR, respectively. When calculating the spectral CC, we down-sampled the fused results to the same spatial resolution of Hyperion and calculated the CCs between the down-sampled fusion results and the Hyperion image for all bands. For the spatial CCs, they are calculated between the corresponding bands of fused results and the SPOT5 MS image. The position of SPOT5 MS and Hyperion bands in the spectrum is shown in Fig. 9. Since the fused HSHS data has the same spectral information as Hyperion, their spectral position with SPOT5 MS is the same as Hyperion. The average spectral CC on 167 bands, the average spatial CC for the corresponding 44 bands in Fig. 7, I_λ, I_s and QNR are shown in Table 3. From the comparisons of spectral CC and I_λ, we can conclude that SRSR is the worst and SUM and DPLM perform very similarly at preserving spectral information; from the comparisons of spatial CC and I_s, we can conclude that DPLM is the best among three methods at preserving spatial information.

Table 3. Quantitative evaluation comparisons of fusion results on Hyperion and SPOT5 MS data without reference. Those in **bold** are the best scores among three methods

Method	Spectral CC	Spatial CC	I_λ	I_s	QNR
SUM	**0.8752**	0.5358	0.1310	0.0463	0.8288
SRSR	0.8412	0.8151	0.1428	0.0278	0.8333
DPLM	0.8689	**0.9429**	**0.0760**	**0.0181**	**0.9072**

5. Conclusion

We proposed a spatial and spectral fusion method based on dictionary-pair learning. This method is devised for the fusion of two categories of remote sensing data: one category possesses coarse spatial details, wide spectrum coverage and more spectral bands, termed as data with low spatial resolution and high spectral resolution (LSHS); while the other category data is characterized by fine spatial details, narrow spectrum coverage and less spectral bands, termed as data with high spatial resolution and low spectral resolution (HSLS). By fusing the spectral information of LSHS data and the spatial information of HSLS data, this method generates synthetic data with both high spatial and high spectral resolutions. Based on sparse non-negative matrix factorization and relevant sparse coding techniques, this method includes two stages: in the first stage, two spectral dictionaries with high and low spectral resolutions (dictionary-pair) are learned to represent the spectral bases of LSHS data and HSLS data, respectively; in the second stage, the fused data is derived by ensuring the LSHS and HSLS data have the same representation coefficients with respect to their dictionaries, thereby incorporating the spatial details from the HSLS data into the LSHS data. The good performance of the proposed method on both simulation data and satellite images validated its effectiveness in fusing multispectral and hyperspectral (or high-spectral) data. When compared with the classic spatial unmixing based method and the sophisticated super-resolution method, the superiority of the proposed method in preserving both spatial information and spectral information is demonstrated.

By representing each pixel vector as a linear combination of the learned spectral bases, the proposed method earns a more consistent prediction with accurate spatial details than the spatial unmixing based method. Comparing the fusion results on simulation data and satellite data, we can observe that our proposed method performs better on simulation data than on satellite data at preserving spectral information, which is caused by the relative poor quality of satellite data. Comparing the fusion results on two satellite data settings, we can conclude that the proposed method performs better at preserving spectral information when the given LSHS image has a larger number of spectral bands; while a smaller scale factor in spatial resolution between LSHS and HSLS data is favorable for our method in preserving spatial details. Since the spatial details are directly extracted from HSLS data for SUM and our method, good geometrical registration between LSHS and HSLS data is required for both methods. However, from the existence of noise bands in fusion of Hyperion and SPOT5 data, we can conclude that the dictionary learning method has a capability of anti-noise. That means if the registration errors are deemed as noise, the proposed method has a certain degree of tolerance for these registration errors. This error tolerance capability is increased when there are sufficient training samples.

References

[1] D.J. Weydahl, F. Bretar, P. Bjerke, Comparison of RADARSAT-1 and IKONOS satellite images for urban features detection, Informat. Fusion 6 (2005) 243–249.

[2] D. Landgrebe, Hyperspectral image data analysis, IEEE Signal Process. Mag. 19 (2002) 17–28.

[3] B. Aiazzi, S. Baronti, F. Lotti, M. Selva, A comparison between global and context-adaptive pansharpening of multispectral images, IEEE Geosci. Remote Sens. Lett. 6 (2009) 302–306.

[4] I. Amro, J. Mateos, M. Vega, R. Molina, A.K. Katsaggelos, A survey of classical methods and new trends in pansharpening of multispectral images, EURASIP Journal on Advances in Signal Processing, open access (2011).

[5] B. Zhukov, D. Oertel, F. Lanzl, G. Reinhackel, Unmixing-based multisensor multiresolution image fusion, IEEE Trans. Geosci. Remote Sens. 37 (1999) 1212–1226.

[6] A. Minghelli-Roman, M. Mangolini, M. Petit, L. Polidori, Spatial resolution improvement of MeRIS images by fusion with TM images, IEEE Trans. Geosci. Remote Sens. 39 (2001) 1533–1536.

[7] R. Zurita-Milla, J. Clevers, M.E. Schaepman, Unmixing-based Landsat TM and MERIS FR data fusion, IEEE Geosci. Remote Sens. Lett. 5 (2008) 453–457.

[8] N. Mezned, S. Abdeljaoued, M.R. Boussema, A comparative study for unmixing based Landsat ETM + and ASTER image fusion, Int. J. Appl. Earth Obs. Geoinf. 12 (2010) 131–137.

[9] J. Amorós-López, L. Gómez-Chova, L. Alonso, L. Guanter, J. Moreno, G. Camps-Valls, Regularized Multiresolution Spatial Unmixing for ENVISAT/MERIS and Landsat/TM Image Fusion, IEEE Geosci. Remote Sens. Lett. 8 (2011) 844−848.

[10] R.C. Hardie, M.T. Eismann, MAP estimation for hyperspectral image resolution enhancement using an auxiliary sensor, IEEE Trans. Image Process. 13 (2004) 1174–1184.

[11] M.T. Eismann, R.C. Hardie, Application of the stochastic mixing model to hyperspectral resolution enhancement, IEEE Trans. Geosci. Remote Sens. 42 (2004) 1924–1933.

[12] M.T. Eismann, R.C. Hardie, Hyperspectral Resolution Enhancement Using high-resolution multispectral imagery with arbitrary response functions, IEEE Trans. Geosci. Remote Sens. 43 (2005) 455–465.

[13] Y. Zhao, J. Yang, Q. Zhang, L. Song, Y. Cheng, Q. Pan, Hyperspectral imagery super-resolution by sparse representation and spectral regularization, EURASIP Journal on Advances in Signal Processing 2011(1) (2011) 1-10.

[14] N. Keshava, J.F. Mustard, Spectral Unmixing, IEEE Signal Processing Magazine 19 (2002) 44–57.

[15] A. Plaza, P. Martínez, R. Pérez, J. Plaza, A quantitative and comparative analysis of endmember extraction algorithms from hyperspectral data, IEEE Geosci. Remote Sens. Lett. 42 (2004) 650−663.

[16] M.-D. Iordache, J.M. Bioucas-Dias, A. Plaza, Sparse unmixing of hyperspectral data, IEEE Trans. Geosci. Remote Sens. 49 (6) (2011) 2014-2039.

[17] I. Toši´c, P. Prossard, Dictionary Learning, IEEE Signal Processing Magazine 28 (2011) 27-38.

[18] M. Aharon, M. Elad, A. Bruckstein, K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation, IEEE Transactions on Signal Processing 54 (2006) 4311–4322.

[19] R. Rubinstein, A. Bruckstein, M. Elad, Dictionaries for sparse representation modeling, IEEE Proceedings – Special Issue on Applications of Sparse Representation & Compressive Sensing 98 (2010) 1045–1057.

[20] J.A. Tropp, Greed is good: Algorithmic results for sparse approximation, IEEE Trans. Inform. Theory 50 (2004) 2231–2242.

[21] R. Tibshirani, Regression shrinkage and selection via the lasso, J. R. Stat. Soc. Ser. B (Method.) 58 (1996) 267–288.

[22] J.A. Tropp, S.J. Wright, Computational methods for sparse solution of linear inverse problems, Proc. IEEE 98 (2010) 948–958.

[23] J. Yang, J. Wright, T. Huang, Y. Ma, Image super-resolution via sparse representation, IEEE

Transactions on Image Processing 19 (2010) 2861−2873.

[24] R. Zeyde, M. Elad, M. Protter, On Single Image Scale-Up using Sparse-Representations, Curves & Surfaces, Avignonm, France, 2010, pp. 711−730.

[25] R. Rubinstein, A.M. Bruckstein, M. Elad, Dictionaries for sparse representation modeling, IEEE Proceedings 98 (2010) 1045−1057.

[26] D.D. Lee, H.S. Seung, Learning the parts of objects by nonnegative matrix factorization, Nature 401 (1999) 788–791.

[27] P.O. Hoyer, Non-negative Matrix Factorization with Sparseness Constraints, Journal of Machine Learning Research 5 (2004) 1457–1468.

[28] R.A. Schowengerdt, Remote Sensing: Models and Methods for Image Processing, Academic Press, New York, 2007.

[29] H. Song, B. Huang, Spatiotemporal satellite image fusion through one-pair image learning, IEEE Trans. Geosci. Remote Sens. 51 (4) (2013) 1883–1896.

[30] C.-I. Chang, Q. Du, Estimation of number of spectrally distinct signal sources in hyperspectral imagery, IEEE Trans. Geosci. Remote Sens. 42 (2004) 608–619.

[31] M.M. Khan, L. Alparone, J. Chanussot, Pansharpening quality assessment using the modulation transfer functions of instruments, IEEE Trans. Geosci. Remote Sens. 47 (2009) 3880−3891.

[32] R.H. Yuhas, A.F.H. Goetz, J.W. Boardman, Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm, in Proc. Summaries 3rd Annu. JPL Airborne Geosci. Workshop (1992) 147–149.

[33] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image Quality Assessment: From Error Visibility to Structural Similarity, IEEE Transactions on Image Processing 13 (2004) 600−612.

[34] L. Alparone, B. Aiazzi, S. Baronti, A. Garzelli, F. Nencini, M. Selva, Multispectral and Panchromatic Data Fusion Assessment without Reference, Photogrammetric Engineering & Remote Sensing 74 (2008) 193–200.

[35] F. Yasuma, T. Mitsunaga, D. Iso, S.K. Nayar, Generalized assorted pixel camera: Postcapture control of resolution, dynamic range, and spectrum, IEEE Trans. IP. 19 (2010) 2241–2253.

[36] G. Chen, S-E Qian, Denoising of Hyperspectral Imagery Using Principal Component Analysis and Wavelet Shrikage, IEEE Trans. Geosci. Remote Sens. 49 (2011) 973−980.