

# Spatial and Spectral Image Fusion Using Sparse Matrix Factorization

Bo Huang, *Associate Member, IEEE*, Huihui Song, Hengbin Cui, Jigen Peng, and Zongben Xu

**Abstract**—In this paper, we present a novel spatial and spectral fusion model (SASFM) that uses sparse matrix factorization to fuse remote sensing imagery with different spatial and spectral properties. By combining the spectral information from sensors with low spatial resolution (LSaR) but high spectral resolution (HSeR) (hereafter called HSeR sensors), with the spatial information from sensors with high spatial resolution (HSaR) but low spectral resolution (LSeR) (hereafter called HSeR sensors), the SASFM can generate synthetic remote sensing data with both HSeR and HSeR. Given two reasonable assumptions, the proposed model can integrate the LSaR and HSeR data via two stages. In the first stage, the model learns from the LSaR data a spectral dictionary containing pure signatures, and in the second stage, the desired HSeR and HSeR data are predicted using the learned spectral dictionary and the known HSeR data. The SASFM is tested with both simulated data and actual Landsat 7 Enhanced Thematic Mapper Plus (ETM+) and Terra Moderate Resolution Imaging Spectroradiometer (MODIS) acquisitions, and it is also compared to other representative algorithms. The experimental results demonstrate that the SASFM outperforms other algorithms in generating fused imagery with both the well-preserved spectral properties of MODIS and the spatial properties of ETM+. Generated imagery with simultaneous HSeR and HSeR opens new avenues for applications of MODIS and ETM+.

**Index Terms**—High spatial resolution (HSeR), high spectral resolution (HSeR), Landsat, matrix factorization, Moderate Resolution Imaging Spectroradiometer (MODIS), sparse coding, spatial and spectral fusion model (SASFM), spatial-spectral fusion.

## I. INTRODUCTION

**D**UE TO technical and budget constraints, in satellite remote sensing systems, there is a tradeoff between spatial resolution and spectral resolution. As a result, it is at present uncircumventable to obtain a remote sensing scene that covers a large area at both high spatial and high spectral resolutions (HSeR). Medium resolution sensors like the Medium Resolution Imaging Spectrometer (MERIS) and the Moderate Resolution Imaging Spectroradiometer (MODIS) provide high spectral resolution (HSeR) but low spatial resolution (LSaR)

(varying from 250 to 1200 m (hereafter called HSeR sensors). HSeR sensors were designed to detect and monitor changes in the Earth's surface on a frequent basis and they can also be used to derive such measurements as the leaf area index [1], aerosol optical depth [2], and so on. For example, MODIS' ability to collect data over large covered areas, its short revisit intervals (1 or 2 days), and its relatively wide spectral coverage (36 spectral bands ranging in wavelength from 0.4 to 14.4  $\mu\text{m}$ ) make it a popular sensor for various remote sensing applications. However, due to the LSaR of the MODIS instrument (250 m to 1 km), it is not suitable for applications that require high spatial details.

In contrast, high spatial resolution (HSeR) sensors like the Landsat Enhanced Thematic Mapper Plus (ETM+), Advanced Spaceborne Thermal Emission and Reflectance Radiometer, and Système Pour l'Observation de la Terre High Resolution Visible provide remote sensing data with spatial resolutions ranging from 10 to 30 m, which allow us to monitor heterogeneous areas with a finer pixel footprint. For instance, the Landsat instruments, as the longest satellite observation systems, have proven extremely useful in monitoring land cover/land use changes, monitoring ecosystem dynamics, and retrieving biogeochemical parameters [3]. However, these sensors have relatively lower spectral resolutions (LSeRs) and cover a narrower spectral range. For example, Landsat covers the spectrum in wavelengths from 0.45 to 2.35  $\mu\text{m}$ , divided into eight spectral bands.

One cost-effective solution to the HSeR tradeoff is to use image fusion to combine the spatial information of the HSeR data with the spectral information of the HSeR data. Traditionally, spatio-spectral image fusion focuses on the blending of one HSeR panchromatic band with multispectral LSaR bands, which is also called pan-sharpening [4], [5]. Numerous methods have been developed for Pan-sharpening, including the classic intensity-hue-saturation transformation [6], principal component substitution [7], “à trous” wavelet transform [8], the multiresolution analysis method [9], and so on. These methods have achieved a reasonable level of success, especially in enhancing the spatial resolution of multispectral bands, but their applications are limited in the following two aspects: 1) there must exist a panchromatic band to provide high-frequency spatial information and 2) the wavelength of the involved multispectral bands is usually located in the visible and near-infrared (NIR) range of the electromagnetic spectrum to match the corresponding panchromatic band. Apparently, the coupling of two multispectral sensors, namely, Landsat ETM+ (with eight spectral bands) and MODIS (with 36 spectral bands), as in our case, goes beyond the fusion

Manuscript received June 12, 2012; revised March 12, 2013; accepted March 13, 2013. This work was supported in part by 863-Hightech Program under Grant 2009AA122004 and the Hong Kong Research Grant Council under Grant CU Hong Kong Research Grant Council under Grant CUHK 444612.

B. Huang and H. Song are with the Chinese University of Hong Kong, Hong Kong (e-mail: bohuang@cuhk.edu.hk; shsherry@cuhk.edu.hk).

H. Cui, J. Peng, and Z. Xu are with Xi'an Jiaotong University, Xi'an 710049, China (e-mail: xjcuihb@stu.xjtu.edu.cn; jgpeng@mail.xjtu.edu.cn; zbxu@mail.xjtu.edu.cn).

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/TGRS.2013.2253612

of a panchromatic band with multispectral bands. MODIS also contains many bands outside the spectral range that can be covered by ETM+. Consequently, the traditional spatio-spectral fusion methods, such as the pan-sharpening method, are not suitable for the spatio-spectral fusion of Landsat ETM+ and MODIS data.

Another category of spatio-spectral fusion methods is based on spatial unmixing (SaU) [10]–[13]. For example, Minghelli-Roman *et al.* [11], [12] have improved the spatial resolution of the MERIS data by merging them with the Landsat TM data. This fusion process consists of three steps: 1) unsupervised classification of the TM data; 2) determination of class spectra in MERIS using the fractional abundances of endmembers derived in step 1; and 3) substitution of each classified pixel in MERIS with its spectral profile. This unmixing-based algorithm performs well in regions with large homogeneity. However, it may lead to a loss of local variability for the same land cover type due to the implementation across the whole image at once. To mitigate this drawback, Zurita-Milla *et al.* [13] employed a sliding window on the MERIS image that uses the neighborhood pixels to increase the class variability among windows and tested it with a land cover map for monitoring vegetation seasonal dynamics [14]. However, this approach lacks intraclass variability inside a window and can obtain only the spectral signatures within the TM spatial resolution. Additionally, the unmixing-based fusion algorithms require a high accuracy of geometric coregistration of the data; this is a strong but not unrealistic requirement for modern coregistration techniques [12].

Sparsity-induced regularization is an effective technique for the spectral unmixing of hyperspectral data [15]. It aims to find in a spectral library an optimal combination of pure spectral signatures that match each mixed pixel in a hyperspectral image, such as the one captured by NASAs airborne visible infrared imaging spectrometer and hyperion. The spectral library is usually known in advance (e.g., spectra collected on the ground by a field spectroradiometer or by airborne spectrometers). Then, the measured spectrum can be sparsely represented as a linear combination of the pure spectral signatures in the library. The risk in using library endmembers is that these spectra are rarely acquired by satellite instruments (such as Landsat ETM+ and MODIS) under the same conditions as the collection instruments. In this paper, we used the sparsity regularization, but built the spectral library (also called the spectral dictionary) directly from the satellite data themselves to ensure consistency between the spectral library and the data to be estimated.

In this paper, we developed a new spatial and spectral fusion model (SASFM) based on sparse matrix factorization. The model combines the spatial information from sensors with HsAR, with the spectral information from sensors with HSeR. We introduced a two-stage algorithm to combine these two categories of remote sensing data. In the first stage, we obtained an optimal set of atoms (i.e., the spectral dictionary) from data with LSaR and HSeR (i.e., HSeR) to represent the spectral signatures of various materials in the scene. Given the simple observation that there are probably only a few land surface materials contributing to each pixel in the LSaR image,

we cast this problem as a sparse matrix factorization problem. In the second stage, by using the spectral dictionary developed in the first stage, together with data with HsAR and LSaR (i.e., HsAR), we reconstructed the spectrum of each pixel to produce a HsSaSeR image via a sparse coding technique. Given the assumptions of spectral sparsity and of a linear combination relationship between the bands of an HsAR image and the desired HsSaSeR data, the proposed method can significantly improve the spatial resolution of the HSeR data, while at the same time, maintain accurate spectrum estimates.

The remainder of this paper is organized as follows. Section II presents the theoretical basis of our model. Our proposed method for the fusion of HSeR data and HsAR data is presented in Section III. Section IV presents the experimental results for both the simulated and the actual satellite data. We conclude in Section V.

## II. THEORETICAL BASIS

Given a signal matrix  $X \in R^{S \times N}$ , sparse matrix factorization seeks to decompose it into a basis matrix  $D \in R^{S \times M}$  (also called a dictionary) and a coefficient matrix  $\Lambda \in R^{M \times N}$ , so that each column of  $X$  can be sparsely represented by the linear combinations of basis vectors in  $D$  [16]. This decomposition process can be expressed in the following function:

$$\min_{D, \Lambda} \|\Lambda\|_0, \quad \text{s.t. } X = D\Lambda \quad (1)$$

where  $\|\Lambda\|_0$  stands for the number of nonzero elements in  $\Lambda$ . Methods for solving the  $l_0$ -norm constrained sparseness problem [as formulated in (1)] include greedy algorithms, such as basis pursuit [17] and orthogonal matching pursuit (OMP) [18]. OMP is an iterative greedy algorithm that selects at each step the column in the dictionary that is most correlated with the current residues. This column is then added into the set of selected columns. This algorithm updates the residues by projecting the observation onto the linear subspace spanned by the columns that have already been selected. These processes are repeated until the signal matrix is satisfactorily decomposed.

However, the problem in (1) is NP hard [19], which indicates that the problem is combinatorial and very complex to solve. Fortunately, this problem can be approximated by replacing the  $l_0$ -norm with the  $l_1$ -norm (sum of magnitudes) in (1) when the coefficient vectors in  $\Lambda$  are sufficiently sparse and the atoms in  $D$  are different [20]

$$\min_{D, \Lambda} \|\Lambda\|_1, \quad \text{s.t. } X = D\Lambda. \quad (2)$$

As representation errors may exist in (2), this equation can be expressed in the following relaxed form:

$$\min_{D, \Lambda} \|\Lambda\|_1, \quad \text{s.t. } \|X - D\Lambda\|_2^2 \leq \varepsilon \quad (3)$$

where  $\|\cdot\|_2^2$  denotes the  $l_2$ -norm (sum of atom squares) and  $\varepsilon$  is a small constant. Due to the good properties of  $l_1$ -norm minimization and its approximation to  $l_0$ -norm in preserving sparseness, many  $l_1$ -norm-based sparse coding techniques have been developed, such as Lasso, the gradient projection, homotopy, iterative shrinkage thresholding [21], and so on.

TABLE I  
BANDWIDTH COMPARISON BETWEEN LANDSAT 7 ETM+ AND MODIS

ETM+ Band	Bandwidth ( $\mu\text{m}$ )	MODIS Band	Bandwidth ( $\mu\text{m}$ )	MODIS Band	Bandwidth ( $\mu\text{m}$ )	MODIS Band	Bandwidth ( $\mu\text{m}$ )	MODIS Band	Bandwidth ( $\mu\text{m}$ )	MODIS Band	Bandwidth ( $\mu\text{m}$ )
1	0.45–0.52	1	0.62–0.67	9	0.44–0.45	17	0.89–0.92	25	4.48–4.54	33	13.1–13.4
2	0.52–0.60	2	0.84–0.87	10	0.48–0.49	18	0.93–0.94	26	1.36–1.39	34	13.4–13.7
3	0.63–0.69	3	0.46–0.48	11	0.52–0.53	19	0.91–0.96	27	6.53–6.89	35	13.7–14.1
4	0.76–0.90	4	0.54–0.56	12	0.54–0.55	20	3.66–3.84	28	7.17–7.47	36	14.1–14.3
5	1.55–1.75	5	1.23–1.25	13	0.66–0.67	21	3.92–3.98	29	8.40–8.70		
6	10.4–12.5	6	1.62–1.65	14	0.67–0.68	22	3.92–3.98	30	9.58–9.88		
7	2.09–2.35	7	2.10–2.15	15	0.74–0.75	23	4.02–4.08	31	10.7–11.2		
8	0.52–0.90	8	0.41–0.42	16	0.86–0.87	24	4.43–4.49	32	11.7–12.2		

To identify the dictionary  $D$  from the training samples  $X$ , several dictionary learning algorithms have been developed. These algorithms can be divided into three categories [20]: 1) probabilistic learning methods; 2) parametric learning methods; and 3) clustering- or vector-quantization-based methods. Probabilistic learning methods build on the assumptions of both distributions of representation errors (usually Gaussian distribution) and representation coefficients (usually Laplacian distribution). For example, the maximum likelihood dictionary learning method [22], falling under this category, maximizes the likelihood that an image matrix  $X$  will have efficient and sparse representations in a redundant dictionary  $D$  by iteratively updating the sparse coefficients and dictionary atoms. The parametric dictionary learning methods mainly deal with the applications that do not necessitate the general forms of dictionary atoms by building the dictionary from a set of parametric functions. For example, one can enforce specific desired dictionary properties during the learning process such as dictionary coherence, shift invariance, or multiscale characteristics of the atoms [19].

In contrast, the clustering- or vector-quantization-based methods optimize a dictionary on a set of image patches by first grouping patterns such that their distance to a given atom is minimal, and then updating the atom such that the overall distance in the group of patterns is minimal [20]. This process assumes that each image patch can be represented by a single atom in the dictionary, and this reduces the learning procedure to a K-means clustering. A generalization of this method for dictionary learning is the K-singular value decomposition (K-SVD) algorithm [23], which represents each patch by using multiple atoms with different weights. In this algorithm, the coefficient matrix and basis matrix are updated alternatively. Suppose that the coefficient matrix  $\Lambda = [\alpha_1, \alpha_2, \dots, \alpha_N]$  has been updated through an OMP algorithm. Then, the update of dictionary  $D$ , with the  $k$ th atom  $d_k$  as an example, is as follows.

- 1) Define the groups of signal vectors that use this atom,  $\omega_k = \{i \mid 1 \leq i \leq N, \alpha_k^T(i) \neq 0\}$ , where  $\alpha_k^T$  is the  $k$ th row of  $\Lambda$ .
- 2) Compute the overall representation error matrix  $E_k$  by

$$E_k = X - \sum_{j \neq k} d_j \alpha_j^T. \quad (4)$$

- 3) Obtain  $E_k^R$  by choosing only the columns corresponding to  $\omega_k$ .

- 4) Apply SVD decomposition to  $E_k^R$

$$E_k^R = U \Delta V^T. \quad (5)$$

- 5) Update the dictionary column  $d_k$  to be the first column of  $U$ .

We adopted this K-SVD algorithm for dictionary learning in our method owing to its simplicity and efficiency during the training process. Accordingly, we adopted the OMP algorithm to solve the sparse coefficients in the image estimation procedure. The MATLAB toolbox for these algorithms can be found in [24].

### III. PROPOSED METHODOLOGY

Given the spatial and spectral properties of MODIS and Landsat as introduced in Section I, it is apparent that these two kinds of sensors are complementary. We therefore take the fusion of MODIS and Landsat as an example to illustrate the spatial and spectral fusion process of our algorithm. The spectral comparison between MODIS and Landsat is shown in Table I. It is clear that several bands of Landsat and MODIS (labeled with the same color) correspond with each other. However, some of the MODIS bands that are not within the Landsat bandwidth can play very important roles in many applications, for example, in monitoring air and marine environments. They could be even more useful if their spatial resolution could be largely improved, especially for those bands that currently have 1-km spatial resolution. For example, bands 8–11 can be used to calculate chlorophyll parameters; band 16 is informative about aerosols; and bands 20–23 are useful for monitoring temperature changes on land or ocean surfaces. To extend the applications of MODIS data, we propose to integrate the high spectral information of MODIS and the high spatial information of Landsat to produce data with both HSaR and HSeR.

For a given study area, we assume that MODIS data and Landsat data could be acquired on closer dates. After radiometric calibration, atmospheric correction, and geometric rectification, the surface reflectances of these two sensors are comparable and correlated with each other on their corresponding bands. However, due to the differences in sensor systems, systematic biases may still exist. The main purpose of our algorithm is to make use of the correlation by training a spectral dictionary from the known MODIS data and then predicting the HSaSeR image from the spectral dictionary

and the Landsat data, while at the same time, minimizing the system biases. As the Terra platform crosses the equator at around 10:30 AM local solar time, roughly 30 min later than Landsat 7, and their orbital parameters, viewing angles (near-nadir), and solar geometries are close to each other, the imaging environments of MODIS and Landsat ETM+ can usually be assumed to be similar. However, the imaging conditions could vary at any time and there are differences between the spectral response functions (SRFs) of MODIS and ETM+, resulting in a difference in the spectral signatures of these two sensors. We alleviate this problem by matching the histogram of ETM+ data to that of MODIS data in the preprocessing step. Histogram matching is a relative detector calibration technique to balance the responses of different detectors. It is commonly used to normalize two images acquired at similar local illumination over the same location, but by different sensors, atmospheric conditions, or global illumination. By changing only the contrast of reflectance, the histogram matching operation barely affects the spatial information of ETM+ data and keeps the spectral information of MODIS data. Because the SRF of MODIS is relatively continuous and narrow compared with that of ETM+, we can approximate the spectral signatures of Landsat ETM+ with a linear combination of pure spectral signatures of MODIS. This indicates that the spectral dictionary derived from MODIS data can be used to represent the Landsat data by integrating over the corresponding spectral range.

For convenience of description, we will refer to the MODIS data as images with HSeR and the Landsat data as images with HSaR. We define the input HSeR image as  $X_{lh} \in R^{w \times h \times S}$ , the input HSaR image as  $X_{hl} \in R^{W \times H \times s}$ , and the desired HSaSeR image as  $X_{hh} \in R^{W \times H \times S}$ . The subscripts of  $X$ , l, and h, stand for “low” and “high,” respectively, and their order denotes the spatial and spectral properties, respectively.  $W$  (or  $w$ ),  $H$ , (or  $h$ ) and  $S$  (or  $s$ ) denote the image width, the image height, and the number of bands, respectively ( $w < W$ ,  $h < H$ , and  $s < S$ ). We made two reasonable assumptions in this paper according to the following two aspects. On the one hand, reconstructing  $X_{hh}$  from  $X_{lh}$  and  $X_{hl}$  is a highly underconstrained problem without further assumptions, because the total number of observations from the input data is much smaller than that of our desired measurements, that is,  $whS + WHs \ll WHS$ . On the other hand, modeling the relationship between the desired variable and the given variables needs proper assumptions. For the following two assumptions, the first one aims to bring up the sparsity constraint for both spectral dictionary and representation coefficients; and the second one aims to link the desired HSaSeR image matrix with the given HSeR and HSaR image matrices.

#### A. Assumptions

1) *Assumption 1:* The mixed pixels can be expressed in the form of linear combinations of a small number of pure spectral signatures in the trained dictionary.

Mixed pixels are common on a HSeR image due to its LSaR. To unmix mixed pixels, a linear mixing model and a nonlinear

mixing model have been proposed [25]. The linear mixing model assumes the minimal secondary reflections or multiple scattering effects in the data collection procedure, thus the measured spectra can be expressed as a linear combination of the spectral signatures of the materials present in the mixed pixel. In contrast, nonlinear unmixing generally requires prior knowledge of the object geometry and the physical properties of the observed objects [25]. Due to its computational tractability and its flexibility for different applications [25], we used a linear mixing model in this paper. As the number of endmembers participating in a mixed pixel is usually very small compared with the total number of material categories in an image, a mixed pixel can be expressed by a linear combination of pure spectral signatures from the dictionary with a sparse fractional abundance vector. If we assume that the number of distinct materials in the scene is  $M$  and denote the measured spectrum of a pixel by  $x \in R^S$ , the spectral library as  $D \in R^{S \times M}$ , and the fractional abundance vector as  $\alpha \in R^M$ , then the linear model can be written as follows:

$$x = D\alpha + n \quad (6)$$

where  $n$  is an  $S \times 1$  vector containing the errors affecting the measurements at each spectral band.

2) *Assumption 2:* The mappings from the desired HSaSeR image to the input HSeR image and HSeR image are both linear, which means that the pixel vectors of both HSeR and HSeR images can be obtained by the linear combinations of the pixel vectors in the HSaSeR image.

If the pixels in an HSaSeR image are considered pure pixels, then the pixels in an HSeR image can be obtained via a linear mixing model as in Assumption 1. If we concatenate all pixel vectors into a matrix, the mapping from the desired HSaSeR image to the HSeR image can be linearly obtained. On the other hand, the digital number of remote sensing data is determined by the total radiation rate of each band, which can be obtained by means of integrating the solar spectrum function  $L(\lambda)$  and the SRF  $f(\lambda)$  of sensors over the band spectral range  $\Omega$

$$DN = \int_{\Omega} L(\lambda) f(\lambda) d\lambda + \varepsilon \quad (7)$$

where  $\lambda$  is the wavelength and  $\varepsilon$  is the noise caused by the dark signal. For the same scene imaged by different sensors,  $L(\lambda)$  is the same, whereas  $f(\lambda)$  is dependent on the sensors. Therefore, the surface reflectance discrepancy between sensors is determined by their SRFs. Considering that the band spectrum of HSeR is narrow and relatively continuous and the band spectrum of HSaR is wide, we approximated the SRF of HSaR with the weighted sum of the SRFs of HSeR. For example, the SRF of the eighth band (i.e., the panchromatic band) in Landsat 7 ETM+ can be approximated by the weighted sum of the SRFs of bands 1–2 and 11–17. The spectral response curves (SRCs) of these bands are shown in Fig. 1, where the values of the  $x$ -axis represent the normalized reflectances and the values of the  $y$ -axis the wavelengths in micrometers. Other bands of ETM+, except for bands 5 and 7, which have corresponding bands in MODIS (bands 6 and 7,

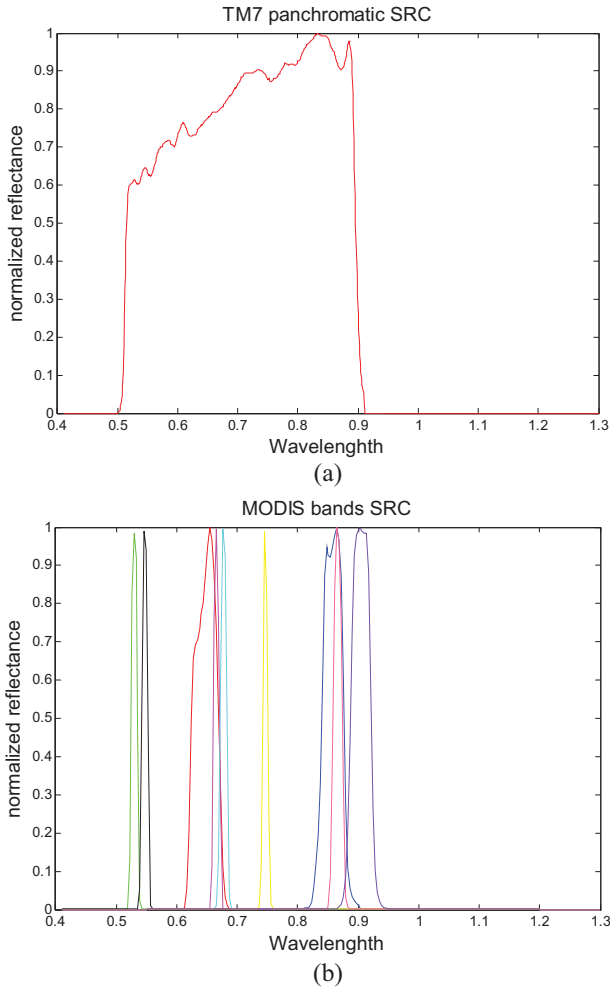


Fig. 1. Comparison of SRCs between the panchromatic band of Landsat and the bands 1–2 and 11–17 of MODIS. (a) SRCs of Landsat. (b) SRCs of MODIS.

respectively), can be approximated by two or three related bands in MODIS that can be found in Table I.

As the SRF of the HSaSeR image is consistent with that of the HSeR, the DN value of one band in the HSaR image can be approximated as the weighted sum of DN values of those HSaSeR bands whose spectra are overlapped with this HSaR band. For a location  $(i, j)$  in the HSaR image, if we put the pixels of all the bands into a spectral vector  $x_{hl} \in R^S$  and denote its counterpart spectral vector in the HSaSeR image as  $x_{hh} \in R^S$ , then  $x_{hl}$  can be linearly approximated by  $x_{hh}$  by introducing a mapping matrix  $\Phi \in R^{S \times S}$  formula as follows:

$$x_{hl} = \Phi x_{hh}. \quad (8)$$

As the mapping matrix  $\Phi$  in (8) is valid for other spectral vectors, (8) can also be expressed in the matrix form:  $X_{hl} = \Phi X_{hh}$ , where each column of  $X_{hl}$  and  $X_{hh}$  stands for a pixel vector.

### B. Model Formulation

According to Assumption 1, the desired HSaSeR image can be decomposed as  $X_{hh} = D\Lambda$ , where the  $i$ th column of  $X_{hh} \in R^{S \times WH}$ ,  $D \in R^{S \times M}$ , and  $\Lambda \in R^{M \times WH}$  denote the  $i$ th spectral

vector, the pure spectral signature of the  $i$ th material in the scene, and the sparse fractional abundance of the  $i$ th spectral vector in respect of all spectral signatures in the dictionary  $D$ , respectively. In the pixel vector form, the  $i$ th spectral vector of  $X_{hh}$  in location  $(i, j)$  can be represented as

$$x_{hh}(i, j) = D\alpha_{hh} \quad (9)$$

where  $\alpha_{hh} \in R^M$  is the fractional abundance vector for pixel vector  $x_{hh}(i, j)$  in respect of the spectral dictionary  $D$ . In line with the linear mapping process from the HSaSeR image to HSeR image in Assumption 2, the pixel vector  $x_{lh}(i, j) \in R^S$  of  $X_{lh}$  can be obtained by spatial aggregation within a local window

$$x_{lh}(i, j) = \sum_{(r,c) \in W_{i,j}} \theta_{rc} x_{hh}(r, c) \quad (10)$$

where  $W_{i,j}$  is an appropriate window of the HSaSeR image  $(r, c)$  denotes the pixel location in this window, and  $\theta_{rc}$  are the weighting coefficients in the window. By combining (9) and (10),  $x_{lh}(i, j)$  can be denoted as

$$x_{lh}(i, j) = D \sum_{r,c \in W_{i,j}} \theta_{rc} \alpha_{hh}(r, c) = D\varphi(i, j) \quad (11)$$

where  $\varphi(i, j) \in R^M$  is the summation of all fractional coefficients in window  $W_{i,j}$ . Since  $\alpha_{hh}$  in each location is very sparse with respect to dictionary  $D$ , the spatial aggregated coefficient  $\varphi(i, j)$  is also sparse with respect to dictionary  $D$ . If we concatenate (9) for all pixels  $(i, j)$ , the matrix equation can be derived

$$X_{lh} = D\Gamma \quad (12)$$

where  $X_{lh} = [x_{lh}(1, 1), \dots, x_{lh}(w, h)] \in R^{S \times wh}$  and  $\Gamma = [\varphi(1, 1), \dots, \varphi(w, h)] \in R^{M \times wh}$ .

The spectral dictionary and the fractional abundance matrices are initially unknown in (12). However, for scenes that obey our above assumptions, we know that the  $M$  columns of  $D$  stand for the pure spectral signatures of  $M$  categories of materials in the scene, and that fractional abundance vectors are sparse. Hence, the two matrices can be solved by using the sparse matrix factorization method discussed in Section II. Although there are many possible decompositions of (10), under fairly mild conditions of the sparsest coefficients, such decompositions are unique [26]. We learn the dictionary  $D$  by using the K-SVD algorithm to solve the following objective function:

$$\min_{D, \Gamma} \left\{ \|X_{lh} - D\Gamma\|_2^2 \right\} \quad \text{s.t.} \quad \forall i, \|\varphi_i\|_0 \leq M_0 \quad (13)$$

where  $\Gamma = [\varphi_1, \varphi_2, \dots, \varphi_{wh}]$  and  $M_0$  is a fixed and predetermined parameter.

Once the spectral dictionary  $D$  is obtained from (13), we can reconstruct the HSaSeR image  $X_{hh}$  by using the sparsity as a guide. Assuming that  $D$  has been correctly estimated, each pixel vector of the target image should admit a sparse representation in terms of the columns of  $D$ :  $x_{hh} = D\alpha$ . By combining (8), that is,  $x_{hl} = \Phi x_{hh}$ , we seek the sparsest coefficients  $\alpha$  that satisfy the measurement equation  $x_{hl} = \Phi x_{hh} = \Phi D\alpha$

$$\hat{\alpha} = \arg \min_{\alpha} \|\alpha\|_0, \quad \text{s.t.} \quad \|x_{hl} - \Phi D\alpha\|_2^2 \leq \varepsilon \quad (14)$$

where  $\varepsilon$  is a small constant, denoting the accepted error tolerance during the estimation process. After solving this sparse coding problem by means of the OMP algorithm, we can then use the estimated fractional coefficients  $\hat{a}$  to form our estimate of  $x_{hh}$

$$x_{hh} = D\hat{a}. \quad (15)$$

As the image  $X_{hl}$  is known, all pixel vectors of  $X_{hh}$  can be predicted in the same process.

### C. Estimation of the Mapping Matrix $\Phi$

From (14), we know that the mapping matrix  $\Phi$  from the HSASeR image to the HSAr image plays an important role in predicting our desired image. One straightforward way to estimate it is through a linear regression between the SRFs of the HSAr and HSeR sensors. However, given the system biases that occur during the data capture procedures, we predict the mapping matrix directly from the known data themselves to obtain a more accurate and robust estimation.

From (8), we can see that this mapping process is executed on two datasets with the same spatial extent and different spectral extent, whereas the two known images,  $X_{hl} \in R^{s \times WH}$  and  $X_{lh} \in R^{S \times wh}$ , are different both in spatial extent and spectral extent. Thus, we downsample the image  $X_{hl}$  in the spatial extent to render it the same spatial extent as  $X_{lh}$ . If we denote the downsampled version of  $X_{hl}$  as  $X_{ll} \in R^{s \times wh}$ , then  $X_{lh}$  and  $X_{ll}$  can be connected by the mapping matrix  $\Phi$  with the following method:

$$X_{ll} = \Phi X_{lh} \quad (16)$$

where  $\Phi \in R^{s \times S}$ . Consequently, the mapping matrix  $\Phi$  can be obtained ( $X_{lh}$  has full row rank)

$$\Phi = X_{ll}X_{lh}^+ = X_{ll}X_{lh}^T (X_{lh}X_{lh}^T)^{-1}. \quad (17)$$

### D. Algorithm

Based on the above assumptions and the estimation model, we illustrate the overall algorithm of the SASFM in Algorithm 1. The algorithm consists of two stages, that is, deriving the spectral dictionary from the HSeR image in the first stage and estimating the HSASeR image by representing the HSAr image using the spectral dictionary in the second stage. From this procedure, we can observe that the extraction of spectral information from the HSeR image and the extraction of spatial information from the HSAr image are implemented separately. Thus, there is no high requirement on geometric registration between the input images. However, the estimation of mapping matrix  $\Phi$  requires the geometric coregistration between the input HSAr and HSeR images because they are used simultaneously in (16). If we estimate the mapping matrix  $\Phi$  through a linear regression between the SRFs of the HSAr and HSeR sensors, there would be no need to establish a direct relationship between the input HSAr and HSeR images. This indicates a relaxation of the requirement of an excellent geometric coregistration between the input HSAr and HSeR images.

---

### Algorithm 1 Overall Spatio-Spectral Fusion Algorithm

---

#### Input:

- . HSAr but low spectral resolution image  $X_{hl}$ .
- . LSAr but HSeR image  $X_{lh}$ .

#### Deriving the spectral dictionary $D$ :

- . Derive the spectral dictionary from  $X_{lh}$  via K-SVD by solving (13).

#### Estimating the mapping matrix $\Phi$ :

- . Derive it by downsampling  $X_{hl}$  and solving (16) and (17).

#### Sparse coding:

- . Compute the sparse fractional abundances of  $X_{hl}$  with respect to spectral dictionary  $D$  through the sparse coding technique of OMP as in (14).

#### Reconstruction:

- . Reconstruct the HSASeR image from the sparse fractional abundances of  $X_{hl}$  and the spectral dictionary  $D$  as in (15).

#### Output

- . High spatial and HSeR image  $X_{hh}$ .
- 

## IV. EXPERIMENTAL RESULTS AND COMPARISONS

To test the performance of our proposed method, we applied it to both the simulated data and actual remote sensing data from Hong Kong, and compared it with several other spatial and spectral fusion methods, including the component substitution method (CSM) [27], local principal component analysis (L-PCA) [28], and the well-known SaU method developed by Zurita-Milla *et al.* [13]. In our algorithm, we set the noise level  $\varepsilon$  as  $1.0e-5$ , and the number of atoms  $M$  was tuned based on the experimental data. All the experiments were conducted on the MATLAB R2010a platform on a personal computer with a 2.10-GHz CPU and a 2.00-GB RAM. For the simulated data, the HSASeR image was already given and the input HSAr and HSeR images were aggregated from the given HSASeR image. We could thus evaluate the fusion results by comparing them with the given HSASeR image based on the following four commonly used quality metrics. The first one is relative average spectral error (RASE) [29], which characterizes the average performance of a method on all hyperspectral bands as follows:

$$\text{RASE} = \frac{100}{\mu} \sqrt{\frac{1}{S} \sum_{i=1}^S \text{RMSE}(X_i)^2} \quad (18)$$

where  $\mu$  is the mean value of the  $S$  spectral bands ( $X_i$ ) of the original multispectral image. The rule for this index is, the lower the value, the better the method. The second metric is the peak signal noise ratio (PSNR) [30], which is commonly used for measuring the restoration degree of images, with a higher value indicating a better performance of the algorithm. To measure the spatial details of the estimated image, we utilized the third metric, universal image quality index (UIQI) [31], which is defined as

$$\text{UIQI} = \frac{4\sigma_{\hat{X}\hat{X}}\mu_X\mu_{\hat{X}}}{(\sigma_X^2 + \sigma_{\hat{X}}^2)(\mu_X^2 + \mu_{\hat{X}}^2)} \quad (19)$$

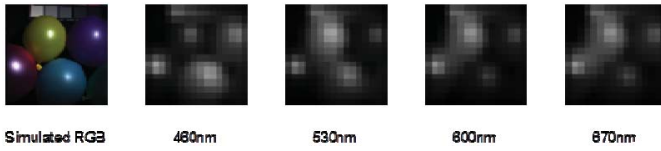


Fig. 2. Input images: simulated RGB image ( $512 \times 512$  pixels) and some scenes of the hyperspectral image ( $16 \times 16$  pixels).

where  $\sigma_{X\hat{X}}$  is the covariance of the original image  $X$  and the estimated image  $\hat{X}$ ;  $\mu_X$  and  $\sigma_X$  are the mean and variance of  $X$ , respectively;  $\mu_{\hat{X}}$  and  $\sigma_{\hat{X}}$  are the mean and variance of  $\hat{X}$ , respectively. The dynamic range of UIQI is  $[-1, 1]$ , with larger value indicating a better estimation of the image. The fourth metric for the simulated data is spectral information divergence (SID) [32] that measures the spectral similarity between each actual pixel vector  $x$  and its corresponding estimated pixel vector  $\hat{x}$

$$\text{SID}(x, \hat{x}) = \sum_{s=1}^S x_s \log\left(\frac{x_s}{\hat{x}_s}\right) + \sum_{s=1}^S \hat{x}_s \log\left(\frac{\hat{x}_s}{x_s}\right) \quad (20)$$

where  $x$  and  $\hat{x}$  are normalized vectors and  $S$  is the number of bands of image  $X$ . The rule for SID is, the lower the value, the better the method. For the actual remote sensing images (i.e., Landsat ETM+ and MODIS in our case), if we used the original MODIS image (with a spatial resolution of 250 m–1 km) as the ground truth and generated the input HSaR and HSeR images by aggregating ETM+ and MODIS, respectively, as in pan-sharpening for result validation, it would be uncircumventable to extract the endmembers or representation atoms and to compare the superiority of algorithms in improving the spatial resolution. We thus evaluated the fusion results of actual remote sensing images by using the spatial correlation coefficient (CC) with the input HSaR image and the spectral correlation coefficient with the input HSeR image. It should be noted that the average metric values of all the bands were used to compare the proposed algorithm with other algorithms.

#### A. Experiments With Simulated Data

Our experiments used the spectral image database described in [33], which provides the reflectance of the materials in the scene from 400 to 700 nm, at 10-nm intervals (31 bands in total). Each band of this dataset is stored in 16-bit grayscale png format, with an image size of  $512 \times 512$  pixels.

To generate a hyperspectral dataset with LSaR, we down-sampled the 31 bands with a scale factor of 32 (i.e.,  $16 \times 16$  pixels). At the same time, the RGB data with HSaR and LSeR were simulated by the three bands, 650–660 nm, 520–530 nm, and 440–450 nm, which are the closest of the 31 bands to the RGB spectrum. Some images from these two simulated dataset are shown in Fig. 2.

To implement our proposed method on the simulated dataset, we fine-tuned the number of atoms  $M$  to 20 and

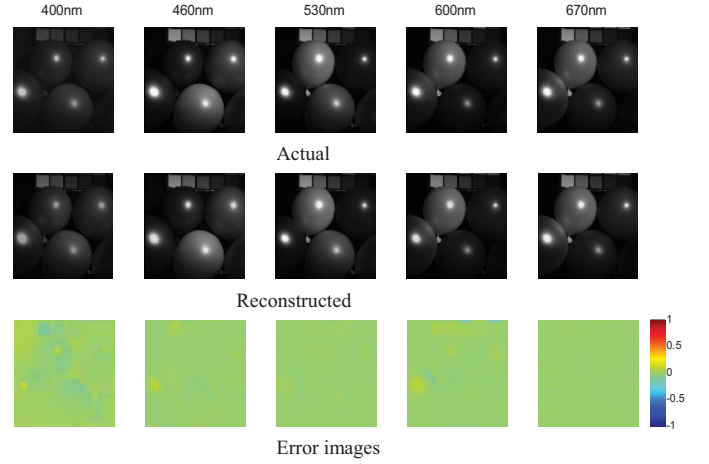


Fig. 3. Reconstructed results and error images.

TABLE II  
EVALUATION RESULT BASED ON SOME QUALITY METRICS

	RASE	PSNR	UIQI	SID
Ideal Value	0	$+\infty$	1	0
CSM	67.36	20.34	0.6666	0.126
L-PCA	11.26	35.31	0.9921	0.232
SaU	27.47	28.13	0.9672	5.994
SASFM	<b>10.95</b>	<b>36.12</b>	<b>0.9926</b>	<b>0.073</b>

generated the following mapping matrix:

$$F = \begin{bmatrix} \dots & \overbrace{0}^5 & \dots & \overbrace{0}^{13} & \dots & \overbrace{1}^{27} & \dots \\ \dots & 0 & \dots & 1 & \dots & 0 & \dots \\ \dots & 1 & \dots & 0 & \dots & 0 & \dots \end{bmatrix}_{3 \times 31}.$$

Five of the reconstructed images generated by our method and the corresponding error images are shown in Fig. 3. Comparing the reconstructed images with the actual images, there seems to be almost no difference. From the error images, we can find that the reconstruction error is small. Additionally, it is easy to see that the errors are larger around specular pixels that change rapidly with respect to the image resolution. We also compared our algorithm with other representative algorithms (i.e., CSM, L-PCA, and SaU) using the quality metrics (i.e., RASE, PSNR, UIQI, and SID) (Table II). For the SaU method, the window size and the number of classes were fine-tuned to be 15 and 20, respectively. The comparison results show that our method is the best on all indices.

#### B. Experiments With Actual Satellite Data

In this section, we apply our proposed algorithm to the fusion of Terra/Aqua MODIS and Landsat 7 ETM+ datathat were acquired on November 17–24, 2001 (composite image) and November 20, 2001, respectively. Due to the short interval, they were considered to have the similar acquisition time. We employed the composite image of the MOD09A1 product (MODIS surface reflectance 8-day L3 global 500 m) and Level 1 product. After converting the Level 1 product to reflectance values by using the MODTRAN5 model, these two products were registered together geometrically. The Landsat



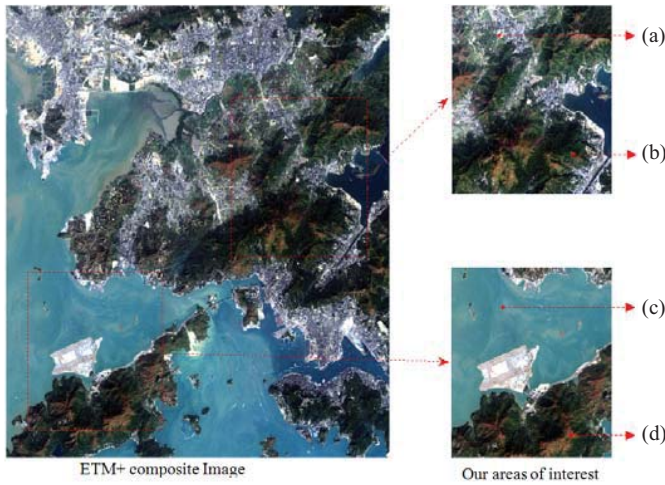


Fig. 4. Left: original ETM+ composite image with bands 3-2-1 as red-green-blue. Right: two areas of interest. (a)–(d) Four representative pixels used in Fig. 8.

ETM+ image was radiometrically and atmospherically corrected using the MODIS 6S approach. Then, the MODIS data were geometrically registered to the Landsat ETM+ data with the coregistration error of three pixels and the histogram of ETM+ data was matched to that of the MODIS data. The study region covers an area of  $15 \times 15$  km in Hong Kong. The Landsat scene with bands 3-2-1 as red-green-blue composite is shown in Fig. 4 that illustrates our two study areas.

The MODIS provides 36 bands covering the spectrum of visible, NIR, 352086318SWIR, MWIR, and TIR with a spatial resolution of 250 m (1–2 bands), 500 m (3–7 bands), and 1 km (the other 29 bands). The Landsat ETM+ sensor provides six multispectral bands, one panchromatic band, and one thermal infrared band with spatial resolutions of 30, 15, and 60 m, respectively. To generate data with the spatial resolution of Landsat and the spectral resolution of MODIS, we fused these two datasets using the proposed method and compared it with other representative methods. For processing convenience, all MODIS bands were upsampled to be 240 m, using bicubic interpolation; similarly, the panchromatic and thermal Landsat bands were preprocessed to be 30 m.

For our method, the mapping matrix was computed according to the downsampling strategy introduced in Section III-C, and the number of atoms  $M$  was set as 30. For the SaU method, the window size and the number of classes were fine-tuned to be 25 and 30, respectively. For our method, we first analyzed the spectral dictionary extracted from the MODIS image. Consider the first area of interest in Fig. 4 as an example. The spectral profiles of 30 normalized dictionary atoms are shown in Fig. 5(a) with the order from left to right and then from top to bottom. From Fig. 5(a) we can observe that the spectral shapes of the atoms from bands 1–15 show a larger variance compared with bands 16–36. To demonstrate the linear combination procedure of dictionary atoms during the fusion process, we take one vegetation pixel in the fused image [see Fig. 5(b) with NIR-red-green as the R-G-B composite] as an example. With the corresponding input Landsat pixel in (14), the sparse

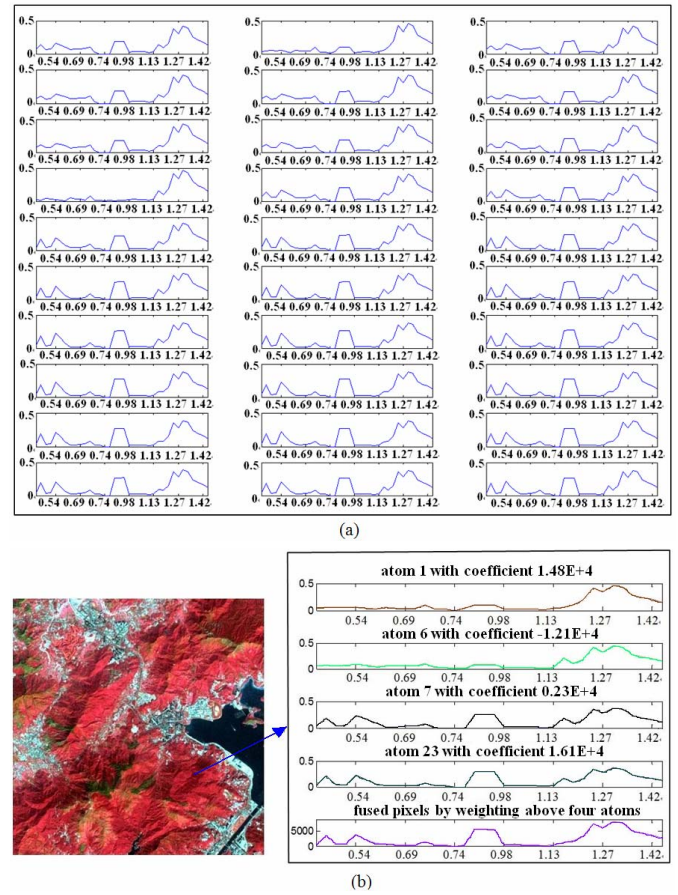


Fig. 5. Illustration of (a) spectral dictionary extracted from the MODIS image (the x-axis represents the spectral wavelength with units of micrometers) and (b) linear combination procedure of dictionary atoms during the fusion process.

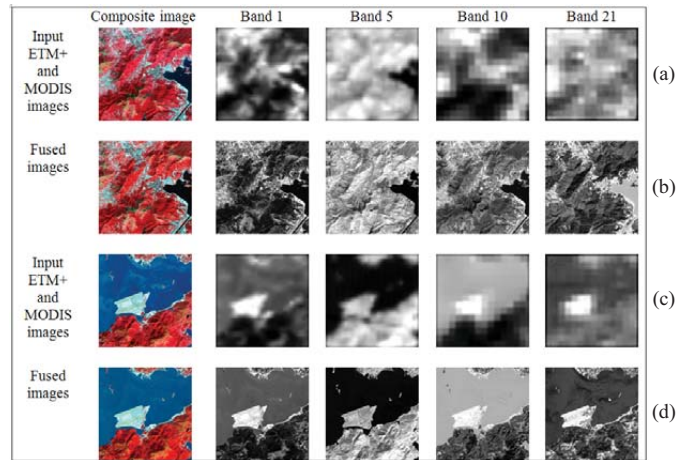


Fig. 6. Inputs and fused results of the two areas. (a) and (c) Input images including Landsat ETM+ (3 out of 8 bands) and MODIS (4 out of 36 bands). (b) and (d) Fused results. The composite image is bands 2-1-4 as R-G-B.

coefficients with respect to the above spectral dictionary can be derived by the OMP algorithm. Then, dictionary atoms 1, 6, 7, and 23 are derived with appropriate coefficients to represent the corresponding Landsat pixel after spectral downsampling via mapping matrix  $\Phi$  [see (14)]. Consequently, the fused



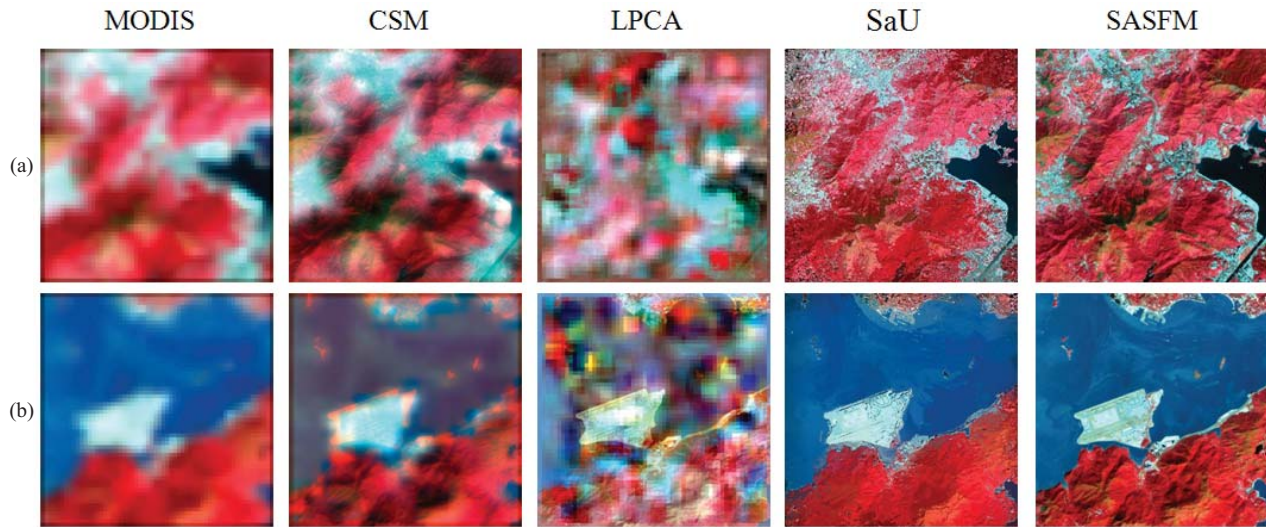


Fig. 7. Comparison of fusion results on the two areas of interest. (a) AOI\_1. (b) AOI\_2. From left to right, the composite images are: the MODIS image and the fusion results of CSM, L-PCA, SaU, and our method.

TABLE III

COMPARISON OF SPATIAL CCs AND SPECTRAL CCs ON AOI\_1 AND AOI\_2, RESPECTIVELY (BOLD NUMBERS INDICATE THE BEST RESULTS)

AOI	Method	Spatial CC			Spectral CC				
		B2	B1	B4	B1	B5	B10	B21	Mean (36 bands)
AOI_1	CSM	0.6750	0.4684	0.4452	<b>0.8271</b>	0.6055	<b>0.7123</b>	<b>0.5234</b>	<b>0.7283</b>
	LPCA	0.3044	0.3738	0.3249	0.1179	0.2138	0.2003	0.1011	0.2204
	SaU	0.4586	0.4893	0.5935	0.6649	0.6755	0.6551	0.3360	0.4325
	SASFM	<b>0.9163</b>	<b>0.8731</b>	<b>0.9768</b>	0.7288	<b>0.6926</b>	0.5911	0.3521	0.4529
AOI_2	CSM	0.8804	0.0701	0.3041	<b>0.8183</b>	0.7311	<b>0.6223</b>	<b>0.8731</b>	<b>0.7652</b>
	LPCA	0.3821	0.5820	0.5792	0.6697	0.4329	0.1031	0.3227	0.4531
	SaU	0.1791	0.1485	0.2714	0.7196	0.8148	0.5179	0.7124	0.6231
	SASFM	<b>0.8956</b>	<b>0.7994</b>	<b>0.9724</b>	0.7208	<b>0.8227</b>	0.4814	0.6932	0.6345

pixel can be predicted by linearly combining these four atoms with their coefficients, which are shown in the right part of Fig. 5(b), via (15). From Fig. 5(b), we can find that the atoms possessing the major spectral properties of the fused pixels are assigned larger coefficients automatically by the OMP algorithm, for example, atoms 1 and 23.

Using the proposed method, the reconstructed bands 1, 5, 10, and 21 are shown in Fig. 6 for the two areas of interest. Due to space limitations, only the representative bands are shown here. Band 21 is outside of ETM+'s spectral coverage. From the comparison of the input Landsat composite image (the first one in row one and row three) with the fused composite image (the first one in row two and row four), we can see that there is almost no difference in either the spatial or the spectral properties. From the comparison between the input MODIS bands and the fused bands, we can conclude that the spatial resolution is greatly improved with almost no spectral distortion.

The fusion results from our method, CSM, L-PCA, and SaU are shown in Fig. 7. It can be found that our method is the best in preserving spatial details. This is mainly because the spatial features of HsAR data are better preserved by sparsely

representing the spectral signatures with respect to the learned spectral dictionary. To quantitatively evaluate the fusion result, the spatial CCs and the spectral CCs were computed for the fusion result of both methods. Since the spectral information of the fused image is derived from the MODIS image, the spectral band profiles of the fused image are consistent with those of the MODIS image. The spatial CCs were calculated between the fusion result and the input ETM+ image on their corresponding bands and the spectral CCs were calculated between the fused bands and the resampled MODIS bands. The spatial CCs between bands 2, 1, and 4 of the fusion result and bands 4, 3, and 2 of the ETM+ image (see Fig. 7) and the spectral CCs between all the bands of the fusion result and the resampled MODIS image (see Fig. 6) were calculated. The comparison of the spatial and spectral CCs is shown in Table III on both areas of interest (AOI\_1 and AOI\_2 as shown in Fig. 4). For spectral CCs, bands 1, 5, 10, and 21 serve as examples, and the mean values of all the 36 bands are listed for comparisons. It can be found that our method outperforms other algorithms on preserving spatial details over all the bands. However, it is weaker than CSM in preserving spectral information because the CSM

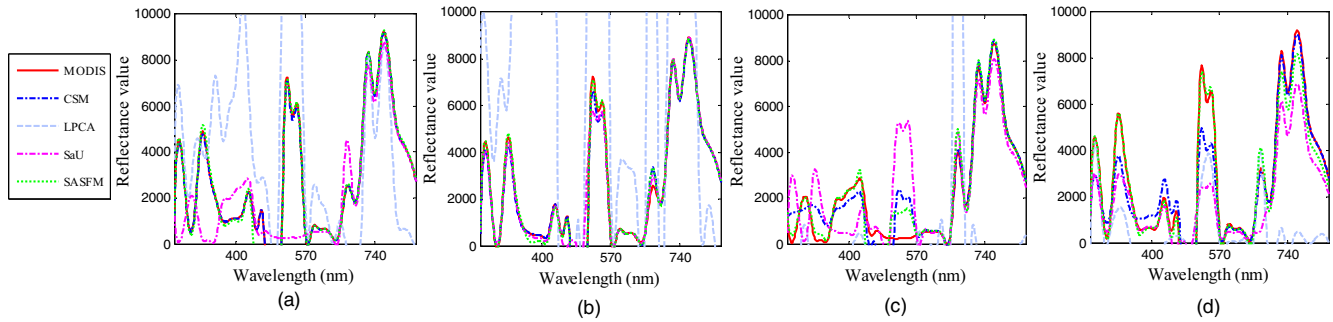


Fig. 8. Comparison of spectral profiles generated using different methods. From left to right, (a)–(d) show the spectral profiles of pixels (a)–(d) in Fig. 4, respectively.

method replaces the high frequency part of the MODIS data directly by that of the ETM+ data in the PCA transform space. It should be noted that the performed preprocessing operations might have negatively affected the unmixing data fusion results of the SaU method. To further evaluate the performance of our method, we chose four representative endmembers in the Landsat ETM+ image [see the four points represented by red dots in Fig. 4(a)–(d)], and then plotted spectral profiles for the four points between the fused bands and the input MODIS bands; these are shown in Fig. 8. Clearly, the spectral profile of our method is closer to that of MODIS.

## V. CONCLUSION

We developed an efficient SASFM based on the sparse matrix factorization theory. This model is applicable to the fusion of two categories of remote sensing data: one category has L<sub>SaR</sub>, wide spectrum coverage, and H<sub>SeR</sub>, and the other category has H<sub>SaR</sub>, narrow spectrum coverage, and L<sub>SeR</sub>. Our model generates synthetic images with the spectral resolution of the first category data and the spatial resolution of the second category data. Under two reasonable assumptions, the SASFM learns the spectral dictionary by representing the given L<sub>SaR</sub> data sparsely with respect to the dictionary, and then predicts the desired H<sub>SaSeR</sub> data by establishing a linear representation relationship with the given H<sub>SaR</sub> data. Comparisons with other methods using both simulated data and actual Landsat/MODIS data validated the effectiveness and superiority of this method.

Although the SASFM was used to fuse Landsat 7 ETM+ and MODIS data, it can be applied to other remote sensing instruments with complementary spatial and spectral resolutions, such as MERIS and Landsat data. Our method still has room for improvement. One issue in the spectral dictionary learning process of our method is that the spectrum of the H<sub>SaR</sub> data may not be completely covered by that of the H<sub>SeR</sub> data. This may cause spectral distortion in the fused H<sub>SaSeR</sub> data. Certainly, when the spectrum of the given H<sub>SeR</sub> data has a high level of overlap with that of the given H<sub>SaR</sub> data, this problem can be alleviated. Another issue with our sparse matrix factorization method is that the computational complexity is relatively high compared with that of other

nonlearning based methods. Future improvements in the speed of dictionary learning and sparse coding algorithms and the use of parallel computing are expected to solve this problem. In addition, the preprocessing of medium spatial resolution imagery (see [34]) is of importance to the fusion accuracy. Thus, an analysis of the preprocessing effects on the fusion results may be our another future work.

## REFERENCES

- [1] F. Deng, J. M. Chen, S. Plummer, M. Z. Chen, and J. Pisek, "Algorithm for global leaf area index retrieval using satellite imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 8, pp. 2219–2229, Aug. 2006.
- [2] C. Li, A. Lau, J. Mao, and D. Chu, "Retrieval, validation and application of the 1-km aerosol optical depth from MODIS measurements over Hong Kong," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 11, pp. 2650–2658, Nov. 2005.
- [3] B. L. Markham, J. C. Storey, D. L. Williams, and J. R. Irons, "Landsat sensor performance: History and current status," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 12, pp. 2691–2694, Dec. 2004.
- [4] L. Alparone, L. Wald, J. Chanussot, C. Thomas, P. Gamba, and L. Mann 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.
- [5] Z. Wang, D. Ziou, and C. Armenakis, "A comparative analysis of image fusion methods," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 6, pp. 1391–1402, Jun. 2005.
- [6] W. J. Carper, T. M. Lilles, and R. W. Kiefer, "The use of intensity-hue-saturation transformations for merging SPOT panchromatic and multispectral image data," *Photogram. Eng. Remote Sens.*, vol. 56, no. 4, pp. 459–467, 1990.
- [7] V. K. Shettigara, "A generalized component substitution technique for spatial enhancement of multispectral images using a higher resolution data set," *Photogram. Eng. Remote Sens.*, vol. 58, no. 5, pp. 561–567, 1992.
- [8] D. A. Yocky, "Multiresolution wavelet decomposition image merger of landsat thematic mapper and SPOT panchromatic data," *Photogram. Eng. Remote Sens.*, vol. 62, no. 9, pp. 1067–1074, 1996.
- [9] 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.
- [10] B. Zhukov, D. Oertel, F. Lanzl, and G. Reinhackel, "Unmixing-based multisensor multiresolution image fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1212–1226, May 1999.
- [11] A. Minghelli-Roman, M. Mangolini, M. Petit, and L. Polidori, "Spatial resolution improvement of MeRIS images by fusion with TM images," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 7, pp. 1533–1536, Jul. 2001.

- [12] A. Minghelli-Roman, L. Polidori, S. Mathieu-Blanc, L. Loubersac, and F. Cauneau, "Spatial resolution improvement by merging MERIS-ETM images for coastal water monitoring," *IEEE Trans. Geosci. Remote Sens. Lett.*, vol. 3, no. 2, pp. 227–231, Apr. 2006.
- [13] R. Zurita-Milla, J. G. P. W. Clevers, and M. E. Schaepman, "Unmixing-based landsat TM and MERIS FR data fusion," *IEEE Trans. Geosci. Remote Sens. Lett.*, vol. 5, no. 3, pp. 453–457, Jul. 2008.
- [14] R. Zurita-Milla, G. Kaiser, J. G. P. W. Clevers, W. Schneider, and M. E. Schaepman, "Downscaling time series of MERIS full resolution data to monitor vegetation seasonal dynamics," *Remote Sens. Environ.*, vol. 113, no. 9, pp. 1874–1885, 2009.
- [15] M.-D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2014–2039, Jun. 2011.
- [16] S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 1999.
- [17] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM Rev.*, vol. 43, no. 1, pp. 129–159, 2001.
- [18] Y. Pati, R. Rezaifar, and P. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. 27th Annu. Asilomar Conf. Signals, Syst. Comput.*, vol. 1, Nov. 1993, pp. 40–44.
- [19] M. Elad, M. A. T. Figueriredo, 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.
- [20] R. Rubinstein, A. Bruckstein, and M. Elad, "Dictionaries for sparse representation modeling," *Proc. IEEE*, vol. 98, no. 6, pp. 1045–1057, Jun. 2010.
- [21] A. Yang, A. Ganesh, Y. Ma, and S. Sastry, "Fast  $\ell_1$ -minimization algorithms and an application in robust face recognition: A review," in *Proc. IEEE 17th Int. Conf. Image Process.*, Sep. 2010, pp. 1849–1852.
- [22] B. Olshausen and D. Field, "Sparse coding with an overcomplete basis set: A strategy employed by v1?" *Vis. Res.*, vol. 37, no. 23, pp. 3311–3325, Dec. 1997.
- [23] 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.
- [24] M. Elad. (May 2012). *Dictionary Learning and Sparse Representation Software Code*. [Online]. Available: <http://www.cs.technion.ac.il/~elad/software>
- [25] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 44–57, Jan. 2002.
- [26] J. Romberg, "Imaging via compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–31, Mar. 2008.
- [27] P. S. Chavez and A. Y. Kwarteng, "Extracting spectral contrast in Landsat Thematic Mapper image data using selective principle component analysis," *Photogram. Eng. Remote Sens.*, vol. 55, no. 3, pp. 339–348, 1989.
- [28] R. Kawakami, J. Wright, Y.-W. Tai, Y. Matsushita, and M. K. Ikeuchi, "High-resolution hyperspectral imaging via matrix factorization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 2329–2336.
- [29] 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.
- [30] Q. Huynh-Thu and M. Ghanbari, "Scope of validity of PSNR in image/video quality assessment," *Electron. Lett.*, vol. 44, no. 13, pp. 800–801, 2008.
- [31] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Mar. 2002.
- [32] C. I. Chang, "Spectral information divergence for hyperspectral image analysis," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, vol. 1, Aug. 1999, pp. 509–511.
- [33] F. Yasuma, T. Mitsunaga, D. Iso, and S. K. Nayar, "Generalized assorted pixel camera: Postcapture control of resolution, dynamic range, and spectrum," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2241–2253, Apr. 2010.
- [34] L. Gomez-Chova, R. Zurita-Milla, L. Alonso, J. Amorós-Lopez, L. Guanter, and G. Camps-Valls, "Gridding artifacts on medium-resolution satellite image time series: MERIS case study," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 7, pp. 2601–2611, Jul. 2011.



**Bo Huang** (A'12) received the Ph.D. degree in remote sensing and mapping from the Institute of Remote Sensing Applications, Chinese Academy of Sciences, Beijing, China, in 1997.

He is currently a Professor with the Department of Geography and Resource Management, The Chinese University of Hong Kong, Shatin, Hong Kong, where he is also the Associate Director of the Institute of Space and Earth Information Science. Prior to this, he held faculty positions at the Schulich School of Engineering, University of Calgary, Calgary, AB, Canada, from 2004 to 2006, and the Department of Civil Engineering, National University of Singapore, Singapore, from 2001 to 2004. His current research interests include broad, covering most aspects of geo information science, specifically unified image fusion for environmental monitoring, spatial/spatiotemporal statistics for land-cover/land-use change modeling, and multi objective spatial optimization for sustainable land-use planning. He is currently exploring precision remote sensing, convinced that this new paradigm will revolutionize the way multisensor remotely sensed data are fused and exploited to improve the performance and quality of various applications in the future.

Dr. Huang serves as the Executive Editor of *Annals of GIS* and on the editorial boards of several international journals, including the *International Journal of Geographical Information Science* and the *Journal of Remote Sensing*.



**Huihui Song** received the B.S. degree in technology and science of electronic information from Ocean University, Qingdao, China, in 2008, and the M.S. degree in communication and information system from the University of Science and Technology of China, Beijing, China, in 2011. She is currently pursuing the Ph.D. degree with the Department of Geography and Resource Management, Chinese University of Hong Kong.

Her current research interests include remote sensing image processing and image fusion.



**Hengbin Cui** received the B.Sc. degree in information and computer science from Xi'an Jiaotong University, Xi'an, China, in 2010. He is currently pursuing the M.Sc. degree in applied mathematics with the Institute for Information and System Science, School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, China.

His current research interests include sparse modeling and latent factor analysis.



**Jigen Peng** received the B.S. degree in mathematics from Jiangxi University, Jiangxi, China, in 1989, and the M.Sc. and Ph.D. degrees in applied mathematics and computing mathematics from Xi'an Jiaotong University, Xi'an, China, in 1992 and 1998, respectively.

He is currently a Professor with the School of Mathematics and Statistics, Xi'an Jiaotong University. His current research interests include nonlinear functional analysis and applications, data set matching theory, the machine learning theory, and sparse

information processing.



**Zongben Xu** received the M.Sc. degree in mathematics from Northwest University, Shaanxi, China, and the Ph.D. degree in applied mathematics from Xi'an Jiaotong University, Xi'an, China, in 1981 and 1987, respectively.

He is a mathematician and an expert in signal and information processing, and a member of the Chinese Academy of Sciences. He is currently a Professor of mathematics and computer science, the Director of the Institute for Information and System Sciences, and the Vice President of Xi'an Jiaotong

University. He is a Chief Scientist of the National Basic Research Program of China (973 Project). His current research interests include intelligent information processing, machine learning and theories in numerical modeling.

Dr. Xu delivered a sectional talk at International Congress of Mathematicians (ICM 2010) upon the invitation of the congress committee. He was a recipient of the National Natural Science Award of China in 2007 and the CSIAM Su Buchin Applied Mathematics Prize in 2008.