

Recovering Quantitative Remote Sensing Products Contaminated by Thick Clouds and Shadows Using Multitemporal Dictionary Learning

Xinghua Li, Huanfeng Shen, *Senior Member, IEEE*, Liangpei Zhang, *Senior Member, IEEE*, Hongyan Zhang, *Member, IEEE*, Qiangqiang Yuan, and Gang Yang

Abstract—With regard to quantitative remote sensing products in the visible and infrared ranges, thick clouds and accompanying shadows are an inevitable source of noise. Due to the absence of adequate supporting information from the data themselves, it is a formidable challenge to accurately restore the surficial information underlying large-scale clouds. In this paper, dictionary learning is expanded into the multitemporal recovery of quantitative data contaminated by thick clouds and shadows. This paper proposes two multitemporal dictionary learning algorithms, expanding on their KSVD and Bayesian counterparts. In order to make better use of the temporal correlations, the expanded KSVD algorithm seeks an optimized temporal path, and the expanded Bayesian method adaptively weights the temporal correlations. In the experiments, the proposed algorithms are applied to a reflectance product and a land surface temperature product, and the respective advantages of the two algorithms are investigated. The results show that, from both the qualitative visual effect and the quantitative objective evaluation, the proposed methods are effective.

Index Terms—Compressed sensing (CS), dictionary learning, land surface temperature (LST), multitemporal, quantitative remote sensing (QRS) product, reflectance, shadows, thick clouds.

I. INTRODUCTION

WITH the ongoing development of both economy and technology, quantitative products of remote sensing (RS) have been attracting more and more attention in all kinds of practical applications due to their physical significance. However, cloud cover and cloud shadow are two of

the most common noise sources for the majority of RS data in the range of the visible and infrared spectra. As a result, recovering the data contaminated by clouds and the attendant shadows is of great significance for many users' demands, such as target recognition, classification, segmentation, feature extraction, etc.

There has been a great deal of cloud removal algorithms developed for RS images. For example, several inpainting methods (just utilizing the information in the images themselves), including the techniques of nearest neighbor interpolation and partial differential equations [1], three different strategies for propagating the spectrogeometrical information [2], and bandelet transform [3], have all been applied to this cumbersome task. In addition, the methods designed for dead or missing pixels in [4]–[7] can also be used for cloud removal. Unfortunately, a drawback inherited by the inpainting-based nature is their sensitivity to the size of the missing area [2]. To break through this bottleneck, some scholars have resorted to multitemporal processing approaches such as mosaicking the cloud-free areas of multitemporal images [8], information cloning [9], fitting the temporal relationship [10], etc. In order to guarantee a smooth transition from cloud-contaminated area to cloud-free area, color adjustment, which brings about a certain distortion, is usually necessary.

There are, however, fewer cloud removal algorithms for quantitative remote sensing (QRS) products than for RS images. Kriging interpolation, a simple and mature method, has been used in this field by direct interpolation [11] and fitting [12]. However, it is only suitable for low-albedo areas such as water and dense vegetation areas. The Savitzky–Golay (SG) filter [13], a simplified least-squares-fit convolution for smoothing and computing derivatives of a set of consecutive values, has been used for the recovery of cloud contamination of the normalized difference vegetation index (NDVI) [14]. Harmonic analysis, also termed spectral analysis or Fourier analysis, decomposes a time-dependent periodic phenomenon into a series of sinusoidal functions, each defined by unique amplitude and phase values [15]. This approach has been widely applied to the recovery of missing QRS data [16]–[18]. Roerink *et al.* [19] reconstructed cloud-free NDVI using the harmonic analysis of time series (HANTS) algorithm. The SG filter and harmonic analysis can be used to interpolate missing data for multitemporal analysis, especially for long time-series and periodic data, which is also their intrinsic limitation for

Manuscript received October 11, 2013; revised December 4, 2013 and January 19, 2014; accepted February 18, 2014. Date of publication April 1, 2014. This work was supported by the National Natural Science Foundation of China under Grant 41271376, by the National Major Basic Research Development Program of P.R. China (973 Program) under Grant 2011CB707105, by the Program for Changjiang Scholars and Innovative Research Team in University under Grant IRT1278, and by the Wuhan Municipal Science and Technology Bureau under Grant 2013072304010825.

X. Li, H. Shen, and G. Yang are with the School of Resource and Environmental Sciences, Wuhan University, Wuhan 430079, China (e-mail: lixinghua5540@whu.edu.cn; shenhf@whu.edu.cn; love64080@163.com).

L. Zhang and H. Zhang are with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan 430079, China (e-mail: zlp62@whu.edu.cn; zhanghongyan@whu.edu.cn).

Q. Yuan is with the School of Geodesy and Geomatics, Wuhan University, Wuhan 430079, China (e-mail: qqyuan@sgg.whu.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.2014.2307354

most practical applications. In other words, if the time series is not long enough, these approaches cannot derive satisfactory results.

As the traditional methods are easily affected by the size of the cloud area, the multitemporal mosaicking/fitting algorithms easily cause distortions, and the harmonic analysis approaches need a large amount of temporal data, the approaches for retrieving data beneath thick clouds and shadows need improvements based on a novel framework. In recent years, the flourishing and popular compressed sensing (CS) theory, or the so-called compressed sampling, as proposed by Donoho [20], Candès and Tao [21], and Romberg *et al.* [22], has started a revolution in the signal/image processing field. By means of this theory, it is possible to reconstruct a compressible or sparse signal in a fixed transform domain with a very high accuracy from a small number of random measurements by solving a linear program [21]. Exploiting this advantage, Lorenzi *et al.* [23] proposed a CS solution for missing-area reconstruction in multispectral images under a compressive sensing perspective through a formulation within a multiobjective genetic optimization scheme. However, they only utilized two temporal images (a contaminated image and a cloud-free image), and the latent temporal correlations in multitemporal data were not utilized. Additionally, cloud removal for QRS products was not investigated.

Specifically, in our proposed methods, we restore the information beneath the thick clouds and shadows of QRS data using a series of multitemporal quantitative products. Two kinds of expanded multitemporal dictionary learning methods in respective KSVD (the method proposed in [24]) and Bayesian (the method used in [25]) frameworks are proposed. This paper makes several novel contributions. First, we introduce CS theory into the recovery of QRS products, and to the best of our knowledge, it is the first time that the CS theory has been used in the recovery of multitemporal cloudy quantitative data. Second, to make the best use of the temporal correlations in the time-series data, we propose to seek an optimized temporal path and temporal transformation for KSVD, and we adaptively trade off their weights in the Bayesian framework. Put more broadly, they have different advantages, depending on the amount of data available.

The remainder of this paper is organized as follows. In the next section, we describe the recoveries of QRS products based on multitemporal dictionary learning, including multitemporal KSVD and multitemporal Bayesian dictionary learning. This is followed by the quantitative evaluation experiments and real experiments with QRS products contaminated by thick clouds and shadows in Section III. Finally, Section IV draws the conclusion.

II. MULTITEMPORAL DICTIONARY LEARNING METHODOLOGY

As stated previously, signals/images/data can be sparse or compressible in some transform domain, which is exactly the basis of many applications of dictionary learning (see [24]–[27]). Multitemporal RS data are no exception [23]; they reflect the physical information from the same area and from different

periods. Therefore, they contain a great deal of redundant information and are more consistent with the sparse characteristic in the particular transform domain. In other words, the supporting information from the other temporal data can be exploited by dictionary learning. In this paper, two kinds of multitemporal dictionary learning are proposed to recover the QRS data's missing information that is occluded by the thick clouds and attendant shadows. The two proposed methods are based on the frameworks of KSVD and Bayesian dictionary learning.

KSVD dictionary learning, which was initially proposed by Aharon *et al.* [24], is the abbreviation of K singular value decomposition (SVD) and the derivative and generalization of the K -means algorithm. It consists of alternating between a sparse coding stage and a dictionary updating stage.

Bayesian dictionary learning [called beta process factor analysis (BPFA)] [25] is another increasingly hot topic in the sparse representation community and is also a kind of Bayesian compressive sensing [28]. There has recently been some research into nonparametric Bayesian dictionary learning as applied to image processing [25], [29]. In the context of CS, the beta-Bernoulli process is employed as a prior for learning the dictionary and for inferring an appropriate dictionary size.

In general, KSVD and BPFA can both be directly applied to the recovery of multitemporal QRS data. However, they do not take the temporal correlations between the multitemporal data into sufficient consideration. Accordingly, in the respective framework of KSVD and BPFA, we propose two algorithms to make the best use of the temporal correlations. In order to avoid any confusion, the proposed algorithms are called multitemporal KSVD (MT-KSVD) and multitemporal BPFA (MT-BPFA), respectively. Further details are presented in Sections II-B and II-C.

A. CS Theory

CS, breaking through the limitation of the traditional Nyquist–Shannon sampling theorem, which (Nyquist) specifies that a signal can be perfectly reconstructed from an infinite sequence of samples if the sampling rate is two times greater than the bandwidth, provides an approach to reconstruct a signal acquired by a projection matrix with incomplete and inaccurate measurements. Supposing that x is an unknown noise-free column vector in \mathbb{R}^n and Ψ is an orthogonal basis in $\mathbb{R}^{n \times n}$, then x is expressed as

$$x = \Psi\alpha \quad (1)$$

where α is the basis coefficients in \mathbb{R}^n . When the number of nonzero entries of α is far less than n , x is called sparse to Ψ . The CS measurement $y \in \mathbb{R}^m$ is derived by a projection matrix $\Phi \in \mathbb{R}^{m \times n}$ ($m < n$) as

$$y = \Phi x \quad (2)$$

where Φ is called the CS matrix [30], sensing matrix [31], or measurement matrix [32]. On the premise of $m < n$, it is clearly an ill-posed inverse problem to recover x from y . Fortunately, when the product of Φ and Ψ satisfies the

condition of the restricted isometry property [22], [32], it will be equivalently translated into a convex optimization

$$\min \|\alpha\|_1 \quad \text{s.t.} \quad y = \Phi\Psi\alpha \quad (3)$$

where $\|\bullet\|_1$ signifies the ℓ_1 norm of the entry, which means the sum of the absolute value of each component of the vector. To date, several heuristic algorithms for CS recovery have been developed, such as BP [33], GBP [34], OMP [35], ROMP [36], CoSaMP [37], SP [38], and FOCUSS [39].

In most cases, it is not necessary to know both Φ and Ψ in CS theory to solve out the sparse representation α . For brevity, $D \in \mathbb{R}^{m \times n}$ is usually selected for substituting for the combination of Φ and Ψ in (1) and (2), respectively; then

$$y = D\alpha. \quad (4)$$

In fact, D amounts to a “black box” of CS. Although its components Φ and Ψ are unknown, its representation α can also be approximately derived. Duarte *et al.* [40] considered D as the holographic basis (the original used another symbol, not D), and Donoho [20] named it the “information operator.” From another point of view, D can also be thought of as a dictionary, with its columns amounting to dictionary atoms. As D is a flat matrix, it is known as an overcomplete dictionary or a redundant dictionary. Interestingly, the process of hunting for D and α has a vivid appellation: dictionary learning [41]. At present, there are several popular dictionary learning methods, such as KSVD [24], [26] and Bayesian dictionary learning [25], [29]. Usually, a typical dictionary learning is an iterative process which alternates between the sparse representations of the given signal or image with the current dictionary and the update of the dictionary atoms.

B. Recovery Using MT-KSVD

MT-KSVD consists of three parts: 1) temporal transformation; 2) temporal permutation; and 3) dictionary learning. Temporal transformation is in an attempt to strengthen the temporal correlation of the multitemporal QRS data. Temporal permutation is adopted to find an optimized path for the successive dictionary learning. Generally speaking, the first two parts are both constructive preprocessing steps of dictionary learning, which will be verified in the experiments.

First, the temporal transformation is presented. In fact, some QRS data [e.g., land surface temperature (LST)] may vary greatly in a short time period. The temporal correlation can be strengthened by a reasonable way. To this end, a transformation (linear or nonlinear) from other temporal data to the corrupted data is adopted. Given a series of QRS data (in the visible and infrared ranges), $\{x_t\}_{t=1}^T \in \mathbb{R}^{n_1 \times n_2}$, registered on the same geographical region and from different periods, where x_s ($1 \leq s \leq T$) is contaminated by thick clouds and shadows (for convenience of description, we just suppose that, in the time series, there is only one temporal data item corrupted by thick clouds, and the others are intact). Let Ω and Ω^- be the cloud-free region and the cloud-contaminated region of x_s , respectively (see Fig. 1), and the identical spatial regions of other x_t ($t \neq s$). Our goal is to transform x_t ($t \neq s$) into the

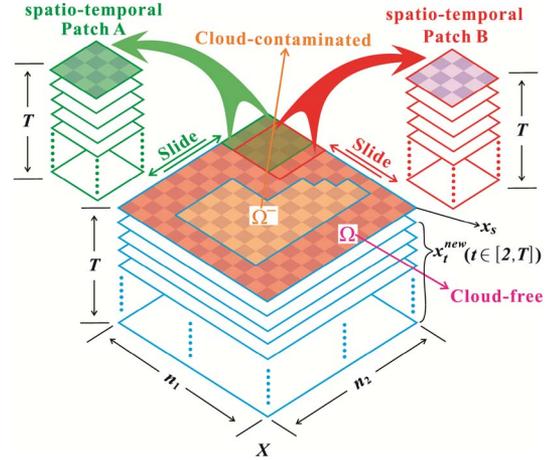


Fig. 1. Reordered multitemporal data overlapping and spatiotemporal patch extraction.

temporal space of x_s . First of all, it is necessary to obtain the transformation relationship. We assume that the cloud-free regions of all of the quantitative data from the different periods satisfy

$$x_s^\Omega = f_t(x_t^\Omega) + \varepsilon_t^\Omega, \quad t \in [1, T] \quad \text{and} \quad t \neq s \quad (5)$$

where x_s^Ω denotes the cloud-free region of x_s , x_t^Ω denotes the same region of x_t , ε_t^Ω denotes the corresponding residual, and $f_t(\bullet)$ denotes the function relationship, which can be linear or nonlinear. In the experiments, a simple linear function is set as the transformation relationship (e.g., $x_s^\Omega = a_t x_t^\Omega + b_t$, where a_t and b_t are the corresponding parameters for different temporal data). Based on the least squares solver, the basic idea is to minimize the sum of the squared residuals

$$\sum_{\Omega} (\varepsilon_t^\Omega)^2, \quad t \in [1, T] \quad \text{and} \quad t \neq s. \quad (6)$$

Once the function is determined, the previous transformation is successively undertaken as

$$x_t^n = f_t(x_t), \quad t \in [1, T] \quad \text{and} \quad t \neq s. \quad (7)$$

Here, the temporal permutation follows. After the previous temporal transformation, the transformed versions of the data $\{x_t^n\}_{t=1}^T \in \mathbb{R}^{n_1 \times n_2}$ ($t \neq s$) and x_s are obtained. These data will then consist of a data set $X_0 \in \mathbb{R}^{n_1 \times n_2 \times T}$, according to their original time sequence. Usually, the time sequence does not represent the optimal overlapping order. Ram *et al.* [42] proposed a shortest spatial path ordering of the image patches. Inspired by this, we seek a reasonable temporal path, ordering the time-series quantitative data to make better use of the temporal correlations. Our basic idea is to ensure that the data with stronger correlations are closer. The reordering permutation is the optimized temporal path. In other words, this kind of temporal permutation is similar to fused lasso [43], [44], which is designed for problems with features that can be ordered in some meaningful way. According to the features of correlation, we estimate a new order to permute the chronological order of the multitemporal data. The temporal permutation encourages

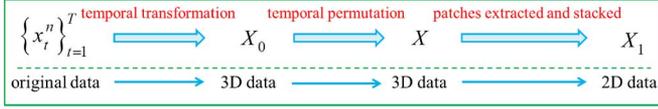


Fig. 2. QRS data changing process before the dictionary learned in the framework of MT-KSVD.

the local constancy of the sparse dictionary coefficients when the dictionary learning is being undergone. Additionally, the metric of correlation is picked as the correlation coefficients (CCs) between the quantitative data contaminated by clouds and the other intact data, only for the cloud-free region data. The CC is calculated by

$$CC_t = \frac{\sum_{\Omega} \left((x_t^n)^{\Omega} - \mu_{(x_t^n)^{\Omega}} \right) \left(x_s^{\Omega} - \mu_{x_s^{\Omega}} \right)}{\sqrt{\sum_{\Omega} \left((x_t^n)^{\Omega} - \mu_{(x_t^n)^{\Omega}} \right)^2} \sqrt{\sum_{\Omega} \left(x_s^{\Omega} - \mu_{x_s^{\Omega}} \right)^2}},$$

$$t = 1, 2, \dots, T \quad (8)$$

where CC_t is the CC between the cloud-free region of the t th data item and x_s^{Ω} , and Ω denotes the cloud-free area of x_s and the corresponding identical area for the other temporal data, i.e., $(x_t^n)^{\Omega}$ represents the cloud-free data of x_t^n ; $\mu_{(x_t^n)^{\Omega}}$ and $\mu_{x_s^{\Omega}}$ represent the average values of $(x_t^n)^{\Omega}$ and x_s^{Ω} , respectively. Note that $(x_s^n)^{\Omega} = x_s^{\Omega}$.

Once the CCs are calculated, the permutation order of the multitemporal data is obtained. This order is sorted from the highest modulus of CCs to the lowest, accompanied with a new sequence $\{x_t^{new}\}_{t=1}^T \in \mathbb{R}^{n_1 \times n_2}$. Suppose that $X \in \mathbb{R}^{n_1 \times n_2 \times T}$ denotes the data set with a new overlapping order. It can be seen that the cloud-contaminated data will be reordered at the top (see Fig. 1). By means of this permutation, the distance between multitemporal data items with stronger correlations is artificially shortened. As a result, the temporal correlation is made better use of.

In some special circumstances, x_s is completely obscured by thick clouds. Therefore, the CC between x_s and x_t is unavailable. Correspondingly, in our framework, the average of all of the temporal data x_s^* is considered as the approximate replacement of x_s . Expressions (5)–(8) can then be continued. In other words, a temporal transformation and temporal permutation is embarked on, in which the average of all of the existing data is set as the reference data.

After the two processing stages (temporal transformation and temporal permutation), an operator P is utilized to extract as many spatiotemporal patches as possible. In other words, P amounts to a cropping operator which extracts the specified patch from the data to be processed. For more details about P , please see [26] and [45]. Fig. 1 shows the extraction process. Suppose that the size of the spatiotemporal patches is $l \times l \times T$. We extract the patches column by column, starting from the top leftmost patch. As a result, we derive $N = (n_1 - l + 1)(n_2 - l + 1)$ patches. We make these patches column stacked, and X will become a $Tl^2 \times N$ matrix X_1 . Fig. 2 shows the QRS data changing process before the dictionary learned in the framework of MT-KSVD, implying that the original data are transformed from 3-D to 2-D. Let $(PX)^{\text{stack}}$ denote $X_1 [(\bullet)^{\text{stack}}$

represents the column stack operator], $(P_k X)^{\text{stack}}$ ($k \in [1, N]$) represents the k th column-stacked patch, and P_k is the k th component of P , which extracts the k th patch. Finally, recovering the missing data obstructed by thick clouds is based on the powerful KSVD dictionary learning [24]. Generally speaking, for our problem, it is set up as follows:

$$\left\{ \{\hat{\alpha}_k\}_{k=1}^N, \hat{Y}_1 \right\} = \arg \min_{Y_1, \{\alpha_k\}} \lambda \|MY_1 - X_1\|_2^2 + \sum_k \mu_k \|\alpha_k\|_0 + \sum_k \|D\alpha_k - Y_{1k}\|_2^2. \quad (9)$$

This expression purports to show an approximation between the column-stacked patches X_1 of the multitemporal QRS data and the unknown ideal counterpart $Y_1 = (PY)^{\text{stack}}$ (Y is the ideal version of X), only for the valid or existing pixels. M is a mask labeling the location that is corrupted by the thick clouds and shadows in the multitemporal QRS products. $D \in \mathbb{R}^{Tl^2 \times h}$ ($Tl^2 < h$) is an overcomplete dictionary, and $\alpha_k \in \mathbb{R}^{h \times 1}$ is the dictionary coefficients of the k th patch. $Y_{1k} = (P_k Y)^{\text{stack}}$ is the k th column of Y_1 . λ and μ_k are both regularization parameters. The second and third terms of (9) are the local sparse priors of the data, which guarantee that every patch of the recovered data has a sparse representation meeting the requirements. On the whole, the quantitative data also have a sparse representation.

In fact, Elad [45] decoupled the minimization of (9) into two subproblems of solving $\hat{\alpha}_k$ and \hat{Y}_1 . Given an initial D , the sparse representations α_k are first sought out; then, the columns of D are updated one after another. In general, these two processes operate alternately. For more details, please refer to [45]. Ultimately, the sparse representation of every spatiotemporal patch is obtained by dictionary learning. However, the resulting Y_1 consists of column-stacked patches of size $Tl^2 \times N$, and therefore, we should undertake the inverse transformation of P and “stack” $Y = P^{-1}Y_1^{\text{stack}^{-1}}$. The missing information will then be recovered. If the temporal transformation and temporal permutation are both cancelled, MT-KSVD amounts to KSVD.

From the aforementioned descriptions, we assume that there is only one data item contaminated by clouds. However, in practice, there may be a number of data items contaminated. Accordingly, we can process these corrupted data items one after another using MT-KSVD, and the previous result is the basis of the subsequent recovery.

C. Recovery Using MT-BPFA

Here, MT-BPFA is applied to the recovery of cloud-contaminated QRS data. MT-BPFA does not seek the optimized temporal path, as in MT-KSVD, but adaptively weights the temporal correlations between the QRS data. As stated previously, given a number of multitemporal QRS data items over the same area $\{x_t\}_{t=1}^T \in \mathbb{R}^{n_1 \times n_2}$, x_s is contaminated by thick clouds and shadows. We first put them into a 3-D cube $X_0 \in \mathbb{R}^{n_1 \times n_2 \times T}$ in chronological order. Subsequently, we extract the $l \times l \times T$ spatiotemporal patches, sliding with the operator P (as with MT-KSVD), and then turn them into column-stacked versions to get a $Tl^2 \times N$ matrix X^b . Fig. 3 shows the QRS

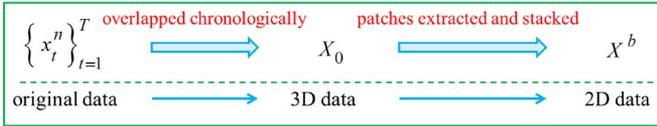


Fig. 3. QRS data changing process before the dictionary learned in the framework of MT-BPFA.

data changing process before the dictionary learned in the framework of MT-BPFA. For the k th column of X^b , with the dictionary $D \in \mathbb{R}^{Tl^2 \times h}$ ($Tl^2 < h$) and additive noise ε_k , the observation model is as follows:

$$X_k^b = D\alpha_k + \varepsilon_k. \quad (10)$$

Interestingly, some impactful priors are necessarily taken into account in the Bayesian framework. Regarding the noise, it is first imposed on the Gaussian prior as

$$\varepsilon_k \sim N(0, \gamma_\varepsilon^{-1} I_Q) \quad (11)$$

where γ_ε^{-1} is the variance of the normal distribution and I_Q is an identity matrix with the size of $Q \times Q$ ($Q = Tl^2$). Furthermore, the conjugate hyperprior is imposed on γ_ε

$$\gamma_\varepsilon \sim \text{Gamma}(a, b) \quad (12)$$

where $\text{Gamma}(\bullet, \bullet)$ means the gamma distribution and a and b are the shape parameter and scale parameter, respectively.

It is well known that temporal correlations exist between multitemporal data, which can be beneficial to the data recovery when the supporting information from the data themselves is deficient. To make a better use of the correlations, we propose to adaptively weight the temporal correlations when the dictionary learning is being undertaken. Briefly, the stronger the correlation is, the heavier the weight is. This idea is accomplished by applying a prior to the dictionary atoms. Towards this end, d_i , the i th atom (column) of D , obeys another normal distribution

$$d_i \sim N(0, (QLw)^{-1} I_Q). \quad (13)$$

In this expression, L is a constant which is used to adjust the joint correlation property of the dictionary atoms, implying a large deviation bound for the sums of the random variables [35]. w denotes the weights of the corresponding dictionary atoms and is a diagonal matrix. In the experiments, the CC between x_t ($t \neq s$) and x_s is picked as the weight indicator, only for the shared cloud-free regions of the quantitative data. Suppose that CC_t denotes the weight of x_t , which is similarly calculated by (8), and all CC_t 's consist of a vector CC^b . Given an extracted patch, for the corresponding components of x_t , CC_t will continuously appear l^2 times in the weight. This is because, in the extracted components of x_t , there are l^2 elements. w is calculated by

$$w = \text{diagonal}(\text{span}(\text{CC}^b)) \quad (14)$$

where $\text{span}(\bullet)$ is an operator making every element of the entry (CC^b) serially appear l^2 times and $\text{diagonal}(\bullet)$ is an operator making a vector be a diagonal. In fact, w is a diagonal matrix

of size $Tl^2(Q)$. When $L = 1$ and w is an identity matrix, MT-BPFA is equivalent to BPFA.

Similar to MT-KSVD, once x_s is completely obscured by thick clouds, the average of all of the temporal data x_s^* is considered as the approximate replacement of x_s . The calculation of CC_t can then be continued. At the same time, 5% of x_s^* is randomly sampled as the initial of x_s . This is because, if there is no initial constraint to the data to be recovered, MT-BPFA will get a very poor result.

To be sparser, the coefficients α_k of the dictionary are broken down into the pointwise product of z_k and s_k [25] as

$$\alpha_k = z_k \circ s_k \quad (15)$$

where z_k is a binary set acting as a mask on s_k and the signifier (\circ) means the pointwise product. On one hand, z_k obeys a Bernoulli distribution

$$z_k \sim \prod_{i=1}^K \text{Bernoulli}(\pi_i). \quad (16)$$

Note that the notation Π denotes that each component of z_k is drawn independently from the distributions of the same form [25]. Hierarchically, π has another conjugate hyperprior

$$\pi \sim \prod_{i=1}^K \text{Beta}\left(\frac{c}{K}, \frac{d(K-1)}{K}\right) \quad (17)$$

where c and d are the model parameters and Π has a similar meaning to (16).

On the other hand, s_k is Gaussian distributed

$$s_k \sim N(0, \gamma_s^{-1} I_h) \quad (18)$$

where I_h is an identity matrix with the size of $h \times h$ (h is the number of atoms). Similar to γ_ε , γ_s has a gamma prior

$$\gamma_s \sim \text{Gamma}(e, f) \quad (19)$$

where e and f are the shape parameter and the scale parameter, respectively.

With the aforementioned ample hierarchical and conjugate prior distributions, the posterior update equations of the Bayesian system are analytical. Under such a framework, the process of the inference and solution is implemented via Gibbs sampling, as detailed in [25]. The consecutive conjugacy in the hierarchical model has a great advantage over the posterior inferences. In fact, it is in line with the typical maximum *a posteriori* optimization

$$\begin{aligned} & -\log p(\Theta | X^b, H) \\ &= \frac{\gamma_\varepsilon}{2} \sum_{k=1}^N \|M_k D(z_k \circ s_k) - X_k^b\|_2^2 + \frac{QLw}{2} \sum_{i=1}^h \|d_i\|_2^2 \\ &+ \frac{\gamma_s}{2} \sum_{k=1}^N \|s_k\|_2^2 - \log \text{Gamma}(\gamma_\varepsilon | H) \\ &- \log \text{Gamma}(\gamma_s | H) - \log f_p(\{z_k\}_{k=1}^N; H) + C \quad (20) \end{aligned}$$

where Θ means all of the model parameters, $X^b = \{X_k^b\}_{k=1}^N$, H represents the hyperparameter, f_p is the beta and Bernoulli priors in (16) and (17), and C is a constant. Note that M_k is the mask of the clouds and shadows in Section II-B. Finally, as with the inverse transformation of MT-KSVD, the recovery result is transformed into the original 3-D version in chronological order.

Differing from settling out the missing information in one spatiotemporal patch at a time using MT-KSVD dictionary learning, MT-BPFA operates all of the patches at the same time. After repeating a number of cycles, the final result is obtained by averaging the results of each occasion. Even though the patches of the same cycle do not overlap, in the whole cycle, they are overlapped like MT-KSVD. The learning process is finished by the continuous Gibbs sampling, according to the inferred posterior, which greatly reduces the time consumption and operating cost, benefitting from the conjugate priors.

III. EXPERIMENTS AND ANALYSES

To allow a comprehensive assessment of the proposed multitemporal dictionary learning methods, experiments are undertaken to recover the QRS products contaminated by thick clouds and shadows, accompanied by a series of qualitative and objective analyses. In general, the proposed methods can be applied to all kinds of quantitative products of optical and infrared RS. In terms of the QRS data, one class performs with stable data features (e.g., reflectance product), and the other class fluctuates wildly (e.g., LST product). Therefore, the two products are appropriate as the experiment data to validate the different aspects of our methods. The cloud detection is not our interest; therefore, we assume that the cloud mask is interactively accessed by hand. For simplicity, the 250-m resolution reflectance product of MODIS L1B and the daily 1-km resolution LST products of MODIS L3 were chosen as the experiment data and were directly downloaded from the NASA website (<http://ladsweb.nascom.nasa.gov/data/search.html>). Note that these products were georeferenced to acquire data in the same geographic area and at different times. The reflectance products were georeferenced by the MODIS conversion toolkit (MCTK), which was downloaded from the NSDIC website (<http://nsdic.org/data/modis/tools.html>), and the LST products were georeferenced by the MODIS reprojection tool (MRT), which was downloaded from the NASA website (https://lpdaac.usgs.gov/tools/modis_reprojection_tool).

Without any special instructions, the size of the spatiotemporal patches is $2 \times 2 \times T$ (T denotes the number of data items), $h = 256$, and D is a flat matrix of size $4T \times 256$, which is initialized by discrete cosine transform (DCT). In MT-KSVD, the transformation function is linear, $\sigma = 0.005$, and the convergence standard is that the quadratic sum of the difference between the adjacent two iterations is less than $4T\sigma^2$ (as with KSVD in [24]). In MT-BPFA, $a = b = e = f = 1 \times 10^{-6}$, $c = d = 1$, $L = 80$, and $Q = 4T$.

A. Quantitative Comparisons With Existing Methods

As the Aqua satellite's afternoon observations combine with the Terra satellite's morning observations, we have access to

TABLE I
EXPERIMENTAL DATA FOR THE QUANTITATIVE COMPARISONS

Reflectance products	LST products
MOD02QKM.A2012182	MOD11A1.A2012067
MOD02QKM.A2012183	MOD11A1.A2012068
MOD02QKM.A2012184	MOD11A1.A2012069
MYD02QKM.A2012185	MOD11A1.A2012070
MOD02QKM.A2012186	MOD11A1.A2012071
MYD02QKM.A2012187	MOD11A1.A2012072
MOD02QKM.A2012188	MOD11A1.A2012073

daily MODIS quantitative products. For the simulated experiments, we downloaded a group of consecutive seven-day reflectance products (see Table I). After being georeferenced, one group of subregions, with the size of 300 pixels by 300 pixels, which are on the Arabian Peninsula [see Fig. 4(a)], was cropped from them. In terms of the MODIS LST products, we also downloaded a group of consecutive seven-day products (see Table I). After georeferencing, we cropped one group of subregions of the size of 300 by 300 [see Fig. 6(a)].

We use mean absolute error (MAE), mean squared error (MSE), mean relative error (MRE), and CC [calculated as (8)] as the evaluation indicators. Given x and y as the data (for brevity, we suppose that both are vectors here) to be compared, then MAE is defined by the following expression:

$$\text{MAE}(x, y) = \frac{1}{M} \sum_{i=1}^M |x_i - y_i| \quad (21)$$

where M is the length of the vector. MSE is calculated by

$$\text{MSE}(x, y) = \frac{1}{M} \sum_{i=1}^M (x_i - y_i)^2. \quad (22)$$

Supposing that y is the reference data, MRE is computed by

$$\text{MRE}(x, y) = \frac{1}{M} \sum_{i=1}^M \frac{|x_i - y_i|}{y_i}. \quad (23)$$

First of all, the capacity of the multitemporal dictionary learning to recover the contamination of large-area thick clouds is tested. Fig. 4 shows the simulated recovery results of the MODIS reflectance products on the Arabian Peninsula, including HANTS [19], MT-BPFA, and MT-KSVD (in consideration of the visual similarity of BPFA and MT-BPFA, and KSVD and MT-KSVD, to save space, we do not show the results of BPFA and KSVD). Note that the corrupted data items are from the 186th day of 2012, which is the fifth temporal data item in the group. From the subjective point of view, HANTS introduces false edges to the result, as shown in the yellow circled regions in Fig. 4(c), and MT-BPFA and MT-KSVD both keep pace with the original reflectances and succeed with satisfactory outcomes. To the best of our knowledge, the existing findings point to HANTS as having great potential for long time-series data analysis. However, in our case, there are only seven temporal reflectances. As a result, HANTS does not get as good an effect as MT-BPFA and MT-KSVD.

Subsequently, the same (fifth) temporal reflectances of Fig. 4 are completely corrupted by simulated thick clouds (labeled as

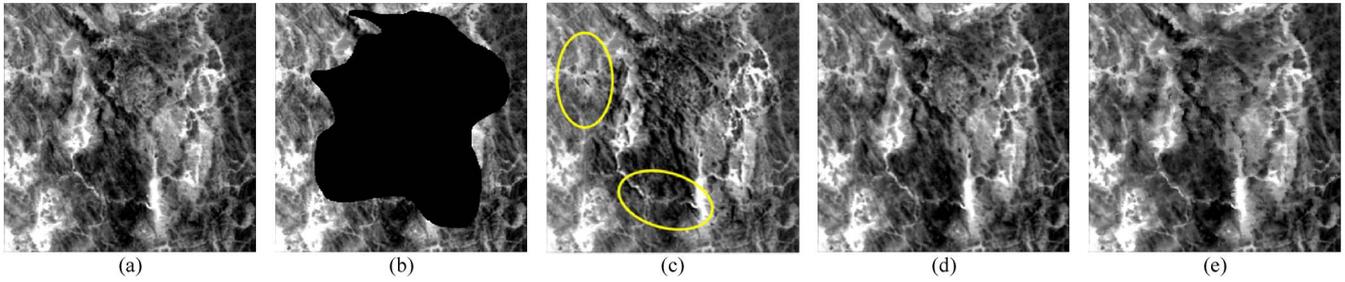


Fig. 4. Simulated thick cloud corruption and the corresponding recovery of the reflectances of MODIS. (a) Original. (b) Simulated thick cloud corruption. (c) HANTS recovery. (d) MT-BPFA recovery. (e) MT-KSVD recovery.

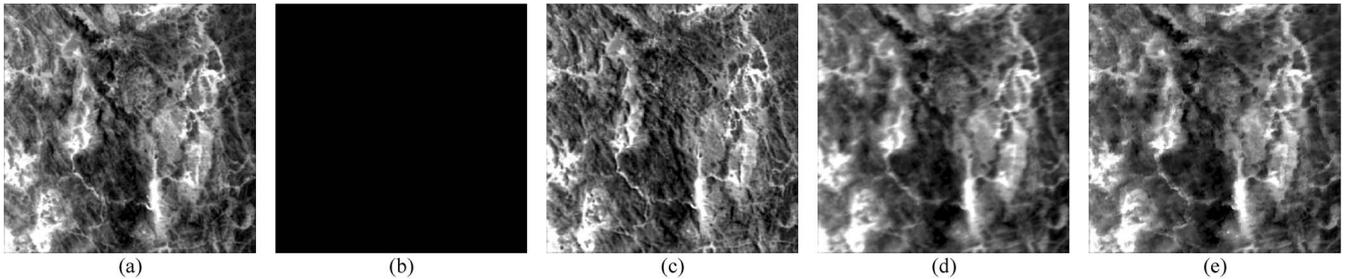


Fig. 5. Simulated thick cloud complete corruption and the corresponding recovery of the reflectances of MODIS. (a) Original. (b) Simulated thick cloud corruption. (c) HANTS recovery. (d) MT-BPFA recovery. (e) MT-KSVD recovery.

zeros), as shown in Fig. 5(b), and are recovered by HANTS and the proposed two methods (again, the results of BPFA and KSVD are omitted). As seen from the figures, the results of MT-BPFA display a certain degree of smoothness, and those of MT-KSVD are alleviated a little. In terms of this kind of smoothness, they are very close to the average of all of the multitemporal data because of the average being set as the reference data of MT-BPFA and MT-KSVD. Additionally, as MT-KSVD is solved patch by patch, the smoothness is weakened. As far as the visual effect is concerned, HANTS performs the best and benefits from the temporal and quantitative restrictions of the harmonic analysis. However, for the quantitative applications, the visual effect is of secondary importance, and the primary concern is the objective evaluation, which is stated in the following.

After the previous visual assessments, the quantitative comparisons of Figs. 4 and 5 for HANTS, BPFA, MT-BPFA, KSVD, and MT-KSVD are made by means of MAE, MSE, MRE, and CC, as shown Table II. On the whole, MT-BPFA outperforms MT-KSVD, which benefits from the global solving solution, and MT-KSVD is solved block by block, which ignores the global constraint; HANTS, which needs considerable temporal data, obtains the worst result. Since MT-BPFA adaptively takes the temporal correlations into consideration, it achieves a more pleasing effect than BPFA. With the help of the temporal transformation and temporal permutation, MT-KSVD finds such an optimized temporal path that the temporal correlations play a more important role. Naturally, MT-KSVD outstrips KSVD. As far as the quantitative indicators are concerned, the improved effect from KSVD to MT-KSVD is more obvious than that from BPFA to MT-BPFA. The reason for this is that BPFA itself achieves such a good effect that the improved capacity of MT-BPFA is limited; however, KSVD

TABLE II
RECOVERIES OF THE REFLECTANCE PRODUCTS
BY THE DIFFERENT METHODS

Data	Methods	MAE/ 10^{-3}	MSE/ 10^{-5}	MRE/ %	CC	Time/ min
Fig. 4	HANTS	10.919	40.812	5.789	0.7983	1.75
	BPFA	3.989	2.612	2.170	0.9846	5.43
	MT-BPFA	2.543	1.212	1.380	0.9926	8.80
	KSVD	6.318	10.212	3.417	0.9381	10.64
	MT-KSVD	4.821	5.012	2.589	0.9703	8.35
Fig. 5	HANTS	21.558	77.212	11.448	0.6528	1.67
	BPFA	7.693	13.812	4.121	0.9169	5.31
	MT-BPFA	7.686	9.512	4.102	0.9574	8.70
	KSVD	10.875	18.812	5.937	0.8846	10.58
	MT-KSVD	8.337	11.212	4.469	0.9419	8.38

does not achieve as good an effect as BPFA, so MT-KSVD has more room for improvement. In addition, when the reflectances are only partly contaminated by thick cloud (Fig. 4), no matter whether for HANTS, BPFA, MT-BPFA, KSVD, or MT-KSVD, the results are better than those when they are completely contaminated (Fig. 5). In terms of the computational time, HANTS shows an overwhelming advantage over MT-BPFA and MT-KSVD, and it benefits from a model that is simpler than multitemporal dictionary learning. Table II confirms that the proposed MT-BPFA and MT-KSVD are both effective expanded versions of BPFA and KSVD, respectively.

Similarly, we also undertake simulated experiments using another group of LST products of MODIS (see Table I). Fig. 6 shows the recovery results with the LST products by the three methods (for the same reason, the results of BPFA and KSVD are again not shown), cropped from the aforementioned data (Table I). Note that the corrupted data are from the 69th day of 2012, which is the third temporal data item in the group. As shown in the figures, MT-KSVD results in more obvious

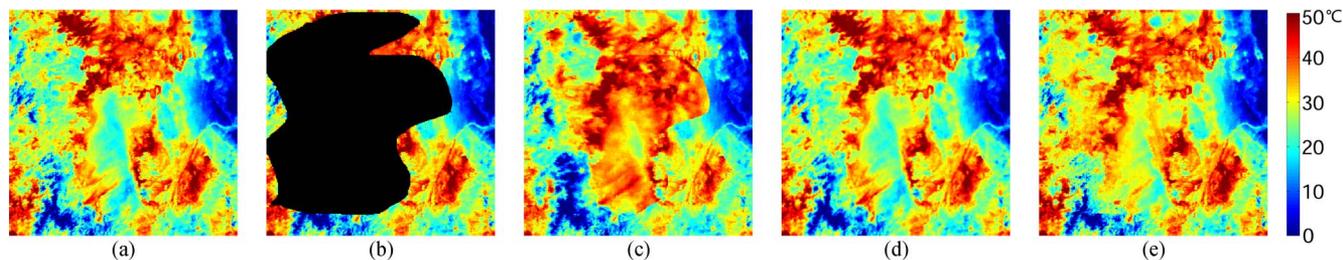


Fig. 6. Simulated thick cloud corruption and the corresponding recovery of the LSTs of MODIS. (a) Original. (b) Simulated thick cloud corruption. (c) HANTS recovery. (d) MT-BPFA recovery. (e) MT-KSVD recovery.

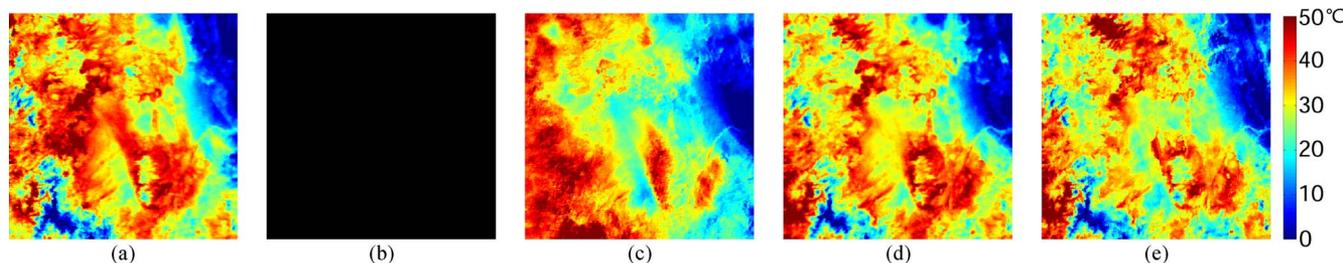


Fig. 7. Simulated thick cloud complete corruption and the corresponding recovery of the LSTs of MODIS. (a) Original. (b) Simulated thick cloud corruption. (c) HANTS recovery. (d) MT-BPFA recovery. (e) MT-KSVD recovery.

distortions than MT-BPFA, and HANTS produces the most artifacts. Moreover, HANTS also introduces some false edges, as with the previous reflectance experiments. If the multitemporal LST data vary greatly or their latent correlations are not strong enough, the recovery results become worse. Fig. 7 shows the different recoveries of complete corruption of one temporal LST, using the same multitemporal data as Fig. 6. In Fig. 7, the corrupted data item is the sixth temporal data item. Even though HANTS can get a better visual effect than MT-BPFA and MT-KSVD, its quantitative assessment is the worst (see Table III), which is the primary consideration in practical applications. In most cases, LST fluctuates greatly with the weather conditions on different days, and it is not as stable as the reflectance. Therefore, the recovery results are not as good as the reflectance. In other words, the temporal correlations of LST are not so strong that the results are inferior to the reflectance product. Table III shows the objective assessments. For the partial corruption (Fig. 6), both MT-BPFA and MT-KSVD outperform their original versions, BPFA and KSVD. However, for the complete corruption (see Fig. 7), the CCs of MT-BPFA and MT-KSVD are superior to BPFA and KSVD. In terms of the other indicators, MT-BPFA and MT-KSVD do not show obvious advantages and even show worse effects. The reason for this is that LST varies dynamically. When the data are completely missing, the temporal constraint no longer exists. Correspondingly, the evaluation indicators appear locally inconsistent. Overall, the change rule is roughly in line with that of the previous reflectance experiments.

B. Intensive Comparison Between MT-KSVD and MT-BPFA

In the aforementioned experiments, six extra temporal data items (seven in total, including the corrupted temporal data item itself) are put in to recover the specific temporal data item contaminated by thick clouds. It is explored how the different

TABLE III
RECOVERIES OF THE LST PRODUCTS BY THE DIFFERENT METHODS

Data	Methods	MAE/ °C	MSE	MRE/ %	CC	Time/ min
Fig. 6	HANTS	1.033	2.857	2.445	0.8647	1.67
	BPFA	0.377	0.278	0.891	0.9789	11.46
	MT-BPFA	0.214	0.082	0.507	0.9936	14.39
	KSVD	1.310	5.242	3.036	0.7187	15.16
	MT-KSVD	0.467	0.606	1.103	0.9533	14.78
Fig. 7	HANTS	5.119	35.992	12.913	0.5085	1.63
	BPFA	1.419	4.200	3.489	0.8015	6.85
	MT-BPFA	1.505	3.506	3.707	0.9002	14.45
	KSVD	2.374	8.211	5.862	0.7160	14.78
	MT-KSVD	1.498	3.675	3.691	0.8334	14.85

numbers of temporal data items exercise an influence on the recovery result in the following content. Table IV shows the quantitative assessments of the recovery results with an extra 1–6 temporal data items, respectively, not including the cloudy data themselves. As seen from this table, when the extra data items are few in number, MT-KSVD can get a better result; as the amount increases, the superiority gradually weakens; when the number arrives at a certain point, MT-KSVD is surpassed by MT-BPFA. However, this does not mean that the more data there is, the better the effect is. As the data become more, the redundant information will also become more, and the result may become worse. In other words, MT-BPFA needs more complementary and auxiliary data to get a satisfactory result; however, this extravagant demand means that its practicality and efficiency are significantly reduced. In general, as the data increase, the time cost also becomes higher.

Overall, MT-KSVD gets a better result with less temporal data. To allow an intuitive understanding of this rule, the MRE variation diagram is shown in Fig. 8, with different numbers of temporal data items for both MT-BPFA and MT-KSVD. In the figures, a lower point means a lower MRE, i.e., a better

TABLE IV
IMPACT ON THE RECOVERY OF DIFFERENT NUMBERS OF TEMPORAL DATA ITEMS

Data	Indicators	Methods	Extra 1	Extra 2	Extra 3	Extra 4	Extra 5	Extra 6
Fig. 4	MAE/ 10^{-3}	MT-BPFA	10.197	10.183	2.980	2.685	2.657	2.543
		MT-KSVD	7.002	7.254	4.760	4.490	4.536	4.821
	MSE/ 10^{-5}	MT-BPFA	32.312	30.612	2.112	1.412	1.312	1.212
		MT-KSVD	13.212	13.912	5.112	4.312	4.312	5.012
	MRE/%	MT-BPFA	5.392	5.348	1.595	1.457	1.442	1.380
		MT-KSVD	3.781	3.897	2.554	2.404	2.433	2.589
	CC	MT-BPFA	0.8446	0.8531	0.9880	0.9918	0.9921	0.9926
		MT-KSVD	0.9179	0.9134	0.9704	0.9749	0.9756	0.9703
	Time/min	MT-BPFA	4.28	5.19	6.14	6.79	7.97	8.80
		MT-KSVD	8.03	7.82	8.10	8.39	9.49	8.35
Fig. 6	MAE/ $^{\circ}\text{C}$	MT-BPFA	1.595	1.455	0.964	0.639	0.237	0.214
		MT-KSVD	0.825	0.679	0.457	0.444	0.449	0.467
	MSE	MT-BPFA	6.685	5.630	2.386	0.885	0.106	0.082
		MT-KSVD	2.011	1.371	0.591	0.574	0.558	0.606
	MRE/%	MT-BPFA	3.679	3.366	2.260	1.501	0.563	0.507
		MT-KSVD	1.892	1.577	1.074	1.048	1.064	1.103
	CC	MT-BPFA	0.6440	0.6636	0.8424	0.9493	0.9918	0.9936
		MT-KSVD	0.8556	0.8898	0.9541	0.9557	0.9580	0.9533
	Time/min	MT-BPFA	4.63	6.45	8.04	9.35	12.12	14.39
		MT-KSVD	8.69	9.22	11.12	12.61	14.77	14.78

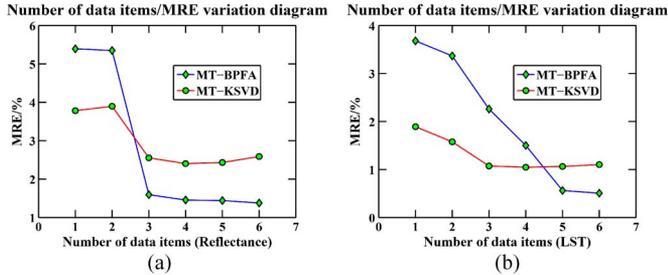


Fig. 8. MRE variation diagram of the recoveries of MT-BPFA and MT-KSVD, with different numbers of temporal data items. (a) For Fig. 4, reflectance. (b) For Fig. 6, LST.

effect. As can be seen, when the temporal data items are few in number, MT-KSVD gets a better effect.

Additionally, Figs. 9 and 10 show the scatter plots of the reflectance (for Fig. 4) and LST (for Fig. 6) products, respectively, between the original and recovered results using different numbers of extra data items. For brevity, scatter plots with only one and six extra data items are shown. As can be seen from the figures, when there is only one extra data item, the points in the scatter plots of MT-KSVD are more concentrated to the diagonal than those of MT-BPFA, which means that MT-KSVD is superior to MT-BPFA. However, when there are six extra data items, the points in the scatter plots of MT-KSVD are more discrete than those of MT-BPFA, which means that MT-KSVD is inferior. In terms of one method, when extra data items are added, the points become closer to the diagonal, which means that the effect becomes better. These plots fully demonstrate that MT-KSVD has more advantages over MT-BPFA when the data are insufficient and vice versa.

C. Influence of the Initial Dictionary on MT-KSVD and MT-BPFA

Here, we investigate if the initial dictionary makes a difference to the results of the methods based on dictionary learning and sparse representation. Table V shows the effect

on the recovery results of different initial dictionaries (bases). DCT is generated by discrete cosine transform; USE represents a uniform spherical ensemble, uniformly distributed on the unit sphere; Fourier is a partial Fourier ensemble (with unit Euclidean length); RST denotes a partial real Fourier ensemble; Hadamard is a partial Hadamard ensemble, generated by the Hadamard matrix; URP signifies a uniform random projection ensemble, generated by SVD in our case; and QR is a random orthobasis ensemble, generated by QR decomposition (sometimes referred to as orthogonal matrix triangularization). For more details, please see SparseLab (<http://sparselab.stanford.edu/>). As far as MT-BPFA is concerned, the influence of the initial dictionary on the results is very weak. This is because the process of dictionary learning is undertaken through the Gibbs sampling, which gets rid of the influence of initials in which the learned dictionary tends toward stability after simulated annealing. In spite of this, the initials have different influences on the resulting recovery of MT-KSVD. For the reflectances, DCT, Fourier, and Hadamard used as the initial dictionaries can arrive at the best effect, and for the LSTs, DCT and Hadamard get the best result. This experiment demonstrates that the initial dictionary makes a difference to the recovery result of MT-BPFA and MT-KSVD. In MT-BPFA, the result is only weakly influenced by the initial dictionary. In MT-KSVD, different initial dictionaries have different abilities to accommodate data features. On the whole, DCT has the advantage of stability and a favorable restoration, and most of the other initial dictionaries also get reasonable results.

D. Applications to Real Quantitative Products

Based on the same methods as those in the quantitative comparison experiments, real experiments with the quantitative products of reflectance and LST (shown in Table VI) are undertaken in this section. Since cloud detection is not our interest, the cloud mask is artificially delineated. In addition, thick clouds and shadows may contaminate several items of the

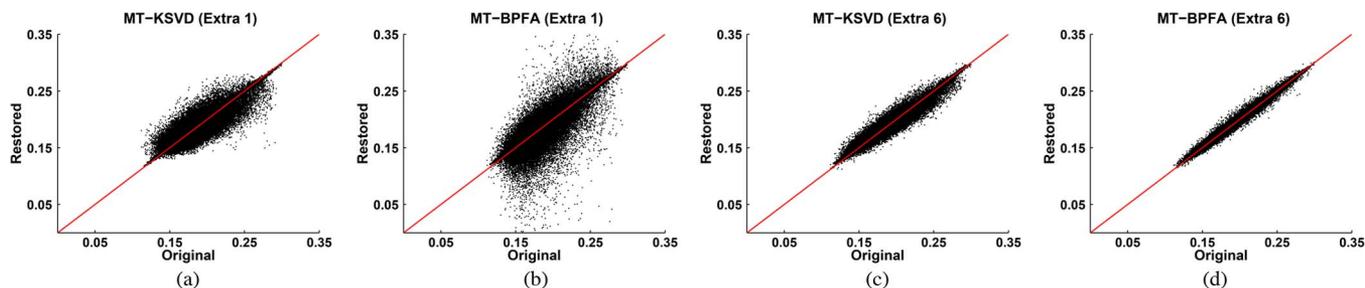


Fig. 9. Scatter plots of the reflectance product (for Fig. 4) between the original and recovery using different numbers of extra data items. (a) MT-KSVD recovery with one extra data item. (b) MT-BPFA recovery with one extra data item. (c) MT-KSVD recovery with six extra data items. (d) MT-BPFA recovery with six extra data items.

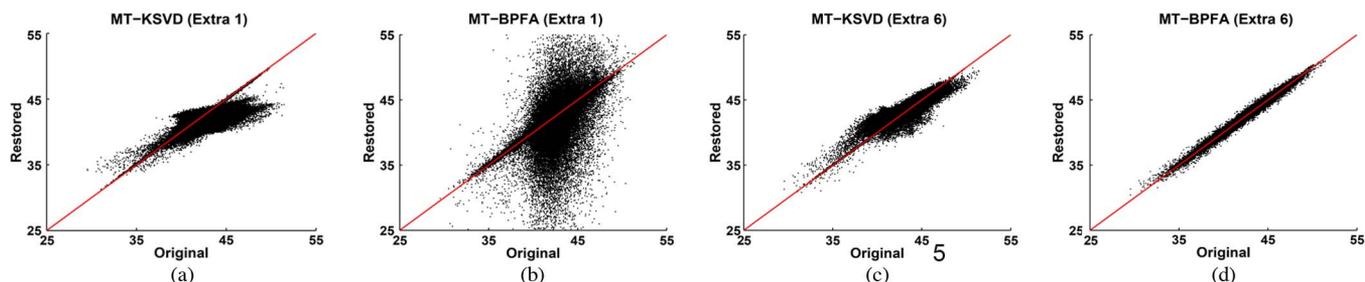


Fig. 10. Scatter plots of the LST product (for Fig. 6) between the original and recovery using different numbers of extra data items. (a) MT-KSVD recovery with one extra data item. (b) MT-BPFA recovery with one extra data item. (c) MT-KSVD recovery with six extra data items. (d) MT-BPFA recovery with six extra data items.

TABLE V
IMPACT ON THE RECOVERY OF DIFFERENT INITIAL DICTIONARIES

Methods	Data	Indicators	DCT	USE	Fourier	RST	Hadamard	URP	QR
MT-BPFA	Fig. 4	MAE/ 10^{-3}	2.543	2.657	2.669	2.744	2.606	2.568	2.633
		MSE/ 10^{-5}	1.212	1.312	1.312	1.412	1.312	1.312	1.312
		MRE/%	1.380	1.442	1.448	1.489	1.411	1.394	1.424
		CC	0.9926	0.9921	0.9920	0.9916	0.9923	0.9924	0.9921
		Time/min	8.80	8.32	8.53	8.47	8.66	8.85	8.45
MT-BPFA	Fig. 6	MAE/ 10^{-3}	0.214	0.247	0.210	0.214	0.214	0.226	0.223
		MSE/ 10^{-5}	0.082	0.115	0.083	0.088	0.087	0.098	0.090
		MRE/%	0.507	0.587	0.499	0.508	0.509	0.538	0.528
		CC	0.9936	0.9912	0.9937	0.9935	0.9934	0.9925	0.9931
		Time/min	14.39	12.60	14.00	13.37	13.79	13.87	13.41
MT-KSVD	Fig. 4	MAE/ 10^{-3}	4.821	9.780	4.112	7.608	4.003	5.103	9.780
		MSE/ 10^{-5}	5.012	26.912	2.912	19.712	3.212	7.412	26.712
		MRE/%	2.589	5.202	2.249	4.045	2.183	2.713	5.241
		CC	0.9703	0.8498	0.9834	0.8896	0.9817	0.9556	0.8761
		Time/min	8.35	12.04	10.98	14.90	10.45	12.94	12.74
MT-KSVD	Fig. 6	MAE/ $^{\circ}$ C	0.467	0.952	0.605	1.266	0.568	1.082	1.020
		MSE	0.606	2.534	1.029	4.223	1.068	2.824	2.417
		MRE/%	1.103	2.266	1.411	2.981	1.325	2.506	2.374
		CC	0.9533	0.8295	0.9185	0.7637	0.9160	0.8369	0.8144
		Time/min	14.78	17.04	20.93	15.23	17.11	16.93	13.37

quantitative data in the time series. In this situation, the corrupted data items are recovered one by one using MT-KSVD and MT-BPFA. However, due to the absence of ideal cloud-free original data, it is impossible to make an objective assessment. First, the experiments are conducted with a group of reflectances of MODIS. To further validate the proposed methods, we do not use daily reflectance data as before but a longer time interval. Fig. 11 shows a group of the original and recovery results of the MODIS reflectances with a size of 300 by 300 pixels, according to HANTS, MT-BPFA, and MT-KSVD, respectively (as before, the results of BPFA and KSVD are not shown). As can be seen from the figures, although the reflectances are acquired over quite a long time interval, the

large-area thick clouds and consequent shadows are effectively recovered. In fact, the time interval can be longer on the condition that the geographical features do not change significantly. Note that, since there are some overlapping regions corrupted by clouds, these pixels cannot be processed by HANTS because of the shortage of data, so they are just replaced by the mean values of the valid pixels of all of the temporal data. Therefore, HANTS gets better results than those of the simulated experiments and also benefits from the high correlations of the reflectances, to some degree. Additionally, if the thick clouds and shadows are on the same (overlapping) region in all of the temporal data, the recovery outcome is poor because of the disappearance of ample supporting and complementary

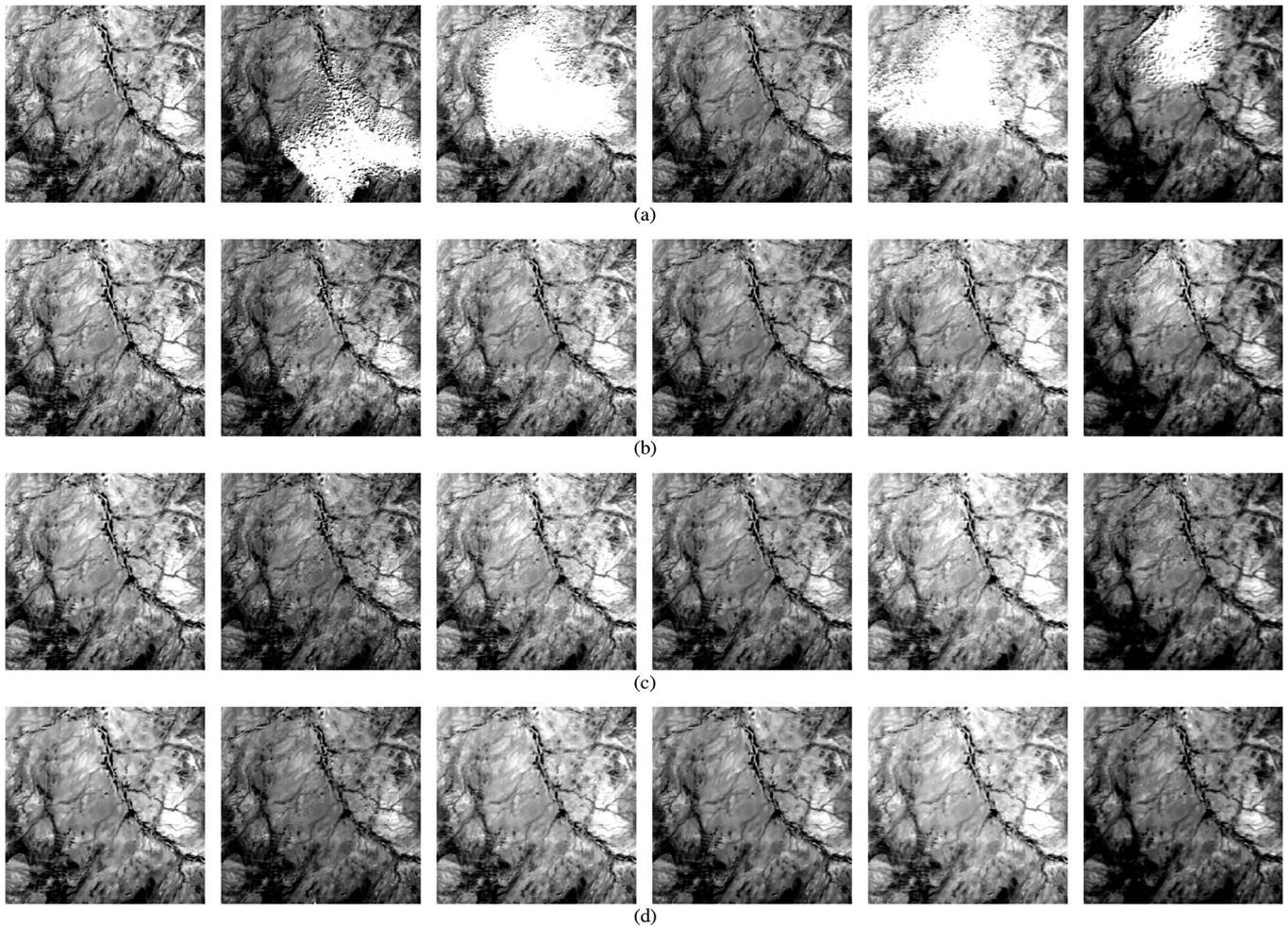


Fig. 11. Real recoveries of the reflectances of MODIS contaminated by thick clouds. (a) Original. (b) HANTS recovery. (c) MT-BPFA recovery. (d) MT-KSVD recovery.

TABLE VI
EXPERIMENTAL DATA IN THE REAL APPLICATIONS

Reflectance products	LST products
MOD02QKM.A2008353	MOD11A1.A2012074
MOD02QKM.A2008360	MOD11A1.A2012076
MOD02QKM.A2008362	MOD11A1.A2012077
MOD02QKM.A2009003	MOD11A1.A2012079
MOD02QKM.A2009005	MOD11A1.A2012081
MOD02QKM.A2009006	MOD11A1.A2012083
MOD02QKM.A2008353	MOD11A1.A2012084

information. In this situation (shown in Fig. 11), MT-BPFA may produce some artifacts on the overlapping region, and HANTS and MT-KSVD can both generate some false edges on the overlapping region. Nevertheless, if the other temporal data items provide sufficient information, this issue will be effectively alleviated.

Another group of MODIS LST products is also used in the real experiments. Similarly, the consecutive daily LST products are not used in the experiments. As LST can vary greatly with time, the time interval of the product series is not so long. The experimental data are shown in Table VI. Like the previous experiment for reflectances, this group of multitemporal LSTs are also contaminated by thick clouds and shadows of varying degrees. Fig. 12 shows the original and cloud recovery results

with HANTS, MT-BPFA, and MT-KSVD, respectively, with a size of 300 by 300 pixels. Fig. 12(a) is on the 79th day of 2012 and is the fourth temporal data item in the group. On the whole, all three of the methods successfully recover the data corrupted by thick clouds. As a result of the dramatic changes and the weak correlations of the LSTs, the results of HANTS, MT-BPFA, and MT-KSVD are all different from each other. Based on the visual effects, MT-BPFA and MT-KSVD perform better than HANTS. However, due to the lack of cloud-free LST, it is impossible to make an objective evaluation of the results.

IV. CONCLUSION

This paper has proposed two expanded multitemporal dictionary learning (MT-BPFA and MT-KSVD) algorithms for the recovery of the quantitative products of optical and infrared RS that are contaminated by large-area thick clouds and the consequent cloud shadows. Based on BPFA, MT-BPFA uses CC between the time-series quantitative data as the weights of the independent dictionary atoms to trade off their latent temporal correlations; based on KSVD, MT-KSVD first makes the auxiliary data a linear/nonlinear temporal transformation to the space of the data to be recovered and then finds an optimal temporal path through the strength of their CCs when

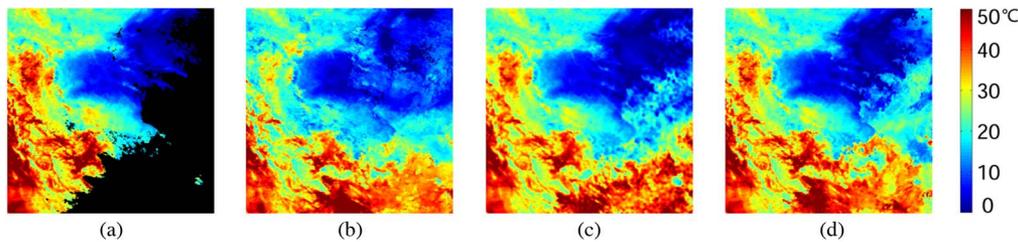


Fig. 12. Real recovery of the LSTs of MODIS contaminated by thick clouds. (a) Original. (b) HANTS recovery. (c) MT-BPFA recovery. (d) MT-KSVD recovery.

dictionary learning is undertaken. In general, both methods can effectively get rid of the harmful influence of clouds on the subsequent applications with these quantitative products. However, MT-BPFA achieves more effective results than MT-KSVD on the condition that more valid data are available. It is this point that restricts its practical values and industrial potential. In other words, MT-KSVD gets better effects when using less data. If the objective condition is sufficient, it will be a wise choice to resort to MT-BPFA when faced with thick clouds. Additionally, in terms of MT-BPFA, a gamma prior for the variance of dictionary may be imposed so that it can be automatically determined from the data, which is our future research direction.

REFERENCES

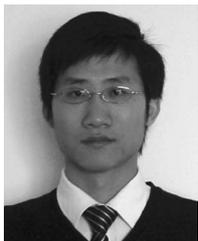
- [1] A. C. Siravenha, D. Sousa, A. Bispo, and E. Pelaes, "Evaluating inpainting methods to the satellite images clouds and shadows removing," *Signal Process., Image Process. Pattern Recognit.*, vol. 260, pp. 56–65, Dec. 2011.
- [2] L. Lorenzi, F. Melgani, and G. Mercier, "Inpainting strategies for reconstruction of missing data in VHR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 5, pp. 914–918, Sep. 2011.
- [3] A. Maalouf, P. Carré, B. Augereau, and C. Fernandez-Maloigne, "A bandelet-based inpainting technique for clouds removal from remotely sensed images," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 7, pp. 2363–2371, Jul. 2009.
- [4] H. Shen and L. Zhang, "A MAP-based algorithm for destriping and inpainting of remotely sensed images," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 5, pp. 1492–1502, May 2009.
- [5] Q. Cheng, H. Shen, L. Zhang, and P. Li, "Inpainting for remotely sensed images with a multichannel nonlocal total variation model," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 175–187, Jan. 2014.
- [6] X. Li, H. Shen, L. Zhang, H. Zhang, and Q. Yuan, "Dead pixel completion of aqua MODIS band 6 using a robust M-estimator multiregression," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 4, pp. 768–772, Apr. 2014.
- [7] C. Zeng, H. Shen, and L. Zhang, "Recovering missing pixels for Landsat ETM + SLC-off imagery using multi-temporal regression analysis and a regularization method," *Remote Sens. Environ.*, vol. 131, pp. 182–194, Apr. 2013.
- [8] D.-C. Tseng, H.-T. Tseng, and C.-L. Chien, "Automatic cloud removal from multi-temporal SPOT images," *Appl. Math. Comput.*, vol. 205, no. 2, pp. 584–600, Nov. 2008.
- [9] C.-H. Lin, P.-H. Tsai, K.-H. Lai, and J.-Y. Chen, "Cloud removal from multitemporal satellite images using information cloning," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 232–241, Jan. 2013.
- [10] F. Melgani, "Contextual reconstruction of cloud-contaminated multitemporal multispectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 2, pp. 442–455, Feb. 2006.
- [11] C. Yu, L. Chen, L. Su, M. Fan, and S. Li, "Kriging interpolation method and its application in retrieval of MODIS aerosol optical depth," in *Proc. 19th Int. Conf. Geoinformat.*, Shanghai, China, 2011, pp. 1–6.
- [12] A. Limaïem, A. Nassef, and H. A. El-Maraghy, "Data fitting using dual kriging and genetic algorithms," *CIRP Annals-Manuf. Technol.*, vol. 45, no. 1, pp. 129–134, Jan. 1996.
- [13] A. Savitzky and M. J. Golay, "Smoothing and differentiation of data by simplified least squares procedures," *Anal. Chem.*, vol. 36, no. 8, pp. 1627–1639, Jul. 1964.
- [14] J. Chen, P. Jönsson, M. Tamura, Z. Gu, B. Matsushita, and L. Eklundh, "A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky–Golay filter," *Remote Sens. Environ.*, vol. 91, no. 3/4, pp. 332–344, Jun. 2004.
- [15] M. E. Jakubauskas, D. R. Legates, and J. H. Kastens, "Harmonic analysis of time-series AVHRR NDVI data," *Photogramm. Eng. Remote Sens.*, vol. 67, no. 4, pp. 461–470, Apr. 2001.
- [16] H. Carrão, P. Goncalves, and M. Caetano, "A nonlinear harmonic model for fitting satellite image time series: Analysis and prediction of land cover dynamics," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 4, pp. 1919–1930, Apr. 2010.
- [17] E. B. Brooks, V. A. Thomas, R. H. Wynne, and J. W. Coulston, "Fitting the multitemporal curve: A Fourier series approach to the missing data problem in remote sensing analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 9, pp. 3340–3353, Sep. 2012.
- [18] M. E. Jakubauskas, D. R. Legates, and J. H. Kastens, "Crop identification using harmonic analysis of time-series AVHRR NDVI data," *Comput. Electron. Agri.*, vol. 37, no. 1–3, pp. 127–139, Dec. 2002.
- [19] G. Roerink, M. Menenti, and W. Verhoef, "Reconstructing cloudfree NDVI composites using Fourier analysis of time series," *Int. J. Remote Sens.*, vol. 21, no. 9, pp. 1911–1917, Sep. 2000.
- [20] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- [21] E. Candes and T. Tao, "Near optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.
- [22] E. Candes, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, Aug. 2006.
- [23] L. Lorenzi, F. Melgani, and G. Mercier, "Missing-area reconstruction in multispectral images under a compressive sensing perspective," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 7, pp. 3998–4008, Jul. 2013.
- [24] 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.
- [25] M. Zhou, H. Chen, J. Paisley, L. Ren, L. Li, Z. Xing, D. Dunson, G. Sapiro, and L. Carin, "Nonparametric Bayesian dictionary learning for analysis of noisy and incomplete images," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 130–144, Jan. 2012.
- [26] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.
- [27] H. Shen, X. Li, L. Zhang, D. Tao, and C. Zeng, "Compressed sensing-based inpainting of aqua moderate resolution imaging spectroradiometer band 6 using adaptive spectrum-weighted sparse Bayesian dictionary learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 2, pp. 894–906, Feb. 2014.
- [28] J. Shihao, X. Ya, and L. Carin, "Bayesian compressive sensing," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2346–2356, Jun. 2008.
- [29] Z. Xing, M. Zhou, A. Castrodad, G. Sapiro, and L. Carin, "Dictionary learning for noisy and incomplete hyperspectral images," *SIAM J. Imag. Sci.*, vol. 5, no. 1, pp. 33–56, Jan. 2012.
- [30] G. Xu and Z. Xu, "Compressed Sensing Matrices from Fourier Matrices 2013," arXiv preprint arXiv:1301.0373.
- [31] J. M. Duarte-Carvajalino and G. Sapiro, "Learning to sense sparse signals: Simultaneous sensing matrix and sparsifying dictionary optimization," *IEEE Trans. Image Process.*, vol. 18, no. 7, pp. 1395–1408, Jul. 2009.
- [32] R. G. Baraniuk, "Compressive sensing [lecture notes]," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 118–121, Jul. 2007.
- [33] S. Chen and D. Donoho, "Basis pursuit," in *Proc. 28th ASILOMAR*, 1994, pp. 41–44.
- [34] P. S. Huggins and S. W. Zucker, "Greedy basis pursuit," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3760–3772, Jul. 2007.

- [35] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.
- [36] D. Needell and R. Vershynin, "Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit," *Found. Comput. Math.*, vol. 9, no. 3, pp. 317–334, Jun. 2009.
- [37] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, pp. 301–321, May 2009.
- [38] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2230–2249, May 2009.
- [39] I. F. Gorodnitsky and B. D. Rao, "Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm," *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 600–616, Mar. 1997.
- [40] M. F. Duarte, M. A. Davenport, M. B. Wakin, and R. G. Baraniuk, "Sparse signal detection from incoherent projections," in *Proc. IEEE ICASSP*, Toulouse, France, 2006, pp. III(305)–III(308).
- [41] I. Tosic and P. Frossard, "Dictionary learning," *IEEE Signal Process. Mag.*, vol. 28, no. 2, pp. 27–38, Mar. 2011.
- [42] I. Ram, M. Elad, and I. Cohen, "Image processing using smooth ordering of its patches," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2764–2774, Jul. 2013.
- [43] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight, "Sparsity and smoothness via the fused lasso," *J. Roy. Stat. Soc., Ser. B (Stat. Methodol.)*, vol. 67, no. 1, pp. 91–108, Feb. 2005.
- [44] A. Rinaldo, "Properties and refinements of the fused lasso," *Ann. Stat.*, vol. 37, no. 5B, pp. 2922–2952, Oct. 2009.
- [45] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. New York, NY, USA: Springer-Verlag, 2010.



Xinghua Li received the B.S. degree in geographical information system from Wuhan University, Wuhan, China, in 2011, where he is currently working toward the Ph.D. degree in the School of Resource and Environmental Sciences.

His current research interests focus on image inpainting and compressed-sensing-based image processing.



Huanfeng Shen (M'10–SM'13) received the B.S. degree in surveying and mapping engineering and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2002 and 2007, respectively.

In July 2007, he joined the School of Resource and Environmental Sciences, Wuhan University, where he is currently a Full Professor. He has been supported by several talent programs, including the New Century Excellent Talents by the Ministry of Education of China (2011) and the Hubei Science Fund for

Distinguished Young Scholars (2011). He is the Principal Investigator (PI) of three National Natural Science Foundation of China projects and a subproject of the Major State Basic Research Development Program of China (973 program). He is also the co-PI of a project funded by the National High Technology Research and Development Program (863 program). He has published more than 80 research papers. His research interests include image processing (for quality improvement), remote sensing application, and data fusion and assimilation.



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

He is currently the Head of the Remote Sensing Division, State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University. He is also a "Chang-Jiang Scholar" Chair Professor appointed by the Ministry of Education of China. He is currently a Principal Scientist for the China State Key Basic Research Project (2011–2016) appointed by the Ministry of National Science and Technology of China to lead the remote sensing program in China. He has more than 310 research papers. He is the holder of five patents. His research interests include hyperspectral remote sensing, high-resolution remote sensing, image processing, and artificial intelligence.

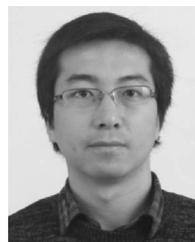
Dr. Zhang is a Fellow of the IEE, Executive Member (Board of Governor) of the China National Committee of International Geosphere–Biosphere Programme, Executive Member of the China Society of Image and Graphics, etc. He regularly serves as a Cochair of the series SPIE Conferences on Multispectral Image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



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

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

Dr. Zhang is a Reviewer of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, and *Journal of Applied Remote Sensing*.



Qiangqiang Yuan received the B.S. degree in surveying and mapping engineering and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2006 and 2012, respectively.

In July 2012, he joined the School of Geodesy and Geomatics, Wuhan University, where he is currently an Assistant Professor. His research interests focus on image restoration and remote sensing image processing and application.



Gang Yang received the M.S. degree in geographical information system from Hunan University of Science and Technology, Xiangtan, China, in 2012. He is currently working toward the Ph.D. degree in the School of Resource and Environmental Sciences, Wuhan University, Wuhan, China.

His current research interests focus on image inpainting and remote sensing time-series product temporal reconstruction.