Spatiotemporal Reflectance Fusion via Sparse Representation

1 author:

Bo Huang
The Chinese University of Hong Kong
295 PUBLICATIONS **8,389** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project   Photogrammetric Engineering & Remote Sensing-2009/8/1 View project

Project   Evaluating and Characterizing Urban Vibrancy by Fusing Multi-Source Spatial Big Data View project

Spatiotemporal Reflectance Fusion via Sparse Representation

Bo Huang, *Associate Member, IEEE*, and Huihui Song

Abstract—This paper presents a novel model for blending remote sensing data of high spatial resolution (HSR), taken at infrequent intervals, with those available frequently but at low spatial resolution (LSR) in the context of monitoring and predicting changes in land usage and phenology. Named "SParse-representation-based SpatioTemporal reflectance Fusion Model" (SPSTFM), the model has been developed for predicting HSR surface reflectances through data blending with LSR scenes. Remarkably, this model forms a unified framework for fusing remote sensing images with temporal reflectance changes, phenology change (e.g., seasonal change of vegetation), or type change (e.g., conversion of farmland to built-up area), by establishing correspondences between structures within HSR images of given areas and their corresponding LSR images. Such corresponding relationship is achieved by means of the sparse representation, specifically by jointly training two dictionaries generated from HSR and LSR difference image patches and sparse coding at the reconstruction stage. SPSTFM was tested using both a simulated data set and an actual data set of Landsat Enhanced Thematic Mapper Plus—Moderate Resolution Imaging Spectroradiometer acquisitions. It was also compared with other related algorithms on two types of data: images primarily with phenology change and images primarily with land-cover type change. Experimental results demonstrate the superiority of SPSTFM in capturing surface reflectance changes on both categories of images.

Index Terms—Image fusion, land-cover change, phenology change, sparse representation, spatial and temporal resolution.

I. INTRODUCTION

WITH growing international concern over global environmental change as a result of human activities, monitoring changes on the Earth's surface has been becoming increasingly important [1], [2]. Although a large number of orbiting remote sensing instruments with different spatial, temporal, and spectral characteristics have been launched, resulting in a dramatic improvement in the ability to acquire images of the Earth's surface, a key technological challenge confronting such instruments is the tradeoff that is still required between spatial resolution and temporal coverage. For instance,

B. Huang is with the Department of Geography and Resource Management and the Institute of Space and Earth Information Science, The Chinese University of Hong Kong, Shatin, Hong Kong (e-mail: bohuang@cuhk.edu.hk).

H. Song is with the Department of Geography and Resource Management, The Chinese University of Hong Kong, Shatin, Hong Kong (e-mail: shhsherry@cuhk.edu.hk).

most instruments with fine spatial resolution (e.g., Système Pour l'Observation de la Terre (SPOT) and Landsat with 10–30-m spatial resolution) can only revisit the same location on Earth at intervals of half to one month, while other instruments with coarse spatial resolution (e.g., Moderate Resolution Imaging Spectroradiometer (MODIS) and SPOT Vegetation with 250–1000-m spatial resolution) can make repeated observations in one day [3]. As a result, there is so far still no sensor that can provide both high spatial resolution (HSR) and frequent temporal coverage. One possible cost-effective solution is to explore data integration methods that can blend the two types of images from different sensors to generate high-resolution synthetic data in both space and time [4]–[6], thereby enhancing the capability of remote sensing for monitoring land surface dynamics, particularly in rapidly changing areas.

While a broad range of image fusion methods has been developed in an attempt to combine remotely sensed data from various sensors at different resolutions [7]–[12], most of them have focused on the fusion of panchromatic and multispectral images that combine the HSR of the panchromatic band with the spectral characteristics of multispectral observations acquired simultaneously at low spatial resolution (LSR), with little attention paid to the simultaneous enhancement of spatial resolution and temporal coverage. Only in recent years have we witnessed the emergence of spatiotemporal reflectance fusion models for predicting an unknown HSR surface reflectance, based on its LSR counterpart at the prediction date (e.g., t_2) and other HSR and LSR reflectance pairs at adjacent dates (e.g., t_1 and t_3, $t_1 < t_2 < t_3$) [7]–[9].

One critical problem that should be addressed by a spatiotemporal reflectance fusion model is that of detecting the temporal change of reflectance over different pixels during an observation period. In general, such change encompasses both phenology change (e.g., seasonal change of vegetation) and type change (e.g., conversion of bare soil to concrete surface), and it is considered more challenging to capture the latter than the former in a fusion model [6]. Ideally, both changes should be incorporated precisely in the fused (predicted) product. Gao *et al.* [4] proposed a spatial and temporal adaptive reflectance fusion model (STARFM) algorithm to blend Landsat (30-m spatial resolution and 16-day revisit rate) with MODIS (500-m spatial resolution and daily revisit rate) data for the purpose of generating daily Landsat surface reflectance. STARFM detects reflectance changes from the MODIS data and predicts the Landsat reflectance through similar neighboring pixels weighted by spectral, temporal, and spatial distances. These features have rendered STARFM particularly well suited to cope with seasonal changes in large-scale distribution of

vegetation, although adopting neighboring pixels may introduce blurs in the predicted image which, in turn, incurs the loss of high-frequency details. As a pioneering model for spatiotemporal data blending, STARFM has since rapidly gained popularity, leading to enhanced STARFM (ESTARFM) [5], which was devised to improve on STARFM in predicting Landsat surface reflectance in more heterogeneous regions using subpixels' information conditional on the change rate of reflectance of each endmember being stable during the observation period. In another STARFM-based work, Hilker *et al.* [6] developed a new fusion model, Spatial and Temporal Adaptive Algorithm for mapping Reflectance CHange (STAARCH), which uses the Landsat images at the two endpoints of an observation period and a dense series of spatially coincident MODIS imagery as input. One of the images at the two ends may finally be chosen according to the closeness to the change occurrence time that is detected by the series of MODIS reflectance data. Among these algorithms, only STAARCH has considered the landscape disturbances, whereas others assume that land-cover types remain unchanged. Relying on the so-called "tasseled cap transformations" [10], the utility of STAARCH may, however, be limited to forest disturbances, rather than for broad types of land-cover change (e.g., transition from vacant land to built-up area). More importantly, it faces the same difficulty as STARFM in delineating and characterizing the disturbances precisely when estimating the Landsat surface reflectance. Consequently, there is a clear need to develop a more generic spatiotemporal reflectance model to systematically account for significant type changes that challenge current methods of monitoring land surface dynamics, particularly in rapidly changing areas. Meanwhile, phenology change should also be thoroughly addressed as it might make up the greatest proportion of change over a short period.

This paper seeks to develop a novel data fusion model, called SParse-representation-based SpatioTemporal reflectance Fusion Model (SPSTFM), that can account for all the reflectance changes during an observation period, whether type or phenology change, in a unified way. Sparse representation is exploited to support single-frame image superresolution that has not been incorporated in STARFM [4] or similar methods [5], [6] because it allows for learning the structure primitives of signals via an overcomplete dictionary and reconstructing signals through sparse coding [14]. Given the similarities among the sequence of remote sensing images, it would be more efficient to predict a difference image, which can then be easily added back to the reference image to produce the predicted HSR scene. Hence, SPSTFM learns the differences between two HSR images and their corresponding LSR acquisitions from a different instrument via sparse signal representation, specifically an analogous dictionary-pair learning method, which was also adopted by Yang *et al.* [15] for natural image superresolution. In [15], the high-resolution images and their two-to-four-times downsampled low-resolution scenes are correlated by two dictionaries, which are generated from the high- and low-resolution image patches and trained in the "feature" domain. However, the magnification factor in our case is 8–16 given the Landsat and MODIS as representatives of HSR and LSR images, respectively; thus, SPSTFM focuses on building up the

corresponding relationship in the "change" domain of HSR and LSR images.

STARFM-based methods estimate the change or change trend from the LSR images (i.e., MODIS image). However, detailed change information is lost in the LSR images, which, in turn, leads to deviation in projecting the changes in the HSR image. By contrast, the proposed model establishes the correspondences between HSR and LSR difference image patches and, hence, can directly predict the changing trend of HSR images in patches. By jointly training the dictionaries of high- and low-resolution images in the form of patches, SPSTFM can predict the high-resolution difference image (HRDI) more accurately than searching similar neighbors for every pixel because it considers the structural similarity (SSIM), particularly for land-cover type changes. Rather than supposing a linear change of reflectance in the ESTARFM method, sparse representation can obtain the change prediction in an intrinsic nonlinear form because sparse coding is a nonlinear reconstruction process through selecting the optimal combination of signal primitives [16]. Equipped with the aforementioned properties, the proposed model has been tested on both simulated and actual data sets that carry phenology and type changes. Results show that it has made significant improvements over STARFM and ESTARFM in capturing both the phenology change and type change precisely when predicting an HSR surface reflectance.

The remainder of this paper is organized as follows. In Section II, sparse representation and sparse coding are briefly introduced, and then, the proposed methodology is presented in Section III. The experimental results with both simulation data and actual data are shown in Section IV which also includes the comparisons with STARFM. Finally, this paper is concluded in Section V.

II. SPARSE REPRESENTATION

The notion of sparsity in signal representation arose in the 1970s; however, the methods such as fast Fourier transform and Karhunen–Loeve transform at that time were all based on linear transforms. Since the emergence of flexible nonlinear formulation in the 1990s, sparse representation has made a big step forward. At the same time, dictionaries and sparse coding (DSC) began to take over the role of transforms because of their localization, adaptivity, geometric invariance, and over-completeness [14]. Localization means that DSC allow more flexible representations based on the local signal characteristics, reducing the side effects of irregularities. Adaptivity denotes that DSC can adapt themselves to the signal content to obtain further sparsity. Finally, dictionaries can achieve geometric invariance when representing signals by containing redundant atoms (overcompleteness).

Suppose that there is a series of signals $\{x_1, x_2, \ldots, x_N\}$, $x_i \in R^n$. In this paper, specifically, each such signal represents a $\sqrt{n} \times \sqrt{n}$ image patch, obtained by lexicographically stacking the pixel values. Sparse representation assumes that there exists a matrix $D \in R^{n \times m}$, $n < m$, where each column stands for a basic atom, m stands for the number of atoms in D, and each signal is supposed to be sparsely represented by the linear combination of a few atoms in D. Thus, an image signal can

be represented by $x = D\alpha$, where $\alpha \in R^m$ denotes the sparse representation coefficients of x with respect to dictionary D. Our aim is to design an efficient dictionary and then find x's representation coefficients with the fewest nonzero elements via a sparse coding algorithm.

The aforementioned problem can be mathematically solved by optimizing the following equation:

$$\hat{\alpha} = \arg\min \|\alpha\|_0 \qquad \text{s.t.} \quad x = D\alpha \qquad (1)$$

where $\|\alpha\|_0$ denotes the number of nonzero elements in α. Considering the existence of representation error and noise in the original image signal, the constrained equation in (1) can be expressed in the following relaxed form:

$$\hat{\alpha} = \arg\min \|\alpha\|_0 \qquad \text{s.t.} \quad \|x - D\alpha\|_2^2 < \varepsilon \qquad (2)$$

where ε is a small constant which is relevant with noise level of the source image. However, the optimization of (2) is an Non-deterministic Polynomial-time hard problem [17] due to the presence of l_0-norm. Fortunately, it has been proven that the l_0-norm problem can be approximated by l_1-norm when the representation coefficients are sufficiently sparse and the atoms of dictionary are different [14], i.e.,

$$\hat{\alpha} = \arg\min \|\alpha\|_1 \qquad \text{s.t.} \quad \|x - D\alpha\|_2^2 < \varepsilon. \qquad (3)$$

Several fast l_1-minimization algorithms have been developed with both relatively high reconstruction precision and quick convergence speed [18]. Here, we adopt the gradient projection sparse representation algorithm in [19], which seeks a sparse representation α along a specified gradient direction. This algorithm reformulates the original problem (3) as an unconstrained one through a Lagrangian formulation

$$\alpha^* = \arg\min_{\alpha} f(\alpha) = \arg\min_{\alpha} \frac{1}{2}\|x - D\alpha\|_2^2 + \lambda\|\alpha\|_1 \qquad (4)$$

where the parameter λ balances the sparsity of the solution and fidelity of the approximation to x. In order to formulate the aforementioned equation as a standard quadratic programming (QP) problem, the coefficients α are separated into positive coefficients α_+ and negative ones α_-, and (4) is rewritten as

$$\min_{\alpha} F(\alpha) = \frac{1}{2}\|x - [D, -D][\alpha_+; \alpha_-]\|_2^2 + \lambda I^T(\alpha_+ + \alpha_-),$$
$$\text{s.t.} \quad \alpha_+ \geq 0; \quad \alpha_- \geq 0. \qquad (5)$$

Thus, (5) can be rewritten in the standard QP form

$$\min_{\alpha} F(\beta) \doteq c^T\beta + \frac{1}{2}\beta^T E\beta, \qquad \text{s.t.} \quad \beta \geq 0 \qquad (6)$$

where $\beta = [\alpha_+; \alpha_-]$, $c = \lambda I + [-D^T x; D^T x]$, and

$$E = \begin{pmatrix} D^T D & -D^T D \\ -D^T D & D^T D \end{pmatrix}. \qquad (7)$$

Furthermore, the gradient of $F(\beta)$ can be obtained as

$$\nabla_\beta F(\beta) = c + E\beta. \qquad (8)$$

The steepest descent algorithm can be applied to search along the negative gradient $-\nabla F(\beta)$ for each iteration β^k as

$$\beta^{k+1} = \beta^k - s^k \nabla F(\beta^k) \qquad (9)$$

where s^k is the step size that can be updated during the iteration process.

III. METHODOLOGY

In our model, the Landsat Enhanced Thematic Mapper Plus (ETM+) surface reflectance data are selected as the image examples with HSR–low temporal resolution, while the MODIS surface reflectance data are selected as the image examples with LSR–high temporal resolution. In the following, we will first show how the image superresolution can be integrated into the process of spatiotemporal image fusion. Then, we will present the dictionary-pair learning process. Subsequently, we will show how to predict the difference image of HSR data using the learned dictionary pair and finally demonstrate the procedure of constructing an HSR image.

A. From Image Superresolution to Spatiotemporal Fusion

Generally, there exist three classes of methods for image superresolution: interpolation-, reconstruction-, and learning-based methods [15], [20]. Since the interpolation-based methods cannot take full advantage of prior information, they could fail in recovering the missing high-frequency components. The reconstruction-based methods require multiple low-resolution images of the same scene, which are aligned with subpixel accuracy. This type of method reconstructs the high-resolution image by optimizing a function including an image degradation model and some prior constraints. When the operators in the degradation model (such as the distortion matrix and the blur matrix) are accurately estimated, these approaches can perform well. However, they cannot perform satisfactorily when the resolution difference between high- and low-resolution images is larger than usual (four times) because of lacking sufficient prior information. To obtain adequate prior information, learning-based methods for superresolution have emerged, which can add more prior information from other source images. Freeman et al. [21] first proposed a learning-based method for low-level vision problems. This method tackles the image superresolution problem by sampling a large amount of image patch pairs and using a Markov network to establish a corresponding relationship between sharp (high-resolution) and blurred (low-resolution) images. Yang et al. [15] presented another superresolution algorithm based on sparse representation. By jointly training two dictionaries of high-frequency feature patches of low- and high-resolution images, this algorithm can enforce the similarity of sparse representations between the low- and high-resolution patch pair with respect to their own dictionaries. Compared to the previous sample-based dictionaries [21], the learned dictionary pair is a more compact representation of the patch pairs, which can generate more competitive results. Such a dictionary-pair learning method is adapted to the difference images in HSR and LSR in this paper.
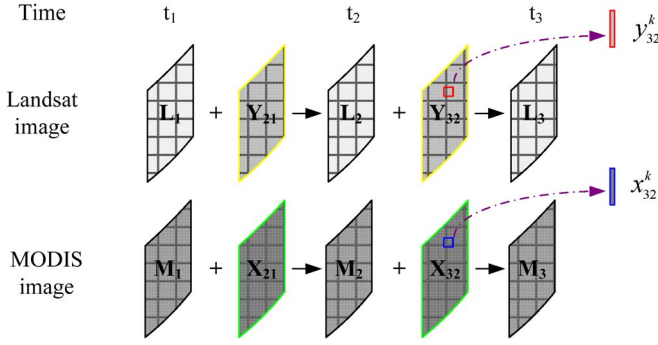
Fig. 1. Relationship diagram for the images and difference images, where y_{32}^k and x_{32}^k stand for the kth patches of Y_{32} and X_{32}, respectively.

Our task is to utilize two Landsat–MODIS image pairs on t_1 and t_3 dates and one MODIS image on t_2 date ($t_1 < t_2 < t_3$) to accurately predict the Landsat image on t_2. Apparently, there exists an analogous low- and high-resolution image pair which is available to build a dictionary pair: the difference image of MODIS data [i.e., the low-resolution difference image (LRDI)] from t_1 to t_3 and its corresponding difference image of Landsat data (i.e., the HRDI) for the same period. In the method of Yang *et al.* [15], the dictionary pair is trained on numerous natural images which are irrelevant to the image to be reconstructed, yet all these natural images (including the reconstructed image) are supposed to possess common elementary structures (primarily the high-frequency information, e.g., edges). In our problem, however, on the one hand, the difference images (LRDI and HRDI) themselves have high-frequency details; on the other hand, the difference images (i.e., between t_1 and t_2, t_2 and t_3, and t_1 and t_3) are closely related. Thus, predicting the HRDI between t_1 and t_2, and t_2 and t_3, from the known HRDI between t_1 and t_3 is reasonable. Specifically, our desired difference image patches can be predicted through the linear combination of the structure primitives extracted from the known HRDI.

B. Dictionary-Pair Learning

We denote the Landsat image and the MODIS image as L_i and M_i on t_i date, respectively, where the MODIS images are extended to have the same size as Landsat via bilinear interpolation. Let Y_{ij} and X_{ij} represent the HRDI and LRDI between t_i and t_j, respectively, and their corresponding patches are x_{ij} and y_{ij}, which are formed by putting patches into column vectors. The relationship diagram for these variables is shown in Fig. 1. L_2 can then be predicted as follows:

$$L_2 = W_1 * (L_1 + \hat{Y}_{21}) + W_3 * (L_3 - \hat{Y}_{32}) \tag{10}$$

where $\hat{Y}_{21}$ and $\hat{Y}_{32}$ are the predicted HRDIs and W_1 is the weighting parameter for the predicted image on t_2 using the Landsat reference image on t_1, which is similar for W_3. The determination of the weighting parameters will be illustrated in Section III-D.

In order to estimate $\hat{Y}_{21}$ and $\hat{Y}_{32}$ in (10), we first need to formulate the dictionary pair D_l and D_m, which are trained using the HRDI and LRDI patches between t_1 and t_3, respec-

tively. Based on the theory of sparse representation and sparse coding in Section II, the dictionary pair can be trained using the following optimization equation:

$$\{D_l^*, D_m^*, \Lambda^*\}$$
$$= \arg \min_{D_l, D_m, \Lambda} \left\{ \|Y - D_l \Lambda\|_2^2 + \|X - D_m \Lambda\|_2^2 + \lambda \|\Lambda\|_1 \right\} \tag{11}$$

where Y and X are the column combination of lexicographically stacking image patches, sampled randomly from Y_{13} and X_{13}, respectively. Similarly, Λ is the column combination of representation coefficients corresponding to every column in Y and X. Here, we name the learned dictionaries D_l and D_m being the HSR dictionary and LSR dictionary, respectively. In the algorithm by Yang *et al.* [15], the training samples are image edge structures, and the two dictionaries are learned through concatenation by optimizing the following function:

$$\{D^*, \Lambda^*\} = \arg \min_{D, \Lambda} \left\{ \|Z - D\Lambda\|_2^2 + \lambda \|\Lambda\|_1 \right\} \tag{12}$$

where $Z = [Y; X]$ and $D = [D_l; D_m]$. Considering the difference between HRDI and LRDI in amplitude and variance, we adopt the alternate update mode to solve D_l, D_m, and Λ in (11); see [22] for the detailed process of dictionary column update. The training process is summarized in Fig. 2. The patch size of the difference image here is chosen to be 7×7, the number of the dictionary atoms is 256, and the number of training samples is 2000. Considering the reflectance difference between Y and X and among different bands, Y and X of every band in step 1 are normalized as follows:

$$Y = \frac{Y - \mu_x}{sigma_x} \qquad X = \frac{X - \mu_x}{sigma_x} \tag{13}$$

where μ_x and $sigma_x$ are the mean and variance of the original X, respectively.

C. HRDI Reconstruction

Since the sparse coefficients are enforced to be the same in the dictionary training process, if obtaining the representation coefficients of the LRDI patch with respect to the dictionary D_m, the corresponding HRDI patch can then be reconstructed using the same representation coefficients and the corresponding dictionary D_l. With x_{21}^k denoting the kth $\sqrt{n} \times \sqrt{n}$ image patch (in the column form) of X_{21}, we can estimate the representation coefficients by minimizing the following l_1-norm problem:

$$\alpha^* = \arg \min_{\alpha} f(\alpha) = \arg \min_{\alpha} \frac{1}{2} \left\| x_{21}^k - D_m \alpha \right\|_2^2 + \lambda \|\alpha\|_1. \tag{14}$$

The HRDI patch counterpart $\hat{y}_{21}^k$ can then be solved

$$\hat{y}_{21}^k = D_l \alpha^*. \tag{15}$$

The implementation details of the reconstruction process are summarized in Fig. 3, which takes the $\hat{Y}_{21}$ reconstruction

Algorithm 1. Training dictionary-pair

Input:
- m and N are the dimension and number of samples.
- High-resolution training samples
  $Y = \{y^1, y^2, ..., y^N\}, y^i \in R^m$ and low-resolution training
  samples $X = \{x^1, x^2, ..., x^N\}, x^i \in R^m$.
- K is the number of atoms in the dictionary; J is the maximum iteration number.

Initialize:
- $D_l \leftarrow \{d_l^1, d_l^2, ..., d_l^K\}$, $D_m \leftarrow \{d_m^1, d_m^2, ..., d_m^K\}$, where d_l^k and d_m^k are randomly sampled from Y and X, respectively.
- j=0; Set the initializing error of the objective function between two consecutive iterations as $E^0 = 0$ and the acceptable error is ε (set to be 0.3 in our experiment).

While j<J and $E^j < \xi$:

1. Fix D_l and D_m, update each column of sparse coefficients Λ using GPSR algorithm (introduced in section II) by optimizing the following equation:
   $\arg\min_{\{\Lambda\}} \{\|S - D\Lambda\|_2^2 + \lambda \|\Lambda\|_1\}$, where $S = [Y; X]$, $D = [D_l; D_m]$

2. Fix Λ and D_m, update D_l, now the objective function is reduced to $\arg\min_{\{D_l\}} \{\| Y - D_l\Lambda \|_2^2\}$.
   Updated the dictionary by column and fix other columns when updating one column .

3. Fix Λ and D_l, update D_m, the objective function is similar to step 2, $\arg\min_{\{D_m\}} \{\|X - D_m\Lambda\|_2^2\}$. The update process is also similar to step 2.

4. Compute E^j.

Output:
- High-resolution dictionary D_l and low-resolution dictionary D_m.

Fig. 2.   Dictionary-pair learning process.

Algorithm 2. Reconstructing HRDI

1: Given dictionaries D_l, D_m, and the LRDI X_{21} between t_1 and t_2.

2: *For* each 7×7 patch x_{21}^k of X_{21} taken starting from the upper-left corner with 2 pixels overlap in each direction, do

3:   Solve the optimization problem with x_{21}^k and D_m defined in Eq. (14);

4:   Generate the corresponding HRDI patch $\hat{y}_{21}^k$ using Eq. (15);

5: *End*

6: The output is HRDI $\hat{Y}_{21}$, which is the column combination of all $\hat{y}_{21}^k$.

Fig. 3.   Reconstruction process of HRD.

process as an example, and $\hat{Y}_{23}$ can be estimated in the same procedure.

D. Landsat Surface Reflectance Reconstruction

Considering the inhomogeneity of local changes in remote sensing images, we utilize different local weightings for predicting the Landsat surface reflectance at t_1 and t_3, respectively, i.e., each local reconstructed patch is weighted differently. Similar to (10), the kth patch of L_2 can be predicted

$$l_2^k = w_1^k * (l_1^k + \hat{y}_{21}^k) + w_3^k * (l_3^k - \hat{y}_{32}^k). \qquad (16)$$

Equation (16) is the corresponding patch form (in column) of (10). After predicting all l_2^k patches, we then put them into the Landsat surface reflectance L_2 (taking the average value in overlapped regions). w_1^k and w_3^k in (16) are local weighting parameters for different reference images. Rather than directly using the reflectance value to compute these weighting parameters, the sum of normalized difference vegetation index and normalized difference built-up index [23] U is employed to assist in measuring the degree of change between two MODIS images. If v_i^k denotes the absolute average change of U in the kth patch of M_i and M_2, then w_i^k is in inverse proportion to it and can be computed as follows:

$$w_i^k = \frac{1/v_i^k}{1/v_1^k + 1/v_3^k}. \qquad (17)$$

However, considering the disturbance events in land-cover type changes (e.g., burns and other human activities), we adopt two weighting strategies. When the absolute difference between v_1^k and v_3^k is greater than a threshold δ (set to be 0.2 empirically), i.e., l_2^k is more like one of them (i.e., l_1^k and l_3^k) than another, we only choose the more similar reference image to make the prediction. If taking the case of $v_1^k < v_3^k$ as an example, then the weighting parameter formulation (17) is rewritten as

$$w_i^k = \begin{cases} 1(\text{or } 0 \text{ when } i = 3), & \text{if } (v_3^k - v_1^k) > \delta \\ \frac{1/v_i^k}{1/v_1^k + 1/v_3^k}, & \text{else.} \end{cases} \qquad (18)$$

IV. EXPERIMENTAL RESULTS AND COMPARISONS

In this section, SPSTFM was compared with the well-known STARFM algorithm in terms of three measures which are detailed hereinafter using both a simulated data set and two actual data sets of Landsat-7 ETM+ and MODIS images. Before applying the two algorithms, the Landsat-7 ETM+ images were radiometrically and atmospherically corrected using the MODIS 6S approach, while the MODIS images were obtained directly from the Land Processes Distributed Active Archive Center (https://lpdaac.usgs.gov/lpdaac) and then resampled to the same size of the Landsat ETM+ image using a bilinear interpolation method.

A. Quality Assessment Indices for Image Fusion Result

To evaluate the quality of fusion results quantitatively and visually, several representative metrics are employed in this

study. The average absolute difference (AAD) between the predicted reflectance and the actual reflectance is shown to directly reflect the deviation of the predicted reflectance. Root mean square error (rmse)—widely applied in quantitative assessment of image qualities—was chosen as the second index for our assessments. To reflect the error fluctuations of the fusion result, the variance of errors (VOE) was adopted as our third index. Also, we utilized the *erreur relative global adimensionnelle de synthèse* (ERGAS) [24] to evaluate the overall fusion result. A better fusion is achieved if ERGAS is smaller and closer to zero.

In addition to quantitative assessment, a visual assessment index, i.e., SSIM [25], was also employed to measure the similarity, on the overall structure, between the predicted and actual images. SSIM is obtained by computing the mean and variance of the predicted–actual comparison images, i.e.,

$$SSIM(L, \hat{L}) = \frac{(2\mu_L \mu_{\hat{L}} + C_1)(2\sigma_{L\hat{L}} + C_2)}{\left(\mu_L^2 + \mu_{\hat{L}}^2 + C_1\right)\left(\sigma_L^2 + \sigma_{\hat{L}}^2 + C_2\right)} \quad (19)$$

where μ_L and $\mu_{\hat{L}}$ stand for the means of the comparison images, σ_L and $\sigma_{\hat{L}}$ are the variances of them, $\sigma_{L\hat{L}}$ is the covariance of $\hat{L}$ and L, and C_1 and C_2 are small constants to avoid instability when $\mu_L^2 + \mu_{\hat{L}}^2$ or $\sigma_L^2 + \sigma_{\hat{L}}^2$ is very close to zero. The comparison rule is as follows: The much closer the SSIM to one, the more similar the $\hat{L}$ to L.

Finally, the scatter plots of the predicted against actual images for each band were also chosen, which provide an intuitive comparison between the estimated and actual reflectances on the approximation extent of distribution.

B. Experiments With Simulated Data

Simulated reflectance data allow us to examine the behavior and performance of the proposed algorithm on both phenology and type changes of land cover. In this simple case, there are only three objects (i.e., a lake represented by a circle, a grassland represented by the upper right rectangle, and a built-up area represented by the lower right square), as shown in Fig. 4. Fig. 4(a)–(c) shows three Landsat-like images with 30-m spatial resolution, and Fig. 4(d)–(f) shows their corresponding MODIS-like images with 500-m spatial resolution, which are aggregated from the Landsat-like images and then are added by white Gaussian noise with a signal-to-noise ratio of 35 dB. It is assumed that the three objects change from t_1 to t_2 to t_3 in the following ways: The circular object keeps a constant reflectance and reduces its radius from 1800 to 1000 m and then to 500 m, the rectangular object remains the same size and changes its reflectance from 20 to 150 and then to 220, and the square object changes its reflectance from 220 to 130 and then to 40 and changes its side length from 1000 to 2000 m and then to 4000 m simultaneously. We used the two Landsat-like images in Fig. 4(a) and (c) and the three MODIS-like images in Fig. 4(d)–(f) to predict the Landsat-like image in Fig. 4(b) and then compared the prediction results with the original image. Fig. 5 shows the prediction result using SPSTFM in comparison with STARFM. Apparently, the image predicted using our algorithm [Fig. 5(c)] is more similar to the original image for all objects, particularly in object shapes and
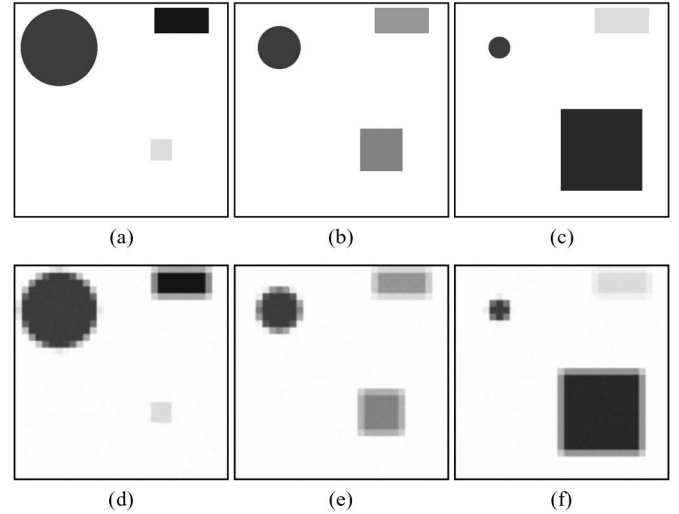

Fig. 4. Three pairs of simulated objects with three kinds of changes. (a)–(c) show the Landsat-like images, and (d)–(f) show their corresponding MODIS-like images. From left to right, three pairs were acquired on t_1, t_2, and t_3, respectively.
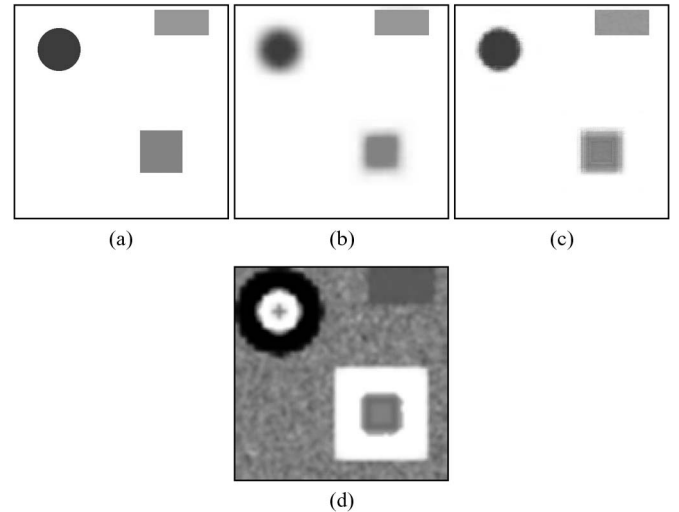

Fig. 5. Comparisons between the original image and predicted results for the simulated data. (a) Original Landsat-like image at t_2. (b) Predicted image at t_2 using STARFM. (c) Predicted image at t_2 using SPSTFM. (d) Distribution of weights for the reference image at t_1.

reflectances, whereas in the image predicted by STARFM, there appears serious blurring in the regions with shape changes [the circular and square objects in Fig. 5(b)]. Fig. 5(d) shows the distribution of the weights (white denotes one and black denotes zero) for the reference image at t_1, from which we can see that the weights are inversely proportional to the changes between the MODIS-like data at t_1 and t_2.

C. Application to Phenology Changes

In this experiment, we used the same Landsat-7 ETM+ (30-m spatial resolution) and MODIS (500-m spatial resolution) surface reflectances as in [4] but clipped the central region equivalent to 500 × 500 Landsat pixels out of the original data to remove the clouds close to the boundary. The bands used are 2, 3, and 4 of Landsat and 4, 1, and 2 of MODIS accordingly.
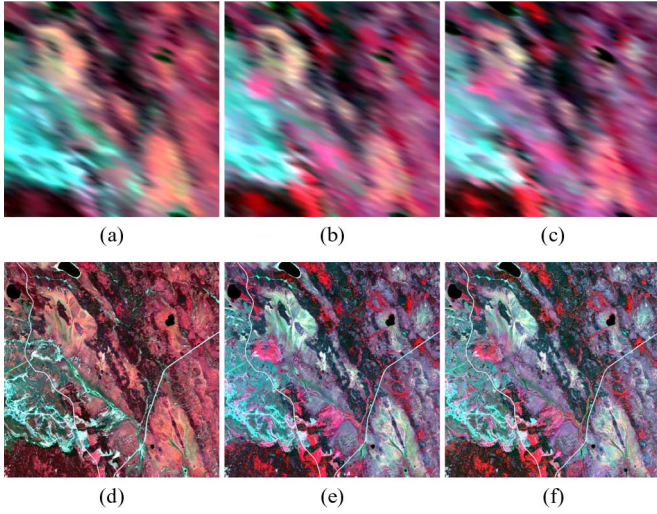
Fig. 6. (Upper row) MODIS composited surface reflectance and (lower row) Landsat composited surface reflectance (500 × 500 pixels) on May 24, July 11, and August 12 in 2001, respectively, from left to right.
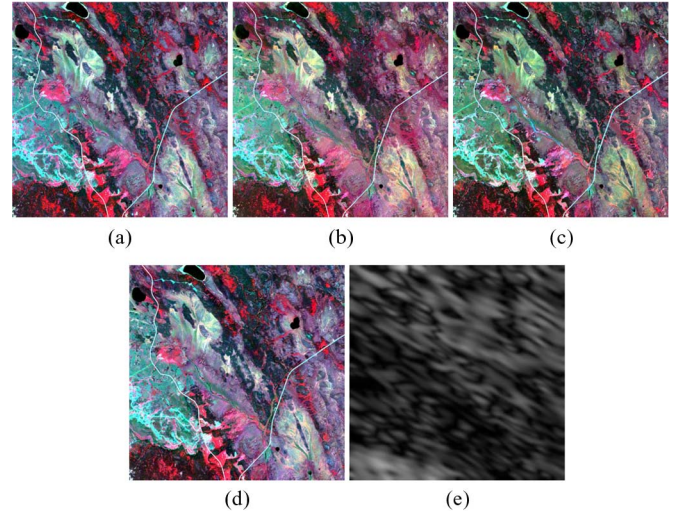

Fig. 7. Comparisons between actual and predicted surface reflectances with focus on seasonal changes. (a) Actual surface reflectance. (b) Predicted surface reflectance using STARFM. (c) Predicted surface reflectance using ESTARFM. (d) Predicted surface reflectance using SPSTFM. (e) Distribution of weights for the reference image on May 24, 2001.

This data set is characterized by seasonal changes and consists of three pairs of images acquired on May 24, July 11, and August 12 in 2001, respectively. Fig. 6 shows the scenes with a near-infrared (NIR)–red–green as red–green–blue composite for both MODIS (upper row) and Landsat (lower row) surface reflectances.

Using STARFM, ESTARFM, and SPSTFM, respectively, we reconstructed the Landsat-7 ETM+ surface reflectance on July 11, 2001, given the two surface reflectance pairs on May 24, 2001 [Fig. 6(a) and (d)], and August 12, 2001 [Fig. 6(c) and (f)], and its MODIS counterpart [Fig. 6(b)]. For fair comparisons, we adjusted the window size and used two given Landsat/MODIS pairs for STARFM and ESTARFM to achieve the best overall prediction effect. Fig. 7 shows the STARFM [Fig. 7(b)], ESTARFM [Fig. 7(c)], and SPSTFM [Fig. 7(d)] predictions in comparison with the actual Landsat ETM+ surface reflectance [Fig. 7(a)] and the distribution of the weights (white denotes one and black denotes zero) for the reference image on May 24, 2001 [Fig. 7(e)]. It can be easily found that the predicted surface reflectance of ESTARFM is better than that of STARFM and the outcome of SPSTFM is the best among the three methods, in terms of the overall spectral colors and structural details. As mentioned in Section I, this can be attributed to the STARFM's weighted sum of the surrounding pixels' reflectance to calculate the central pixel's reflectance, thereby causing blurs upon the predicted image. From Fig. 7(e), we can see that the weights are small for the regions with apparent changes between the prediction and reference images and vice versa. Since the image on July 11 is less similar to the one on May 24 in comparison with the one on August 12, the overall weights are low.

The comparison on three bands (red–green–NIR) between STARFM, ESTARFM, and SPSTFM in terms of AAD, rmse, VOE, ERGAS, and SSIM is listed in Table I. The average AAD values of the three bands for STARFM, ESTARFM, and SPSTFM are 0.007, 0.0071, and 0.006, respectively, and the average rmse values of the three bands for the three models are 0.0099, 0.0100, and 0.0076, respectively. These indicate

that SPSTFM can reconstruct the Landsat surface reflectance more precisely than STARFM and ESTARFM. The average VOE values of the three bands for STARFM, ESTARFM, and SPSTFM are 1.47, 1.48, and 0.60 ($\times 10^{-4}$), respectively, and the average SSIMs of the three bands for the three methods are 0.85, 0.86, and 0.90, respectively, which indicate that SPSTFM can retrieve more structural details on the surface reflectance than STARFM and ESTARFM with smaller reflectance deviations. Also, the ERGAS value reveals that the overall quality of the SPSTFM fusion result is better than that of STARFM and ESTARFM (0.8001 versus 1.0272 versus 0.9678). The scatter plots in Fig. 8 show the reflectance distribution between the predicted and actual surface reflectances of the three bands: red, green, and NIR. Fig. 8(a)–(c) shows the difference images between August 12 and May 24 for the red, green, and NIR bands, respectively. Fig. 8(d)–(f), (g)–(i), and (j)–(l) shows the scatter plots (scale factor is 10 000) of predicted reflectance using the three methods against the observed reflectance in the red, green, and NIR bands, respectively. From Fig. 8(a)–(c), we can see that there are relatively intense variations in the NIR band [Fig. 8(c)], and the comparison between the second, third, and fourth rows of scatter plots shows that SPSTFM achieves better effects than both STARFM and ESTARFM in all three bands, particularly in band NIR [Fig. 8(f), (i), and (l)]. This reveals that the proposed SPSTFM is more robust than STARFM and ESTARFM in dealing with complex changes.

D. Application to Type Changes

Detecting land-cover type change plays an important role in monitoring Earth's surface dynamics as a result of disturbances. Here, we used the Landsat-7 ETM+ and MODIS images of an area of the size of 15 km × 15 km in Shenzhen, China, to examine the performance of our algorithm in incorporating land-cover type changes in the prediction of a Landsat-like surface reflectance. This data set contains three Landsat/MODIS

TABLE I
QUANTITATIVE COMPARISON OF SPSTFM WITH OTHER METHODS IN IMAGES PRIMARILY WITH SEASONAL CHANGES

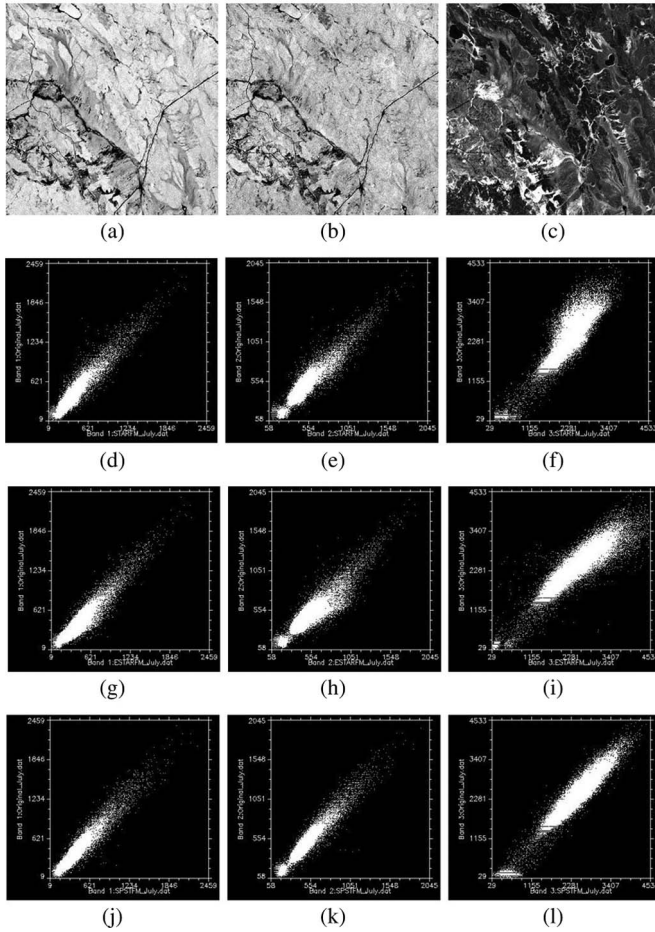| Method	AAD			RMSE			VOE (×10⁻⁴)			SSIM			
	Red	Green	NIR	Red	Green	NIR	Red	Green	NIR	Red	Green	NIR	ERGAS
STARFM	0.0032	0.0041	0.0137	0.0044	0.0055	0.0199	0.19	0.26	3.96	0.85	0.84	0.86	1.0272
ESTARFM	0.0037	0.0036	0.0140	0.0053	0.0049	0.0199	0.26	0.24	3.96	0.84	0.88	0.86	0.9678
SPSTFM	**0.0028**	**0.0034**	**0.0119**	**0.0037**	**0.0045**	**0.0148**	**0.20**	**0.13**	**1.49**	**0.88**	**0.93**	**0.88**	**0.8001**

Fig. 8. Difference images and scatter plots of predicted against reflectances for red–green–NIR bands from left to right. (a)–(c) Show the difference images between August 12 and May 24. (d)–(f) Show the scatter plots of the predicted reflectance using STARFM against the actual reflectance. (g)–(i) Show the scatter plots of the predicted reflectance using ESTARFM against the actual reflectance. (j)–(l) Show the scatter plots of the predicted reflectance using SPSTFM against the actual reflectance.
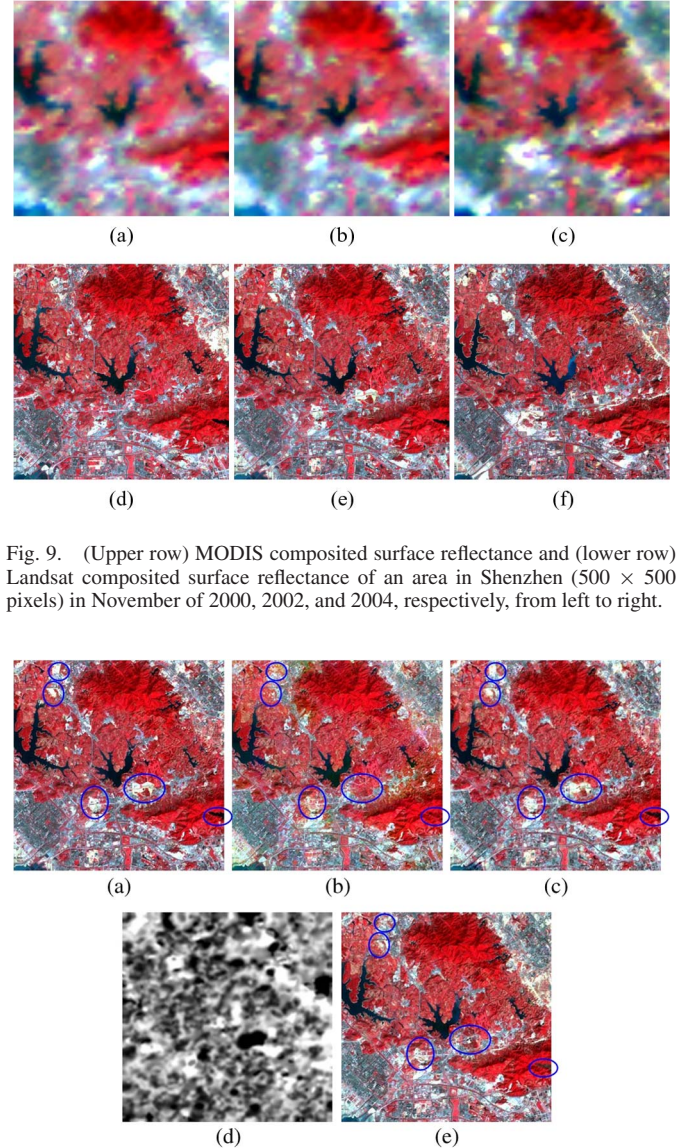

Fig. 9. (Upper row) MODIS composited surface reflectance and (lower row) Landsat composited surface reflectance of an area in Shenzhen (500 × 500 pixels) in November of 2000, 2002, and 2004, respectively, from left to right.

Fig. 10. Comparisons between actual surface reflectance and predicted surface reflectances with focus on type changes. (a) Actual surface reflectance in 2002. (b) Predicted surface reflectance in 2002 using STARFM. (c) Predicted surface reflectance in 2002 using SPSTFM. (d) Distribution of weights for the reference image at 2000. (e) Predicted surface reflectance in 2002 using only WFs.

pairs acquired in the same month but in different years, i.e., November 1, 2000, November 7, 2002, and November 8, 2004, respectively. We chose bands 2, 3, and 4 for a Landsat surface reflectance, all with the spatial resolution of 30 m and the image size of 500 × 500 pixels. Their corresponding MODIS bands are 4, 1, and 2, and the spatial resolution and image size are 250, 250, and 500 m and 60 × 60, 60 × 60, and 30 × 30 pixels, respectively. The MODIS bands with 250-m spatial resolution were aggregated to 500-m spatial resolution for convenience of processing. Fig. 9 shows the scenes with a NIR–red–green as red–green–blue composite for both MODIS (upper row) and Landsat (lower row) surface reflectances. It

can be observed that most of the vegetation regions remained similar from 2000 to 2002 to 2004; however, some vegetation regions were developed into built-up areas or vice versa from 2000 to 2002. Similar land-cover type changes took place from 2002 to 2004.

TABLE II
QUANTITATIVE COMPARISON OF SPSTFM WITH OTHER METHODS IN IMAGES PRIMARILY WITH LAND-COVER TYPE CHANGES

Method	AAD			RMSE			VOE ($\times 10^{-4}$)			SSIM			
	Red	Green	NIR	Red	Green	NIR	Red	Green	NIR	Red	Green	NIR	ERGAS
WF	0.019	0.016	0.012	0.027	0.025	0.017	6.49	5.45	2.18	0.82	0.75	0.82	1.4521
STARFM	0.017	0.013	0.009	0.024	0.022	0.014	6.11	5.09	2.16	0.78	0.75	0.83	1.4085
SPSTFM	**0.014**	**0.012**	**0.008**	**0.021**	**0.019**	**0.013**	**4.73**	**3.75**	**1.73**	**0.84**	**0.78**	**0.84**	**1.2097**

Similar to Section IV-C, we used the image pairs on November 1, 2000, and November 8, 2004, and the MODIS image on November 7, 2002, to predict the Landsat surface reflectance on November 7, 2002. To examine and compare the effects of weighting and sparse representation in our method, the fusion result through a weighting function (WF) only (i.e., the sum of the two weighting matrices multiplied by their corresponding high-resolution images in t_1 and t_3, respectively) is also given. Since ESTARFM assumes that the proportion of each endmember remains stable over the observation period, its performance is even worse than STARFM in this experiment. Thus, it is not chosen here to be compared with SPSTFM. The predictions using SPSTFM, STARFM, and WF are shown in Fig. 10, from which we can see that SPSTFM can detect almost all the change regions while WF and STARFM can hardly achieve that (see the blue circled regions in Fig. 10). However, there are slight deviations in the edges of the change areas in our predicted surface reflectance, which are mainly caused by the large resolution differences and slight geometric mismatch between Landsat ETM+ and MODIS images as well as the complex change structures in this data set. The comparisons in terms of AAD, rmse, VOE, ERGAS, and SSIM are listed in Table II. The average AAD values of the three bands for WF, STARFM, and SPSTFM are 0.016, 0.013, and 0.011, respectively, and the average rmse values of the three bands for the three models are 0.023, 0.020, and 0.017, respectively. These indicate that SPSTFM can reconstruct the Landsat surface reflectance more precisely than WF and STARFM. The average VOE values of the three bands using WF, STARFM, and SPSTFM are 4.70, 4.45, and 3.40 ($\times 10^{-4}$), respectively, and the average SSIMs of the three bands for the three methods are 0.79, 0.78, and 0.82, respectively, which show that SPSTFM retrieved more structural details on the surface reflectance than WF and STARFM with smaller reflectance deviations. Also, the ERGAS value indicates that the overall quality of the SPSTFM fusion result is better than that of the WF and STARFM fusion results. The scatter plots (scale factor is 10 000) of the predicted surface reflectances using SPSTFM and STARFM in the three bands, NIR–red–green, are shown in Fig. 11, respectively. Fig. 11(a)–(c) (upper row) shows scatter plots between the predicted reflectance using STARFM and the actual reflectance in NIR–red–green bands, respectively, while Fig. 11(d)–(f) (lower row) shows those between the predicted reflectance using SPSTFM and the actual reflectance in the corresponding three bands. The comparison between the upper and lower rows of the scatter plots in Fig. 11 shows that SPSTFM achieves a better fit to the 1 : 1 line in all three bands, indicating that SPSTFM captures more land-cover type changes than STARFM in predicting the Landsat surface reflectance in
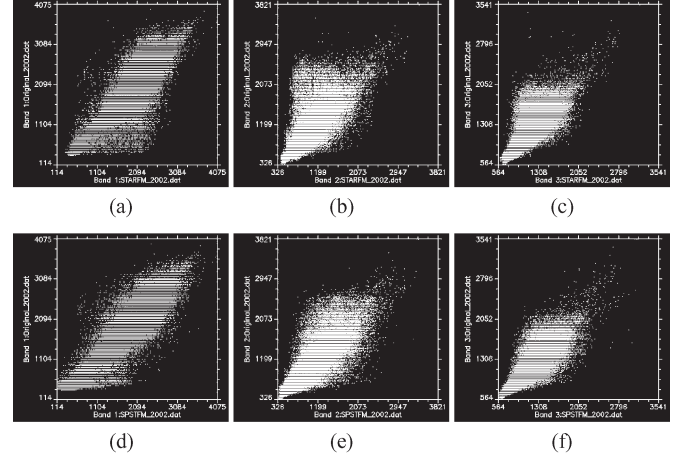

Fig. 11. Scatter plots of predicted reflectance against actual reflectance in 2002. (a)–(c) show scatter plots of the predicted reflectance using STARFM against the actual reflectance in NIR, red, and green bands, respectively. (d)–(f) show scatter plots of the predicted reflectance using SPSTFM against the actual reflectance in the same three bands.

2002. This should be attributed to the dictionary-pair learning and sparse coding mechanisms inside the SPSTFM algorithm, which result in a better characterization and delineation of type changes.

V. CONCLUSION

This paper has presented a sparse-representation-based learning method, i.e., SPSTFM, for spatiotemporal image fusion which provides a unified framework for detecting and characterizing phenology change and type change of land cover on the Earth surface. By establishing a correspondence between the difference images of high- and low-resolution images through the dictionary-pair learning of patches, the amplitude and structure of changes (e.g., shape and texture) can be captured more precisely with SPSTFM. In other words, the phenology and type changes can both be incorporated appropriately in the predicted surface reflectance. Also, using a patch rather than a pixel as the processing unit in SPSTFM guarantees the reconstruction of change structures to consider the neighborhood relationship between the central pixel and its neighboring pixels. Although STARFM-based algorithms also take into account the neighborhood relationship when predicting the central pixel's reflectance, they do not tackle it from the perspective of the change trend, although it might be discontinuous. Consequently, these facilities enable SPSTFM to outperform STARFM and similar algorithms, particularly in delineating and characterizing the changes.

While SPSTFM can deal with both phenology change (with less complex change structure) and type change (with more

complex change structure) of land cover, detection of type change and incorporating it properly in the "fused" image still remain a highly challenging task. Our experiments show that SPSTFM performs better on phenology change than type change. This can be interpreted in terms of the theory of sparse representation, i.e., more representation errors usually arise when there are more complex signals to be represented. In addition, compared with STARFM-based algorithms, our learning-based method takes longer computation time if we include the time for training the dictionary pair, although post-training computation in SPSTFM is as efficient as STARFM-based methods. To this end, our future work will continue along the line of improving the estimation accuracy of land-cover type change and reducing computational complexity.

ACKNOWLEDGMENT

REFERENCES

[1] J. Townshend, C. Justice, W. Li, C. Gurney, and J. McManus, "Global land cover classification by remote sensing: Present capabilities and future possibilities," *Remote Sens. Environ.*, vol. 35, no. 2/3, pp. 243–255, Feb./Mar. 1991.

[2] J. E. Vogelmann, S. M. Howard, L. Yang, C. R. Larson, B. K. Wylie, and J. N. Van Driel, "Completion of the 1990s national land cover data set for the conterminous United States from Landsat Thematic Mapper data and ancillary data sources," *Photogramm. Eng. Remote Sens.*, vol. 67, no. 6, pp. 650–662, Jun. 2001.

[3] J. C. Price, "How unique are spectral signatures?" *Remote Sens. Environ.*, vol. 49, no. 3, pp. 181–186, Sep. 1994.

[4] F. Gao, J. Masek, M. Schwaller, and F. Hall, "On the blending of the Landsat and MODIS surface reflectance: Predicting daily Landsat surface reflectance," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 8, pp. 2207–2218, Aug. 2006.

[5] X. Zhu, J. Chen, F. Gao, X. H. Chen, and J. G. Masek, "An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions," *Remote Sens. Environ.*, vol. 114, no. 11, pp. 2610–2623, Nov. 2010.

[6] T. Hilker, M. A. Wulder, N. C. Coops, J. Linke, J. McDermid, J. G. Masek, F. Gao, and J. C. White, "A new data fusion model for high spatial- and temporal-resolution mapping of forest based on Landsat and MODIS," *Remote Sens. Environ.*, vol. 113, no. 8, pp. 1613–1627, Aug. 2009.

[7] C. Pohl and J. L. van Genderen, "Multisensor image fusion in remote sensing: Concepts, methods and applications," *Int. J. Remote Sens.*, vol. 19, no. 5, pp. 823–854, 1998.

[8] Y. Zhang, "Understanding image fusion," *Photogramm. Eng. Remote Sens.*, vol. 70, no. 6, pp. 657–661, Jun. 2004.

[9] Z. J. Wang, D. Ziou, C. Armenakis, D. R. Li, and Q. Q. Li, "A comparative analysis of image fusion methods," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 6, pp. 1391–1402, Jun. 2005.

[10] S. Li and B. Yang, "A new pan-sharpening method using a compressed sensing technique," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 2, pp. 738–746, Feb. 2011.

[11] A. Garzelli, F. Nencini, and L. Capobianco, "Optimal MMSE pan sharpening of very high resolution multispectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 1, pp. 228–236, Jan. 2008.

[12] C. Thomas, T. Ranchin, L. Wald, and J. Chanussot, "Synthesis of multispectral images to high spatial resolution: A critical review of fusion methods based on remote sensing physics," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1301–1312, May 2008.

[13] S. P. Healey, W. B. Cohen, Z. Q. Yang, and O. N. Krankina, "Comparison of tasseled cap-based Landsat data structures for use in forest disturbance detection," *Remote Sens. Environ.*, vol. 97, no. 3, pp. 301–310, Aug. 2005.

[14] R. Rubinstein, A. M. Bruckstein, and M. Elad, "Dictionaries for sparse representation modeling," *Proc. IEEE*, vol. 98, no. 6, pp. 1045–1057, Jun. 2010.

[15] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, Nov. 2010.

[16] J. A. Tropp, "Just relax: Convex programming methods for subset selection and sparse approximation," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1030–1051, Mar. 2006.

[17] M. Elad, M. A. T. Figueiredo, and Y. Mak, "On the role of sparse and redundant representation in image processing," *Proc. IEEE*, vol. 98, no. 6, pp. 972–982, Jun. 2010.

[18] D. Donoho and Y. Tsaig, "Fast solution of L1-norm minimization problems when the solution may be sparse," Inst. Comput. Math. Eng., Stanford, CA, 2006. [Online]. Available: http://www.Stanford.edu/~tsaig

[19] M. Figueiredo, R. Nowak, and S. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 4, pp. 586–597, Dec. 2007.

[20] H. F. Shen, L. P. Zhang, B. Huang, and P. X. Li, "A MAP approach for joint motion estimation, segmentation and super resolution," *IEEE Trans. Image Process.*, vol. 16, no. 2, pp. 479–490, Feb. 2007.

[21] W. T. Freeman, E. C. Pasztor, and O. T. Carmichael, "Learning low-level vision," *Int. J. Comput. Vis.*, vol. 40, no. 1, pp. 25–47, Oct. 2000.

[22] 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.

[23] C. H. He, P. J. Sh, D. Y. Xie, and Y. Y. Zhao, "Improving the normalized difference built-up index to map urban built-up areas using a semiautomatic segmentation approach," *Remote Sens. Lett.*, vol. 1, no. 4, pp. 213–221, Dec. 2010.

[24] M. M. Khan, L. Alparone, and J. Chanussot, "Pansharpening quality assessment using the modulation transfer functions of instruments," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 11, pp. 3880–3891, Nov. 2009.

[25] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

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 in the Schulich School of Engineering, University of Calgary, Calgary, AB, Canada (2004–2006), and the Department of Civil Engineering, National University of Singapore, Singapore (2001–2004). He 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*. His research interests are broad, covering most aspects of geoinformation science, specifically spatiotemporal image fusion for environmental monitoring, spatial/spatiotemporal statistics for land-cover/land-use change modeling, and multiobjective spatial optimization for sustainable land-use planning. He is currently exploring along the line of precision remote sensing, convinced that this new paradigm will revolutionize the way how multisensor remotely sensed data are fused and exploited to improve the performance and quality of various applications in the future.
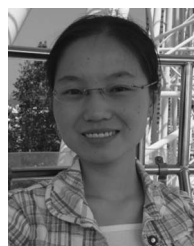
Huihui Song received the B.S. degree in technology and science of electronic information from the Ocean University of China, Qingdao, China, in 2008 and the M.S. degree in communication and information system from the University of Science and Technology of China, Hefei, China, in 2011. She is currently working toward the Ph.D. degree in the Department of Geography and Resource Management, The Chinese University of Hong Kong, Shatin, Hong Kong.

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