


Article

Spatio-Temporal Series Remote Sensing Image Prediction Based on Multi-Dictionary Bayesian Fusion

Chu He ^{1,2} , Zhi Zhang ¹, Dehui Xiong ¹, Juan Du ^{3,*} and Mingsheng Liao ^{2,4}

¹ Electronic and Information School, Wuhan University, Wuhan 430072, China; chuhe@whu.edu.cn (C.H.); zhizhang@whu.edu.cn (Z.Z.); dhui.xiong@gmail.com (D.X.)

² State Key Laboratory for Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China; liao@whu.edu.cn

³ Remote Sensing and Information Engineering School, Wuhan University, Wuhan 430079, China

⁴ Collaborative Innovation Center of Geospatial Technology, 129 Luoyu Road, Wuhan 430079, China

* Correspondence: dujuan_rs@whu.edu.cn; Tel.: +86-133-0716-2028

Received: 13 October 2017; Accepted: 15 November 2017; Published: 21 November 2017

Abstract: Contradictions in spatial resolution and temporal coverage emerge from earth observation remote sensing images due to limitations in technology and cost. Therefore, how to combine remote sensing images with low spatial yet high temporal resolution as well as those with high spatial yet low temporal resolution to construct images with both high spatial resolution and high temporal coverage has become an important problem called spatio-temporal fusion problem in both research and practice. A Multi-Dictionary Bayesian Spatio-Temporal Reflectance Fusion Model (MDBFM) has been proposed in this paper. First, multiple dictionaries from regions of different classes are trained. Second, a Bayesian framework is constructed to solve the dictionary selection problem. A pixel-dictionary likelihood function and a dictionary-dictionary prior function are constructed under the Bayesian framework. Third, remote sensing images before and after the middle moment are combined to predict images at the middle moment. Diverse shapes and textures information is learned from different landscapes in multi-dictionary learning to help dictionaries capture the distinctions between regions. The Bayesian framework makes full use of the priori information while the input image is classified. The experiments with one simulated dataset and two satellite datasets validate that the MDBFM is highly effective in both subjective and objective evaluation indexes. The results of MDBFM show more precise details and have a higher similarity with real images when dealing with both type changes and phenology changes.

Keywords: spatio-temporal fusion; multi-dictionary learning; sparse representation; MODIS; Landsat; Bayes; maximum a posterior

1. Introduction

MODerate Resolution Imaging Spectroradiometer (MODIS) with a short repeated observation cycle (one to eight days) [1] has a spatial resolution ranging from 250 m to 1000 m [2]. The low spatial resolution of MODIS confines its application in fragmented and extremely heterogeneous landscapes [3]. SPOT and surface satellite instruments (such as Landsat) with a spatial resolution ranging from 10 m to 30 m are exceptional sources of satellite information [4]. However, cloud impact and long revisit rate (16 days to a month) [5] limit the detection of changes caused by human activities or rapid surface changes caused by disturbances. The need to better observe and analyze the changes in surface features motivates the emergence of the spatio-temporal fusion technology. This technique combines images with a short repeated observation cycle and images with a rich spatial information

to construct images with both high spatial and high temporal resolution [6]. This technique has also been successfully used to predict high-resolution images in different applications, such as forest change detection [7], urban heat island [8], vegetation indices [9], surface temperature [10] and evapotranspiration [11]. During the observation period, two main changes, namely, phenology changes and type changes, are included in the remote sensing images. The change caused by the season and climate is called phenology changes (e.g., seasonal change of vegetation), which leads to the spectral reflectance change in the image. The type changes are not just in terms of overall spectral reflectance but also in terms of shape and size, such as the shrinking of the lake and the expanding of the urban area. The type changes are considered more challenging to capture than the phenology changes.

Proposed by Gao et al. in 2006, the Spatial and Temporal Adaptive Reflectance Fusion Model (STARFM) [12] is a classical spatiotemporal fusion model. Many improved models have been proposed based on this model. STARFM predicts the center pixel value through the similarity pixel value while considering the distance, the spectral similarity, and the time difference in order to greatly improve the accuracy of the fusion result. Given that STARFM are based on the assumption that the ratio of different land cover types remains constant over the period of observation, Hilker et al. proposed the Spatial Temporal Adaptive Algorithm for Mapping Reflectance Change (STAARCH) [13], which identifies spatial and temporal changes in the landscape with a high level of detail and produces synthetic Landsat images for each available date of MODIS imagery by using Tasseled Cap transformations of both Landsat TM/ETM and MODIS reflectance data. Compared with STARFM, the Enhanced Spatial and Temporal Adaptive Reflectance Fusion Model (ESTARFM) [14] proposed by Zhu et al. in 2010 is more suitable for non-homogeneous landscapes by using a conversion factor between surface satellite and MODIS data. In 2013, Huang et al. [8] combined the STARFM and bilateral filter to generate the land surface temperature with simultaneous spatial and temporal resolutions. In 2014, Weng, Fu, and Gao [10] improved and modified STARFM by considering the annual temperature cycle and urban thermal landscape heterogeneity. Michishita et al. [15] proposed a customized blending model in 2015, which was developed based on ESTARFM.

In 2010, Yang et al. proposed an image super-resolution method based on sparse representation [16] that describes the relationship between high- and low-resolution images with a dictionary pair by applying sparse representation theory. Huang et al. applied sparse representation to fuse remote sensing images, and proposed the Sparse Representation-Based Spatio-Temporal Reflectance Fusion Model (SPSTFM) in 2012 [17]. This model processes well both on phenology changes and type changes. A Super-Resolution Convolutional Neural Network (SRCNN) proposed by Dong et al. in 2014 directly learns an end-to-end mapping that is represented as a deep Convolutional Neural Network (CNN) between low and high resolution images [18]. Recently, Zhu et al. [19] proposed A Flexible Spatio-Temporal Data Fusion Model (FSDAF) based on spectral unmixing analysis and a thin plate spline interpolator.

However, the results of STARFM based model of the land cover type changes area are unsatisfactory because of the assumption that the ratio of land cover type remains unchanged during the observation period. The reconstruction result of SPSTFM is limited by the validity of the dictionary for this model, which only learns from one unified dictionary. Even if the dimension of the dictionary is set to large enough, various types of information still cannot be fully learned by training single dictionary pairs. Although the capability of FSDAF in predicting type changes has been validated, there still remain open topics in preserving better spectral fidelity and spatial details of the predictions.

In order to improve the performance of the model when dealing with type changes, more information, especially edge information of the change area, must be captured. This paper proposes the Multi-Dictionary Bayesian Spatio-Temporal Reflectance Fusion Model (MDBFM) to address this problem. The Bayesian framework fully utilizes the prior information and the multi-dictionary is more targeted than one dictionary.

The rest of the paper is organized as follows. The sparse representation in spatio-temporal fusion is explained in Section 2. Section 3 provides the details of the proposed model. Section 4 shows the validations on three datasets, including one simulated dataset and two actual datasets. The advantages and some future work are discussed in Section 5. The conclusion is provided in Section 6. In Section 3, the process of MDBFM is as follows. First, different classes of dictionary pairs are trained. Second, the Bayesian framework is applied to the reconstruction process by using the maximum a posterior probability estimation to classify pixels into dictionary classes. Third, the high-resolution image is reconstructed from the sparse representation.

2. Preliminaries

2.1. Sparse Representation in Spatio-Temporal Fusion

The aim of spatio-temporal fusion is to construct high spatial resolution images (e.g., Landsat) from low spatial resolution images (e.g., MODIS). Sparse representation is introduced to spatial-temporal fusion because of its ability to generate more competitive results by joint training high- and low-resolution dictionary pairs. Sparse representation assumes that signals can be represented linearly by several atoms in the over-complete dictionary matrix. Here, each column of the dictionary matrix stands for a basic atom. After applying sparse representation to spatio-temporal fusion, the signals can be represented by high- or low-resolution image patches. Accordingly, the dictionary matrix can be represented by high- or low-resolution dictionary.

$$L = D_l \cdot \alpha \quad (1)$$

$$H = D_h \cdot \alpha \quad (2)$$

In Equations (1) and (2), H represents the training sample of the high-resolution dictionary, L represents the training sample of low-resolution dictionary, D_l is the low-resolution dictionary, D_h is the high-resolution dictionary, and α is the sparse representation coefficient for both the high- and low-resolution dictionaries, that is to say, the sparse representation coefficient is common for different image resolutions at the same time. The dictionaries can be learned by optimizing Equation (3):

$$\{D^*, \alpha^*\} = \arg \min_{D, \alpha} \{ \|I - D \cdot \alpha\|_2^2 + \lambda \|\alpha\|_1 \} \quad (3)$$

where $D = [D_h; D_l]$, $I = [H; L]$, D^* is the learned dictionary, α^* is the learned sparse representation coefficient, and the dictionaries are jointly trained by the K-SVD [20] algorithm. The steps of the K-SVD algorithm are shown in Table 1. Some approximating algorithms can efficiently solve Equation (3), such as basis pursuit [21] and Focall Underdetermined System Solver (FOCUSS) [22]. In this paper, the sparse representation coefficient α is calculated by the Orthogonal Matching Pursuit (OMP [23]) optimization algorithm.

Table 1. K-SVD Algorithm.

Require: sample matrix Z , dictionary dimension S , iterations J

Initialization: initialize dictionary D by randomly selecting S columns in sample matrix Z

1. Apply OMP to compute sparse coefficients.
 2. Update each column s in dictionary, and stop until convergence:
 - (1) Multiply row s in the sparse matrix with column s in the dictionary to obtain α_T^s
 - (2) Compute the overall representation error matrix $E_s = Z - \alpha_T^s$
 - (3) Restrict E_s by the remaining columns corresponding to column s in dictionary, and then obtain E_s^R
 - (4) Apply SVD to E_s^R , $E_s^R = U\Delta V^T$
 - (5) Replace column k in the dictionary with the first column of U , and then update α_T^s by $V(:, 1) * \Delta(1, 1)$
 - (6) Increase iterations J
-

2.2. Diagram of SPSTFM

Huang et al. proposed the Sparse Representation-Based Spatio-Temporal Reflectance Fusion Model (SPSTFM) [17] in 2012, for predicting Landsat images through data blending with MODIS.

The block diagram of SPSTFM is shown in Figure 1. X_i and Y_i represent MODIS and Landsat image on date t_i , L_{ij} is the low-resolution difference image obtained by subtracting X_j from X_i , and H_{ij} is the high-resolution difference image obtained by subtracting Y_j from Y_i . The difference image is used as input in order to extract the change area and high frequency information.

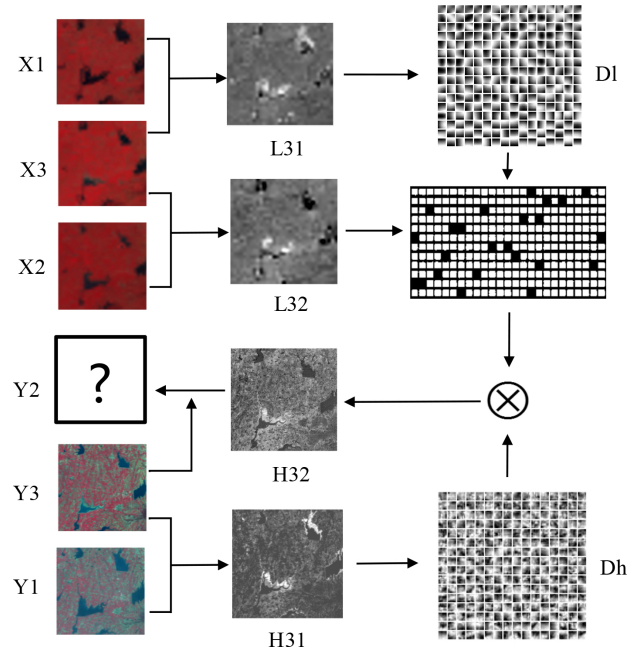


Figure 1. Block diagram of SPSTFM.

The difference images L_{31} and H_{31} are divided into patches, and each patch is stacked in columns to forming a training sample matrix. The high-resolution and corresponding low-resolution dictionary are trained from two training sample matrixes by using the KSVD method. Given that the low-resolution dictionary D_l has been learned, the OMP method is applied to calculate the sparse representation coefficient α of input difference image L_{32} . The high-resolution difference image corresponding to the input low-resolution image is denoted as $D_h \cdot \alpha$. H_{32} can be constructed in the same way when the input image is L_{21} . The final predicted Landsat image on t_2 is calculated by Equation (4), which is a synthesis of two high-resolution difference images corresponding to two input low-resolution difference images on $t_2 - t_1$ and $t_3 - t_2$.

$$Y_2 = W_1 \cdot (H_{21} + Y_1) + W_3 \cdot (Y_3 - H_{32}) \quad (4)$$

where H_{ij} is the high-resolution difference image on $t_i - t_j$, Y_i is the original high-resolution image on t_i , and W_1 and W_3 are weight coefficients. Many methods can be used to establish the weight coefficient. In this paper, the method in [17] is applied.

$$W_i = \begin{cases} 1 & \text{if } (V_3 - V_1) > \delta \text{ and } i = 1, \\ 0 & \text{if } (V_3 - V_1) > \delta \text{ and } i = 3, \\ \frac{1/V_i}{1/V_1 + 1/V_3} & \text{else.} \end{cases} \quad (5)$$

where V_i is calculated by the normalized difference built-up index [24], which measures the change degree between MODIS on t_i and MODIS on t_2 . V_i is calculated in each patch.

Although SPSTFM can deal with both phenology changes and type changes, the experiments show that this model performs better on phenology changes than type changes. In other words, the image with type changes has a more complex structure than that with phenology changes, thereby causing more representation errors in sparse representation [17]. We seek a new fusion model called Multi-Dictionary Bayesian Spatio-Temporal Reflectance Fusion Model (MDBFM), that can capture more effective information, especially the edge information of change area in images with type changes. The Bayesian framework makes full use of class information by constructing a pixel-dictionary likelihood function and a dictionary-dictionary prior function. In each class, a high- and low- resolution dictionary pair is trained. Multiple dictionary pairs that are trained within class are more specific and targeted than one dictionary pairs that are trained from the whole image, in which the detailed structure of every class is learned in the dictionary pair corresponding to this class. MDBFM is described in detail in Section 3.

3. Proposed Method

Similar to SPSTFM, our proposed algorithm focuses on the MODIS difference image to reconstruct the Landsat difference image. MDBFM consists of a training phase and a reconstruction phase. The first step in the training phase is pre-processing, which includes normalization and reshaping. The training samples are then classified, and class-dependent dictionaries are trained for each class. In the reconstruction phase, the test image is initially classified pixel-by-pixel given its MODIS difference value. Afterward, according to the class label of the central pixel, high-resolution patches are constructed by the dictionary pair in the same class. Recovered high-resolution patches are merged by averaging in the overlap area to create the resulting image. The resulting image from t_1, t_2 is combined with that from t_2, t_3 to obtain the Landsat-like image.

3.1. Preprocessing

Considering the reflectance difference between Landsat Y and MODIS X , and the reflectance difference among different bands, Y and X are normalized in each band. In order to simplify the processing, X is scaled-up to the size of Y via bicubic interpolation. We do not directly train dictionaries from the high- and low-resolution images because we focus the training on high-frequency details. The low frequency is removed by computing the difference images as follows:

$$L_{31} = X_3 - X_1 \quad (6)$$

$$H_{31} = Y_3 - Y_1 \quad (7)$$

where X_3, X_1 represents the MODIS image on t_3, t_1 , Y_3, Y_1 represents the MODIS image on t_3, t_1 , and L_{31}, H_{31} is the difference images for dictionary training. Instead of working directly with difference images, we rely on the patches sampled from L_{31} and H_{31} . For class k , the training patches of size $n \times n$ are extracted by the operator R_k from the difference images in location i . The low-resolution patch $p_l^{i,k}$ and high-resolution patch $p_h^{i,k}$ can be represented as follows:

$$p_l^{i,k} = R_k L_{31} \quad (8)$$

$$p_h^{i,k} = R_k H_{31} \quad (9)$$

The patches $p_l^{i,k}$ and $p_h^{i,k}$ are stacked into columns to form the training matrices Z_l^k and Z_h^k . Columns in Z_l^k and Z_h^k correspond with each other, and we will use them to train low- and high-resolution dictionaries.

3.2. Multi-Dictionary Training

The traditional super-resolution based on sparse representation involves single-dictionary pair learning, which is limited when dealing with complex images (the shape, size and reflectance of the object both changed during the observation time). In MDBFM, the dictionary pair is trained in every dictionary class. Therefore, the sparse representation in MDBFM focuses on multi-dictionary learning, which makes full use of the abundant texture, structure, and shape information in various classes.

Suppose that there are K dictionaries $D = [D_1, D_2, \dots, D_k]$ corresponding to the K variables of classes. Training samples for the k th class are extracted to train the k th dictionary. We manually label the samples during the training. Based on the sparse representation in Section 2.1, the k th dictionary pair is learned by optimizing the following equation:

$$\{D_l^{k*}, D_h^{k*}, \alpha^{k*}\} = \arg \min_{D_l^k, D_h^k, \alpha^k} \{\|Z_l^k - D_l^k \cdot \alpha^k\|_2^2 + \|Z_h^k - D_h^k \cdot \alpha^k\|_2^2 + \lambda \|\alpha^k\|_1\} \quad (10)$$

where D_l^k is the k th low-resolution dictionary, D_h^k is the k th high-resolution dictionary, α^k is the sparse representation coefficient corresponding to the k th dictionary pair, and Z_l^k and Z_h^k are the low- and high-resolution sample matrixes of class k . D_l^{k*} , D_h^{k*} and α^{k*} are the learned low resolution dictionary, high resolution dictionary and sparse representation coefficient. Suppose the dictionary dimension is S , the dictionary D is initialized by randomly selecting S columns in sample matrix Z . The alternate update mode is adopted to solve Equation (10). See Table 1 for the detailed process of dictionary training.

3.3. Classification of Test Image

In order to reconstruct the Landsat difference image with dictionaries of different classes, the input MODIS image must be classified given its difference value. Let $J(C, V)$ represent the world state of the label image C and the MODIS difference image V . To get the label image C given its value V , the following Bayes formula is used:

$$P(C/V) = \frac{P(V/C) \cdot P(C)}{P(V)} \quad (11)$$

$P(V/C)$ denotes the probability of the value to be V given the class label C , which is called the likelihood probability. $P(C)$ is the prior probability of the class label C . $P(V)$ represents the priori probability that the value is V . $P(C/V)$ is the posterior probability. Given that $P(V)$ remains the same, this parameter can be removed for the calculation of MAP probability. After the logarithm, MAP can be formulated as

$$J(C, V) = -\log(P(V/C)) - \log(P(C, \lambda)) \quad (12)$$

$$C_s = \arg \min_C J(C, V) \quad (13)$$

where $P(C, \lambda)$ is the prior probability on C , and λ is the parameter of modeling C . $J(C, V)$ is minimized to obtain the classification result C_s .

Figure 2 shows the block diagram of Bayes. Each pixel in the image corresponds to four values, and an image can be thought of as four layers, including likelihood probability, gray scale, dictionary classes, and priori probability. Assuming that the class of the surrounding pixel is known, our objective is to get the class label of the central pixel. Step (1) in Figure 2 obtains the likelihood probability of the center pixel. In Step (2), the prior probability of the central pixel is modeled according to the distribution of the surrounding pixel. In Step (3), the posterior probability and the likelihood probability are both considered to determine the label of the central pixel.

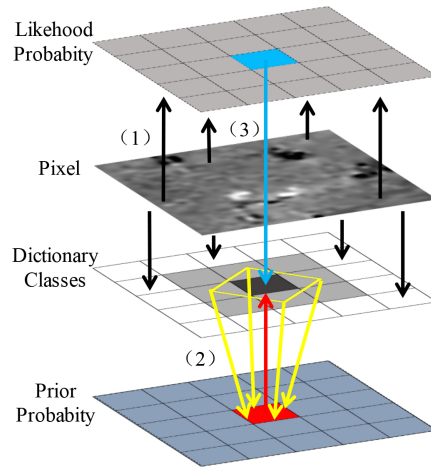


Figure 2. Diagram of Bayesian Framework.

Various likelihood and prior terms can be adopted. In this paper, we use Markov Random Fields (MRF) and Potts model [25] to describe the prior terms and use the Gaussian function to model the likelihood term. $P(C, \lambda)$ can be represented by the Potts model as follows:

$$P(C = c, \lambda) = \alpha \exp\{\lambda \sum V_{pq}(C)\} \tag{14}$$

$$V_{pq}(c_p, c_q) = \begin{cases} 1 & \text{if } c_p \neq c_q, \\ 0 & \text{otherwise.} \end{cases} \tag{15}$$

where p, q denotes the neighborhood pixels, and $V_{pq}(c_p, c_q)$ defines the relation of two label sites in the 2 rand clique set illustrated in Figure 3.

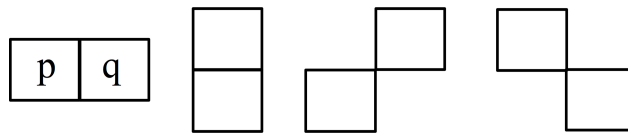


Figure 3. The 2 rand clique set.

The likelihood probability is modeled by the Gaussian function. For each class, consider that each pixel is conditionally independent and that every pixel follows the Gaussian distribution. Therefore, we can build the Gaussian probability-density function for each class. Figure 4 shows the process of obtaining the likelihood function. Graph Cuts [26] is used as a global optimization algorithm to minimize $J(C, V)$.

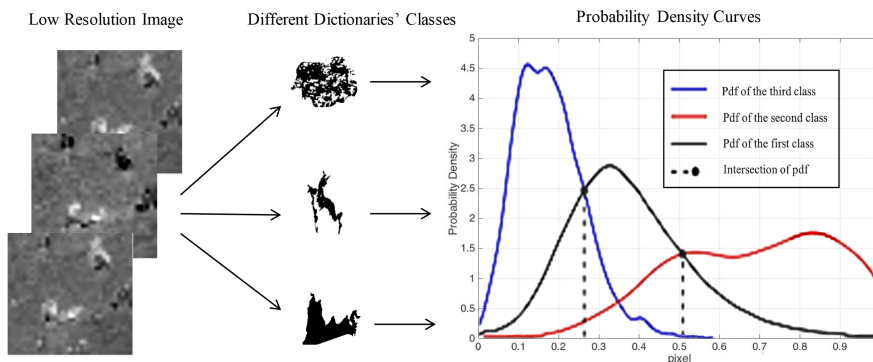


Figure 4. Process of the Likelihood Function.

3.4. Reconstruction

The difference MODIS image L_{21} is scaled up to the same size of the Landsat difference image and is classified pixel-by-pixel. For each patch, the label of the central pixel represents the class of the patch. For class k , operation R_i^k extracts patches from location i . These patches are reshaped to form the input matrix Z_i^k . The sparse representation of Z_i^k is expressed as follows:

$$\alpha^{k*} = \arg \min \|\alpha^k\|_1 \quad \text{s.t.} \quad \|Z_i^k - D_i^k \alpha^k\|_2^2 \leq \epsilon \quad (16)$$

where ϵ is a small constant, and α^{k*} is the learned sparse representation coefficient. The OMP algorithm is applied to obtain the sparse representation vectors α^k . The high-resolution matrix Z_h^k can be obtained by multiplying the high-resolution dictionary D_h^k and the sparse representation vectors α^k .

$$Z_h^k = D_h^k \alpha^k \quad (17)$$

where Z_h^k will be reshaped to high-resolution patches p_h^k . These patches construct the high-resolution difference image Y_{21} by using the following equation:

$$Y_{21} = \left[\sum_i (R_i^k)^T R_i^k \right]^{-1} \sum_i (R_i^k)^T p_h^k \quad (18)$$

This equation means that the patch p_h^k is placed in location i and that the average operation is adopted in the overlap regions. Similar to Y_{21} , Y_{32} can be reconstructed from X_{32} . We will fuse Y_{21} and Y_{32} to construct the final Landsat-like image L_2 by applying Equation (4).

3.5. Diagram of MDBFM

Figure 5 shows the block diagram of MDBFM, and the specific implementation steps are as follows:

1. The MODIS image is scaled up to the size of Landsat and the difference image L_{31}, H_{31} is obtained after preprocessing.
2. L_{31}, M_{31} is classified manually into classes (e.g., three classes, namely, lake, degradation area of lake, and forest in the type changes dataset). For each class, patches are extracted from the difference images and are stacked into training sample matrixes.
3. For every class, the high- and low-resolution sample matrixes are combined to train dictionary pairs. The training process of the dictionary is the same as that of SPSTFM by using the K-SVD method.
4. After preprocessing, X_2 and X_1 are differentiated to obtain the difference image L_{21} . Bayesian is then applied to classify the input difference image pixel-by-pixel.
5. L_{21} is divided into patches according to the classification result. The patches with the same label are stacked to form input matrixes of this class.
6. For each input matrix, the sparse representation coefficients are obtained with the corresponding low-resolution dictionary by using OMP.
7. The sparse representation coefficients of each class are multiplied by the high-resolution dictionaries to obtain the high-resolution matrix.
8. The high-resolution matrix is reshaped, and the high-resolution patches are obtained.
9. Each high-resolution patch is placed in its proper location and is averaged in the overlap regions. The Landsat-like difference image H_{21} is then obtained.
10. Using another difference image L_{32} as input, the other Landsat-like difference image H_{32} can be reconstructed in the same way.
11. Equation (4) is applied to fuse L_{32} and L_{21} as well as to obtain the final reconstruction Landsat-like image.

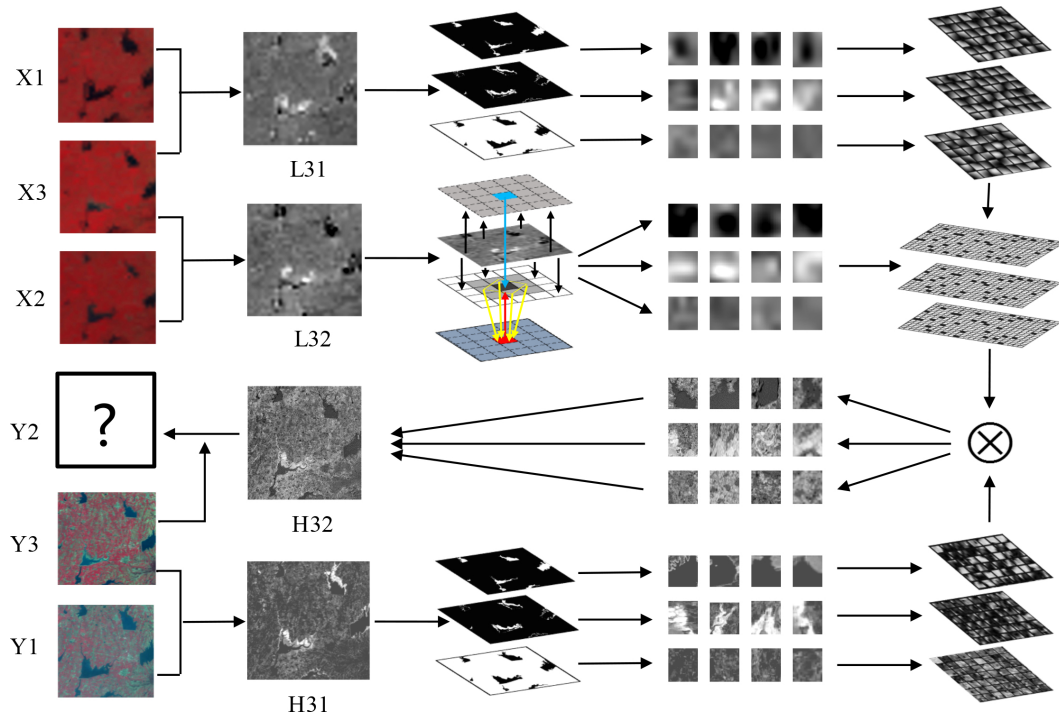


Figure 5. Block diagram of MDBFM.

4. Experiment

A simulated dataset and two remote sensing datasets are selected to examine the performance of MDBFM. The simulated dataset includes both changes in reflectivity and changes in land cover type. One of the actual remote sensing dataset which contains land cover type changes, is obtained in the same month in a different year. The other actual remote sensing dataset which contains phenology changes, is obtained in a different month in the same year. Four algorithms, namely, STARFM, SPSTFM, SRCNN, and FSDAF, are adopted as the comparison algorithms in this paper.

In order to evaluate the results of the four models, four measures are employed. Root mean square error (RMSE) is chosen as the first index since it is the most popular index in quantitative assessment of image qualities. To reflect the deviation between the predicted and the actual image, the average absolute difference (AAD) is adopted as the second index. Also, we utilized the correlation coefficient (R) as the third index to show the linear relationship between the predicted and the actual images. To measure the the overall structure similarity between the predicted and the actual images, structural similarity index measurement (SSIM) [26] is employed as the fourth index. The SSIM is obtained by computing the mean and the variance of the predicted-actual comparison images. For an ideally predicted image, SSIM and R should be one, and the much closer SSIM to one, the more similar the predicted image to the actual image.

4.1. Simulated Data

The simulated dataset is used as a simple case to compare the performance of the four methods when dealing with different changes. As shown in Figure 6, three types of land cover are included in the simulated image. Suppose that the circle represents the lake, the square represents the grass land, and the rectangle represents the urban area. The spatial resolution of the Landsat-like images is 30 m. The Landsat-like images are blurred, down-sampled, and added with white Gaussian noise to produce MODIS-like images. The radius of the lake is reduced from 2400 m to 1500 m and finally to 600 m, and the reflectance remains constant. The area of the urban area remains constant (4200 m × 1800 m), while the reflectance increases from 50 m to 120 m and finally to 170 m. The length of the grass land is increased from 1200 m to 1800 m and finally to 2400 m, at the same time, the reflectance is reduced

from 200 m to 140 m and finally to 170 m. Three types of changes are included in the simulated dataset: (1) only the change in reflectivity; (2) only the change in coverage area; (3) the changes in reflectivity and coverage area.

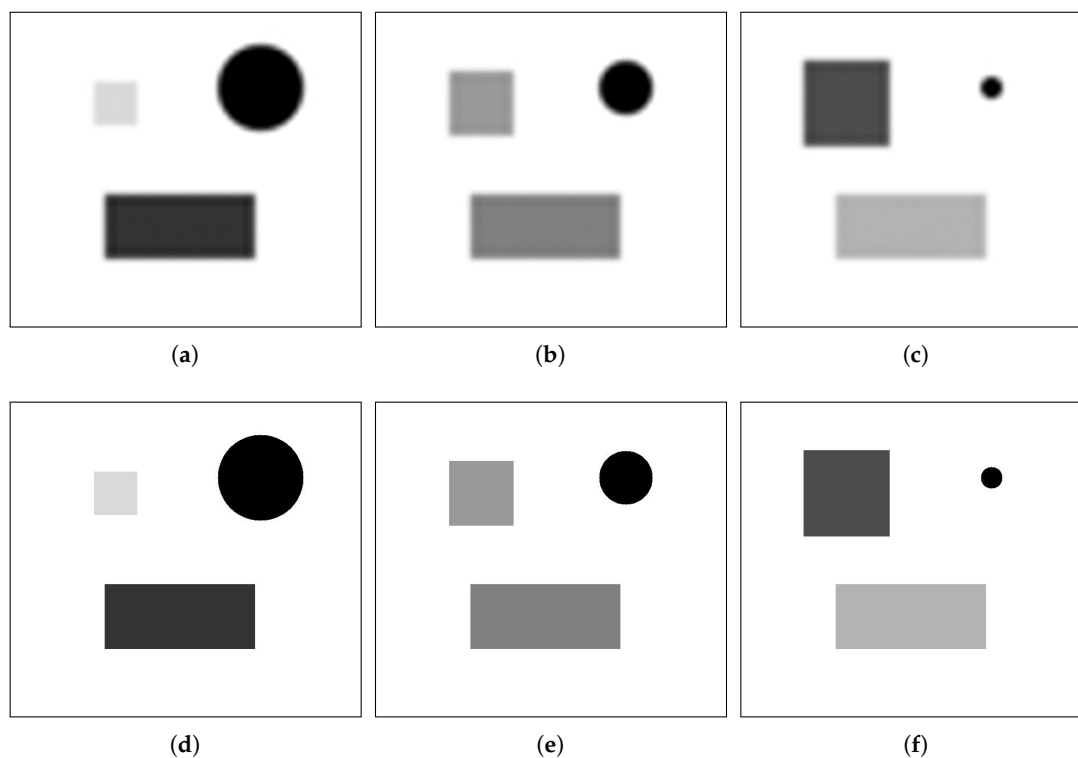


Figure 6. Simulated images: (a) MODIS-like image at t_1 ; (b) MODIS-like at t_2 ; (c) MODIS-like at t_3 ; (d) Landsat-like at t_1 ; (e) Landsat-like at t_2 ; and (f) Landsat-like at t_3 .

The parameters are adjusted to achieve the best reconstruction performances. The STARFM window size is 20×20 pixels, A is set to 100, the spectral similarity pixel threshold is 0.1, $\sigma_l = 0.02$, and $\sigma_m = 0.02$. In SPSTFM, the dictionary dimension is 256, the patch size is 20×20 , the step is set to 5, the allowed maximum error in KSVD is 0.001, and the maximum number of iterations is 250. In MDBFM, the patch size is 7×7 , 7×7 , and 25×25 , the dictionary dimension is 256, 256, and 512, and the other parameters are the same as those in SPSTFM.

The reconstructed results of four algorithms are shown in Figure 7. By comparing Figure 7d with Figure 7f, the results predicted by our method have a smaller error than those predicted by SRCNN when dealing with phenology changes. Compared with STARFM and SRCNN, the SPSTFM and FSDAF produce much better predictions, but slight blurring effects still exist at the edges of the circle because of the large spatial resolution gap between the high- and low-resolution images. The edges of the square and circle reconstructed using our method are clearer, which indicates that MDBFM is more effective than other methods when dealing with land cover type changes. The quantitative metrics in Table 2 also shows that the proposed method has lower prediction error than the other models.

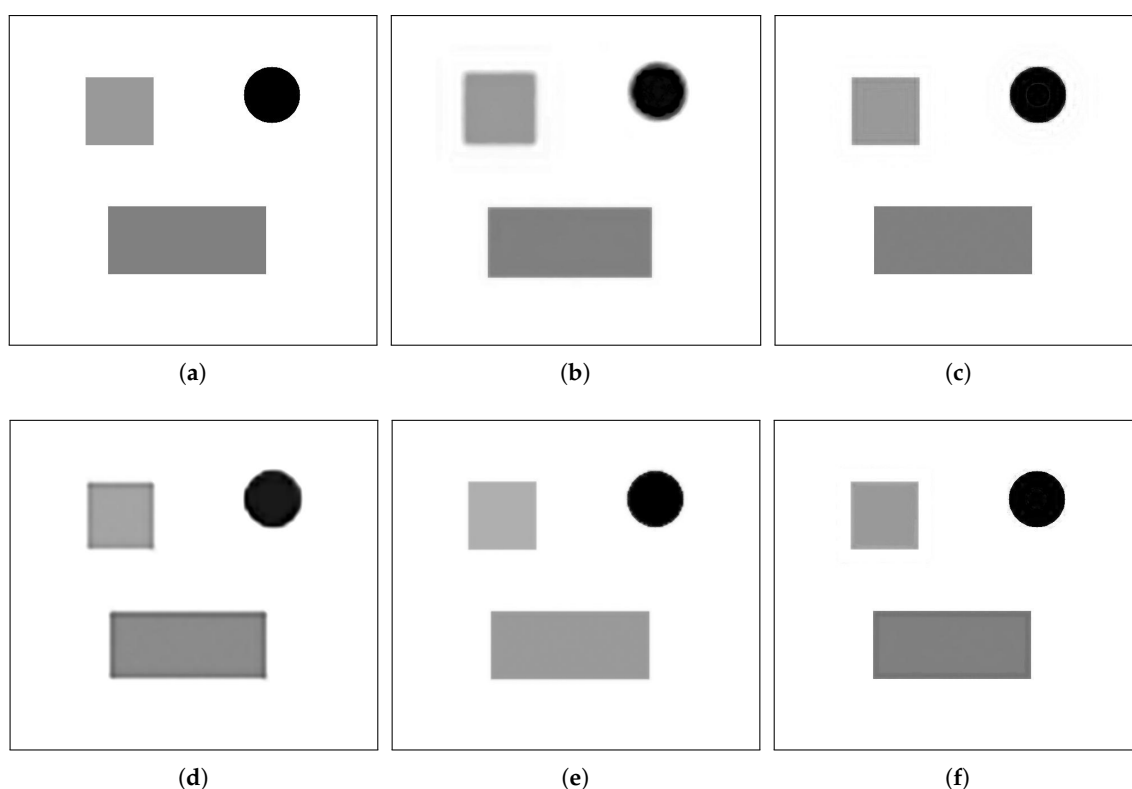


Figure 7. Predicted results of the simulated data: (a) the actual Landsat-like image; (b) predicted Landsat-like image using STARFM; (c) predicted Landsat-like image using SPSTFM; (d) predicted Landsat-like image using SRCNN; (e) predicted Landsat-like image using FSDAF; (f) predicted Landsat-like image using MDBFM.

Table 2. Index analysis of simulated dataset.

Index	STARFM	SPSTFM	SRCNN	FSDAF	MDBFM
RMSE	0.0068	0.0041	0.0052	0.0042	0.0035
AAD	0.0047	0.0013	0.0035	0.0016	0.0013
R	0.8234	0.9856	0.8827	0.9731	0.9894
SSIM	0.8349	0.8866	0.8195	0.8664	0.9151

4.2. Remote Sensing Images with Type Changes

Data Sources

In this experiment, we select a $15 \text{ km} \times 15 \text{ km}$ area at the junction of Hubei Province and Henan Province in China. The coordinates of the area are: ($32^{\circ}20'35'' \text{ N}$, $112^{\circ}49'26'' \text{ E}$), ($32^{\circ}20'35'' \text{ N}$, $112^{\circ}59'00'' \text{ E}$), ($32^{\circ}12'29'' \text{ N}$, $112^{\circ}49'26'' \text{ E}$), and ($32^{\circ}12'29'' \text{ N}$, $112^{\circ}59'00'' \text{ E}$). This area contains several lakes and forest land. The low-resolution image is the terrestrial standard product of MODIS image with a product ID of MOD09, resolution of 500 m, image size of 33×33 pixels, and repeated observation cycle of 8 days. We select 2 (near-infrared, NIR), 1 (red), and 4 (green) band as R, G, B channels. The high-resolution image is the Landsat7 ETM + image with a resolution of 30 m, image size of 500×500 pixels, and repeated observation cycle of 16 days. For Landsat, we select 4 (NIR), 3 (red), and 2 (green) band as R, G, B channels.

Given that the Scan Line Corrector (SLC) failed on 31 May 2003, the Landsat-7 ETM + products show some data gaps. Therefore, we only use the data prior to May 2003. When no MODIS image is

obtained on the same date of Landsat or if the MODIS image is covered by clouds, the most recent date is selected as the corresponding image of Landsat.

Figure 8a–c are the low-resolution images, Figure 8d–f are the known high-resolution images, and Figure 8e is the true image of the high-resolution image that will be reconstructed. Given that these images are obtained in the same month in different years, the overall value of the image slightly changes because the vegetation remains almost the same for all seasons and the climate is similar. As shown in Figure 8a–c, as the years pass the lakes shrink obviously. In the Landsat images, the edge shape and shrinking traces of the lakes are clear. In the MODIS images, the shape of the lake is not depicted in detail; therefore, we fuse the existing image to reconstruct high-resolution images at the middle moment.

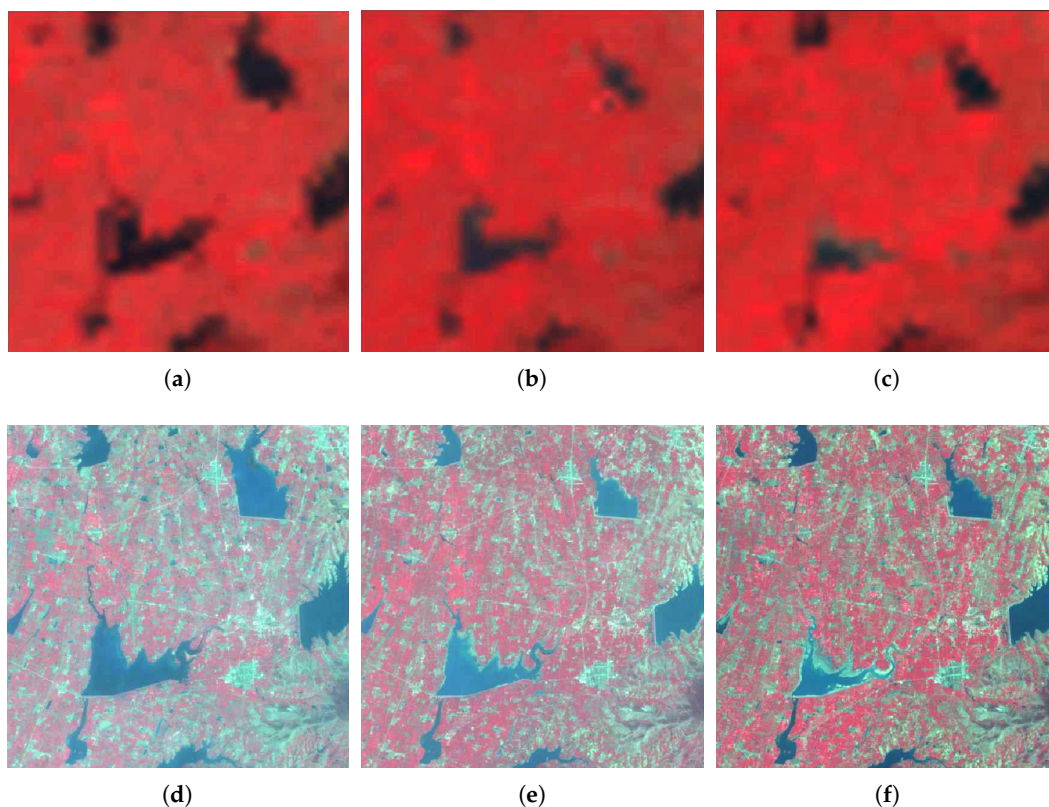


Figure 8. Remote sensing data with type changes: (a) MODIS image on 7 April 2001; (b) MODIS on 7 April 2002; (c) MODIS on 15 April 2003; (d) Landsat on 8 April 2001; (e) Landsat on 11 April 2002; (f) Landsat on 15 April 2003.

In the STARFM algorithm, the weight is $A = 750$, the sliding window size is 21×21 , and the threshold of the pixels with similar characteristics is set to 7. For the visible band (red and green) $\sigma_l = 0.002$, and $\sigma_m = 0.002$. For the NIR band, $\sigma_l = 0.005$, and $\sigma_m = 0.005$. In the SPSTFM algorithm, the MODIS image is bicubic interpolated to the same size as that of the Landsat image. In the K-SVD algorithm, the allowable error is 0.003, the dictionary dimension is 256, the number of iterations is 500, and the patch size is 7×7 . Each signal is represented up to 25 atoms. In the SRCNN algorithm, the sub-images are randomly cropped from the ground truth images to compose the training image. The training images are prepared as 30×30 with a stride of 15, and the resolution difference is 15. We set $f_1 = 8$, $n_1 = 64$, $f_2 = 1$, $n_2 = 32$, and $f_3 = 4$. In MDBFM, the image block size of the lake and the degradation area of lake are 5×5 , while that for forest land is 9×9 . The other parameters are the same as those in the SPSTFM model. We set the parameters in FSDAF following its paper. The original FSDAF is designed for one pair image learning, while our method learns from two pairs.

Therefore, we extended FSDAF to use two pairs of high- and low-resolution images. Our strategy is applied FSDAF to each pair of input to get two separate predictions and then use the weight in Equation (5) to combine the two predictions.

In Figure 9, the result of STARFM is similar to the Landsat images taken on 15 April 2003, which has shrank lakes. However, it is the Landsat images taken on 11 April 2002 need to be reconstructed, and the lakes have not completely shrunk at that time. The images predicted by SPSTFM are closer to the true image than those predicted by STARFM. However, the lake area still shrunk excessively and the shape of the lake is blurred. The SRCNN reconstruction results are more blurred and lack much details. Compared to STARFM, SPSTFM and SRCNN, FSDAF is better at preserving the shapes of small objects. The result of MDBFM is superior to that of the comparison models in terms of higher accuracy and more similarity to the true image. After the subjective evaluation, the results are objectively evaluated using several evaluation parameters.

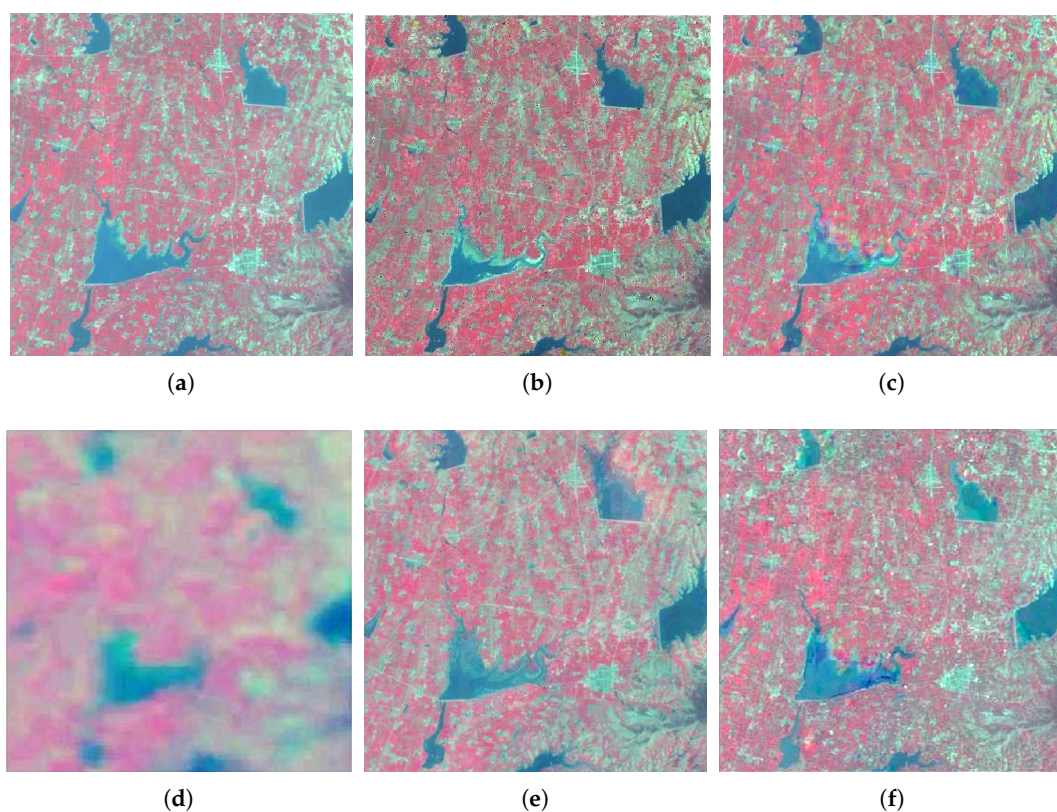


Figure 9. Reconstruction results of type changes images: (a) the actual Landsat image; (b) predicted result of STARFM; (c) predicted result of SPSTFM; (d) predicted result of SRCNN; (e) predicted result of STARF; (f) predicted result of MDBFM.

In Table 3, the mean RMSE and mean AAD of three bands for MDBFM is smaller than the comparison models, which implies that the deviation between the prediction image and the real image is smaller for MDBFM than for the other models. The larger R and SSIM for MDBFM indicates that, the high resolution image reconstructed by our method is more linearly correlated with the true image, has a higher structural similarity, and can recover more precise structural details. It reveals that our proposed method is more robust to tackle with the type change prediction.

Table 3. Indexes analysis of data with type changes.

Index	Channel	STARFM	SPSTFM	SRCNN	FSDAF	MDBFM
RMSE	Red	0.0072	0.0033	0.0151	0.0041	0.0032
	Green	0.0102	0.0055	0.0147	0.0031	0.0046
	NIR	0.0103	0.0068	0.0152	0.0074	0.0047
AAD	Red	0.0041	0.0026	0.0149	0.0028	0.0027
	Green	0.0066	0.0042	0.0142	0.0021	0.0033
	NIR	0.0074	0.0052	0.0146	0.0054	0.0034
R	Red	0.7834	0.8946	0.7346	0.8444	0.9169
	Green	0.7225	0.8998	0.7076	0.8215	0.9321
	NIR	0.7555	0.9421	0.6888	0.9090	0.9479
SSIM	Red	0.7329	0.7803	0.7162	0.7521	0.8506
	Green	0.7058	0.7780	0.6983	0.7363	0.8626
	NIR	0.7175	0.7562	0.6528	0.7235	0.7954

4.3. Remote Sensing Images with Phenology Changes

This dataset contains three pairs of Landsat-MODIS image, which are taken in different months in the same year. Given the short interval, the land cover type does not change significantly, and the main difference lies in the overall reflectance is caused by the season and the climate. We select an area in Shenzhen, China. Figure 10 shows the dataset.

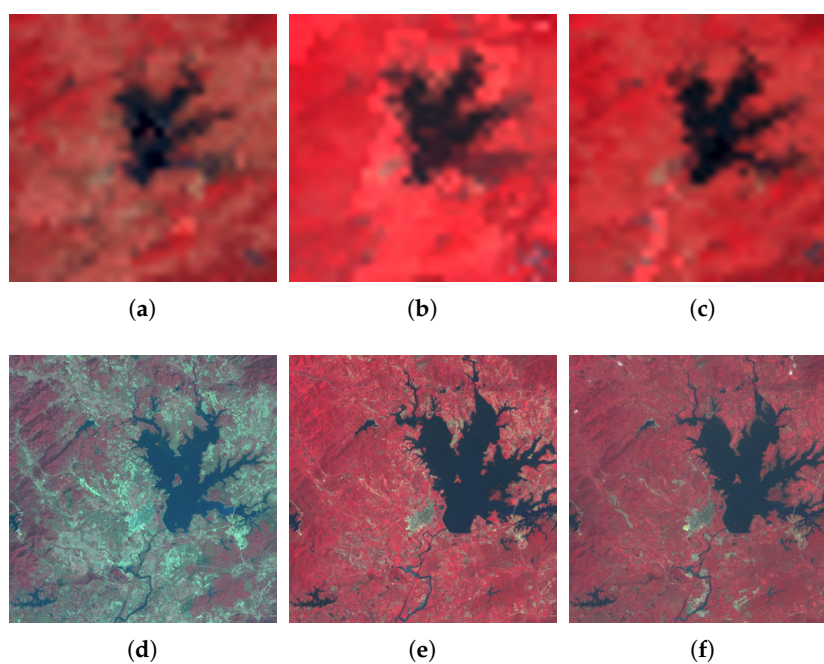


Figure 10. Remote sensing images with phenology changes: (a) MODIS on 11 February 2002; (b) MODIS on 29 August 2002; (c) MODIS on 15 October 2002; (d) Landsat on 17 February 2002; (e) Landsat on 29 August 2002; and (f) Landsat on 15 October 2002.

Similar to the experiment described above, the parameters are adjusted to achieve the best reconstruction results. The results of the quantitative evaluation are shown in Table 4. FSDAF has lower accuracy than SPSTFM in terms of quantitative assessment but it performed much better than STARFM. SPSTFM provided a more accurate prediction with the smaller RMSE, higher R and higher SSIM than STARFM and FSDAF. Among the five methods, MDBFM has the smallest errors and highest similarity to the true image. In Figure 11, the predicted result of SRCNN is worse than that of STARFM,

which can be attributed to the large resolution differences and the lack of sufficient training samples. However, the overall predicted image of STARFM is closer to Landsat on t_3 than to the actual Landsat on t_2 , because the small weight of the area with larger changes limits the performance of STARFM. A visual comparison between the actual observation and predictions shows that all predictions capture the phenology changes from the MODIS data, but among all five algorithms, the results of MDBFM are the most similar to the actual image. Comparing zoom-in area of the predictions, it can be found that, the predicted image of our method is more similar to the original image than the image predicted by FSDAF in regards to spatial details.

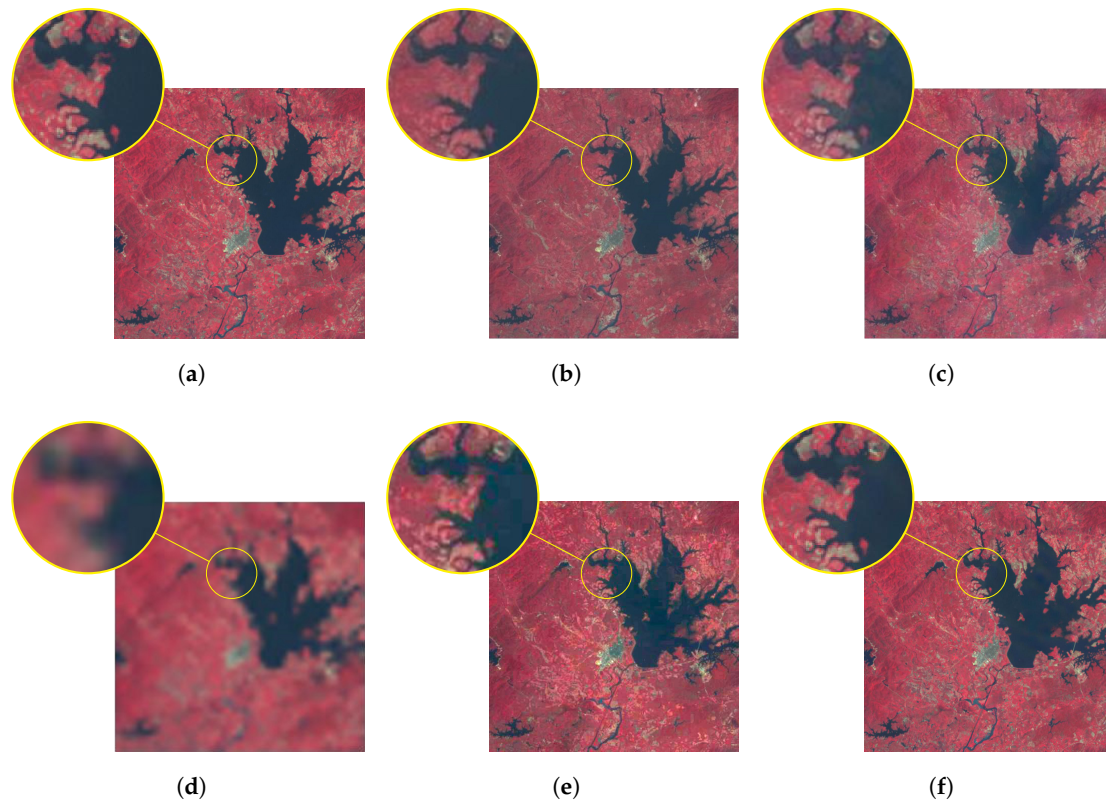


Figure 11. Predicted results of phenology changes images: (a) the actual Landsat image; (b) predicted result of STARFM; (c) predicted result of SPSTFM; (d) predicted result of SRCNN; (e) predicted result of FSDAF; (f) predicted result of MDBFM.

Table 4. Index analysis of data with phenology changes.

Index	Channel	STARFM	SPSTFM	SRCNN	FSDAF	MDBFM
RMSE	Red	0.0062	0.0045	0.0220	0.0046	0.0039
	Green	0.0059	0.0048	0.0196	0.0043	0.0042
	NIR	0.0121	0.0051	0.0281	0.0074	0.0046
AAD	Red	0.0034	0.0022	0.0188	0.0022	0.0015
	Green	0.0032	0.0041	0.0124	0.0020	0.0017
	NIR	0.0043	0.0040	0.0236	0.0041	0.0027
R	Red	0.8842	0.9147	0.8254	0.9103	0.9297
	Green	0.9128	0.8976	0.8328	0.8983	0.8932
	NIR	0.8614	0.8955	0.8132	0.8808	0.9691
SSIM	Red	0.7978	0.8667	0.7410	0.8574	0.8947
	Green	0.8314	0.8850	0.7409	0.8517	0.8798
	NIR	0.7570	0.8806	0.7417	0.8406	0.8904

5. Discussion

To achieve a better spatio-temporal fusion result, additional prior information is incorporated into our fusion model by constructing a multi-dictionary Bayesian framework, where the pixel-dictionary likelihood function and the dictionary-dictionary prior function help dictionaries capture the distinctions between regions. Experiments are conducted on datasets with type changes and phenology changes.

The experiments reveal that the MDBFM shows the best performance on both datasets with type changes and phenology changes compared with the other methods. The prediction results for phenology changes are better than those for type changes because land cover type changes lead to a more complex change structure, while in phenology changes, the shape and type of land cover almost remain invariant, thereby leading to a less complex change in structure.

STARFM shows unsatisfactory results for spatio-temporal fusion with type changes because its accuracy depends on whether a pure MODIS cell can be found in the search window. SPSTFM improves the results through learning sufficient information by training dictionaries. The accuracy of SRCNN can be further increased if the resolution difference is decreased or if sufficient training data are available. MDBFM shows slight deviations in the edges of the change areas as multi-dictionary learning and the Bayesian framework are introduced.

To further improve the accuracy of MDBFM, some aspects should be considered.

When constructing the dictionary-dictionary prior function, we take the class label of the pixel near the center pixel into account. The impact of the point located far away from the central point should be taken into consideration as additional constraint knowledge.

Having a proper number of atoms in the dictionary can lead to a trade-off between accuracy and computational cost. In theory, single dictionary training can achieve the same effect with the multi-dictionary architecture as long as the dictionary atomic number is large enough. However, an increase in the number of atoms in the dictionary leads to a higher computational cost. Having a less number of atoms in the dictionary is required in the multi-dictionary architecture without reducing the accuracy, and all kinds of dictionaries can be trained in parallel.

The division of the land cover type can help better reconstruct the edges of the changing regions. Therefore, when selecting the number and type of classes, the region with shape changes needs further analysis. In this paper, the rough division is taken by binarization. To further improve the accuracy, more efficient and accurate classification methods can be used.

6. Conclusions

In order to take full use of considerable and accessible remote sensing images to predict time series images with both high spatial and temporal resolution. A multi-dictionaries bayesian spatio-temporal reflectance fusion model has been proposed. More edge information is captured in the Bayesian framework, and the multi-dictionary training of different categories is more targeted than one dictionary pair training. This model works well on both type changes and phenology changes. The experimental results highlight the superior performance of MDBFM in describing the edge of the changing areas, in comparison with STARFM, SPSTFM, and SRCNN.

Acknowledgments: This work was partially supported by the National Natural Science Foundation of China (No. 61331016, No. 41371342), and by the National Key Research and Development Program of China (No. 2016YFC0803003-01).

Author Contributions: Chu He and Zhi Zhang conceived and designed the experiments; Zhi Zhang performed the experiments and analyzed the results; Chu He, Zhi Zhang and Juan Du wrote the paper; Dehui Xiong and Mingsheng Liao revised the paper.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

MODIS	Moderate resolution Imaging Spectroradiometer
CNN	Convolutional Neural Network
OMP	Orthogonal Matching Pursuit
MAP	Maximum A Posterior
SLC	Scan Line Corrector

References

- Arvor, D.; Jonathan, M.; Meirelles, M.S.P.; Dubreuil, V.; Durieux, L. Classification of MODIS EVI time series for crop mapping in the state of Mato Grosso, Brazil. *Int. J. Remote Sens.* **2011**, *32*, 7847–7871.
- Notarnicola, C.; Duguay, M.; Moelg, N.; Schellenberger, T.; Tetzlaff, A.; Monsorno, R.; Costa, A.; Steurer, C.; Zebisch, M. Snow cover maps from MODIS images at 250 m resolution, Part 1: Algorithm description. *Remote Sens.* **2013**, *5*, 110–126.
- Shabanov, N.; Wang, Y.; Buermann, W.; Dong, J.; Hoffman, S.; Smith, G.; Tian, Y.; Knyazikhin, Y.; Myneni, R. Effect of foliage spatial heterogeneity in the MODIS LAI and FPAR algorithm over broadleaf forests. *Remote Sens. Environ.* **2003**, *85*, 410–423.
- Swathika, R.; Sharmila, T.S. Multi-resolution spatial incorporation for MODIS and LANDSAT image fusion using CSSTARFM. In Proceedings of the 2016 IEEE Region 10 Conference (TENCON), Marina Bay Sands, Singapore, 22–25 November 2016; pp. 691–696.
- González-Sanpedro, M.; Le Toan, T.; Moreno, J.; Kergoat, L.; Rubio, E. Seasonal variations of leaf area index of agricultural fields retrieved from Landsat data. *Remote Sens. Environ.* **2008**, *112*, 810–824.
- Zhang, Y. Understanding image fusion. *Photogramm. Eng. Remote Sens.* **2004**, *70*, 657–661.
- Hilker, T.; Wulder, M.A.; Coops, N.C.; Seitz, N.; White, J.C.; Gao, F.; Masek, J.G.; Stenhouse, G. Generation of dense time series synthetic Landsat data through data blending with MODIS using a spatial and temporal adaptive reflectance fusion model. *Remote Sens. Environ.* **2009**, *113*, 1988–1999.
- Huang, B.; Wang, J.; Song, H.; Fu, D.; Wong, K. Generating high spatiotemporal resolution land surface temperature for urban heat island monitoring. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 1011–1015.
- Walker, J.; De Beurs, K.; Wynne, R.; Gao, F. Evaluation of Landsat and MODIS data fusion products for analysis of dryland forest phenology. *Remote Sens. Environ.* **2012**, *117*, 381–393.
- Weng, Q.; Fu, P.; Gao, F. Generating daily land surface temperature at Landsat resolution by fusing Landsat and MODIS data. *Remote Sens. Environ.* **2014**, *145*, 55–67.
- Cammalleri, C.; Anderson, M.; Gao, F.; Hain, C.; Kustas, W. A data fusion approach for mapping daily evapotranspiration at field scale. *Water Resour. Res.* **2013**, *49*, 4672–4686.
- Gao, F.; Masek, J.; Schwaller, M.; Hall, F. On the blending of the Landsat and MODIS surface reflectance: Predicting daily Landsat surface reflectance. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 2207–2218.
- Hilker, T.; Wulder, M.A.; Coops, N.C.; Linke, J.; McDermid, G.; Masek, J.G.; Gao, F.; White, J.C. A new data fusion model for high spatial-and temporal-resolution mapping of forest disturbance based on Landsat and MODIS. *Remote Sens. Environ.* **2009**, *113*, 1613–1627.
- Zhu, X.; Chen, J.; Gao, F.; Chen, X.; Masek, J.G. An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions. *Remote Sens. Environ.* **2010**, *114*, 2610–2623.
- Michishita, R.; Chen, L.; Chen, J.; Zhu, X.; Xu, B. Spatiotemporal reflectance blending in a wetland environment. *Int. J. Digit. Earth* **2015**, *8*, 364–382.
- Yang, J.; Wright, J.; Huang, T.S.; Ma, Y. Image super-resolution via sparse representation. *IEEE Trans. Image Process.* **2010**, *19*, 2861–2873.
- Huang, B.; Song, H. Spatiotemporal reflectance fusion via sparse representation. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 3707–3716.
- Dong, C.; Loy, C.C.; He, K.; Tang, X. Learning a deep convolutional network for image super-resolution. In *European Conference on Computer Vision*; Springer: Cham, Switzerland, 2014; pp. 184–199.
- Zhu, X.; Helmer, E.H.; Gao, F.; Liu, D.; Chen, J.; Lefsky, M.A. A flexible spatiotemporal method for fusing satellite images with different resolutions. *Remote Sens. Environ.* **2016**, *172*, 165–177.

20. Aharon, M.; Elad, M.; Bruckstein, A. *rmk-SVD: An algorithm for designing overcomplete dictionaries for sparse representation*. *IEEE Trans. Signal Process.* **2006**, *54*, 4311–4322.
21. Chen, S.S.; Donoho, D.L.; Saunders, M.A. Atomic decomposition by basis pursuit. *SIAM Rev.* **2001**, *43*, 129–159.
22. Gorodnitsky, I.F.; Rao, B.D. Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm. *IEEE Trans. Signal Process.* **1997**, *45*, 600–616.
23. Pati, Y.C.; Rezaifar, R.; Krishnaprasad, P. Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. In Proceedings of the 1993 Conference Record of the Twenty-Seventh Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA, 1–3 November 1993; pp. 40–44.
24. He, C.; Shi, P.; Xie, D.; Zhao, Y. Improving the normalized difference built-up index to map urban built-up areas using a semiautomatic segmentation approach. *Remote Sens. Lett.* **2010**, *1*, 213–221.
25. Wu, F.Y. The potts model. *Rev. Mod. Phys.* **1982**, *54*, 235.
26. Boykov, Y.; Kolmogorov, V. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Trans. Pattern Anal. Mach. Intell.* **2004**, *26*, 1124–1137.



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).