

Article

An Enhanced Single-Pair Learning-Based Reflectance Fusion Algorithm with Spatiotemporally Extended Training Samples

Dacheng Li¹, Yanrong Li^{1,*}, Wenfu Yang^{2,3}, Yanqin Ge¹, Qijin Han⁴, Lingling Ma⁵, Yonghong Chen¹ and Xuan Li¹

- ¹ College of Ming Engineering, Taiyuan University of Technology, Taiyuan 030024, China; lidacheng@tyut.edu.cn (D.L.); yanqin_ge@163.com (Y.G.); shandongchen1992@163.com (Y.C.); pdstr011@163.com (X.L.)
- ² School of Land Science and Technology, China University of Geosciences, Beijing 100083, China; wcyywf@163.com
- ³ Shanxi Coal Geology Geophysical Surveying Exploration Institute, Jinzhong 030600, China
- ⁴ China Centre for Resources Satellite Data and Application, Beijing 100094, China; 15810492859@163.com
- ⁵ Academy of Opto-Electronics, Chinese Academy of Science, Beijing 100094, China; Ilma@aoe.ac.cn
- * Correspondence: li.dennis@hotmail.com

Received: 28 May 2018; Accepted: 19 July 2018; Published: 1 August 2018



Abstract: Spatiotemporal fusion methods are considered a useful tool for generating multi-temporal reflectance data with limited high-resolution images and necessary low-resolution images. In particular, the superiority of sparse representation-based spatiotemporal reflectance fusion model (SPSTFM) in capturing phenology and type changes of land covers has been preliminarily demonstrated. Meanwhile, the dictionary training process, which is a key step in the sparse learning-based fusion algorithm, and its effect on fusion quality are still unclear. In this paper, an enhanced spatiotemporal fusion scheme based on the single-pair SPSTFM algorithm has been proposed through improving the process of dictionary learning, and then evaluated using two actual datasets, with one representing a rural area with phenology changes and the other representing an urban area with land cover type changes. The validated strategy for enhancing the dictionary learning process is divided into two modes to enlarge the training datasets with spatially and temporally extended samples. Compared to the original learning-based algorithm and other employed typical single-pair-based fusion models, experimental results from the proposed fusion method with two extension modes show improved performance in modeling reflectance using the two preceding datasets. Furthermore, the strategy with temporally extended training samples is more effective than the strategy with spatially extended training samples for the land cover area with phenology changes, whereas it is opposite for the land cover area with type changes.

Keywords: sparse learning; single image-pair; reflectance fusion; dictionary training; spatiotemporal extension

1. Introduction

Given the growing application requirements for a variety of refined and high-frequency monographic studies, such as land use and cover change [1], ecological environment monitoring [2], forest and pasture [3], oceanographic survey [4], and disaster monitoring [5], possible solutions for frequent acquisition of high-spatial-resolution remotely sensed data have been widely proposed. One significant attempt among current works presented a radical solution that involves the progressively increasing launch of various high-quality remote sensors, some of which adopt high



spatial (e.g., WorldView-3/4 and Gaojing-1/2 with 0.31 and 0.5 m resolution, respectively), temporal (e.g., Moderate Resolution Imaging Spectroradiometer (MODIS) and other meteorological satellites), and spectral resolutions (e.g., EO-1 Hyperion and Gaofen-3, both with 30 m resolution) or even high quantities (e.g., the Gaojing project, which will send 16 similar optical satellites into space no later than 2020). These remote sensors cannot technologically possess overall attributes because of the inherent conflicts among the spatial, temporal, and spectral characteristics of imaging systems. Given the constrained orbits of carrying platforms, severe climate conditions, and economic costs of mass spatiotemporal data, this challenging problem would not be resolved in the foreseeable future despite the increasing number of satellites.

Other attempts that rely on spatiotemporal reconstruction techniques of remotely sensed data are rapidly being implemented to create a new image with high-quality spatial, temporal, and spectral resolutions. Coupling spatial and temporal factors is particularly urgent. In a broad sense, analogous techniques, such as image restoration [6,7], superresolution [8], and gap filling [9,10], can be simply considered different patterns of image reconstruction. Although acceptable results from these handling methods can be expected under proper application conditions, restrictions in model universality, reconstruction precision, and physical principles of remote sensing significantly affect the depth and scope of their applications. The image fusion strategy, especially in the spatial and temporal dimension, provides another effective way to synthesize an optimized image by combining spatial and temporal information from multi-source remote sensors that separately occupy different spatiotemporal characteristics (e.g., a high-spatial and low-temporal resolution image and a low-spatial and high-temporal resolution image).

Early image fusion frameworks relying on transformation models that devote high-resolution multispectral image retrieval from a high-resolution panchromatic image and a low-resolution multispectral image, such as principal component analysis [11], hue intensity saturation, and [12] wavelet transforms [13] based only on digital number, and a few mathematical models, are inherited from the digital image processing field. The virtual effect of generated images is remarkably enhanced by these traditional approaches, whereas the physical meanings of the fused image itself and application-oriented analysis and validation are generally absent [14]. Thus, this category of fusion strategies is inadequate for enhancing or further parsing user-interested image information. Spatiotemporal fusion, which aims to achieve high-temporal prediction of high-resolution images by blending high-resolution images at observed dates and low-resolution images under relative dates, has been emerging in this research field as a promising way to resolve the previously mentioned problem. Unlike early fusion methods, spatiotemporal fusion models establish spatiotemporal correlations between inputted high- and low-resolution images based on physical parameters in remote sensing, such as radiance and apparent or surface reflectance. However, from another perspective, spatiotemporal fusion methods that rely on geological or physical parameters do not provide such a novel technique as a thinking mode of fusion strategies by which spectral unmixing, spatiotemporal filtering, and sparse learning are currently utilized for an accurate description of the radiometric spectrum changes of surface features.

The unmixing-based fusion methodology that is considered effective for the case without significant seasonal changes was first presented by Fortin et al. [15] and Zhukov et al. [16] and then validated and improved by Minghelli [17], Zurita [18], and Gevaert and Garcia [19]. The difference among the preceding methods is that neighborhood spectral information was not introduced by Fortin et al. [15] and Maselli [20] but was embedded in the works of Zhukov et al. [16] and Cherchali et al. [21]. In addition, the linear unmixing model resolved by least squares or multiple linear regression is preferred due to its simplicity and efficiency. Recently, a flexible spatiotemporal data fusion (FSDAF) method was proposed by combining spectral unmixing analysis and a thin-plate spline interpolator [22] and compared with the algorithm of Zurita [18]. FSDAF demonstrates superior performance in capturing reflectance changes due to land cover conversions.

As a popular spatiotemporal fusion strategy, models based on spatiotemporal filtering assign additional temporal and spectral information to a high-resolution image with the help of ancillary low-spatial and high-temporal resolution images. Typically, the spatial and temporal adaptive reflectance fusion model (STARFM) [23] algorithm provides accurate, efficient, and stable prediction despite various inputting data conditions. From the viewpoint of sensor observation differences between different cover types when calculating their weight contribution to the pending pixel [24] and data optimization [25], two improved versions of STARFM have been proposed. To capture surface changing information with a short-lived fluctuation in the image, the enhanced STARFM (ESTARFM) [26] algorithm maintains the weight function and its contributing rules in STARFM and concentrates on promoting fusion quality for land covers with significant temporal spectrum variation (e.g., vegetation). Although additional detailed spatial change features can be obtained by ESTARFM, the temporal characteristics to be simulated should be similar and even very close to the observed data. Thus, blending high- and low-resolution images at observed date(s) and low-resolution image at a predicted date is theoretically unreasonable for the prediction when a substantial temporal discrepancy exists between these images. Apart from the models that have similar theoretical principles [27] or confined improvements [28] with STARFM and ESTARFM, a reflectance fusion algorithm based on the semi-physical model [29] provides another novel path to build a spatiotemporal correlation between multi-source images. This algorithm has been preliminarily validated in a regional application [30].

Another fusion method derived from sparse learning theory was recently developed by combining super-resolution reconstruction and sparse representation achieved using dictionary learning. Although learning-based models that currently include the single-pair-based method [31] and the two-pair-based method [32] according to the number of inputting training images, hold promises for solving fundamental problems in spatiotemporal fusion [33,34], their performance have been proved to be less stable than reconstruction-based models, such as STARFM (single-pair) and ESTARFM (two-pair). Considering that these sparse-learning fusion strategies are built upon the prior learning process by training insufficient image samples, the dictionary training step in sparse learning models therefore has difficulty providing a redundant expression of the inputting high- and low-resolution images. That means the derived "overcomplete" dictionary is not typical for both acquired data at observed and modeled dates, and the accurate retrieval of transition images, even two-layered fusion results, is difficult. To this end, an enhanced single-pair learning-based fusion scheme with the improved dictionary learning step, and its evaluation method for selecting spatiotemporal extension mode of dictionary training samples are proposed in Section 2. Experimental results are shown in Section 3, and the discussion is presented in Section 4. Conclusions are drawn in Section 5.

2. Methodology

Although two main existing spatiotemporal fusion methods based on sparse learning theory vary in model construction, fusion pattern, and complexity, the primary theoretical basis and its contributing mode for their fusion process are nearly the same. Considering the universality and simplicity of the algorithm with single image pair, an improved fusion scheme from the single-image-pair method is firstly proposed on a basis of an enhanced dictionary training strategy and then evaluated by two remotely sensed datasets.

2.1. Proposed Fusion Scheme with Enhanced Dictionary-Training Process

In the sparse-learning fusion method, remotely sensed images from the same sensor and channel are treated as different sparse "versions" of an invariable overcomplete dictionary D on different acquiring dates. Thereinto, the sparse "version" is generally called the sparse coefficient α and considered as an indicator of the seasonal factor of an acquired remotely sensed image, related to D (mainly indicates spatial and texture features). When no significant discrepancy in texture context occurs, the dictionary D derived from the observed image-pair can be on behalf of the one from the modeled image-pair. The key step of the sparse-learning fusion algorithm is therefore to

retrieve a high-precision overcomplete dictionary *D* through training the high- and the low-resolution image pair at the observed date. If a steadily performing dictionary training algorithm is applied (e.g., coupled K-SVD algorithm), the accuracy of the sparse-learning fusion results is significantly related to the sufficiency of inputting training samples, which is, obviously, not satisfied in the original single-pair-based fusion algorithm.

For the retrieval of high-precision D, an improved sparse-learning fusion method with enhanced dictionary training process is proposed in this study and the overall processing flow is shown in Figure 1. In this method, spatiotemporally extended training samples are utilized to promote the sufficiency of dictionary training operations in both fusion layers of the single-pair learning-based algorithm. Two modes are furthermore designed to increase employed training samples: the spatially extended mode and the temporally extended mode. Specifically, the spatially extended mode increases only the image size (from S_0 to S_1 in Figure 1) of all the inputting training samples (including the low-resolution image and the high–low resolution image-pair) at the observed date (t_1 in Figure 1). By contrast, the temporally extended mode increases the number of inputting training samples, which are all obtained from different acquired dates (t_3 , t_4 to t_n in Figure 1) and have the same image size as the original inputting images.

Considering the case where the single image pair is employed as inputs, assume that H_1 and L_1 denote high-resolution image and low-resolution image at t_1 (observed date), L_2 denotes low-resolution image at t_2 (modeled date), H_2 denotes high-resolution image at t_2 that is to be predicted, and the image size of L_1 , L_2 , M_1 and M_2 is $S_0 \times S_0$. The high-resolution dictionary D_h and the low-resolution dictionary D_l are now derived by minimizing the following improved objective functions:

$$\{D_l, \alpha_1\} = \underset{D_l, \alpha_1}{\operatorname{argmin}} \Big\{ \|X_1^{new} - D_l \alpha_1\|_F^2 \Big\}$$
(1)

$$D_h = \underset{D_h}{\operatorname{argmin}} \|Y_1^{new} - D_h \alpha_1\|_F^2$$
(2)

where α_1 is the sparse coefficients of D_l and D_h at t_1 ; and X_1^{new} and Y_1^{new} , respectively, denote the spatially or temporally extended training sample matrices instead of the original training sample matrices X_1 and Y_1 extracted from the difference image $(H_1 - L_1)$ and the low-resolution image L_1 .

As to the dictionary training strategy with the spatially extended mode, the employed training sample matrices X_1^{new} and Y_1^{new} in Equations (1) and (2) are finally extracted from the enlarged difference image $(H_1^{enl} - L_1^{enl})$ and the enlarged low-resolution image L_1^{enl} both with an spatially extended image size $S_1 \times S_1$ rather than the original image size $S_0 \times S_0$. When no significant seasonal change occurs between t_1 and t_2 (type changes), the completeness of dictionary D is mainly limited by spatial heterogeneity and diversity of surface features. This mode actually intends to address the issue of completeness of spatial features by learning larger image area where more samples of surface features can be found.

Another situation is, when the temporally extended mode is selected, the training sample matrices X_1^{new} and Y_1^{new} can be expressed as feature images respectively derived from the dataset $\{L_1, L_3^{add}, L_4^{add}, \ldots, L_n^{add}\}$ and the dataset $\{(H_1 - L_1), (H_3^{add} - L_3^{add}), (H_4^{add} - L_4^{add}), \ldots, (H_n^{add} - L_n^{add})\}$. Thereinto, $H_3^{add}, H_4^{add}, \ldots, H_n^{add}$ and $L_3^{add}, L_4^{add}, \ldots, L_n^{add}$ are additional high- and low-resolution training images (with the same image size as H_1 and L_1) observed at t_3, t_4, \ldots , and t_n . Under the assumption that seasonal change occurs between t_1 and t_2 , the temporally extended mode of training samples can improve the description of phenology features extracted from training data observed at different dates. Since a single dictionary D is considered to be hard to provide a complete and precise expression for overall seasonal features, it is reasonable to define an approximately "overcomplete" and phenology-based dictionary.

Moreover, to find out the effectivity of two modes mentioned above, an evaluation strategy is presented in Section 3 (Results) to give a convictive selection proposal when both spatially and temporally extended samples are available and execution efficiency is required.



Figure 1. The proposed fusion scheme with spatially or temporally extended samples in dictionary training.

2.2. Assessment Indices of the Proposed Fusion Scheme

To obtain an accurate description of the fusion results, four types of indices are provided from the aspect of spectral errors per band, similarity of the overall structure, spectral distortion, and overall spectral errors, and then applied to the modeled reflectance and the actual reflectance for an all-sided quality evaluation of the fusion results. Five quantitative indices, namely, average absolute difference (AAD), root-mean-square error (RMSE), structure similarity (SSIM) [35], spectral angle mapper (SAM) [36], and Erreur Relative Global Adimensionnelle de Synthèse (ERGAS) [37], which correspond to the indicated aspects of foregoing assessment indices, are gathered to validate the quality of the predicted images from different assessment views. SSIM, SAM, and ERGAS are obtained by computing the following equations:

$$SAM = \cos^{-1} \left(\frac{\sum_{i=1}^{B} \rho_{P_i} \rho_{R_i}}{\sqrt{\sum_{i=1}^{B} \rho_{P_i}^2} \sqrt{\sum_{i=1}^{B} \rho_{R_i}^2}} \right)$$
(3)

$$SSIM_{i} = \frac{(2\mu_{P_{i}}\mu_{R_{i}} + C_{1})(2\sigma_{P_{i}R_{i}} + C_{2})}{(\mu_{P_{i}}^{2} + \mu_{R_{i}}^{2} + C_{1})(\sigma_{P_{i}}^{2} + \sigma_{R_{i}}^{2} + C_{2})}$$
(4)

$$ERGAS = 100 \frac{p}{r} \sqrt{\frac{\sum_{i=1}^{B} (RMSE_i)^2}{B}}$$
(5)

where ρ_{P_i} and ρ_{R_i} are the reflectance in band $i \in [1, B]$ of the modeled image P and the actual image R; (μ_{P_i}, μ_{R_i}) , $(\sigma_{P_i}, \sigma_{R_i})$, and $\sigma_{P_iR_i}$ correspond to the mean value, standard deviation, and covariance in band i of P and R, respectively; $C_1 = (k_1 * L)^2$ and $C_2 = (k_2 * L)^2$; k_1 and k_2 are generally set as 0.01 and 0.03; L is the grayscale of reflectance images; $RMSE_i$ is the RMSE in band i of P and R; and p and r are the spatial resolutions of P and R. Small values of AAD, RMSE, SAM, and ERGAS and a high value of SSIM between the modeled reflectance image and the actual reflectance image indicate a considerable fusion result.

Scatter plots based on the channel-specified reflectance of the modeled data against actual data are provided to supplement the aforementioned quantitative indices with the visualized pattern to provide an intuitive quality assessment of fusion results, and, moreover, the total time-consumption of employed channels used in fusion strategies with spatiotemporally extended training samples is also considered here to give a general description of their efficiencies.

3. Results

In view of only one acquired image-pair (the high- and the low-resolution images) the proposed learning-based algorithm has required, two reconstruction-based spatiotemporal fusion models, the STARFM and the semi-physical reflectance fusion model, are employed in comparison to the original and the improved learning-based fusion algorithms. Specifically, the single-pair version of STARFM with default parameters and the improved algorithm based on semi-physical fusion model (SPFM) [30] are finally adopted to perform the experiment.

3.1. Datasets

In this paper, two datasets, which consist of rural and urban datasets, are employed to perform the fusion strategy that utilizes spatiotemporally extended training samples. The rural dataset uses the same experimental data as in [23], which has been characterized as a study area with phenology changes [32] and comprises Landsat ETM+ images with 30 m spatial resolution and the MODIS daily 500 m surface reflectance product (MOD09GHK) acquired on 24 May, 11 July, and 12 August 2001 (Figure 2). Beijing, which is a typical urban area in China, is selected as the urban dataset to validate the fusion quality with extended spatiotemporal training samples because the sparse learning fusion method is more sensitive to texture and structural features of fused images than others. For the urban dataset listed in Table 1, reflectance products comprised the 20 Landsat-8 OLI (30 m spatial resolution) scenes, and the corresponding MODIS 8-day MOD09-A1 (500 m spatial resolution) and-Q1 (250 m spatial resolution) acquired from 2013 to 2017 are used to perform the fusion strategy described in Figure 1. The Landsat-8 Surface Reflectance product, whose performance is accepted to be either close or better than Landsat TM/ETM+ reflectance products from the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) [38], has been generated from the Landsat Surface Reflectance Code (LaSRC). By contrast, the MODIS reflectance product is provided by combining the green channel of the MOD09A1 and the red and NIR channels of the MOD09Q1 that are directly downloaded from the Land Processes Distributed Active Archive Center (LPDAAC). Notably, only centered sub-images with 500×500 Landsat pixels (covering an area of 15 km \times 15 km) of the preceding two datasets are used in the fusion procedure, and spatially or temporally extended training samples will be merely utilized in the process of dictionary training.

Figures 2 and 3 show that training samples initially cover an area of 15 km \times 15 km (500 \times 500 Landsat pixels) for both datasets and finally reach 36 km \times 36 km (1200 \times 1200 Landsat pixels) for the rural dataset and 60 km \times 60 km (2000 \times 2000 Landsat pixels) for the urban dataset as their maximum sizes in the spatially extended fusion experiment. Between the original and the maximal sizes of training samples in each dataset, a dozen training samples with different image sizes are clipped with a size step of 3 km \times 3 km (100 \times 100 Landsat pixels). As a result, all training samples have been resized as 500 \times 500, 600 \times 600, ..., 1200 \times 1200 Landsat pixels for the rural dataset, and 500 \times 500, 600 \times 600, ..., 2000 \times 2000 for the urban dataset. In consideration of the heterogeneity and the diversities of the spatial extension of surface features in different directions, all the employed training samples yield to the same central position as their original training images, which are also inputted as the observed reflectance for Landsat and MODIS (500×500 Landsat pixels). For instance, the urban study area is positioned at the center of Beijing City proper, which is composed of two main parts, namely, Dongcheng District and Xicheng District. By this way, the fusion strategy with spatial extension of training samples tends to be less sensitive to the variety and texture features of land cover from different study areas. A fair and authentic comparison between the original fusion algorithm and its modified strategy with spatial extension can therefore be expected.

To ensure a consistent comparison against temporal directions, a bi-direction fusion scheme is adopted for the preceding two datasets. The dates 24 May and 11 July 2001 are determined as the bi-direction observed or modeled dates for the rural dataset, while 10 July and 12 September 2017 are selected for the urban dataset. The bi-direction fusion indicates that one of the bi-directional dates acts as the observed time, and the other serves as the modeled time (to be predicted). Original inputting reflectance images from each dataset at the observed dates are spatially replaced by or temporally added to the extended training samples to retrieve a new, enhanced training sample.



Figure 2. Employed Landsat and MODIS reflectance data (NIR/red/green) of the rural dataset: (**a**–**f**) Landsat ETM+ images with a size of 500 \times 500 and 2000 \times 2000 pixels on 24 May, 11 July, and 12 August 2001, respectively; (**g**–**l**) MODIS images correspond to (**a**–**f**).



Figure 3. Landsat and MODIS reflectance data (NIR/red/green) that cover the Beijing urban area: (**a**–**d**) Landsat-8 images observed on 10 July and 12 September 2017 with a size of 500×500 and 2000×2000 pixels; (**e**–**h**) MODIS reflectance products that correspond to (**a**–**d**).

	Landsat-8 OLI		MODIS MOD09A1/Q1							
D	ate	Data Info	D	ate	Data Info					
31 July 2013 1 September 2013 19 August 2014 4 September 2014 22 August 2015 7 September 2015 20 May 2016 11 October 2016 31 January 2017 4 March 2017	21 April 2017 7 May 2017 23 May 2017 10 July 2017 12 September 2017 28 September 2017 30 October 2017 15 November 2017 1 December 2017 17 December 2017	Orbit: 123–32 Band: 3–5 Resolution: 30 m	28 July 2013 29 August 2013 21 August 2014 6 September 2014 21 August 2015 6 September 2015 16 May 2016 7 October 2016 2 February 2017 6 March 2017	23 April 2017 9 May 2017 25 May 2017 12 July 2017 14 September 2017 30 September 2017 1 November 2017 17 November 2017 3 December 2017 19 December 2017	Orbit: 26–04,05 Band: 1, 2 (MOD09Q1) and 4 (MOD09A1) Resolution: 250 m (MOD09Q1) and 500 (MOD09A1)					

Table 1. Employed Landsat-8 OLI and MODIS reflectance products of the urban dataset.

3.2. Experimental Results with the Rural Dataset

3.2.1. Experiments with Spatially Extended Training Samples

In this experiment, the rural dataset is employed to implement the bi-directional fusion scheme for modeling reflectance images on 24 May (Table 2) or 11 July (Table 3) 2001 with different sizes of spatially extended training samples, which are used in the dictionary learning process. The quality assessment of fusion results from the aforementioned bi-direction scheme is shown in Tables 2 and 3 and Figure 4 for a graphical description of the total statistics. Several modeled images are selected from the predicted results and are then validated by scatter plots with actual reflectance (Figure 5).

Table 2. Assessment indices of the spatially extended fusion for modeling reflectance on 24 May 2001 of the rural dataset.

Methods	Training	A	$AD \times 1$	0 ²	R	MSE × 1	10 ²	S	$SIM \times 1$.0 ²	CAM	ERGAS
Methods	Image Size	G	R	NIR	G	R	NIR	G	R	NIR	- SAM	
Original algorithm	500×500	0.46	0.79	1.74	0.66	1.16	2.63	96.16	90.79	79.73	1.8065	18.8364
	600×600	0.43	0.75	1.55	0.63	1.11	2.33	96.53	91.51	83.17	1.8116	17.7550
	700×700	0.42	0.72	1.37	0.60	1.08	2.07	96.83	91.89	86.52	1.8151	16.9564
Proposed	800 imes 800	0.41	0.70	1.26	0.59	1.05	1.84	96.95	92.26	89.18	1.8175	16.3198
algorithm	900×900	0.39	0.68	1.19	0.58	1.02	1.75	97.10	92.66	90.16	1.8198	15.8187
argoritimi	1000×1000	0.39	0.67	1.16	0.57	1.01	1.69	97.13	92.82	90.80	1.8206	15.6352
	1100×1100	0.38	0.66	1.13	0.56	0.99	1.63	97.2	92.95	91.32	1.8215	15.3832
	1200×1200	0.38	0.64	1.11	0.56	0.98	1.63	97.21	93.03	91.37	1.8219	15.2739
STARFM	-	0.42	0.69	1.78	0.60	1.08	2.65	97.01	92.11	88.31	1.8123	17.0671
SPFM	-	0.41	0.71	1.68	0.59	1.10	2.47	96.49	91.99	88.52	1.8163	16.5105

Table 3. Assessment indices of the spatially extended fusion for modeling reflectance on 11 July 2001 of the rural dataset.

Methods	Training	А	$AD \times 1$	0 ²	R	$MSE \times 1$	10 ²	S	SIM imes 1	0 ²	SAM	EDCAS
Methous	Image Size	G	R	NIR	G	R	NIR	G	R	NIR	- SAW	EKGAS
Original algorithm	500×500	0.45	0.67	2.26	0.60	1.00	3.23	96.43	92.29	79.17	1.8050	20.4154
	600×600	0.43	0.62	2.17	0.60	0.88	3.15	96.61	93.93	79.75	1.8102	18.7794
	700×700	0.43	0.61	2.13	0.60	0.87	3.06	96.86	94.06	80.90	1.8130	18.2613
Proposed	800 imes 800	0.40	0.56	1.89	0.55	0.80	2.71	97.36	94.97	84.44	1.8190	16.5935
algorithm	900×900	0.38	0.55	1.82	0.54	0.78	2.61	97.49	95.18	85.52	1.8206	16.1860
argoritimi	1000×1000	0.38	0.54	1.79	0.53	0.77	2.58	97.59	95.34	85.78	1.8215	15.9259
	1100×1100	0.37	0.53	1.77	0.52	0.75	2.52	97.65	95.45	86.29	1.8223	15.6691
	1200×1200	0.37	0.52	1.75	0.52	0.75	2.51	97.66	95.52	86.33	1.8226	15.5429
STARFM	-	0.50	0.68	2.03	0.70	1.06	2.83	96.74	92.54	84.02	1.8172	16.4957
SPFM	-	0.41	0.74	1.91	0.59	1.08	2.79	97.10	92.19	84.81	1.8171	16.5169



Figure 4. Graphical assessment indices of the proposed bi-directional fusion with spatially extended training samples from the rural dataset, (**a**) and (**b**) are respectively for modeling the reflectance on 24 May and 11 July 2001.



Figure 5. Proposed bi-directional fusion results with spatially extended training samples from the rural dataset: (**a**–**d**) the composited fusion results (NIR/red/green) modeled on 24 May and 11 July 2001 with training image sizes of 500×500 pixels and 1200×1200 pixels, respectively; and (**e**–**p**) comparisons among green, red, and NIR bands of the modeled reflectance and the actual reflectance that correspond to (**a**–**d**).

3.2.2. Experiments with Temporally Extended Training Samples

Considering that only three pairs of temporal reflectance images are held in the rural dataset, the image pair acquired on 12 August 2001 is always taken as an additional training sample to model either 24 May or 11 July 2001. The resulting fused images from the bi-directional fusion scheme with temporally extended training samples and their reflectance scatter plots are shown in Figure 6, and the assessment indices are listed in Table 4.



Figure 6. Proposed bi-directional fusion results with temporally extended training samples from the rural dataset: (**a**–**d**) the composited fusion results modeled on 24 May 2001 and the comparison (green, red, and NIR) with the actual reflectance; (**e**–**h**) the composited fusion results modeled on 11 July 2001 and the comparison (green, red, and NIR) with the actual reflectance.

Table 4. Assessment indices of the proposed bi-directional fusion with temporally extended training samples from the rural dataset.

Modeled Dates	A	$AAD imes 10^2$			ASE ×	10 ²	SS	$SIM \times 1$.0 ²	SAM	EDCAS
	G	R	NIR	G	R	NIR	G	R	NIR	- SAM	ENGAS
24 May 11 July	0.38 0.38	0.65 0.53	1.13 1.77	0.57 0.53	1.00 0.76	1.66 2.55	97.15 97.59	92.99 95.38	91.11 85.83	1.8213 1.8217	15.4358 15.7507

3.3. Experimental Results with the Urban Dataset

3.3.1. Experiments with Spatially Extended Training Samples

Along with the temporal-corresponding MODIS reflectance product MOD09A1 and MOD09Q1 (Table 1), only the Landsat-8 surface reflectance products acquired on 10 July and 12 September 2017 (Figure 3a,c) are used as basic experimental data to apply the bi-directional fusion scheme with spatially extended training samples that cover the Beijing urban area. Assessment indices related to both temporal directions are summarized in Tables 5 and 6, with the graphical presentation shown in Figure 7. Several typical modeled results and their scatter plots with actual reflectance from this spatially extended fusion with training image sizes of 500×500 and 1500×1500 pixels are displayed in Figure 8.

Mathada	Training	А	$AD \times 1$	0 ²	R	$MSE \times 1$	10 ²	S	SIM imes 1	0 ²	SAM	EPCAS
Methods	Image Size	G	R	NIR	G	R	NIR	G	R	NIR	- SAM	EKGAS
Original algorithm	500×500	1.76	2.11	3.54	2.41	2.96	4.71	84.48	81.45	77.53	1.7854	26.2220
	600×600	1.71	2.02	3.42	2.29	2.79	4.58	86.66	84.08	79.49	1.7946	24.9515
	700×700	1.70	1.99	3.40	2.27	2.73	4.56	87.34	84.8	79.87	1.7973	24.6278
	800 imes 800	1.69	1.98	3.39	2.26	2.70	4.54	87.56	85.64	80.35	1.7986	24.3276
	900×900	1.68	1.97	3.38	2.25	2.69	4.53	87.75	85.72	80.78	1.8002	24.2388
	1000×1000	1.67	1.96	3.38	2.24	2.68	4.51	87.81	85.88	80.94	1.8010	24.1858
	1100×1100	1.67	1.96	3.37	2.23	2.68	4.50	87.89	85.91	81.09	1.8009	24.1359
Dramagad	1200×1200	1.66	1.95	3.36	2.22	2.66	4.47	87.97	85.94	81.27	1.8025	24.0686
algorithm	1300×1300	1.66	1.94	3.35	2.22	2.66	4.46	87.97	85.99	81.35	1.8024	24.0527
aigoriunn	1400×1400	1.65	1.93	3.35	2.21	2.64	4.46	88.02	86.03	81.42	1.8028	24.0454
	1500×1500	1.65	1.92	3.35	2.21	2.64	4.44	88.02	86.05	81.47	1.8031	23.9964
	1600×1600	1.65	1.92	3.35	2.21	2.64	4.46	88.03	86.05	81.46	1.8030	24.0518
	1700×1700	1.66	1.94	3.37	2.22	2.65	4.51	88.03	86.03	81.44	1.8027	24.0764
	1800 imes 1800	1.65	1.92	3.35	2.22	2.64	4.43	88.03	86.04	81.45	1.8030	24.0452
	1900×1900	1.65	1.92	3.34	2.21	2.63	4.44	88.04	86.07	81.47	1.8032	23.9916
	2000×2000	1.65	1.92	3.35	2.22	2.64	4.45	88.03	86.06	81.45	1.8028	24.0490
STARFM	_	1.66	1.95	3.50	2.22	2.67	4.63	87.96	85.95	78.39	1.8016	24.7963
SPFM	_	1.65	2.00	3.61	2.21	2.95	4.86	88.01	85.81	77.28	1.7953	25.1976

Table 5. Assessment indices of the spatially extended fusion with the urban dataset for modeling reflectance on 10 July 2017.

Table 6. Assessment indices of the spatially extended fusion with the urban dataset for modeling reflectance at 12 September 2017.

Mathada	Training	А	$AD \times 1$	0 ²	R	$MSE \times 1$	10 ²	S	SIM imes 1	0 ²	SAM	EDCAS
Methous	Image Size	G	R	NIR	G	R	NIR	G	R	NIR	SAN	EKGAS
Original algorithm	500×500	1.73	2.00	3.45	2.18	2.64	4.64	87.55	85.01	79.01	1.7850	29.3382
	600×600	1.70	1.97	3.40	2.14	2.59	4.47	88.61	86.35	82.89	1.7922	28.3268
	700×700	1.70	1.96	3.40	2.14	2.56	4.35	88.63	86.47	82.88	1.7924	28.2852
	800 imes 800	1.68	1.94	3.38	2.12	2.53	4.31	88.95	87.12	83.30	1.7943	27.8460
	900×900	1.69	1.95	3.40	2.13	2.54	4.33	88.83	86.79	82.89	1.7886	27.8857
	1000×1000	1.68	1.93	3.39	2.12	2.53	4.31	89.02	87.21	83.49	1.7947	27.7921
Dramagad	1100×1100	1.68	1.92	3.37	2.12	2.53	4.31	89.03	87.44	83.54	1.7947	27.7726
	1200×1200	1.68	1.91	3.37	2.12	2.52	4.30	89.06	87.56	83.62	1.7949	27.6851
algorithm	1300×1300	1.68	1.90	3.35	2.12	2.52	4.29	89.07	87.61	83.63	1.7952	27.5976
aigonum	1400×1400	1.68	1.90	3.34	2.12	2.51	4.28	89.09	87.63	83.67	1.7964	27.5520
	1500×1500	1.68	1.90	3.34	2.12	2.51	4.28	89.10	87.66	83.70	1.7985	27.5435
	1600×1600	1.68	1.90	3.34	2.12	2.51	4.29	89.09	87.64	83.69	1.7980	27.5481
	1700×1700	1.68	1.89	3.32	2.12	2.49	4.21	89.12	87.71	83.78	1.8000	27.4747
	1800 imes 1800	1.69	1.91	3.35	2.12	2.53	4.31	89.10	87.65	83.71	1.7975	27.5554
	1900×1900	1.68	1.90	3.34	2.12	2.51	4.29	89.11	87.66	83.69	1.7989	27.5174
	2000×2000	1.68	1.90	3.33	2.12	2.51	4.25	89.13	87.69	83.75	1.7991	27.4951
STARFM	_	1.70	1.95	3.43	2.16	2.56	4.51	88.51	86.68	82.79	1.7939	28.5247
SPFM	_	1.68	2.01	3.44	2.17	2.63	4.50	87.68	84.83	82.45	1.7901	29.2313



Figure 7. Graphical assessment indices of the proposed bi-directional fusion with spatially extended training samples from the urban dataset, (**a**) and (**b**) are respectively for modeling the reflectance on 10 July and 12 September 2017.



Figure 8. Proposed bi-directional fusion results with spatially extended training samples from the urban dataset: (a-d) the composited fusion results (NIR/red/green) modeled on 10 July and 12 September 2017 with training image sizes of 500 × 500 pixels and 1500 × 1500 pixels, respectively; and (e-p) the scatter plots of (a-d), which indicate the comparison among the green, red, and NIR bands of the modeled and actual reflectance.

3.3.2. Experiments with Temporal Extended Training Data

With regard to the temporal extension of training samples, we first defined an optimized selection mode for temporal training samples by analyzing the fusion quality of the Beijing urban dataset in 2017 and selecting eligible acquired dates from 2013 to 2016 according to Landsat-8 reflectance data, which are accumulated as additional training samples into the dictionary learning process. This bi-directional fusion strategy with the temporally extended training samples is shown in Figure 9.



Figure 9. Proposed bi-directional fusion strategy with temporally extended training samples.

Table 7. Assessment indices from the proposed fusion with the urban data acquired in 2017 for modeling reflectance on 10 July 2017.

Added Training	A	$AAD imes 10^2$			ASE ×	10 ²	SS	$IM \times 1$.0 ²	SAM	FRGAS
Dates	G	R	NIR	G	R	NIR	G	R	NIR	SAM	EKGAS
31 January 2017	1.76	2.08	3.57	2.40	2.90	4.83	84.55	82.25	76.36	1.7857	26.1273
4 March 2017	1.74	2.03	3.48	2.37	2.82	4.63	85.23	83.41	78.58	1.7918	25.4679
21 April 2017	1.75	2.01	3.49	2.37	2.79	4.59	85.2	83.97	79.05	1.7930	25.2690
7 May 2017	1.75	2.03	3.52	2.37	2.83	4.65	85.04	83.24	78.44	1.7910	25.5186
23 May 2017	1.73	1.96	3.46	2.35	2.72	4.74	85.63	84.96	77.24	1.7947	25.1566
28 September 2017	1.73	1.97	3.46	2.36	2.73	4.58	85.35	84.73	78.92	1.7953	25.0129
30 October 2017	1.72	2.00	3.47	2.36	2.78	4.67	85.38	84.10	78.06	1.7935	25.3112
15 November 2017	1.76	2.06	3.46	2.41	2.88	4.61	84.59	82.61	78.75	1.7901	25.7875
1 December 2017	1.73	2.07	3.51	2.37	2.89	4.69	85.05	82.45	77.85	1.7903	25.8093
17 December 2017	1.75	2.05	3.63	2.38	2.84	5.00	85.19	83.22	72.14	1.7888	26.0933

Nearly all 12 reflectance data acquired from 31 January to 17 December (Table 1) were taken as additional training samples for the dictionary learning process, except for two Landsat reflectance images acquired on 10 July and 12 September 2017. The assessment indices are listed in Tables 7 and 8. Although only a small discrepancy exists among the fusion results with additional training samples from different acquisition dates, from which the reflectance data are between or close to the observed date and the modeled date, slightly higher fusion accuracy can be expected (23 May and 28 September

2017 in this experiment). Two reflectance images that satisfy the aforementioned assumption were then selected from each year from 2013 to 2016 and finally used to perform and validate the fusion with temporally accumulated training samples (Tables 9 and 10 and Figure 10).

Table 8. Assessment indices from the proposed fusion with the urban data acquired in 2017 for modeling reflectance on 12 September 2017.

Added Training	A	$AD \times 1$	0 ²	RN	ASE ×	10 ²	SS	$IM \times 1$.0 ²	SAM	FRGAS
Dates	G	R	NIR	G	R	NIR	G	R	NIR	SAM	LKGAS
31 January 2017	1.72	2.02	3.48	2.17	2.65	4.70	87.22	84.73	78.46	1.7836	29.4916
4 March 2017	1.72	1.98	3.69	2.17	2.59	4.86	87.38	85.26	73.24	1.7813	29.4935
21 April 2017	1.71	2.00	3.4	2.15	2.60	4.45	87.95	85.46	81.61	1.7885	28.7471
7 May 2017	1.71	1.97	3.39	2.15	2.56	4.35	87.98	85.95	82.39	1.7900	28.3728
23 May 2017	1.72	1.95	3.40	2.15	2.53	4.35	87.96	86.15	82.38	1.7902	28.2556
28 September 2017	1.70	1.92	3.35	2.12	2.48	4.34	88.30	86.80	82.27	1.7916	27.8873
30 October 2017	1.72	1.98	3.39	2.18	2.59	4.53	87.49	85.58	80.27	1.7869	28.9585
15 November 2017	1.71	2.00	3.46	2.16	2.64	4.69	87.68	84.99	77.91	1.7844	29.3909
1 December 2017	1.73	2.01	4.63	2.19	2.62	6.42	87.32	85.25	38.30	1.7599	32.9495
17 December 2017	1.72	1.99	3.40	2.18	2.61	4.37	87.40	85.15	82.36	1.7878	28.7952

Table 9. Assessment indices from the proposed fusion with the urban data acquired from 2013 to 2016 for modeling reflectance on 10 July 2017.

Added Training	$AAD \times 10^2$			RN	ASE ×	10 ²	S	$SIM \times 1$	10 ²	SAM	EPCAS
Years	G	R	NIR	G	R	NIR	G	R	NIR	SAM	EKGAS
2013	1.73	2.01	3.47	2.36	2.78	4.62	85.43	84.17	78.66	1.7927	25.2327
2013 and 2014	1.69	1.99	3.45	2.29	2.75	4.62	86.58	84.69	78.72	1.7960	24.8839
2013 to 2015	1.66	1.93	3.38	2.25	2.66	4.45	87.53	85.81	80.64	1.8020	24.1563
2013 to 2016	1.66	1.93	3.37	2.25	2.66	4.47	87.58	85.81	80.37	1.8018	24.2050



Figure 10. Proposed bi-directional fusion results with temporally extended training samples (from 2013 to 2016) using the urban dataset: (**a**–**d**) the composited fusion results of modeled reflectance on 10 July 2017 and the comparison with actual reflectance; (**e**–**h**) the composited fusion results of modeled reflectance or reflectance on 12 September 2017 and the comparison with actual reflectance.

Added Training Years	$AAD \times 10^2$			RN	ASE ×	10 ²	S	$SIM \times 1$	10 ²	SAM	EPCAS
	G	R	NIR	G	R	NIR	G	R	NIR	SAM	EKGAS
2013	1.69	1.95	3.38	2.13	2.52	4.4	88.55	86.89	81.97	1.7921	28.1927
2013 and 2014	1.69	1.94	3.36	2.13	2.5	4.3	88.76	87.18	83.32	1.7941	27.8998
2013 to 2015	1.68	1.91	3.34	2.12	2.46	4.25	88.98	87.58	83.86	1.7955	27.6131
2013 to 2016	1.69	1.92	3.36	2.13	2.48	4.41	88.87	87.29	82.07	1.7934	28.0341

Table 10. Assessment indices from the proposed fusion with the urban data acquired from 2013 to 2016 for modeling reflectance on 12 September 2017.

4. Discussion

4.1. Fusion Quality with Spatial Extended Training Samples

The resulting assessment indices from the two datasets indicate good agreement in both temporal directions of the fusion strategy with spatially extended training samples, which varied from 500×500 to 1200×1200 and 2000×2000 pixels (Tables 2, 3, 5 and 6 and Figures 5 and 8). On the one hand, the overall fusion quality increases with larger training image sizes, and only small improvements can be expected when the size of the training images reaches a threshold, which is approximately two and three times the original image size for the rural and the urban datasets, respectively. Besides, the proposed fusion algorithm generally has a better performance than STARFM and SPFM models under the "size threshold" for both the rural dataset (phenology changes) and the urban dataset (type changes). On the other hand, AAD, RMSE, and SSIM indices show increasing errors from the green, red, and NIR bands through all training image sizes (Figures 4 and 7). A reasonable explanation for the threshold size of the training samples is the reduced spatial similarity. Therefore, the features of image structures become less effective with the increasing image size of the training samples.

The different levels of fusion errors over bands yield different standard deviations for each band. For the rural dataset with phenology changes, the standard deviations of a reflectance image used to quantify the spectral amount of variation or dispersion of images for the acquisitions on 24 May and 11 July 2001 are 0.0102, 0.013, and 0.0476 and 0.0108, 0.0165, and 0.0306, respectively. A more integrative description of fusion results has been addressed by SSIM, SAM, and ERGAS indices rather than AAD and RMSE indices. The ERGAS index in particular provides a significant difference between assessment values of AAD and RMSE indices with very small discrepancies in one or more channels. Similarly, the increasing fusion errors from the green and the red bands to the NIR band yield standard deviations of 0.0334, 0.0413, and 0.0611 and 0.031, 0.0374, and 0.0565, for the Beijing urban data acquired on 10 July and 12 September 2017, respectively. The change in the threshold size of training images from two times (rural area) to three times (urban area) is mainly ascribed to the difference in surface features and employed reflectance products (Landsat-7 ETM plus and MODGHK for the rural dataset, and Landsat-8 OLI and MOD09A1/Q1 for the urban dataset). At the threshold size of the urban training images (approximately 1500×1500 pixels), the modeled images in both temporal directions, especially for the reflectance on 12 September 2017, seem to have less noise disturbance than other image sizes (Figure 8). Considering running time of the fusion with spatially extended training samples, the procedure with larger size of training image become more and more time consuming, which is growing not linearly but exponentially.

4.2. Fusion Quality with Temporal Extended Training Samples

Unlike the original bi-directional fusion results, the temporally extended fusion strategy can promote fusion quality and perform better than the spatially extended fusion strategy when an equal number of training images are handled. The training image size in the temporally extended fusion scheme with the rural dataset (Table 4) corresponds to the training image size of 700×700 pixels used in the spatially extended fusion scheme (Tables 2 and 3). The temporal extension scheme is therefore

more efficient than the spatial extension scheme in training the rural dataset with phenology changes. Moreover, the assessment indices from the urban dataset primarily show decreasing fusion error when temporal training samples are added (Figure 11), and disagreement occurs when the two reflectance images acquired in 2016 participate in the training sample set. This phenomenon may be attributed to the large seasonal difference between the acquisition dates in 2016 and the observed–modeled period. In addition, a more effective fusion strategy for the rural dataset (the Beijing area) is used to bring the spatially extended training samples, rather than the temporally extended training samples, into the training set (Figure 12).



Figure 11. Proposed bi-directional fusion strategy in the temporal extension of training samples using the urban dataset, (**a**) and (**b**) are respectively for modeling the reflectance on 10 July and 12 September 2017.



Figure 12. Quality and efficiency of proposed bi-directional fusion with the spatiotemporally extended training samples from the urban dataset, (**a**) and (**b**) are respectively for the spatially extended mode and the temporally extended mode.

The results from the rural dataset, especially for the NIR channel, are more sensitive to the added temporal training image than the urban dataset (Figure 6d,h and Figure 10d,h). Regardless of the seasonal characteristics of the employed temporal training images, the discrepancy in spatial features, such as texture and structure, plays a leading role in the comparison of fused results from two datasets with different change types (primarily the phenology change in the rural dataset and the texture and structural change in the urban dataset). The time-consumption of the fusion with temporally extended training samples intends to be similar with the strategy with spatial extension due to the proportional image size between spatially and temporally extended training samples.

5. Conclusions

An enhanced fusion scheme based on the single-pair sparse learning fusion model is proposed by improving the dictionary training process, and its evaluation strategy is designed by employing the spatially and the temporally extended training samples in this paper. Results from the bi-directional fusion scheme show high agreement in the assessment indices of the fusion quality, which indicates a decrease in prediction errors and an increase in image similarity with the extension of spatial or temporal training samples. This fusion scheme is significantly effective until the spatial threshold size (approximately two to three times the original image size used here) of the training images is reached or one or more temporal training sample(s) with dissimilar acquisition seasons is added. Compared to STARFM and SPFM models, a better fusion quality can also be obtained by the proposed method with an enhanced "threshold" training size. In detail, the fusion strategy with spatially extended training samples obtain better performance than the fusion strategy with temporally extended training samples for the urban dataset, whereas an opposite inference can be derived from the rural dataset. In consideration of the land cover characteristics of the two datasets, in which phenology changes occur in the rural dataset and type changes appear in the urban dataset, a reliable approach is to adopt an adaptive pattern of training samples extended spatially or temporally to promote fusion quality according to the data acquisition condition and the land cover change type of a study area. The results of the temporally extended fusion scheme are significantly affected by additional training samples with different seasonal features. Therefore, the proposed sparse learning-based fusion scheme is more sensitive to temporal changes than to spatial changes in surface features. To promote the efficiency of sparse learning-based fusion methods, a spatial and temporal similarity measure should be designed for filtering and training spatiotemporal samples and then integrated with the fusion procedure after its availability is validated by typical areas with various land cover changes.

Compared to the whole process of the original sparse-learning algorithm cost 3.7 min for an image with 500×500 pixels, which is more efficient than the STARFM (about 4 min) and less than the SPFM (2.3 min), the proposed method with spatiotemporally extended training samples will become far more time consuming if a better fusion result is required. Actually, this issue can be effectively addressed by some updated sparse coding techniques. For instance, the online dictionary learning methods [39,40] can significantly reduce the time consumption of the entire training process to 1 min or so for 500×500 pixels and expect to be more effective with growing image size. By this way, the proposed method has a high potential in processing large scenes (spatial or temporal acquirements) usually with multiple channels and taking them into consideration for reflectance reconstruction.

Author Contributions: D.L. and Y.L. conceived and designed the work of this paper. D.L. wrote the manuscript. W.Y. and Y.G. designed the experiment and analyzed the results. Q.H. and L.M. revised the manuscript and re-recognized the Methodology Section. Y.L. and W.Y. approved the final version. Y.C. and X.L. collected and preprocessed the experimental data.

Funding: This study was supported by National Natural Science Foundation of China: 41501372, National High Technology Research and Development Program of China: 2014AA123202, National Key R&D Program of China: 2018YFB0504800 (2018YFB0504804), Scientific and Technological Innovation Projects of Shanxi, China: 2016144.

Acknowledgments: The authors would like to thank F. Gao, B. Huang, and H. Song for sharing their experimental data and source codes on the Internet.

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

References

- 1. Dewan, A.M.; Yamaguchi, Y. Land use and land cover change in Greater Dhaka, Bangladesh: Using remote sensing to promote sustainable urbanization. *Appl. Geogr.* **2009**, *29*, 390–401. [CrossRef]
- Cohen, W.B.; Goward, S.N. Landsat's role in ecological applications of remote sensing. *AIBS Bull.* 2004, 54, 535–545. [CrossRef]

- Kennedy, R.E.; Yang, Z.; Cohen, W.B. Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendr-Temporal segmentation algorithms. *Remote Sens. Environ.* 2010, 114, 2897–2910. [CrossRef]
- 4. Steinberg, D.K.; Carlson, C.A.; Bates, N.R.; Johnson, R.J.; Michaels, A.F.; Knap, A.H. Overview of the US JGOFS Bermuda Atlantic Time-series Study (BATS): A decade-scale look at ocean biology and biogeochemistry. *Deep Sea Res. Part II Top. Stud. Oceanogr.* **2001**, *48*, 1405–1447. [CrossRef]
- Joyce, K.E.; Belliss, S.E.; Samsonov, S.V.; McNeill, S.J.; Glassey, P.J. A review of the status of satellite remote sensing and image processing techniques for mapping natural hazards and disasters. *Prog. Phys. Geogr.* 2009, 33, 183–207. [CrossRef]
- 6. Richardson, W.H. Bayesian-based iterative method of image restoration. JOSA 1972, 62, 55–59. [CrossRef]
- 7. Elad, M.; Feuer, A. Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images. *IEEE Trans. Image Process.* **1997**, *6*, 1646–1658. [CrossRef] [PubMed]
- 8. Park, S.C.; Park, M.K.; Kang, M.G. Super-resolution image reconstruction: A technical overview. *IEEE Signal Process. Mag.* 2003, 20, 21–36. [CrossRef]
- 9. Tseng, D.C.; Tseng, H.T.; Chien, C.L. Automatic cloud removal from multi-temporal SPOT images. *Appl. Math. Comput.* 2008, 205, 584–600. [CrossRef]
- 10. Chen, J.; Zhu, X.; Vogelmann, J.E.; Gao, F.; Jin, S. A simple and effective method for filling gaps in Landsat ETM+ SLC-off images. *Remote Sens. Environ.* **2011**, *115*, 1053–1064. [CrossRef]
- 11. Kwarteng, P.S.; Chavez, A.Y. Extracting spectral contrast in Landsat Thematic Mapper image data using selective principal component analysis. *Photogramm. Eng. Remote Sens.* **1989**, *55*, 339–348.
- 12. Carper, W.; Lillesand, T.; Kiefer, R. The use of intensity-hue-saturation transformations for merging SPOT panchromatic and multispectral image data. *Photogramm. Eng. Remote Sens.* **1990**, *56*, 459–467.
- 13. Yocky, D.A. Multiresolution wavelet decomposition image merger of Landsat Thematic Mapper and SPOT panchromatic data. *Photogramm. Eng. Remote Sens.* **1996**, *62*, 1067–1074.
- 14. Sun, H.; Dou, W.; Yi, W. Discussion of status, predicament and development tendency in the remotely sensed image fusion. *Remote Sens. Inf.* **2011**, *1*, 104–108.
- 15. Fortin, J.P.; Bernier, M.; Lapointe, S.; Gauthier, Y.; De Sève, D.; Beaudoin, S. *Estimation of Surface Variables at the Sub-Pixel Level for Use as Input to Climate and Hydrological Models*; INRS-Eau: Sainte-Foy, QC, Canada, 1998.
- 16. Zhukov, B.; Oertel, D. Multi-sensor multi-resolution technique and its simulation. *Zeitschrift für Photogrammetrie und Fernerkundung* **1996**, *1*, 11–21.
- 17. Minghelli-Roman, A.; Mangolini, M.; Petit, M.; Polidori, L. Spatial resolution improvement of MeRIS images by fusion with TM images. *IEEE Trans. Geosci. Remote Sens.* **2001**, *39*, 1533–1536. [CrossRef]
- Zurita-Milla, R.; Clevers, J.G.P.W.; Schaepman, M.E. Unmixing-based Landsat TM and MERIS FR data fusion. *IEEE Geosci. Remote Sens. Lett.* 2008, 5, 453–457. [CrossRef]
- 19. Gevaert, C.M.; García-Haro, F.J. A comparison of STARFM and an unmixing-based algorithm for Landsat and MODIS data fusion. *Remote Sens. Environ.* **2015**, *156*, 34–44. [CrossRef]
- 20. Maselli, F. Definition of spatially variable spectral endmembers by locally calibrated multivariate regression analyses. *Remote Sens. Environ.* **2001**, *75*, 29–38. [CrossRef]
- Cherchali, S.; Flouzat, G. Linear mixture modelling applied to AVHRR data for monitoring vegetation. IEEE Geoscience and Remote Sensing Symposium. *IGARSS Surf. Atmos. Remote Sens. Technol. Data Anal. Interpret.* 1994, 2, 1242–1244.
- 22. 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. [CrossRef]
- 23. 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.
- 24. Shen, H.; Wu, P.; Liu, Y.; Ai, T.; Wang, Y.; Liu, X. A spatial and temporal reflectance fusion model considering sensor observation differences. *Int. J. Remote Sens.* **2013**, *34*, 4367–4383. [CrossRef]
- 25. 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. [CrossRef]
- 26. 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. [CrossRef]

- 27. 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. [CrossRef]
- 28. Michishita, R.; Jiang, Z.; Gong, P.; Xu, B. Bi-scale analysis of multitemporal land cover fractions for wetland vegetation mapping. *ISPRS J. Photogramm. Remote Sens.* **2012**, *72*, 1–15. [CrossRef]
- 29. Roy, D.P.; Ju, J.; Lewis, P.; Schaaf, C.; Gao, F.; Hansen, M.; Lindquist, E. Multi-temporal MODIS–Landsat data fusion for relative radiometric normalization, gap filling, and prediction of Landsat data. *Remote Sens. Environ.* **2008**, *112*, 3112–3130. [CrossRef]
- 30. Dacheng, L.; Ping, T.; Changmiao, H.; Ke, Z. Spatial-temporal fusion algorithm based on an extended semi-physical model and its preliminary application. *J. Remote Sens.* **2014**, *18*, 307–319.
- 31. Song, H.; Huang, B. Spatiotemporal satellite image fusion through one-pair image learning. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 1883–1896. [CrossRef]
- 32. Huang, B.; Song, H. Spatiotemporal reflectance fusion via sparse representation. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 3707–3716. [CrossRef]
- 33. Chen, B.; Huang, B.; Xu, B. Comparison of spatiotemporal fusion models: A review. *Remote Sens.* 2015, 7, 1798–1835. [CrossRef]
- 34. Huang, B.; Song, H.; Cui, H.; Peng, J.; Xu, Z. Spatial and spectral image fusion using sparse matrix factorization. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 1693–1704. [CrossRef]
- 35. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef] [PubMed]
- 36. Yuhas, R.H.; Goetz, A.F.H.; Boardman, J.W. Discrimination among Semi-Arid Landscape Endmembers Using the Spectral Angle Mapper (SAM) Algorithm. In *JPL, Summaries of the Third Annual JPL Airborne Geoscience Workshop*; NASA: Pasadena, CA, USA, 1992; Volume 1, pp. 147–149.
- 37. Renza, D.; Martinez, E.; Arquero, A. A new approach to change detection in multispectral images by means of ERGAS index. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 76–80. [CrossRef]
- 38. Vermote, E.; Justice, C.; Claverie, M.; Franch, B. Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. *Remote Sens. Environ.* **2016**, *185*, 46–56. [CrossRef]
- 39. Mairal, J.; Bach, F.; Ponce, J.; Sapiro, G. Online learning for matrix factorization and sparse coding. *J. Mach. Learn. Res.* **2010**, *11*, 19–60.
- Mairal, J.; Bach, F.; Ponce, J.; Sapiro, G. Online dictionary learning for sparse coding. In Proceedings of the 26th Annual International Conference on Machine Learning, Montreal, QC, Canada, 14–18 June 2009; ACM: New York, NY, USA, 2009; pp. 689–696.



© 2018 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/).