Fourier-Transform-Based Unmixing Method for Fusion of Multiresolution Satellite Images

Zheng Lu, *Senior Member, IEEE*, Bozheng Shu, Xiaoqing Wang, Zheng Ge, and Lingxi Guo

***Abstract*—Many applications, such as agriculture and water monitoring, require frequent observations, as well as high spatial resolution. In practice, it is difficult to achieve both high resolution and temporal coverage for satellite sensors. Accordingly, several fusion algorithms for spatial and temporal images have been proposed. Among them, the unmixing-based methods are widely used. However, large-scale changes between categories cannot be detected accurately, as they cannot estimate the variance within individual categories, when the variances in the inhomogeneous areas are large. To solve these problems, a fusion method based on the frequency domain is proposed. It determines the reflectance of the high-resolution (HR) image using the frequency relationship between the HR image and the coarse-resolution one. It is faster and more accurate than the conventional spatial-domain-based fusion methods, as its operations are realized in the frequency domain. Both the large-scale textures and the small-scale textures can be preserved, even in the case of discern sudden or large-scale changes. Finally, experiments over simulated images and real satellite ones, using Landsat thematic mapper image and Sentinel-2 image, are carried out to demonstrate the performance of the proposed approach.**

***Index Terms*—Data fusion, image fusion, point spread function (PSF), spatial and temporal resolution, unmixing-based data fusion.**

I. Introduction

THE satellite imagery with high resolution (HR) is usually obtained within a long revisit time, while high-revisit satellite imagery has relatively low resolution [1], [2], [3], [4], [5]. Many applications of remote sensing, such as agriculture and water monitoring, require frequent observations, as well as high spatial resolution. For instance, the revisit period should ideally be within a week or shorter with spatial resolution of 10 m or higher [6], [7]. However, HR and medium-resolution satellites, such as Spot-5 [8] with the resolution of 2.5 m, and WorldView-3 [9] with the resolution of 0.3 m, have significantly longer revisit periods than those of low-resolution satellites, such as MODIS [10], [11], Landsat-8 (HR satellites with a good revisit, such as Sentinel will have a huge overhead, which may exceed the actual use value, and obtaining practically usable images is difficult due to cloud cover and other reasons). Therefore, several algorithms for the fusion of spatial and temporal images have been proposed in recent decades, with the aim of obtaining HR images with high revisit frequency. Primarily, three types of methods are used for combining spatial and temporal images: 1) weighted-function-based methods, such as the spatial and temporal adaptive reflectance fusion model (STARFM) method [12], spatial temporal adaptive algorithm for mapping reflectance change [13], and enhanced STARFM (ESTARFM) [14] 2) unmixing-based methods, such as the multisensory multiresolution technique (MMT) [15], spatial temporal data fusion approach [16], and spatial and temporal reflectance unmixing model (STRUM) [17], linear mixing growth model (LMGM) [18], spectral variability augmented sparse unmixing method (SVASU) [19], the flexible spatiotemporal data fusion method (FSDAF) [20] and FSDAF2.0 [21], reliable and adaptive spatiotemporal data fusion method (RASDF) [22], blocks-removed spatial unmixing (SU-BR) [23], geographically weighted spatial unmixing method [24], variation-based spatiotemporal data fusion method (VSDF) [25], and Fit-FC [26]; and 3) learning-base methods, including dictionary-pair-learning-based methods [27] and deep-learning-based methods [28], [29], [30], [31].

Most weighted-function-based methods assume no land cover type changes between input and prediction date. As a result, they can successfully predict pixels with changes in attributes like vegetation phenology or soil moisture, because these changes are strongly related to the changes in similar pixels selected from the input imagery. However, current methods are not effective for predicting spectral changes that are sudden or not observed in input imagery, in that the changes are not predictable from pixels that were similar in the input date. These changes include urbanization, deforestation/reforestation, wildfires, floods, and land cover transitions caused by other forces.

Dictionary-pair-learning-based methods only use statistical relationships between fine- and coarse-resolution (CR) images rather than any physical properties of remote sensing signals. Although they can better predict pixels with land cover type changes, they do not accurately maintain the shape of objects, especially when the scale difference between fine- and coarse-resolution images is large.

Manuscript received 29 August 2023; revised 16 November 2023 and 31 January 2024; accepted 2 February 2024. Date of publication 16 February 2024; date of current version 4 March 2024. This work was supported by Beijing Nova Program under Grant Z201100006820103 and in part by the National Key R&D Program of China under Grant 2023YFB3904905. *(Corresponding author: Xiaoqing Wang.)*

Zheng Lu is with the China Academy of Space Technology, Beijing 100094, China.

Bozheng Shu is with the Institute of Microelectronics of the Chinese Academy of Sciences, Beijing 100017, China.

Xiaoqing Wang and Zheng Ge are with the School of Electronic and Communication Engineering, Sun Yat-Sen University, Shenzhen 518107, China, and also with the Pengcheng Laboratory, Shenzhen 518000, China (e-mail: wangxq58@mail.sysu.edu.cn).

Lingxi Guo is with the Science and Technology on Space Physics Laboratory, Beijing 100190, China.

Digital Object Identifier 10.1109/JSTARS.2024.3365823

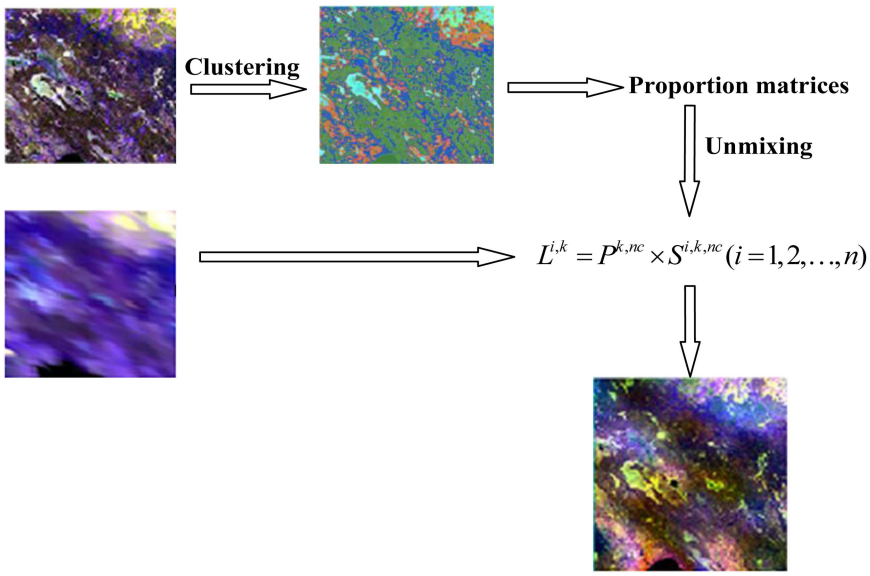

Fig. 1. Theoretical workflow of the unmixing-based data fusion algorithm.

Compared with the afore-mentioned two types of methods, the main advantage of unmixing-based methods is that they do not require high- and medium-resolution data to have corresponding spectral bands. This allows for two additional possibilities. First, unmixing-based data fusion can be used to downscale extra spectral bands and/or biophysical parameters to increase the spectral resolution of the HR datasets. Second, auxiliary datasets such land cover may supplement or replace HR imagery in the grouping of spectrally similar pixels into clusters. So, the unmixing-based methods are popular for the fusion of multiresolution images.

The first unmixing-based method is MMT [15]. It predicts the HR image by classifying the input HR one to define the endmembers at a CR, then calculates the endmember fractions of each coarse pixel. Subsequently, it unmixes the coarse pixels at the time of prediction within a moving window and assigns the reflectance to obtain the HR image.

In recent years, MMT has been modified by several studies to improve its accuracy [17], [32]. STRUM combines the unmixing method with STARFM, which has better performance when there is a sudden change. In the case that there are less previous images, FSDAF [20] and FSDAF2.0 [21] method employs one pair of high- and low-resolution images, and gains better performance than STARFM. LMGM uses the corresponding growth rate to get a better result. Unmixing-based algorithms are still the hot spot in the field of spatiotemporal fusion, and excellent algorithms and strategies, such as Fit-FC [26], SVASU [19], RASDF [22], SU-BR [23], VSDF [25], and have emerged in recent years.

However, existing unmixing-based fusion methods cannot estimate the variance within individual categories (i.e., areas in which the reflectance is fluctuating) when the variances in the inhomogeneous areas are large. It is because the calculated reflectance for each category is regarded as the mean value and the variance among each category is not considered. Therefore, large-scale changes between categories cannot be detected accurately. Conventional weighted-function-based methods, such as STARFM and ESTARFM, predict the HR image by averaging over the spatial domain, while other unmixing-based methods, such as MMT and STRUM, predict the HR image by calculating the mean reflectance of each category. On the other hand, these methods consider the CR image to be the average of the neighborhoods in the HR image, and the point spread function (PSF) [33] of the optical sensor system used for the CR image is not considered, resulting in errors, especially at the junction of two categories. Moreover, sudden or large-scale changes between categories are also difficult to predict using these methods.

To solve the above-mentioned problems, a fusion method based on the frequency domain (FMBFD) is proposed. It exploits the time-frequency relationship with respect to signal processing. The HR image can be regarded as the convolution of the PSF and the CR image in the spatial domain. Pointwise multiplication with the PSF in the Fourier domain is equivalent to convolution with the PSF in the spatial domain. However, the former process is faster and we can also solve the problem of sudden change better in frequency domain, for that product operation in the frequency domain is simpler. Therefore, the proposed method uses the frequency domain relationship between the CR image and the HR image based on the PSF.

The contributions of this work are summarized as follows.

1) The proposed algorithm is less computational than other unmixing-based algorithms.

Most unmixing-based algorithms require a sliding window convolution operation. Since the proposed algorithm is implemented in the frequency domain, it can transform the convolutional operation with large computation into the product operation in the frequency domain with much less computation. The ideal convolution window is the PSF which describes the optical sensor system's ability to resolve point sources. However,

most unmixing-based algorithm use a simple rectangle window to reduce the convolution computational, while the proposed algorithm can adopt a precise PSF window without increasing computational.

2) The real CR image information is fully retained.

Most unmixing-based algorithms are realized in the spatial domain, they predict the reflect value of each category to approximate the real CR image, and then use the predicted reflect value to reproduce the HR image under the assumption of no sudden change. While our algorithm only reproduces the high frequency part of the predicted HR image, the low frequency part still uses that of real CR image. Because the low frequency corresponding to the large-scale texture, the proposed algorithm can preserve the information of the real CR image, especially the sudden change of large-scale texture.

3) The temporal variation model of each category is more accurate.

The temporal change of each category is divided into two parts: the change in the mean value and the change in the variance within individual categories, which allows one to readily determine the variance in the pixel reflectance of the individual categories.

The rest of this article is organized as follows. Section II describes the traditional unmixing-based methods and the theoretical basis of FMBFD. Section III shows the results of validation tests performed to evaluate the effectiveness of FMBFD. These include results of simulated images and real satellite images. Finally, Section IV concludes this article.

II. Algorithm Theory

A. Traditional Unmixing-Based Data Fusion

The traditional unmixing-based image fusion applies four steps to solve the linear mixing model [1], as shown in Fig. 1.

First, the HR image is classified into unsupervised classes using unsupervised method, such as K-means, e.g., five number class (nc) values were used as 10, 20, 30, 40, and 60.

Second, a sliding window is applied in CR image covering $k \times k$ pixels, the classification results in the first step are used to get a proportion matrix, these matrices contain the proportions of each of the nc classes that fall within each CR pixel.

Third, the spectral information of all classes is unmixed using the proportion matrices and their corresponding CR radiance values (each band is solved independently). Each pixel only provides one equation so it is necessary that $k \times k$ is bigger than nc, and then we could solve the equation set.

Finally, each CR pixel could be replaced by corresponding unmixed HR pixels, and we could get the predict HR image. This notation in third step should be interpreted as follows:

$$L^{i,k} = P^{k,nc} \times S^{i,k,nc} \, (i = 1, 2, \ldots, n) \tag{1}$$

where $L^{i,k}$ denotes a $k^2 \times 1$ vector that contains the CR pixel in band i; $P^{k,nc}$ represents a $k^2 \times nc$ matrix that save the proportion of each HR classes that fall inside each of the CR pixel. $S^{i,k,nc}$ is what we want to get, and it is the $nc \times 1$ unknown vector of unmixed spectral information for each class in band i. We could solve this equation set using least mean square (LMS) or others. Considering all the bands and windows, we can get the final predict image whose relevant endmember is assigned to each pixel at HR scale.

B. FMBFD

It is well-known that the frequency-domain equivalent of an image can be obtained by applying the Fourier transform (FT). The low-frequency part of the image represents the large-scale outline, while its high-frequency part represents the small-scale textures. The frequency-domain relationship between an HR image and the corresponding CR image can be used to fuse them. A frequency-domain fusion method based on the FT (FMBFD) for the fusion of multiresolution images is proposed, with a flowchart illustrated in Fig. 2.

1) Frequency Domain Decomposition Model for the HR Image: First, the classification of the original HR image is performed. For this, we use an unsupervised classification method called the k-means clustering. Supervised methods such as the support vector machine method can also be used for this purpose, but it requires us to know the supervision information in advance; k-means clustering is easier to get a result. Assume that there are categories of the HR image. Thus, each category should have a distribution. Hence, each category of the HR image will have a mean reflectance value. The pixel reflectance for each category can be described as the sum of the mean reflectance value and the residual value, where the residual value is normalized by the mean reflectance, which is calculated as given in (2). Therefore, the dividend, which consists of the mean part and the residual part, can be determined. The 2-D frequency-domain equivalent of the mean and residual parts of the distribution can be obtained using the FT. The distribution of the mean and residual parts of each classification result can be represented as follows:

$$\begin{cases} f_{ci}(x,y) = 1, & \text{when} \quad C(x,y) = c \\ f_{ci}(x,y) = 0, & \text{when} \quad C(x,y) \neq c \\ f_{cd}(x,y) = \frac{R_c(x,y) - \bar{R}_c}{\bar{R}_c}, & \text{when} \quad C(x,y) = c \\ f_{cd}(x,y) = 0, & \text{when} \quad C(x,y) \neq c \\ f_H(x,y) = \sum_{c=1}^{K} \bar{R}_c \left[f_{ci}(x,y) + f_{cd}(x,y) \right] \end{cases} \tag{2}$$

where (x, y) denotes the row and column coordinates of the image, $R_c(x, y)$ is the reflectance of the pixel belonging to the cth category at position (x, y) in the HR image, and $\bar{R}_c$ represents the mean value of $R_c(x, y)$.

Furthermore, $f_{ci}(x, y)$ is the category distribution function, which is one when the pixel belongs to the cth category and zero when the pixel does not belong to the cth category. $f_{cd}(x, y)$ represents the residual part for the cth category and can be obtained by subtracting the mean part $f_{ci}(x, y)$ and subsequent normalization using $\bar{R}_c$. $f_H(x, y)$ represents the HR image, which can be expressed as the sum of the K categories's mean part and the residual part multiplied by their mean reflectance value. Fig. 3 is an example to illustrate the meaning of (2).

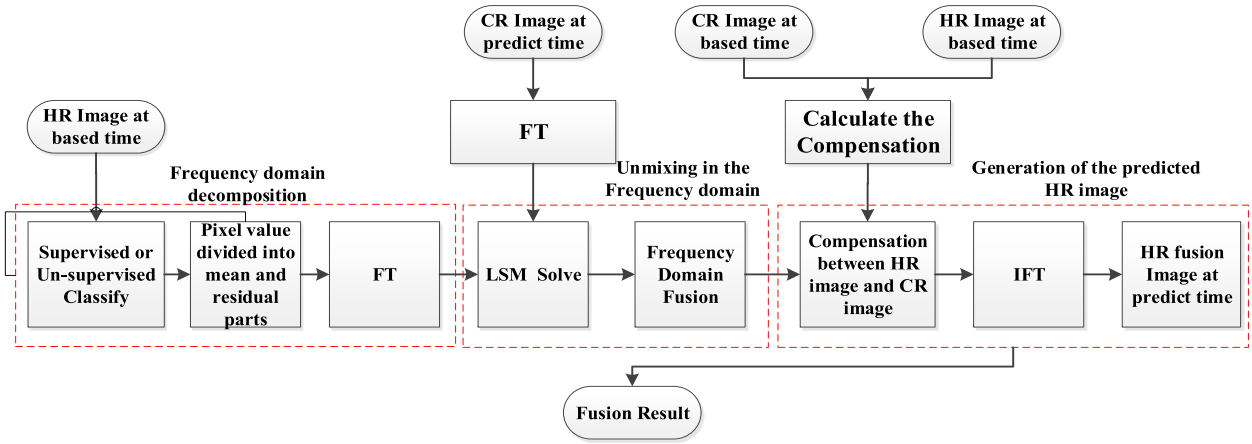

Fig. 2. Flow chart of FMBFD.

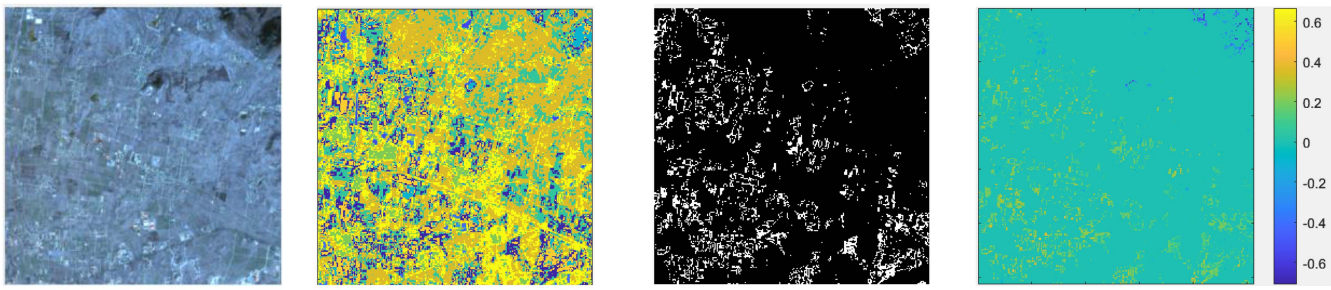

Fig. 3. Example for (2). (a) High resolution image. (b) Classfication map (K = 10). (c) Distribution of 1st class ($f_{1\mathrm{i}}$). (d) Residual part of 1st class ($f_{1\mathrm{d}}$).

The frequency-domain equivalent of the HR image can be obtained using (3). Before performing the fusion procedure, the CR image is resampled at the same spatial resolution as that of the HR image using bilinear or cubic convolution interpolation.

The frequency-domain relationship between the HR image and each category distribution function is shown as follows:

$$\begin{cases} F_{ci}\left(k_x, k_y\right) = FT\left[f_{ci}\left(x, y\right)\right] \\ F_{cd}\left(k_x, k_y\right) = FT\left[f_{cd}\left(x, y\right)\right] \\ F_H\left(k_x, k_y\right) = \sum_{c=1}^{K} \bar{R}_c \left[F_{ci}\left(k_x, k_y\right) + F_{cd}\left(k_x, k_y\right)\right] \end{cases} \tag{3}$$

where (k_x, k_y) is the wavenumber in the x- and y-directions in the frequency domain. Furthermore, $F_{ci}(k_x, k_y)$ is the FT of $f_{ci}(x, y)$ while $F_{cd}(k_x, k_y)$ is the FT of $f_{cd}(x, y)(FT[]$, i.e., the FT operator, which transforms the spatial-domain image into the frequency domain). $F_H(k_x, k_y)$ represents the FT of the HR image, which can be expressed as the sum of the FTs of the K categories's mean part and the residual part multiplied by their mean reflectance value.

2) Unmixing Method in the Frequency Domain: The PSF describes the optical sensor system's ability to resolve point sources and PSF is usually completely determined by the imaging system, and the entire image can be explained by obtaining the optical parameters of the system. This process is usually formulated by a convolution equation, it is representative of the relationship between the reflectance of HR and CR images, which can be represented as follows:

$$\begin{cases} f_{psf}\left(x, y\right) \otimes f_H\left(x, y\right) = f_C\left(x, y\right) \\ F_{psf}\left(k_x, k_y\right) \otimes F_H\left(k_x, k_y\right) = F_C\left(k_x, k_y\right) \end{cases} \tag{4}$$

where $\otimes$ is the convolution operator. The CR image can be represented as the convolution of the PSF and the HR image in the spatial domain or the product of the PSF and the HR image in frequency domain. Here, $F_{psf}(k_x, k_y)$ is the PSF, which, as stated earlier, describes the relationship between the HR and CR images while $f_H(x, y)$ and $f_c(x, y)$ are the HR and CR images, respectively. Furthermore, $F_H(k_x, k_y)$ and $F_C(k_x, k_y)$ are the FTs of $f_H(x, y)$ and $f_c(x, y)$, respectively. Since the CR image is resampled to be the same resolution as the HR image, the frequency-domain ranges of the CR and the HR images are the same. It is easier to determine the product of the PSF and the HR image than to perform the convolution operation in the spatial domain. Hence, the former is used to describe the complex convolution relationship between the CR and HR images.

Next, selecting the CR image at the time of prediction and the HR image at the base time in the same area, the following

equation is formulated:

$$\left[\sum_{c=1}^{K} F_{ci}\left(k_x,k_y\right)\cdot B_{ci}+\frac{\sum_{c=1}^{K} F_{cd}\left(k_x,k_y\right)\cdot B_{cd}}{\left(|k_x|\le C_{kx}\ \ |k_y|\le C_{ky}\right)}\right]\cdot F_{psf}\left(k_x,k_y\right)=\frac{F_C{}^p\left(k_x,k_y\right)}{\left(|k_x|\le C_{kx}\quad |k_y|\le C_{ky}\right)} \tag{5}$$

where $F_C{}^p(k_x,k_y)$ is the FT of the CR image at the time of prediction, B_{ci} is the mean reflectance of the cth category, and B_{cd} is the scale factor which effects the residual part of the reflectance of the cth category at the time of prediction.

Furthermore, $(|k_x|\le C_{kx}\ \ |k_y|\le C_{ky})$ implies that the frequency bin corresponds to the low-frequency part of the resampled CR image which means this equation is solved in the low-frequency to get B_{ci} and B_{cd} for that CR image almost only have low-frequency part, C_{kx} and C_{ky} are thresholds corresponding to the low-frequency area of the resampled CR image, which can be calculated using the following equation:

$$\begin{cases} C_{kx}=\frac{1}{\rho_{cx}} \\ C_{ky}=\frac{1}{\rho_{cy}} \end{cases} \tag{6}$$

where ρ_{cx} and ρ_{cy} are spatial resolutions of the CR image in the x- and y-directions, respectively, before resampling. Equation (5) implies that changes in the mean and residual parts of the reflectance for each category could be different. For example, during the growth of homologous crops, some could grow better than others at the base time, and the growth situation may remain the same at the time of prediction. However, the change percentage of the pixel reflectance compared to the base time would differ. The change in the variance of the pixel reflectance can be estimated by computing the residual part of the reflectance for each category. Both the change in the mean reflectance and the variance between the different categories are determined. The fusion process should reserve the variance between the various categories, as well as the change in the mean pixel reflectance of each category.

Finally, B_{ci} and B_{cd} for each category are determined using the LMS, as shown in (7). The low-frequency bins of the CR image are used for the LMS calculation to determine reflectances B_{ci} and B_{cd}.

To balance the importance of the mean part in the LMS, an undetermined weight factor, β, is introduced, since the variance within each category is smaller than the variance between the different categories. β is generally larger than 1, and can be varied depending on the HR image.

$$B_{ci},\beta,\ B_{cd}=\operatorname{argmin}\sum_{|k_y|\le C_{ky}}\sum_{|k_x|\le C_{kx}}\left|\left[\sum_{c=1}^{K} F_{ci}\left(k_x,k_y\right)\cdot B_{ci}\cdot\beta+\sum_{c=1}^{K} F_{cd}\left(k_x,k_y\right)\cdot B_{cd}\right]\times F_{psf}\left(k_x,k_y\right)-F_C{}^p\left(k_x,k_y\right)\right|^2 \tag{7}$$

where β is the weight factor for striking a balance between the mean part and the residual one. After determining B_{ci} and B_{cd} of each category and the balancing factor β in the low-frequency part, the fusion image based on the fusion of the different categories and the CR image at the time of prediction can be represented using (8) and (9).

$$F_{fit}^{p}\left(k_x,k_y\right)=\sum_{c=1}^{K} F_{ci}\left(k_x,k_y\right)\cdot B_{ci}\cdot\beta+\sum_{c=1}^{K} F_{cd}\left(k_x,k_y\right)\cdot B_{cd}. \tag{8}$$

3) Generation of the Predicted HR Image: The predicted HR image can be generated by the following equation:

$$f_{ht}{}^p\left(x,y\right)=IFT\left\{F_{fit}^{p}\left(k_x,k_y\right)\cdot\left(1-F_{psf}\left(k_x,k_y\right)\right)+F_C{}^p\left(k_x,k_y\right)\right\} \tag{9}$$

where $F_{fit}^{p}(k_x,k_y)$ is the fitted HR image in the frequency domain which has both low-frequency and high-frequency part, $f_{ht}{}^p(x,y)$ is the predicted HR image, and $IFT[]$ represents the inverse FT (IFT) operator.

Generally, $F_{psf}(k_x,k_y)$ is a bell-shaped function. In the low-frequency region, it is very close to 1 (i.e., $|k_x|\le C_{kx}$ and $|k_y|\le C_{ky}$), while in the high-frequency region it is very close to 0 (i.e., $|k_x|\ge C_{kx}$ or $|k_y|\ge C_{ky}$). Low frequency corresponds to smooth variations and constitutes the base of an image while high frequency presents the edge information which gives the detailed information in the image. We add up the true CR image and the high-part of $F_{fit}^{p}(k_x,k_y)$ to get $f_{ht}{}^p(x,y)$. Therefore, (7) implies that the predicted result $f_{ht}{}^p(x,y)$ includes the large-scale textures in the CR image at the time of prediction and the small-scale ones in the fitted HR image. Thus, if there are sudden large-scale changes at the time of prediction, such as a wildfire, deforestation, reforestation, or floods, they will be captured within $f_{ht}{}^p(x,y)$.

To compensate for the system error between the CR and HR optical sensors, additional compensation is performed. The compensation part is calculated by comparing the predicted result and the HR image both at the base time, as shown in the following equation:

$$F_{comp}\left(k_x,k_y\right)=F_h^b\ \left(k_x,k_y\right)-F_{ht}^{b}\left(k_x,k_y\right) \tag{10}$$

where $F_h^b(k_x,k_y)$ is the frequency-domain equivalent of the HR image at the base time, $F_{ht}^{b}(k_x,k_y)$ is the frequency-domain equivalent of the image predicted using (4) based on CR image at the base time (here, the right side of (5) is not $F_c^p(k_x,k_y)$ but the FT of the CR image at the base time), and $F_{comp}(k_x,k_y)$ is the additional compensation in the frequency domain. The compensation is performed in the frequency domain before IFT as follows:

$$f_{hp}^{p}\left(x,y\right)=IFT\left\{F_{ht}^{p}\left(k_x,k_y\right)+F_{comp}\left(k_x,k_y\right)\right\} \tag{11}$$

TABLE I
PARAMETERS USED FOR SIMULATED IMAGES

Parameter	HR image at t_0	HR image at t_1
Mean reflectance of circular object	3500	1500
Mean radius of circular object	4000 m	4000 m
Mean reflectance of square object	1500	2500
Side of square object	3500 m	3500 m
Background noise	30	30
Categories in each object	3	3
Variance in standard deviation in each category	400	600
Reflectance of square object within circular object	N/A	2000
Side of square object within circular object	N/A	1500 m
Scene size	36000 m×36000 m	36000 m×36000 m
Pixel size	30	30

where $f_{hp}^{p}(x, y)$ is the predicted HR image after compensating for the system error between different optical sensors.

If the predicted HR image exhibits the same band features as the HR image, the compensation is necessary to reduce the prediction error caused by different sensors. However, to obtain a super-resolution image, one only needs to improve the spatial resolution of the CR image, and does not need to recover the band of the HR image which also does not exist. Therefore, compensation is not needed in this case.

III. EVALUATION OF PROPOSED METHOD

A. Fusion of Simulated High Spatial and High Temporal Images

1) Simulation Setup and the Evaluation Index: An HR image with base time t_0 and prediction time t_1 were generated using the simulation parameters listed in Table I.

The HR image contained two objects, a circle with a radius of 4000 m and a square with a side of 3500 m. Gaussian white noise with a standard derivation of 30 was added to both generated images. The reflectance of the circular object changed from 3500 to 1500 while that of the square object changed from 1500 to 2500 at t_1. Next, a small square object was introduced within the circular one in the HR image at the prediction time, t_1; the side of this new square was 1500 m and its reflectance was 800. The resolution of the HR images was 30 m. The CR images at t_0 and t_1 were obtained by convoluting the HR images at t_0 and t_1 with the PSF, which was a 2-D Gaussian function with a standard deviation of 500 m. The resolution of the CR images was 240 m. To determine the accuracy of the predictions for each category using the proposed algorithm, a fluctuation with a standard deviation of 400 was introduced in the reflectance of each object in HR image at t_0. The distribution pattern of the fluctuation in the reflectance of objects at time t_1 was the same as that at t_0; however, the standard deviation changed to 600. The simulated HR image and CR image at t_0, and HR image and CR image at t_1, are shown in Figs. 4 and 5, respectively.

By applying the proposed algorithm to the HR image at t_0 and the CR image at t_1, the HR fusion image could be obtained.

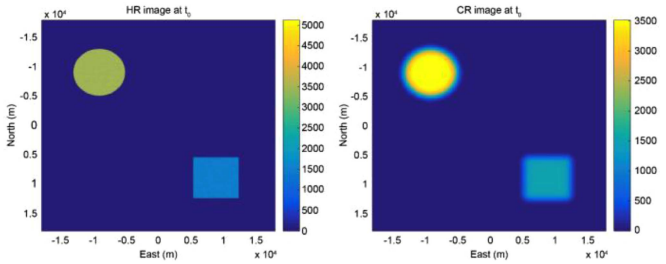

Fig. 4. (a) Simulated HR image t_0. (b) Simulated CR image at t_0.

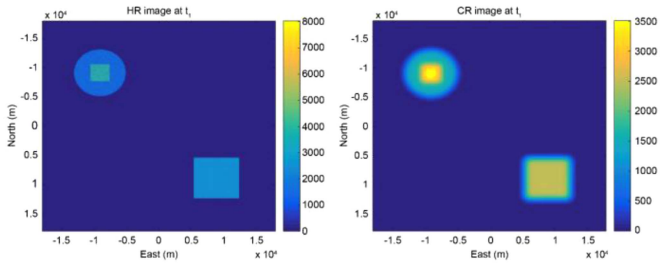

Fig. 5. (a) Simulated HR image at t_1. (b) Simulated CR image at t_1.

The correlation coefficient (CC), average absolute difference (AAD) [34], and root mean square error (RMSE) between the HR image and the HR fusion image were calculated to evaluate the proposed algorithm and compare its performance with those of other algorithms. In addition, the structure similarity (SSIM) [35] index was used to assess the overall similarity between the fusion images and the actual ones. The closer the RMSE and AAD are to zero, the more similar the fusion images are to the actual ones. Similarly, the closer the CC and SSIM were to one, the more similar the fusion images are to the actual ones. The expressions for calculating the AAD, RMSE, and SSIM are as follows:

$$\begin{aligned} AAD &= \Delta\left|\bar{R}\left(x, y\right)\right| \\ RMSE &= \sqrt{\sum\left(\Delta\bar{R}\left(x, y\right)\right)^2} \\ SSIM &= \frac{\left(2\mu_x\mu_y + C_1\right)\left(2\sigma_{XY} + C_2\right)}{\left(\mu_X^2 + \mu_Y^2 + C_1\right)\left(\sigma_X + \sigma_Y + C_2\right)} \end{aligned} \tag{12}$$

where $\Delta R(x, y)$ is the difference in the pixel reflectances of the HR fusion image and the HR image at position (x, y) at the time of prediction. Furthermore, μ_X and μ_Y are the mean pixel reflectances of the original HR image and the HR fusion image, respectively; σ_X and σ_Y are the variances of the original HR image and the HR fusion image, respectively; and σ_{XY} is the covariance between the HR image and the HR fusion image. Finally, C_1 and C_2 are very small constants, which were used to ensure that the fraction does not result in a singular value.

To highlight the advantages of the proposed algorithm, five different simulations were performed using the algorithm, in the five simulations, five different improvement parameters in the method are considered as the following.

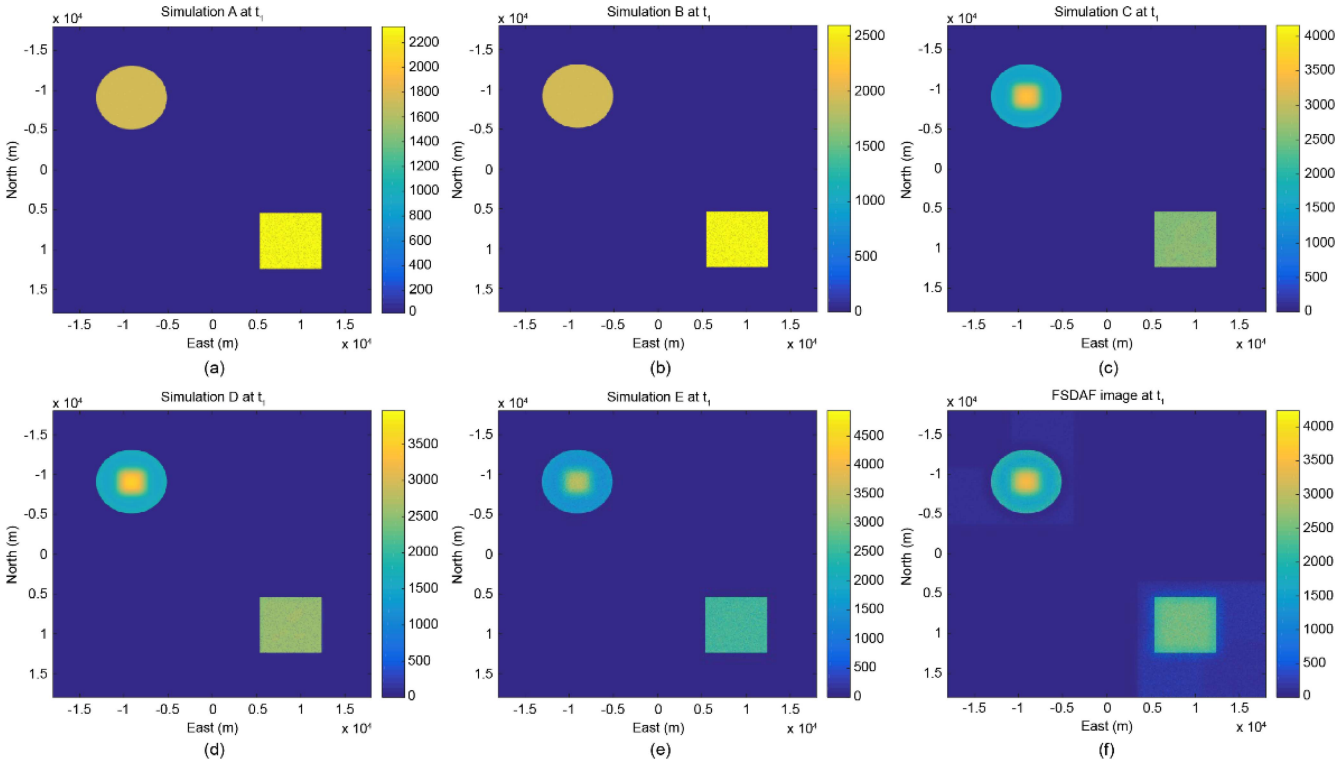

Fig. 6. Fusion results. (a) Simulation A. (b) Simulation B. (c) Simulation C. (d) Simulation D. (e) Simulation E. (f) Fusion results for FSDAF.

Simulation A: Factors, including PSF, the low-frequency part of CR image, the variance within each category, and compensation between HR and CR images were not considered.

Simulation B: The PSF was considered. However, the low-frequency part of CR image was not considered, the variance within each category was not considered, and compensation between the HR and CR images was not performed.

Simulation C: The PSF was considered and so was the low-frequency part of the CR image. However, the variance within each category was not considered, and compensation between the HR and CR images was not performed.

Simulation D: The PSF was considered along with the low-frequency part of the CR image and the variance within each category. However, compensation between the HR and CR images was not performed.

Simulation E: The PSF was considered along with the low-frequency part of the CR image and the variance within each category. Furthermore, compensation was performed between the HR and CR images.

Simulation E employed the entire FMBFD algorithm. In addition, the performance of FMBFD was compared with that of FSDAF, which is one of the results of a recent study about unmixing-method, which used the same classification results as FMBFD. It was to determine whether the classification process affects fusion results. Fusion results and errors of Simulation A $\sim$ E and FSDAF are shown in Figs. 6 and 7, respectively.

2) Simulation Experiment Result and Analysis: FSDAF mainly uses five steps to fuse MODIS images and Landsat images. First, the Landsat image endmembers are extracted and the abundance is calculated. The temporal and spatial results are predicted assuming that the type of ground features have not changed, and then the residuals and distribution are calculated, and the fusion result is finally obtained. There are similarities with the FMBFD algorithm, so this article compares the two algorithms through experiments. The absolute error between the fusion image and the HR image at t_1 is the statistic over all pixels. The histogram of the errors for FMBFD and FDSAF is shown in Fig. 8. It can be seen that FMBFD resulted mostly in small errors, in contrast to FDSAF.

Figs. 6 and 7 show that the performance of FMBFD with respect to predicting the shape, reflection, and large changes was better than that of FSDAF. The results of the two methods for simulations A–E were compared; these results are shown in Table II. It is shown that the introduction of the PSF improved the fusion performance slightly. On the other hand, the use of the low-frequency part of the CR image significantly affected the prediction results when the shape of the object was changed.

TABLE II
RESULTS OF TWO METHODS FOR SIMULATIONS A–E

Simulation	CC	AAD	RMSE	SSIM
A	0.9241	62.15	245.29	0.9098
B	0.9241	61.75	241.29	0.9212
C	0.9478	55.04	201.06	0.9453
D	0.9486	54.98	199.57	0.9460
E	0.9790	24.78	128.33	0.9789
FSDAF	0.9602	55.07	177.846	0.9493

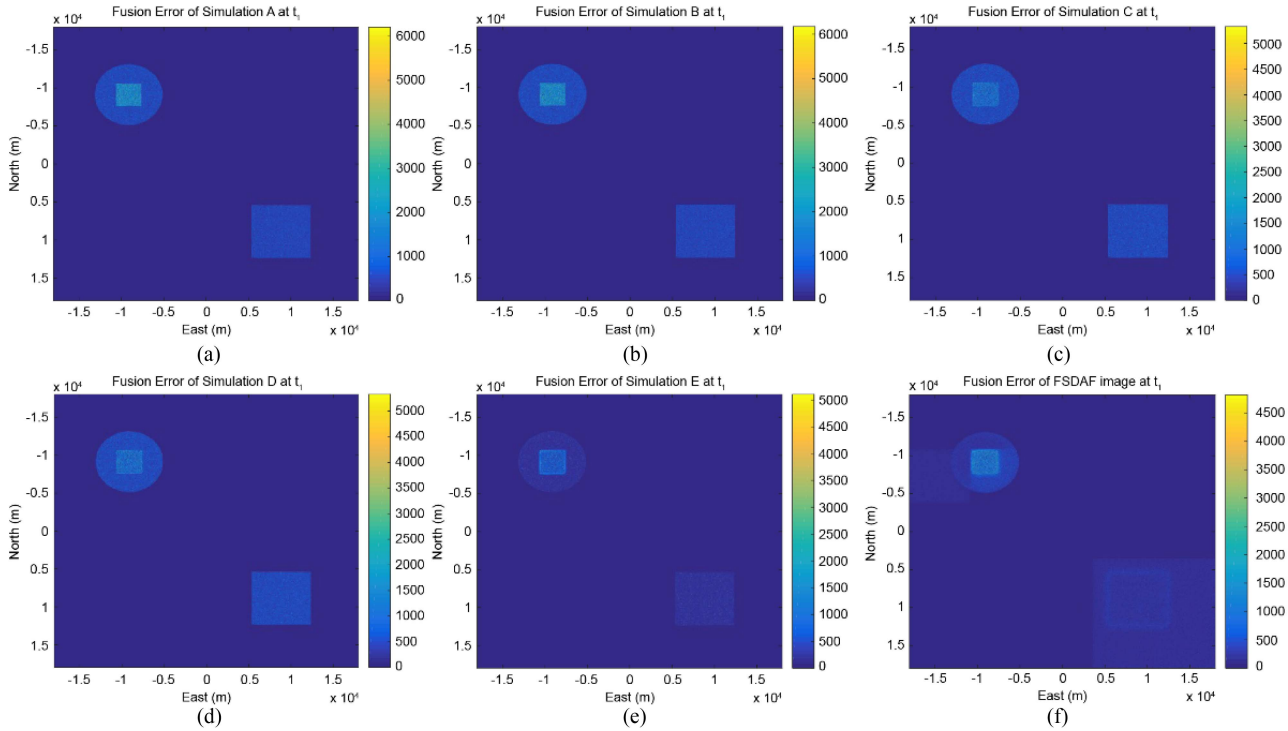

Fig. 7. Fusion errors. (a) Simulation A. (b) Simulation B. (c) Simulation C. (d) Simulation D. (e) Simulation E. (f) Fusion error for FSDAF.

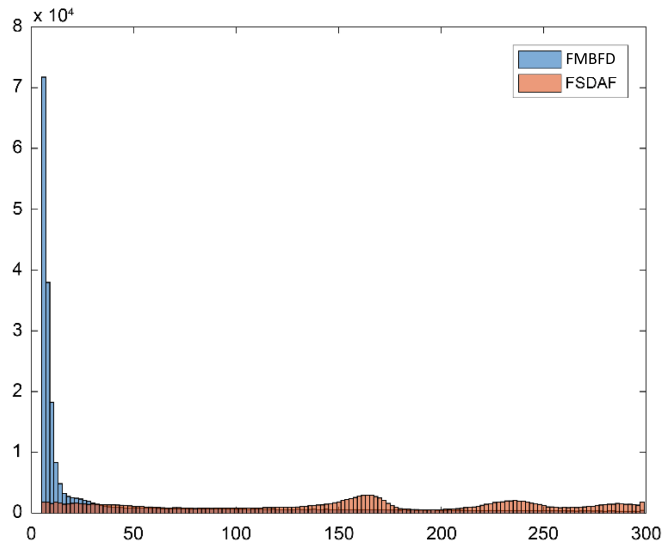

Fig. 8. Histogram of absolute error between predicted image and original HR image for FMBFD and FSDAF.

Furthermore, whether compensation for the HR and CR images was performed or not also affected the prediction results. FMBFD showed better performance than that of FSDAF during simulation E, wherein the former was used in its complete form.

To highlight the difference between the two algorithms, the simulation where a rectangular object was placed within the circular one was performed, which was an enlargement of the circular area in the upper left corner of Fig. 6.

TABLE III
COMPARISON OF PERFORMANCES OF FSDAF AND FMBFD

Simulation	CC	AAD	RMSE	SSIM
A	0.6954	513.92	885.36	0.6339
B	0.6959	536.35	880.38	0.6625
C	0.8333	423.13	677.25	0.8158
D	0.8337	424.29	676.73	0.8163
E	0.9260	243.74	470.32	0.9150
FSDAF	0.8816	369.11	591.65	0.8664

The CC, AAD, RMSE, and SSIM values corresponding to the small areas were calculated, as shown in Table III. The use of the low-frequency part of the CR image, as well as allowing for compensation between the HR and CR images significantly improved the prediction results, as shown in Figs. 9 and 10.

It can be seen that the greater the number of factors considered during the simulation, the better the performance of the proposed algorithm was. The performance of simulation C, during which the low-frequency part of the CR image at the time of prediction was considered, was markedly better than that of simulation B. Thus, the low-frequency part of the CR image at the time of prediction allowed the sharp changes in the circular object to be predicted. The evaluation indices for simulation C were better than those for simulation B. Furthermore, the proposed method exhibited even better performance during simulation E, as reflected in the SSIM, AAD, RMSE, and CC values. Fig. 9 shows the image predicted using FMBFD during simulation E was more similar to the original HR image that the one predicted using FSDAF.

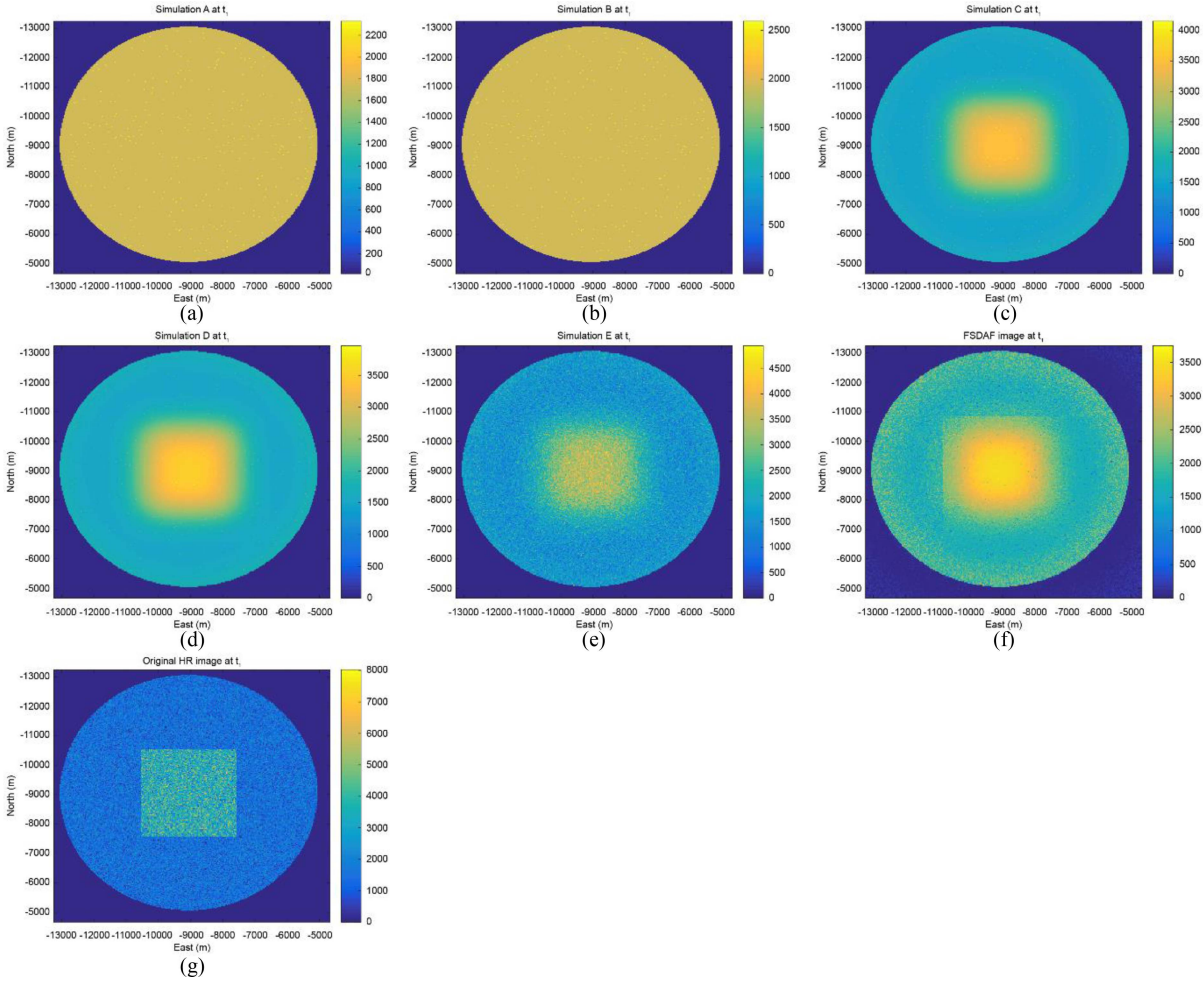

Fig. 9. Fusion result obtained using the proposed method for circular object for (a) simulation A, (b) simulation B, (c) simulation C, (d) simulation D, and (e) simulation E. (f) Fusion result obtained using FSDAF. (g) Original HR image.

B. *Fusion of Real Satellite Spatial and Temporal Images*

1) Fusion Experiment for Landsat and Sentinel-2: To confirm the suitability of the proposed algorithm, a pair of real satellite images was analyzed using the algorithm. A Landsat thematic mapper (TM) image was fused with a Sentinel-2 image because the resolution of TM images is 30 m whereas that of Sentinel-2 images is 10 m.

The images corresponded to the Yellow River (Shanxi province, China) and contain farmland, forest regions, hills, and urban areas. It was used to evaluate the performance of the FMBFD algorithm with respect to complex scenes. The longitude and latitude of the center area were 34°54'2.405"E and 110°14'48.846"N. Furthermore, these images can be employed to evaluate the algorithm's performance for relatively large areas as well as to test whether it is capable of fusing images with different types of areas.

Since FSDAF needs a pair of HR and CR images at the base time, TM and Sentinel-2 image corresponding to the same time were used. The dates on which the TM images were taken were February 24th, 2018 and September 3rd, 2018 (at approximately 12 o'clock, GMT+8), as shown in Figs. 11(a) and 12(a). The corresponding Sentinel-2 images are shown in Figs. 11(b) and 12(b), respectively.

The same classification results were used for both algorithms, to eliminate the effect of the classification process. The number of the categories was 10, and the classifier method used was k-means clustering. The prediction results for the two algorithms in the case of the Sentinel-2 image obtained on September 3rd, 2018 are shown in Fig. 12.

The areas within the red rectangle and the blue circle are illustrated in Fig. 14 for details. Fig. 14(a) and (b) show the prediction results for FSDAF and FMBFD, respectively, for the area within the red rectangle in Fig. 13(a), while the corresponding area in the real satellite image is shown in Fig. 14(c). It can be seen that the FMBFD image shows more details and could predict the shape of the area in the real satellite image with greater accuracy. Similarly, Fig. 14(d) and (e) show the FSDAF and FMBFD prediction results, respectively, for the area within the blue circle in Fig. 12. In this case, FMBFD resulted in a more accurate prediction of the object shape and reflectance than FSDAF did.

The CC, AAD, RMSE, and SSIM values for the two images predicted using the FMBFD and FSDAF algorithms are listed in

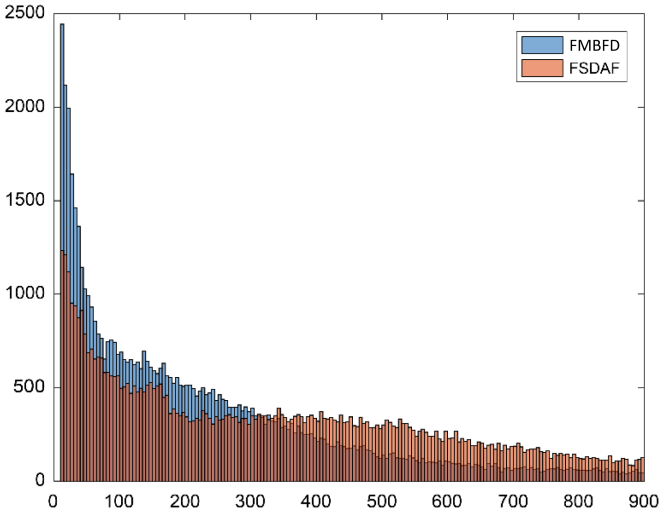

Fig. 10. Histogram of absolute error between predicted image and original HR image of circular object at for FMBFD and FSDAF.

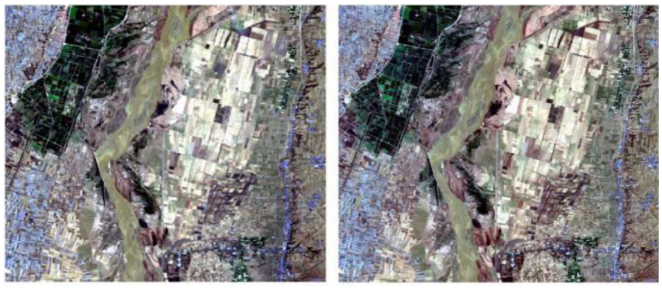

Fig. 11. (a) TM image and (b) Sentinel-2 image at t_0 on February 24, 2018.

Fig. 12. (a) TM image and (b) Sentinel-2 image at t_1 on September 3, 2018.

Fig. 13. Comparison of images predicted at t_1 using (a) FSDAF. (b) FMBFD.

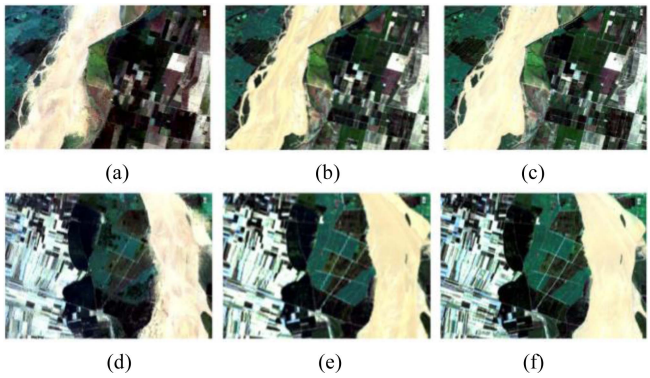

Fig. 14. Comparison of magnified versions of images predicted by (a), (d) FSDAF and (b), (e) FMBFD and (c), (f) original Sentinel-2 image on September 3, 2018 (upper-row images show area within red rectangle in Fig. 13 while lower-row images show area within blue circle in Fig. 13).

TABLE IV
FUSION RESULTS USING LANDSAT AND SENTINEL-2

Band	Method	CC	AAD	RMSE	SSIM
Blue	FSDAF	0.8274	134.44	172.90	0.8156
	FMBFD	0.9334	122.64	151.54	0.8922
Green	FSDAF	0.8388	212.67	258.58	0.8188
	FMBFD	0.9189	130.35	162.32	0.9141
Red	FSDAF	0.9005	228.61	295.56	0.8772
	FMBFD	0.9470	131.75	184.36	0.9430
NIR	FSDAF	0.9220	287.71	400.65	0.9191
	FMBFD	0.9616	204.55	281.95	0.9610

Table IV. The absolute value of the error between the predicted image and the actual one for all four bands (i.e., NIR, R, G, and B) was also determined; the result is shown in Fig. 15. It can be seen from the histogram in the figure that the FMBFD algorithm resulted in a greater number of smaller errors and fewer larger errors than FSDAF did. Hence, FMBFD performed better over all the bands, as shown in Table IV. Figs. 16–19 show the scatter plots for predicted and actual values corresponding to the satellite image for four bands. It can be seen that the values predicted by FMBFD are closer to the actual values than those predicted by FSDAF.

2) Fusion Experiment for the Scene With a Sudden Change: In order to further compare the performance of FMBFD, and test its performance when there is a sudden change, we choose a better fusion algorithm ESTARFM which needs at least two pairs of low- and high-resolution images on prior and posterior dates and one CR image on the predicted date to compare with FMBFD considering all the improvement point. It has a good prediction effect when there are sudden or large changes due to the prior and posterior information.

In this part, a Landsat-5 image was fused with an image from Google map, their resolution is 30 and 2.5 m. The images corresponding to an area that lies along A city in Shantou, China, for it has a sudden change in a short time. The longitude and latitude of the center of the imaged area were 116°31'8.25"E and 23°16'32.13"N. We used two pairs of HR and CR images

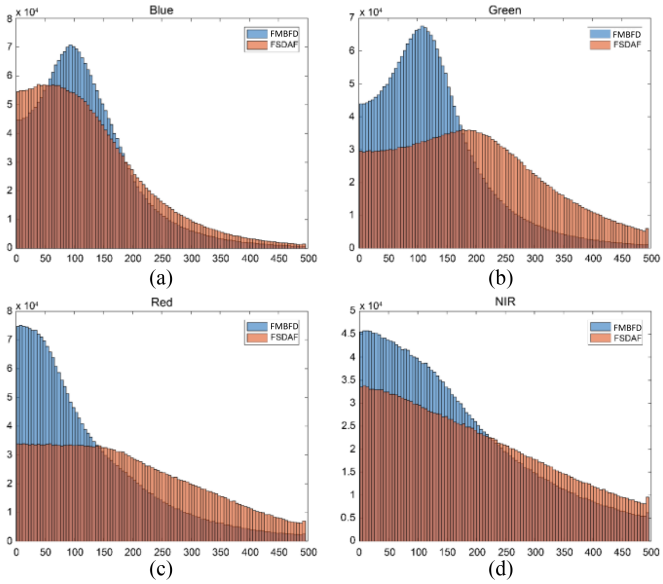

Fig. 15. Distributions of absolute value of error between predicted image and true satellite image for FMBFD and FSDAF. (a) Blue band. (b) Green band. (c) Red band. (d) NIR band.

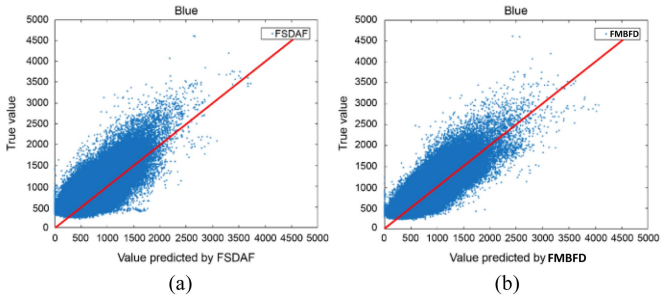

Fig. 16. Scatter plots of actual and predicted values for blue band (darker color indicates higher density of points; line is 1:1 line). (a) FSDAF. (b) FMBFD.

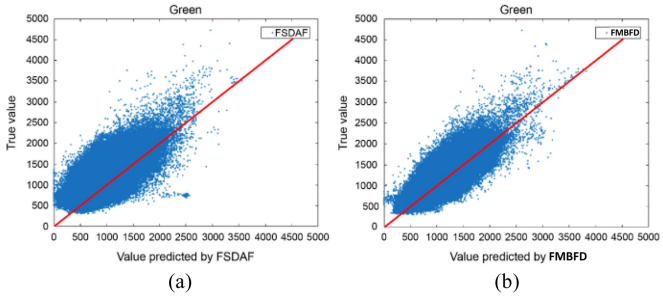

Fig. 17. Scatter plots of actual and predicted values for green band (darker color indicates higher density of points; line is 1:1 line). (a) FSDAF. (b) FMBFD.

at the base time. The prior time is January 7th, 2009, and the posterior time is November 4th, 2012 (at approximately 12 o'clock, GMT+8), the corresponding HR image is shown in Fig. 20(f) and (e) (which is also the HR image of FMBFD), we can see there is a sudden change inside the red box. The predict time is November 28th, 2009, the corresponding HR and CR

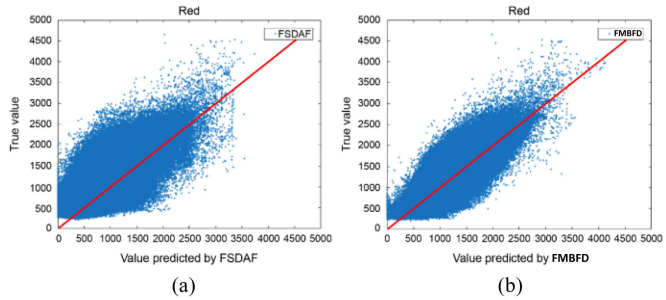

Fig. 18. Scatter plots of actual and predicted values for red band (darker color indicates higher density of points; line is 1:1 line). (a) FSDAF. (b) FMBFD.

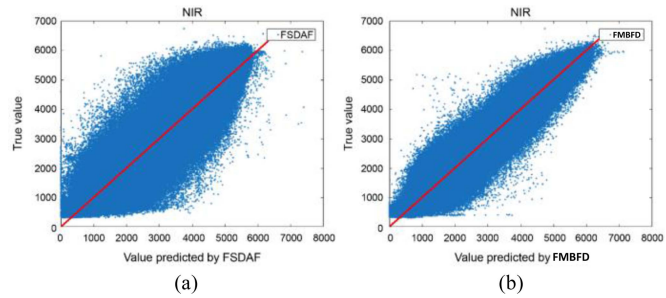

Fig. 19. Scatter plots of actual and predicted values for NIR band (darker color indicates higher density of points; line is 1:1 line). (a) FSDAF. (b) FMBFD.

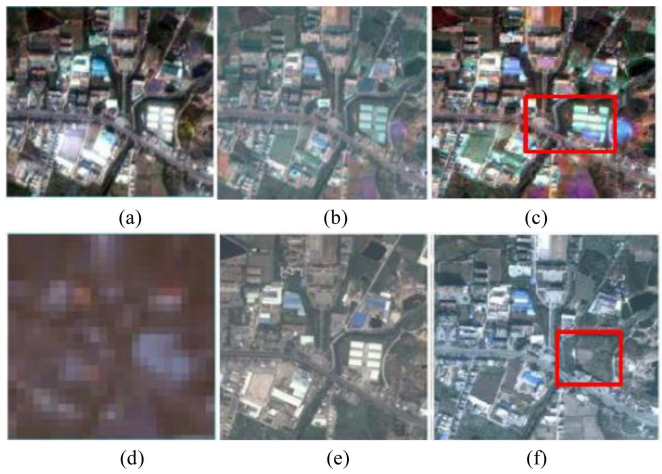

Fig. 20. Base time and fused image. (a) Predict time HR image on November 28, 2009. (b) Fused result used ESTARFM algorithm. (c) Fused result used FMBFD algorithm. (d) Predict time CR image on November 28, 2009. (e) Posterior HR image for ESTARFM on November 4,2012. (f) Prior HR image for ESTARFM on January 7, 2009 (base time of FMBFD).

images are shown in Fig. 19(a) and (d), and the fusion result using FMBFD and ESTARFM are shown in Fig. 20(b) and (c).

As we can see in Fig. 20(c), the FMBFD method which only used one pair image Fig. 20(d) and (f) can predict a sudden change shown in the red frame correctly as well as ESTARFM.

Then, we test the performance of FSDAF and LMGM using the afore-mentioned image (base time image is Fig. 20(f)), and can get the result as Fig. 21. The result is the same as that of the previous experiment between FSDAF and FMBFD, as shown

(a) (b)

Fig. 21. Fusion results. (a) FSDAF. (b) LMGM.

TABLE V
FUSION RESULTS USING REAL SATELLITE IMAGES

Band	Method	CC	AAD	RMSE	SSIM
Blue	ESTARFM	0.8619	41.2279	45.4420	0.8191
	FMBFD	0.8808	40.9915	44.6695	0.8191
Green	ESTARFM	0.9117	62.0354	63.9943	0.8248
	FMBFD	0.9474	62.4871	63.8946	0.8448
Red	ESTARFM	0.8405	52.1081	62.1171	0.7574
	FMBFD	0.8939	58.0626	61.3981	0.7774

in Fig. 21(a), the sudden buildings in the red box cannot be predicted correctly, and the result is very close to the base time HR image, because it only uses two prior image which is also one of the disadvantages of FMBFD, but FMBFD can extract information from the two images better, so it can have a better result. Fusion result using FSDAF is worse than ESTARFM and FMBFD. As shown in Fig. 21(b), LMGM is not able to predict the sudden change, but it still has better results than the original algorithm in the case of differences within categories. LMGM uses unmixing method to estimate the growth rate of each category. Meanwhile, it is simpler and more stable than other algorithms, so its performance is a little worse than others.

Table V shows that FMBFD and ESTARFM perform almost the same, even FMBFD can perform slightly better than ESTARFM in some band. Furthermore, FMBFD has an advantage over EASTARFM. ESTARFM needs two pairs of low- and high-resolution images and one low-resolution image, while FMBFD only needs a pair of images before fusion and requires less computation. During the above-mentioned fusion process, a fused image that contains 337×592 pixels was produced. Using the base time image, FSDAF and LMGM perform relatively worse than the other two algorithms, as shown in Table VI, as they cannot predict the sudden change. However, they also have good effect in some conditions. Though LMGM and FMBFD have similar prior condition, FMBFD performs better.

3) Fusion Experiment for Landsat-8 and Modis: In order to compare the performance and efficiency of the proposed algorithm with other algorithms on the same public dataset, we select the dataset provided by Guo [36], from which eight images L1, L2, M1, M2 and L4, L5, M4 and M5 in the Balle area are selected as two sets of test data. In the two set images, L1 and M1 (L4 and M4) are the Landsat and MODIS images at base time, respectively, L2 and M2 (L5 and M5) are the Landsat and MODIS images at predict time (L2/L5 are the label images for validate the performance of each method).

TABLE VI
FUSION RESULTS USING REAL SATELLITE IMAGES

Band	Method	CC	AAD	RMSE	SSIM
Blue	FSDAF	0.7912	43.1863	49.9861	0.7212
	LMGM	0.7742	42.7654	48.3458	0.7554
Green	FSDAF	0.7517	56.3513	62.5693	0.7025
	LMGM	0.7216	56.1022	61.4570	0.7564
Red	FSDAF	0.7222	47.9429	56.5934	0.6977
	LMGM	0.7212	49.3256	55.3259	0.7045

TABLE VII
FUSION RESULTS USING LANDSAT AND MODIS

Band	Method	CC	AAD	RMSE	SSIM
Blue	RASDF	0.6743	81.479	115.434	0.6636
	VSDF	0.6598	84.638	117.364	0.6467
	Fit-FC	0.6550	70.517	111.707	0.5217
	FMBFD	0.6472	87.822	122.744	0.6416
Green	RASDF	0.7217	99.699	139.137	0.6963
	VSDF	0.7239	102.461	141.088	0.7072
	Fit-FC	0.6788	109.394	150.946	0.5576
	FMBFD	0.6863	111.860	153.494	0.6830
Red	RASDF	0.6924	111.014	168.856	0.6804
	VSDF	0.6909	113.413	168.257	0.6735
	Fit-FC	0.6558	121.989	180.331	0.5167
	FMBFD	0.6639	124.431	182.321	0.6620
NIR	RASDF	0.6617	566.661	768.141	0.6020
	VSDF	0.6989	550.679	740.656	0.6657
	Fit-FC	0.7479	516.065	701.939	0.6346
	FMBFD	0.5899	649.802	854.907	0.5776

TABLE VIII
COMPUTATIONAL EFFICIENCY COMPARISON

	RASDF	VSDF	Fit-FC	FMBFD
Computation Time(s)	287.89	256.77	288.03	11.07

Three recent new unmixing-based algorithm: RASDF [22], VSDF [25], and Fit-FC [26] are selected as comparison methods, whose parameters are the same as the default parameters in its open source code. All methods were performed on a computer with i7-8700 K (3.70 GHz) and 80 GB RAM. The VSDF algorithm runs in Python, while other algorithms, including our algorithm, all run in MATLAB 2022. The fusion results are given in Figs. 22 and 23. In Figs. 22 and 23, the upper line is the comparison between the combination of the fusion results of near-red, green, and blue bands and the reference image, and the bottom line is the comparison between the combination of the fusion results of red, green, and blue bands and the reference image. The CC, AAD, RMSE, and SSIM values of each method were calculated in Table VII and the computational efficiency of these methods is given in Table VIII. From the visual effect, RASDF, VSDF, and our algorithm can maintain relatively clear texture better, while Fit-FC loses some detail

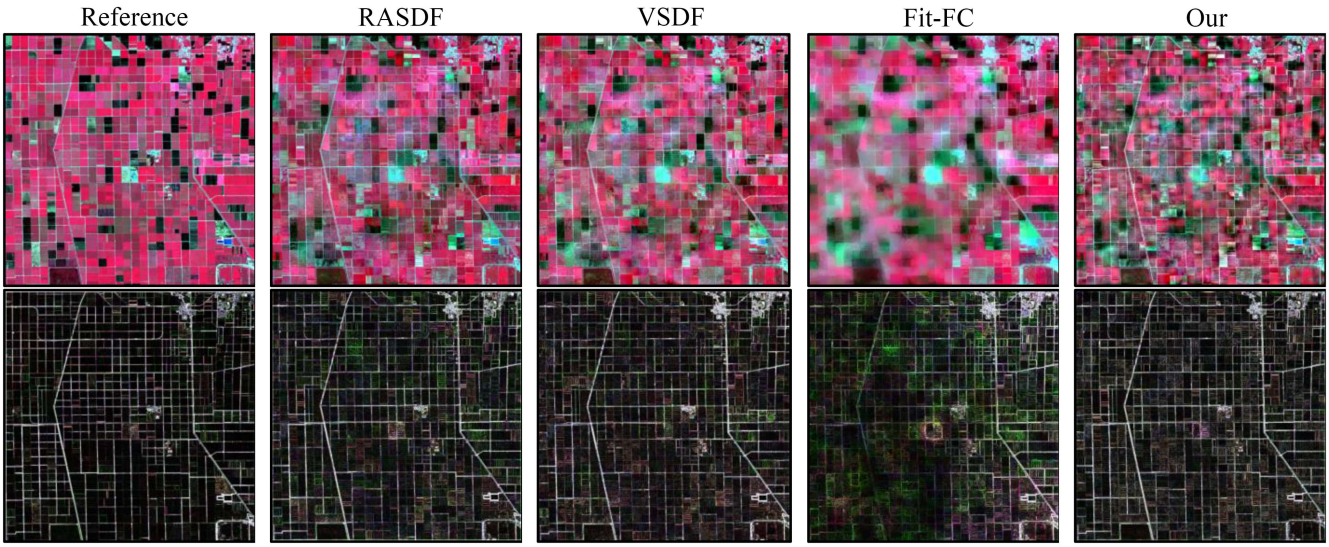

Fig. 22. Fusion results from L1, M1, M2.

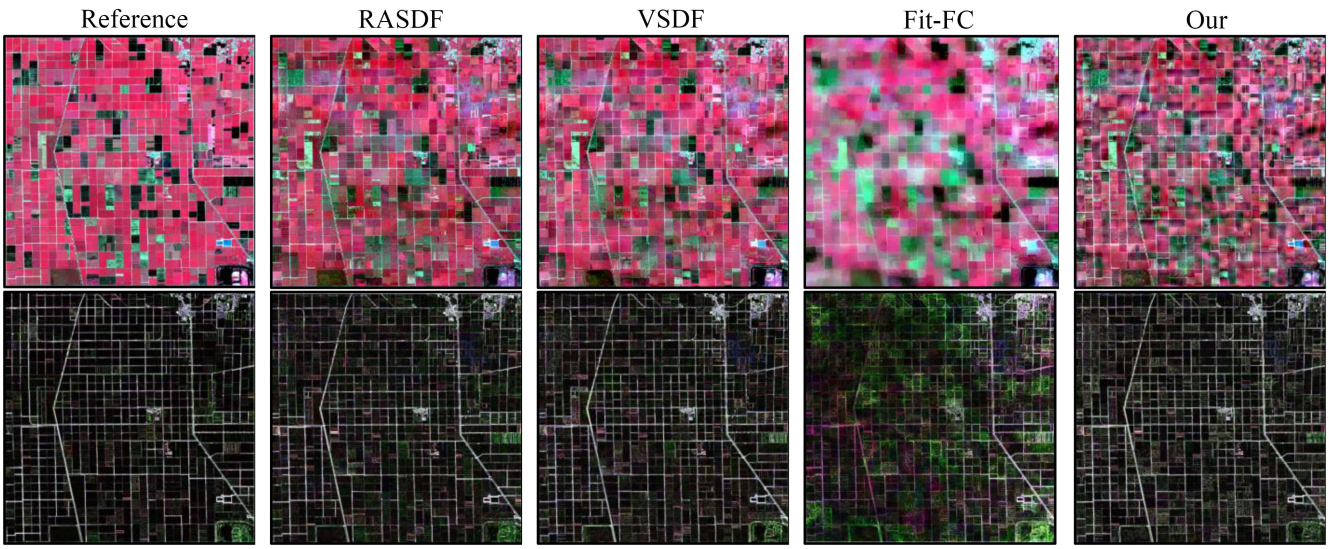

Fig. 23. Fusion results from L4, M4, M5.

texture structure, which can also be reflected in the SSIM in Table VII. In terms of spectral retention (corresponding to CC, AAD, and RMSE in Table VII), the proposed algorithm is indeed slightly worse than the other three algorithms, because our algorithm is characterized by maintaining the truth value of the CR image at the predicted time in the low-frequency part of the frequency domain, however, the resolution of the low-resolution image in this experiment is only 1/16 of that of the HR image, so the frequency domain range that can maintain the truth value is very small. Therefore, this experiment also shows that our algorithm is slightly less applicable when the resolution gap between the high- and low-resolution images is too large. Table VIII shows that our algorithm outperforms the other three algorithms in terms of computational efficiency. The calculation time of the proposed algorithm is less than 1/20 of the other algorithms, Therefore, compared with the other three algorithms, the proposed algorithm greatly improves the calculation efficiency at the cost of a small amount of spectral retention errors.

IV. Conclusion

The fusion of spatial and temporal images is one way of overcoming the problem of poor revisit encountered during HR satellite imaging. However, the traditional unmixing-based fusion methods determine the reflectance of each category as the mean value and do not consider the variance within each category. Moreover, these methods do not consider the PSF.

Finally, some basic algorithm cannot account for sudden, large-scale changes. Therefore, in this study, we developed a new algorithm for the fusion of spatial and temporal satellite images, i.e., FMBFD, which has the following advantages.

1) It is less computational and can employ precise PSF which describes the relationship between HR and CR images, because it is realized in frequency domain in which the computationally intensive convolution operations are transformed to the product operation with less computational.
2) It divides the pixel reflectance for each category into a mean part and a residual part, and estimates the change in the mean and residual part, respectively, which is more precise for the variance among the homogeneous areas.
3) It combines the real CR image and the fitted HR image in the frequency domain to retain the sharp changes within categories.

Experiments over simulated images, real satellite ones, and public dataset are carried out to demonstrate the performance of the proposed approach.

REFERENCES

[1] J. Amorós-López et al., "Multitemporal fusion of Landsat/TM and ENVISAT/MERIS for crop monitoring," *Int. J. Appl. Earth Observ. Geoinformation*, vol. 23, pp. 132–141, 2013.

[2] F. Gao et al., "Fusing Landsat and Modis data for vegetation monitoring," *IEEE Geosci. Remote Sens. Mag.*, vol. 3, no. 3, pp. 47–60, Sep. 2015.

[3] R. Zurita-Milla, J. G. P. W. Clevers, and M. E. Schaepman, "Unmixing-based Landsat TM and MERIS FR data fusion," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 3, pp. 453–457, Jul. 2008.

[4] Q. Wang and P. M. Atkinson, "Spatio-temporal fusion for daily Sentinel-2 images," *Remote Sens. Environ.*, vol. 204, pp. 31–42, 2018.

[5] Y. Zhao, B. Huang, and H. Song, "A robust adaptive spatial and temporal image fusion model for complex land surface changes," *Remote Sens. Environ.*, vol. 208, pp. 42–62, 2018.

[6] B. Chen, B. Huang, and B. Mus., "Comparison of spatiotemporal fusion models: A review," *Remote Sens.*, vol. 7, no. 2, pp. 1798–1835, 2016.

[7] X. Jiang and B. Huang, "Unmixing-based spatiotemporal image fusion accounting for complex land cover changes," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, no. 5, May 2022, Art. no. 5623010.

[8] S. Niculescu, C. Lardeux, I. Grigoras, J. Hanganu, and L. David, "Synergy between LiDAR, RADARSAT-2, and Spot-5 images for the detection and mapping of wetland vegetation in the Danube Delta," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 8, pp. 3651–3666, Aug. 2016.

[9] F. Tong, H. Tong, R. Mishra, and Y. Zhang, "Delineation of individual tree crowns using high spatial resolution multispectral WorldView-3 satellite imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, no. 7, pp. 7751–7761, Jul. 2021.

[10] H. Zhang, J. M. Chen, B. Huang, H. Song, and Y. Li, "Reconstructing seasonal variation of Landsat vegetation index related to leaf area index by fusing with MODIS data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 3, pp. 950–960, Mar. 2014.

[11] S. Corradini, L. Merucci, and A. Folch, "Volcanic ash cloud properties: Comparison between MODIS satellite retrievals and FALL3D transport model," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 2, pp. 248–252, Mar. 2011.

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

[13] T. Hilker et al., "A new data fusion model for high spatial- and temporal-resolution mapping of forest disturbance based on Landsat and MODIS," *Remote Sens. Environ.*, vol. 113, pp. 1613–1627, 2009.

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

[15] B. Zhukov, D. Oertel, F. Lanzl, and G. Reinhäckel, "Unmixing-based multisensory multiresolution image fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1212–1226, May 1999.

[16] M. Wu, C. Wang, and L. Wang, "Use of MODIS and Landsat time series data to generate high-resolution temporal synthetic Landsat data using a spatial and temporal reflectance fusion model," *J. Appl. Remote Sens.*, vol. 6, no. 1, Mar. 2012, Art. no. 23620.

[17] C. M. Gevaert and F. J. García-Haroa, "A comparison of STARFM and an unmixing-based algorithm for Landsat and MODIS data fusion," *Remote Sens. Environ.*, vol. 156, pp. 34–44, Jan. 2015.

[18] Y. H. Rao, X. L. Zhu, J. Chen, and J. M. Wang, "An improved method for producing high spatial-resolution NDVI time series datasets with multi-temporal MODIS NDVI data and Landsat TM/ETM plus images," *Remote Sens.*, vol. 7, no. 6, pp. 7865–7891, 2015.

[19] G. Zhang et al., "Spectral variability augmented sparse unmixing of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, no. 4, Apr. 2022, Art. no. 5527413.

[20] X. Zhu, E. H. Helmer, F. Gao, D. Liu, J. Chen, and M. A. Lefsky, "A flexible spatiotemporal method for fusing satellite images with different resolutions," *Remote Sens. Environ.*, vol. 172, pp. 165–177, 2016.

[21] D. Guo, W. Shi, M. Hao, and X. Zhu, "FSDAF 2.0: Improving the performance of retrieving land cover changes and preserving spatial details," *Remote Sens. Environ.*, vol. 248, 2020, Art. no. 111973.

[22] W. Shi, D. Guo, and H. Zhang, "A reliable and adaptive spatiotemporal data fusion method for blending multi-spatiotemporal-resolution satellite images," *Remote Sens. Environ.*, vol. 268, 2022, Art. no. 112770.

[23] Q. Wang, K. Peng, Y. Tang, X. Tong, and P. M. Atkinson, "Blocks-removed spatial unmixing for downscaling MODIS images," *Remote Sens. Environ.*, vol. 256, 2021, Art. no. 112325.

[24] K. Peng, Q. Wang, Y. Tang, X. Tong, and P. M. Atkinson, "Geographically weighted spatial unmixing for spatiotemporal fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, no. 10, Oct. 2021, Art. no. 5404217.

[25] C. Xu, X. Du, Z. Yan, J. Zhu, S. Xu, and X. Fan, "VSDF: A variation-based spatiotemporal data fusion method," *Remote Sens. Environ.*, vol. 283, 2022, Art. no. 113309.

[26] Q. Wang and P. M. Atkinson, "Spatio-temporal fusion for daily Sentinel-2 images," *Remote Sens. Environ.*, vol. 204, pp. 31–42, 2018.

[27] B. Huang and H. Song, "Spatiotemporal reflectance fusion via sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 10, pp. 3707–3716, Oct. 2012.

[28] H. Zhang, Y. Song, C. Han, and L. Zhang, "Remote sensing image spatiotemporal fusion using a generative adversarial network," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 4273–4286, May 2021.

[29] D. Guo, W. Shi, M. Hao, and X. Zhu, "FSDAF 2.0: Improving the performance of retrieving land cover changes and preserving spatial details," *Remote Sens. Environ.*, vol. 248, Oct. 2020, Art. no. 111973, doi: 10.1016/j.rse.2020.111973.

[30] W. Li, X. Zhang, Y. Peng, and M. Dong, "Spatiotemporal fusion of remote sensing images using a convolutional neural network with attention and multiscale mechanisms," *Int. J. Remote Sens.*, vol. 42, no. 6, pp. 1973–1993, Mar. 2021, doi: 10.1080/01431161.2020.1809742.

[31] J. Chen, L. Wang, R. Feng, P. Liu, W. Han, and X. Chen, "CycleGAN-STF: Spatiotemporal fusion via CycleGAN-based image generation," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5851–5865, Jul. 2021, doi: 10.1109/TGRS.2020.3023432.

[32] R. Zurita-Milla, J. G. P. W. Clevers, and M. E. Schaepman, "Unmixing-based Landsat TM and MERIS FR data fusion," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 3, pp. 453–457, Jul. 2008.

[33] J. Peter, D. Shaw, and J. Rawlins, "The point-spread function of a confocal microscope: Its measurement and use in deconvolution of 3-D data," *J. Microsc.*, vol. 163, no. 2, pp. 151–165, 1991.

[34] B. Huang and H. Zhang, "Spatio-temporal reflectance fusion via unmixing: Accounting for both phenological and land cover changes," *Int. J. Remote Sens.*, vol. 35, pp. 6213–6233, 2014.

[35] H. Song and B. Huang, "Spatiotemporal satellite image fusion through one-pair image learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 4, pp. 1883–1896, Apr. 2013.

[36] D. Guo, W. Shi, H. Zhang, and M. Hao, "A flexible object-level processing strategy to enhance the weight function-based spatiotemporal fusion method," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, no. 10, Oct. 2022, Art. no. 4414811.

Zheng Lu (Senior Member, IEEE) received the B.S. degree in information engineering and the Ph.D. degree in target detection and recognition from the Beijing Institute of Technology, Beijing, China, in 2008 and 2013, respectively.

He is currently a Senior Engineer with the Institute of Remote Sensing Satellite, China Academy of Space Technology, Beijing. He is a Program Manager of the Spaceborne Bistatic SAR Team. He has authored or coauthored three academic books more than 60 refereed journals and conference papers and granted 16 national invention patents of China. His research interests include spaceborne radar systems and signal processing.

Dr. Lu was the recipient of several awards and prizes, including the First Prize of National Surveying-Mapping Science and Technology Progress Award, in 2019 and Beijing Nova Program Award, in 2020.

Bozheng Shu was born in Jiangxi, China, in November 1992. He received the B.S. and master's degrees in information engineering from the Beijing Institute of Technology, Beijing, China, in 2013 and 2016, respectively.

He is currently with the Institute of Microelectronics of the Chinese Academy of Sciences, Beijing. His research interests include the fields of space-borne SAR imaging, LiDAR data processing, and LiDAR system architecture.

Xiaoqing Wang was born in Jiangxi, China, in 1978. He received the B.S. degree in electronic engineering from Xiamen University, Xiamen, China, in 2000, and the Ph.D. degree in signal and information processing from the University of Chinese Academy of Sciences, Beijing, China, in 2005.

From 2005 to 2016, he was an Assistant Researcher and Associate Researcher with the Institute of Electronics, Chinese Academy of Sciences. From 2016 to 2019, he was an Associate Researcher and Researcher with the Institute of Microelectronics, Chinese Academy of Sciences. Since 2019, he has been a Professor with the School of Electronics and Communication, Sun Yat-sen University, Shenzhen, China. His research interests include ocean microwave remote sensing, SAR signal processing, and radar processing.

Zheng Ge was born in China, in 1996. She received the M.S. degree in environmental engineering from Inner Mongolia University, Hohhot, China, in 2022. She is currently working toward the joint Ph.D. degrees in information and communication engineering with the School of Electronics and Communication, Sun Yat-sen University, Shenzhen, China, and Pengcheng Laboratory, Shenzhen, China.

Her research interests encompass image fusion, image generation, deep learning, and information extraction and application in remote sensing image.

Lingxi Guo was born in Hubei Province, China, in 1985. He received the M.S. degree in information perception and confrontation from the Beijing Institute of Technology, Beijing, China, in 2009.

He is currently a Senior Engineer with Science and Technology on Space Physics Laboratory, Beijing. His current research interests include object detection, image content understanding, image simulation, and deep learning.