Journal of Network Intelligence
Taiwan Ubiquitous Information
 ISSN 2414-8105 (Online)
Volume **4**, Number **3**, August **2019**

Spatiotemporal Fusion Algorithm for Single-Time Phase High Resolution Remote Sensing Image Based on Sparse Representationy

Xiaoyi Wang[1], Xiaofei Wang[1,*], Shu-Chuan Chu[2], John F. Roddick[2]

[1]1College of Electrics Engineering,
Heilongjiang University, Harbin 150080, China
773324723@qq.com, nk_wxf@hlju.edu.cn

[2]College of Science and Engineering Flinders University,
Sturt Rd, Bedford Park SA 5042, South Australia
scchu0803@gmail.com, john.roddick@inders.edu.au

Received March 2019, Revision May 2019

Abstract. *High spatiotemporal resolution remote sensing images can provide on landform changes information fast and accurately, which have a wide range of applications and needs in many areas such as agricultural monitoring and urban planning and construction. However, due to the limitations of sensor hardware, remote sensing images have the phenomenon that both high spatial resolution and high temporal resolution are not compatible. In view of the complementary advantages of information between images from different sensors, the fusion of remote sensing images is a very meaningful direction. Based on sparse representation theory, this paper proposes a double-layer spatiotemporal fusion framework suitable for single-phase high-resolution remote sensing images. The Landsat8 OLI and MODIS remote sensing image are used as experimental data to fully analyze the method, and compared with the classical spatiotemporal fusion methods of STARFM, and analyze the impact of down sampling the multiple of down sampled in the double-layer on the fusion results. The experimental results show that our method has higher prediction accuracy, and the experimental results are best when the multiple of the down-sampled is four.*

Keywords: RSparse representation, Single-phase, Spatiotemporal fusion, Landsat8 OLI, MODIS.

1. **Introduction.** With the continuous development of remote sensing technology, remote sensing technology can provide various information about crop ecological environment and crop growth objectively, accurately and timely. It is an important source for accurate field data. However, under the constraints of the hardware technical conditions of existing satellite sensors and the cost of satellite launching, remote sensing satellites cannot obtain remote sensing images with multi-attribute high resolution, which restricts the application of remote sensing images. For example, Landsat satellites obtain images with spatial resolutions in the 30m range, whereas a 16-day return visit cycle limits its use to detect rapid changes in land and, on the other hand, medium resolution imaging spectra carried on Terra/Aqua satellites radiometer (MODIS) provides daily observations, but having a coarse spatial resolution of 250-1000 m is not conducive to monitoring land cover changes in heterogeneous landscapes. Therefore, spatiotemporal fusion has emerged as a method of providing satellite imagery with high temporal and spatial resolution. During recent decades, the method of spatiotemporal fusion has been widely applied and mentioned. Gao et al. (2006) first proposed Spatial And Temporal Adaptive Reflectance Fusion Model (STARFM) [1] for the identification of surface cover types of fractured blocks, which can effectively eliminate singular points and is suitable for detecting the gradual changes of the large-scale range of space. Zhu et al. (2010) proposed Enhanced Spatial and Temporal Adaptive Reflectance Fusion Model (ESTARFM) based on STARFM [2], which is more accurate and efficient for complex and heterogeneous

features. Solved the "time smoothing" problem of STARFM algorithm. Both STARFM and ESTARFM are weight-based spatiotemporal fusion algorithms, which are limited to fine-grained landscapes and will reduce the accuracy of their fused images in fine-grained heterogeneous landscapes. Hilker et al. (2009) proposed a Spatial Temporal Adaptive Algorithm for mapping Reflectance Change(STAARCH) [3] to observe changes in forest vegetation. Although this algorithm can analyze the change of reflectivity of input images and handle the dynamic change of land cover type, but requires two landscape Landsat images, not suitable for areas where image acquisition is difficult. Wu et al. (2012) proposed a Spatial And Temporal Data Fusion Approach (STDFA)[4] based on the time-varying feature of the feature for the extraction of rice area. This algorithm also requires multi-phase image support, which limits the application of the algorithm. With the wide application of machine learning theory in the field of image processing, For example, Eric Ke Wang et al. (2019) proposed a multilayer dense attention model for image caption[5]. Super-resolution reconstruction technology which based on machine learning has gradually been introduced into data reconstruction research. Hong et al. (2012) proposed a Sparse Representation-based Spatiotemporal Reflectance Fusion Model (SPSTFM) [6]. The model uses the difference images of the MODIS and Landsat image pairs of the front and back phases of the predicted phase to train high- and low-resolution dictionaries representing time-varying features, and then uses the MODIS image of the predicted time to generate a Landsat-like fusion image, the fusion accuracy of this model is very high, but it is also limited by the difficulty of data acquisition; Subsequently, Song and Huang. (2013) proposed a sparse representation spatiotemporal reflectance fusion model using only a pair of known high and low spatial resolution image [7]. The model first uses the sparse representation algorithm to enhance the MODIS image to obtain the intermediate transition image, and then uses the high-pass filtering model to fuse the observed Landsat image and the transition image. This model reduces the number of known image pairs that need to be input, so that the algorithm can be applied in the absence of data, and has universal applicability. The feature-based learning-based spatio-temporal fusion method takes into account the spatial information characteristics of the changing image, and can make more accurate predictions of the complex surface reflectance image compared to the weight-based method. Although the fusion accuracy of this model is very high, the requirements for input data quality are high, especially when dealing with high-resolution images, the accuracy and efficiency of dictionary training need to be further improved. Compared with the above method, we present a spatiotemporal fusion algorithm for single-temporal high resolution images based on sparse representation. The prediction results which based on sparse representation theory with higher precision, and the use of single-phase data pairs can also reduce the difficulty in obtaining data in ESTARFM, SRCAAH, SPSTFM and other methods. In the second part of this section, we will briefly introduce the knowledge of theoretical sparse representation and its application to single image super-resolution. Then, in Section 3, the method we propose will be explained in detail. In Section 4, Landsat8 OLI images and MODIS images were used as experimental data, and our method was compared with the experimental results of STARFM. Then, we will conclude this article in Section 5 to discuss our findings.

2. **Sparse Representation and Dictionary Learning Theory.**

The problem of sparse representation of images originated from the "effective coding hypothesis" in the field of neurobiology. Relevant researchers believe that the main function of primary visual cortical neurons is to remove the statistical correlation of input stimuli. The researchers have made this hypothesis based on this hypothesis. A large number of related experiments to verify this hypothesis, the conclusion that the sparse representation of the signal conforms to the human visual perception mechanism. In the field of signal and image processing, this idea of sparse coding is widely used, for example, Zhao et al.(2016)[8] used sparse representation and dictionary learning theory to detect text; Tang et al.(2016)[9] used NSCT sparse representation and pulse coupled neural network to achieve multi-modal medical image fusion;Tang et al.(2018)[10]used sparse representation to achieve automatic classification of chord information in noisy music and multimedia signals.

2.1. **Sparse representation of the image.**

Sparse representation of an image refers to the approximate representation of an image using a linear combination of atoms, which constitute an over-complete dictionary. In order to simplify the solving problem of the image processing process, it is required that the coefficient matrix of the linear combination contains as few non-zero items as possible, so it is called sparse coding. The sparse representation theory holds that the signal can be obtained from an over complete dictionary and its corresponding sparse coding. The expression is as follows:

$$x = Da \tag{1}$$

where $x \in R^n$ is the signal to be processed, $D \in R^{n \times m}(n < m)$is the over complete dictionary matrix, and $x \in R^m$ is the sparse coding matrix.

2.2. **Overcomplete dictionary training.** For the sparse representation of remote sensing images, the image needs to be converted into a raster image, that is, the image is divided into image blocks $\sqrt{n} \times \sqrt{n}$, and these image blocks are stored in the form of column vectors to form a signal to be processed$X = [x_1, x_2, ..., x_n]$, each column represents an atom, the process is the signalization of the image. Generally, the K-SVD algorithm[11] is used to train the overcomplete dictionary . The objective function to be solved can be expressed as:

$$\min_{D,\Lambda} \{\|X - D\Lambda\|_F^2\} \tag{2}$$

Where $\Lambda = [a_1, a_2, ...a_n]$, and $\|a\|_0$represents the number of non-zero elements in , and is the number of non-zero elements preset.Requires that for any one $a_i(1 < i < n)$, the following conditions are met:

$$\|a_i\|_0 \leq K_0 \tag{3}$$

Firstly, an initial dictionary is developed by K-SVD method, and then the signal to be processed is sparsely represented based on the Orthogonal Matching Pursuit (OMP) algorithm[13]. Finally, the optimal solution is approximated by iteratively updating the atoms one by one. For a well-trained overcomplete dictionary and a known pending signal, the next main purpose is to find a sparse representation matrix a with the fewest non-zero elements. This problem can be represented by the following sparse estimate:

$$\hat{a} = \arg\min_a \|a\|_0 \tag{4}$$

Using to represent the tolerance, the following relationship is satisfied:

$$\|x - Da\|_2^2 \leq \varepsilon \tag{5}$$

The above-mentioned sparse coding solution is an NP-hard problem, which can be solved quickly and effectively by the estimation method. The commonly used Matching Pursuit(MP) (Mallat SG et al.,1993)[12] and its enhanced orthogonal Orthogonal Matching Pursuit(OMP) (Davis G et al., 1997)[13], Basis Pursuit(BP)(Chen SSB et al., 2001)[14], and so on.

3. **Methodology.** The spatiotemporal fusion based on a single known data pair is based on the high-resolution image H_1 on t_1 and the low-resolution images L_1 and L_2 on t_1 and t_2, predicting the high resolution image H_2 on $t_2$by using high-pass modulation and super-resolution reconstruction technology based on sparse representation. The main processing steps are as follows.

3.1. **Resolution reconstruction theory.** Image super-resolution needs to find a mapping relationship between high- and low-resolution images, which includes many methods. Among them, Tsu-Yang Wu et al. proposed a mapping method based on logistic[15] and Chebyshev generator[16]. In this paper, the high- and low-resolution dictionary obtained by high- and low-resolution image training has the same sparsity coefficient to achieve the mapping between high- and low-resolution effects. In the field of image processing, low-resolution images are considered to be obtained by blurring and downsampling high-resolution images. The model can be expressed as:

$$Y = SHX + C \tag{6}$$

Where X is a high-resolution image, Y is a low-resolution image, H is a blur filter, S is a downsampled model, and C is random noise.

3.2. **Acquisition of transition images.** It is known that Landsat-MODIS image pairs on t_1 and MODIS images on t_2 $(t_1 < t_2)$, Landsat images with a spatial resolution of 30m are used as high-resolution images, and MODIS images with a spatial resolution of 500m are used as low-resolution images. It is defined that H_i and L_i is Landsat images and MODIS images at a time t_i $(i = 1, 2)$, and define the transition image as T_i. The difference image between high-resolution and low-resolution$Y_1 = H_1 - L_1)$ on $t_1$and the gradient feature image of the low-resolution image L_1 which named $X_1 = (f * L_1)$ (where '$*$' represents the convolution operation) as training samples for dictionary training, gradient feature image in patch form and this article uses 5×5. The low-resolution overcomplete dictionary D_l can be obtained by training X_1 based on the K-SVD joint dictionary pair:

$$\{D_l, a_1\} = \arg\min_{D_l, a_1} \{\|X_1 - D_l a_1\|_F^2\} \tag{7}$$

$$s.t. \forall i, \|a_1\|_0 \leq K_0$$

Because the sparse representation coefficient of the high-low resolution difference image Y_1 by the high-resolution dictionary D_h is the same as the sparse representation coefficient of the low-resolution gradient feature image X_1 by the low-resolution dictionary D_l(that isa_1), the high-resolution dictionary D_h can be obtained by minimizing the approximation error of Y_1 under the coefficient matrix a_1:

$$D_h = \arg\min_{D_h} \|Y_1 - D_h a_1\|_F^2 \tag{8}$$

Under the condition that the coefficient matrix a_1 is full rank, the above formula can directly obtain the high-resolution overcomplete dictionary D_h by using the pseudo-inverse expression as follows:

$$D_h = Y_1(a_1)^+ = Y_1 {a_1}^T ({a_1 {a_1}^T})^{-1} \tag{9}$$

Let the high and low-resolution difference image Y_2 = T_2-L_2 on t_2 and the gradient feature image $X_2 = (f * L_2)$ of the low-resolution image L_2, then after the high-resolution dictionary $D_h$and gradient features image X_2 were be obtained, the sparse coefficient matrixa_2 on t_2 is obtained by the OMP. Further, the difference image Y_2 can be expressed as:

$$Y_2 = D_h a_2 = T_2 - L_2 \tag{10}$$

Therefore, the transition image T_2 can be expressed as:

$$T_2 = Y_2 + L_2 \tag{11}$$

Similarly, the transition image T_1 on t_1 can be obtained.

3.3. **High-pass modulation.** Because of the similarity between T_i and H_i, they can be considered to have the same pixel purity at the same time. If the following linear relationship can express the transition image at different times:

$$T_2 = aT_1 + b \tag{12}$$

Then, the relationship is equally applicable to the high-resolution image H_i, i.e.:

$$H_2 = aH_1 + b \tag{13}$$

Thus, the high-resolution image on t_2 can be obtained by high-pass modulation fusion, expressed as:

$$H_1 = T_2 + \frac{T_2}{T_1}[H_1 - T_1] \tag{14}$$

3.4. **Spatiotemporal fusion based on scale progression.** Traditional fusion studies have shown that the spatial scale difference factor $r \leq 4$ between high-resolution and low-resolution images can achieve optimal results.It is necessary to use resampling and multi-layer progressive spatiotemporal fusion to make the spatial scale difference between high-resolution image and low-resolution image $r = r_l/r_h \leq 4$, due to the spatial resolution difference between Landsat(30m) image and MODIS(500m) image is nearly 16 times, how to choose a reasonable intermediate resolution multiple has become a problem. Considering that the algorithm in this paper is divided into two layers, the first layer establishes the transition image T_i , and the second layer predicts H_{2p}. The specific operation process is shown in Fig.1. As shown in Fig.1., the double-layer spatiotemporal fusion network has the following points:

(1) In the first layer, superscript represents the number of layers; subscript is the phase. To train a dictionary pair in the first layer, the high-resolution image features and low-resolution image features are extracted from difference image space of original $H_1^1 - L_1^1$ and gradient feature space of original L_1^1 in patch form (This article uses 5×5), then use K-SVD algorithm to train dictionary, we get a high-resolution dictionary D_h^1 and low-resolution dictionary D_l^1.

(2) The original low-resolution images L_1^1 and L_2^1 on t_1 and t_2 are input as low-resolution images, and the pixel size of the original high-resolution image H_1^1 on t_1 is resampled by bicubic interpolation $r_l/2$, $r_l/4$ or $r_l/8$ to obtain high-resolution input image.

(3) Transition image T_1^1 and T_2^1on t_1 and t_2 are predicted using the sparse coding technique OMP algorithm and Equation (11).

(4) Using T_1^1 and T_2^1 resampled H_1^1 as input, utilizing Equation (14) to obtain the predicted high-resolution image H_{2P}^1 of the first layer on t_2.

(5) In the second layer, the original H_1^2 is downsampled by bicubic interpolation $r_l/2$,$r_l/4$or $r_l/8$ as L_1^2, and the predicted H_{2p}^1 is taken as L_2^2, which are input as low-resolution images, and the original H_1^2 is input as a high-resolution image.

(6) The high-resolution image features and low-resolution image features are extracted from original H_1^2 and resampled H_1^2 in patch form (This article uses 5×5), then use K-SVD algorithm to train dictionary, we get a high-resolution dictionary D_h^2 and low-resolution dictionary D_l^2.

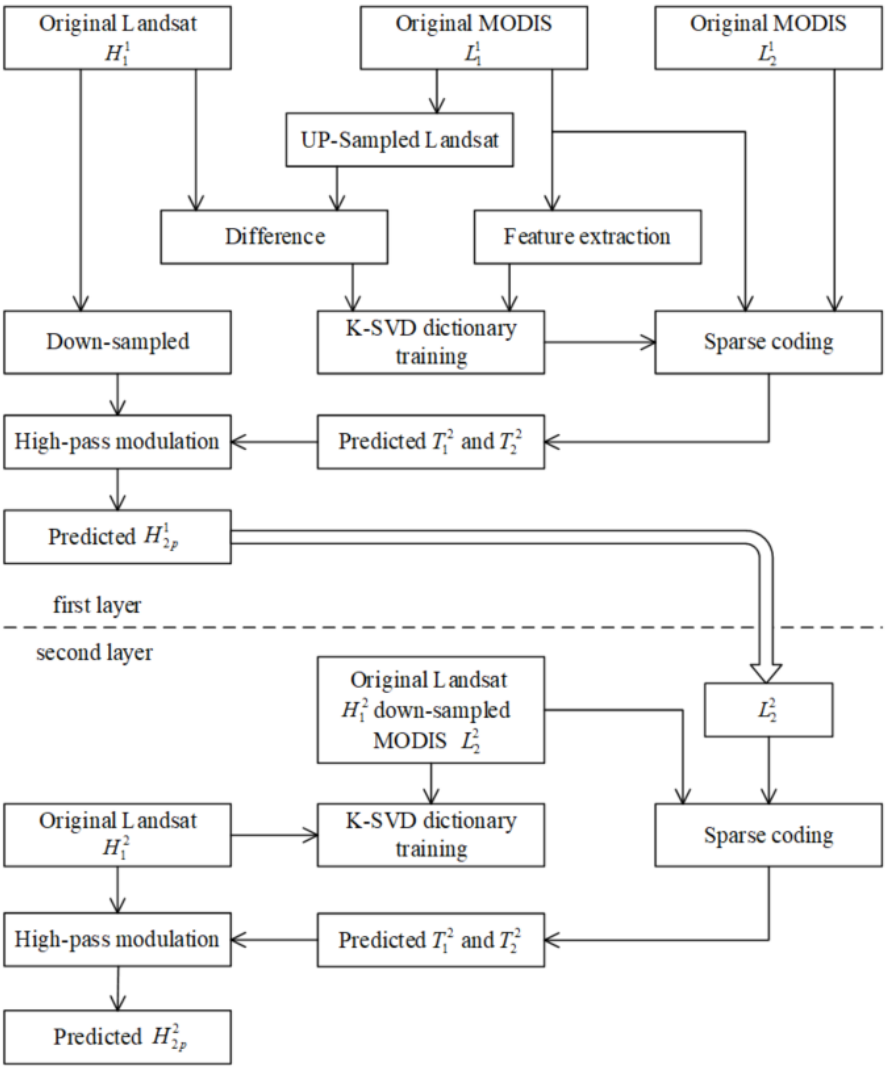

FIGURE 1. Flowchart of the proposed spatiotemporal fusion framework

(7) Transition image T_1^2 and T_2^2on t_1 and t_2 are predicted using the sparse coding technique OMP algorithm and Equation (11).

(8) Using T_1^2 and T_2^2 original H_1^2as input, utilizing Equation (14) to obtain the predicted high-resolution image H_{2P}^2 of the first layer on t_2. H_{2P}^2 is the final result.

4. **Simulation experiments and results analysis.** Three pairs of Landsat 8 and MODIS surface reflectance images covering a 28 km by 28 km area in Coleambally, Australia, were used in this experiment. The type of land cover is mainly farmland. This paper uses the Landsat8 OLI images of the experimental area and the corresponding MODIS images as experimental data. Atmospheric correction is performed on the Landsat 8 OLI image using the ENVI-Flaash Atmospheric Correction Module. Atmospheric corrected images were geometrically corrected using a 1:10000 topographic map and reprojected to UTM-WGS84. The coordinate system has an error of fewer than 0.5 pixels. The pixel area (with a spatial resolution of 30 m) was intercepted as an experimental analysis area. The MODIS reflection raw data is projected by the sinusoidal projection method. Reproject the MODIS data into the same UTM-WGS84 coordinate system as the Landsat8 OLI data using the MODIS Reprojection Tools. For ease of processing, this paper resamples MODIS data to 480m spatial resolution during reprojection. The surface coverage type of this area is stable, and the change in surface reflectance is considered to be determined only by the phenological phenomena of the vegetation. Fig. 2.(a) and (b) show the Landsat and MODIS images on June 3, 2013, respectively, and Fig. 2.(c) and (d) show the Landsat and MODIS images of the experimental area on September 7, 2013, respectively. Both are standard false-color images, select bands

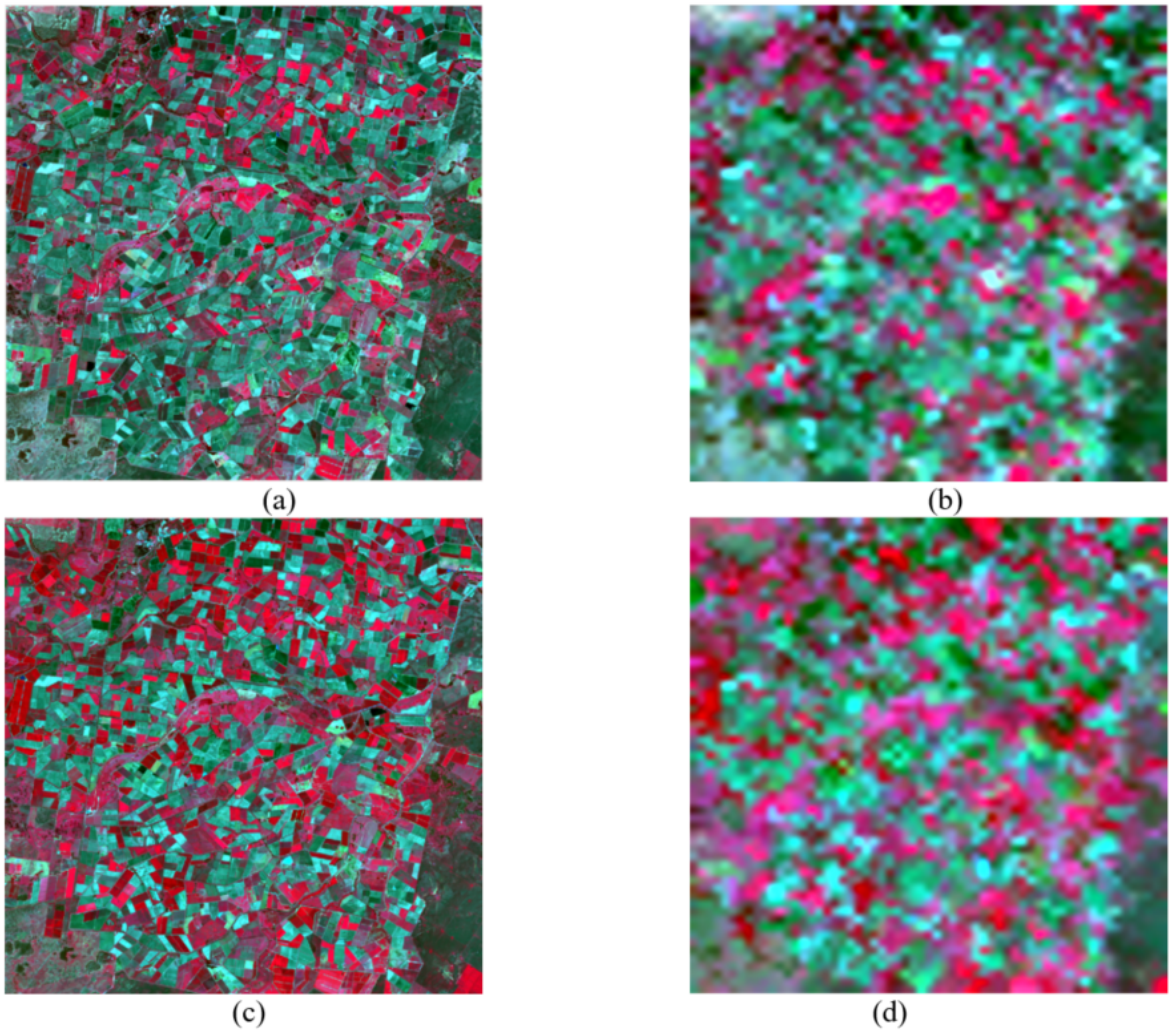

FIGURE 2. (Upper row) Landsat8 OLI composited surface reflectance, and MODIS composited surface reflectance (512×512 pixels) on June 3, 2013;(Lower row) Landsat8 OLI composited surface reflectance, and MODIS composited surface reflectance (512×512 pixels) on September 7, 2013

5, 4, 3 for Landsat8 OLI images and select bands 2, 1, 4 for MODIS images, the combination of Nir, Red and Green bands. It can be seen that the MODIS and Landsat images on the same day are very similar, and both images reflect the change in surface reflectance during this period. In this section, we compare the proposed method with the well-known STARFM algorithm by using the Landsat8 OLI images and MODIS images of the experimental area, using our method to set Landsat downsampling multiples to 2, 4, 8 for three experiments. The predicted image is then compared to the actual observed image, and subjective evaluations and goals are evaluated to assess prediction accuracy. Refer to the objective evaluation indicators for image quality proposed by Qiaoyue Li et al.[17] and Chate Harold et al[18], root mean square error (RMSE), the mean absolute difference (AAD), exponential-structural similarity (SSIM)[19] and ERGAS [20] were chosen as objective quality evaluation indicators.

From the subjective visual observation, it can be seen that in the crop area, compared with the proposed method, the fusion result of the STARFM method has a large fusion error, mainly because the method searches for similar pixels near the center pixel based on the reference time. It is assumed that adjacent similar pixels in the basic phase periods experience a similar phenological change with the center pixel, and the reflectance of the similar pixels is used to obtain the reflectance of the center pixel of the predicted phase. When large phenological changes occur, due to the high heterogeneity of the selected regions, the central pixel and adjacent similar pixels have different degrees of phenological changes in

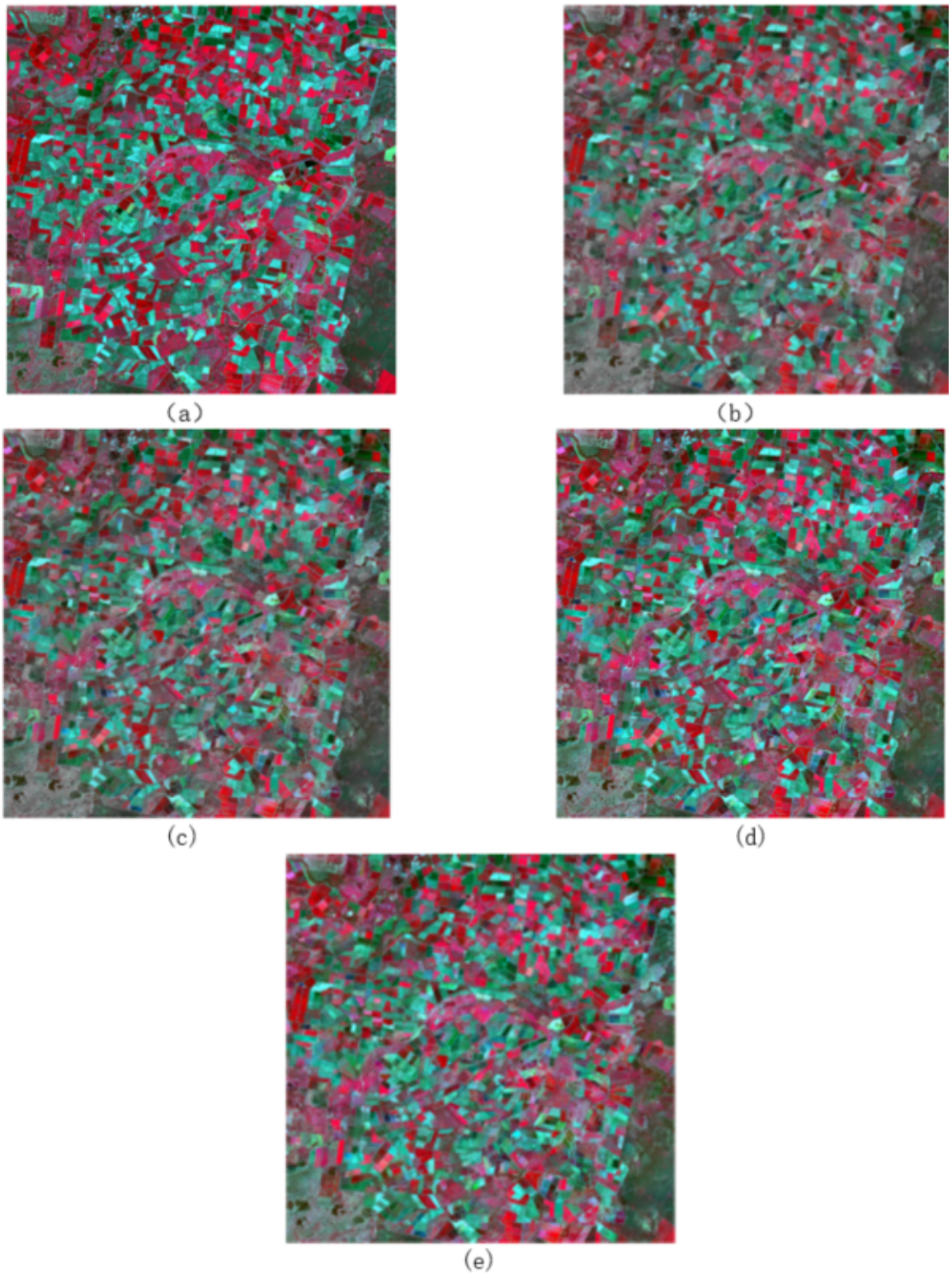

FIGURE 3. Comparisons between the actual and the predicted surface reflectances with focus on seasonal changes: (a) Actual surface reflectance, (b) predicted surface reflectance obtained by STARFM, (c) predicted surface reflectance obtained by our method1, (d) predicted surface reflectance obtained by our method2,(e) predicted surface reflectance obtained by our method3

TABLE 1. Accuracy evaluation of fusion results in experimental areas

Fusion method	AAD			RMSE			SSIM			ERGAS
	Green	Red	Nir	Green	Red	Nir	Green	Red	Nir	
STARFM	0.0181	0.0152	0.0122	0.0144	0.0091	0.0289	0.812	0.814	0.822	0.6518
Our method1	0.0164	0.0143	0.0104	0.0099	0.0084	0.0257	0.829	0.835	0.857	0.6377
Our method2	0.0153	0.0132	0.0090	0.0080	0.0071	0.0212	0.854	0.870	0.885	0.6115
Our method3	0.0160	0.0141	0.0094	0.0085	0.0076	0.0246	0.848	0.859	0.872	0.6265

the cycle, resulting in inaccurate fusion results. The comparisons in terms of AAD, RMSE, ERGAS, and SSIM are listed in Tab.1., the average AAD values of the three bands for STARFM, Our method1, Our method2, and Our method3 are 0.0151, 0.0134, 0.0125 and 0.0132, respectively, and the average RMSE values of the three bands for these models are 0.0174, 0.0147, 0.0121 and 0.0136, respectively. These indicate that our method can reconstruct the Landsat surface reflectance more precisely than STARFM. The average SSIM values of the three bands for these methods are 0.814,0.840,0.869 and 0.860, respectively, and these indicate that our method can retrieve more precise structural details on the surface reflectance than STARFM with smaller reflectance deviations. The ERGAS values for these methods are 0.6518, 0.6377, 0.6115, and 0.6265, respectively, and these indicate that the spatial details and spectral colors of our fusion result are better than this of STARFM. Also, it can be seen from the experimental results that the multiple of down-sample has a great influence on the experimental results. When the multiple is 4, the fusion precision reaches the optimal value, and it can be seen when the multiple is 2, the experimental precision is lower than when the multiple is 8. It indicates that in the first layer for transition image reconstruction, it is not suitable to use larger multiples or smaller multiples.

5. **Conclusions.** This paper proposes a two-layer spatiotemporal fusion model based on sparse representation for single-phase input. Using our method to compare with STARFM, the results show that our method has better fusion accuracy and better reconstruction of spatial details. The multiples of the down-sampled were compared and analyzed, and it was found that too large or too small multiples would reduce the reconstruction accuracy. Spatio-temporal fusion is the complementary advantage of remote sensing image information, making it fully utilized in agricultural monitoring and other fields. In the future research, on the one hand, how to make full use of remote sensing image information to make it more widely used in various fields; on the other hand, attempts to explore spatial information of low-resolution image MODIS more widely, in order to obtain better Landsat predicted images.

Acknowledgment. This work is supported by National Key R&D Program of China (2016YFB0502502), National Natural Science Foundations of China (61871150).

REFERENCES

[1] F. Gao, J.Masek, M.Schwaller, et al. On the blending of the Landsat and MODIS surface reflectance: predicting daily Landsat surface reflectance J. *IEEE Transactions on Geoscience & Remote Sensing*, 2006, vol. 44, no. 8, pp. 2207-2218.

[2] X.L.Zhu, J.Chen, F.Gao, et al. An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions *[J].Remote Sensing of Environment*, 2010, vol. 114, no. 11, pp. 2610-2623.

[3] T.Hilker, M.A.Wulder, N.C.Coops, et al. A new data fusion model for high spatial- and temporal-resolution mapping of forest disturbance based on Landsat and MODIS.*[J]. Remote Sensing of Environment*, 2009, vol. 113, no. 8, pp. 1613-1627.

[4] M.Q.Wu, Z.Niu. C.Y.Wang et al. 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]. Journal of applied Remote Sensing*, 2012,6(1): 063507-1-063507-13.

[5] E. K. Wang, X. Zhang, F. Wang, T.-Y. Wu, C.-M. Chen, Multilayer Dense Attention Model for Image Caption, IEEE Access,vol. 7, pp. 66358-66268, 2019

[6] B. Huang, H. Song. Spatiotemporal Reflectance Fusion via Sparse Representation[J]. *IEEE Transactions on Geoscience & Remote Sensing*, 2012, vol. 50(10): 3707-3716.

[7] H.Song, B.Huang. Spatiotemporal satellite image fusion through one-pair image learning[J]. Geoscience and Remote Sensing, IEEE Transactions on, 2013, 51(4)1883-1896.

[8] C. D.Zhao, Z.F.Zhu, and Y.Zhao.Text Detection Based On Discriminative Dictionary Learning *Journal of Information Hiding and Multimedia Signal Processing*, vol. 7, no. 4, pp. 674-684, July 2016.

[9] L. Tang, L.D.Li, J.S.Qian, J.Y.Zhang, and J.S.Pan. NSCT-Based Multimodal Medical Image Fusion With Sparse Representation and Pulse Coupled Neural Network, *Journal of Information Hiding and Multimedia Signal Processing*, Vol. 7, No. 6, pp. 1306-1316, November 2016.

[10] Z. Y.Rao, C. Y.Feng.Sparse Representation Classification-Based Automatic Chord Recognition For Noisy Music, *Journal of Information Hiding and Multimedia Signal Processing*, vol. 9, no. 2, pp. 400-409, March 2018.

[11] M. Aharon, M. Elad, and A. M. Bruckstein.The K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation, IEEE Trans. Signal Process., vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[12] S. G. Mallat, Z. F. Zhang, Matching pursuits with time-frequency dictionaries, *IEEE Transaction Signal Processing*, 1993, vol. 41, no. 12, pp. 3377-3341.

[13] G. Davis, S. Mallat, and M. Avellaneda. Adaptive greedy approximations.Constructive Approx., vol. 13, no. 1, pp. 57-98, Mar. 1997.

[14] S. Chen, D. Donoho, and M. Saunders. Atomic decomposition by basis pursuit. SIAM, Rev., vol. 43, no. 1, pp. 129-159, 2001.

[15] T.-Y. Wu, Xiaoning Fan, King-Hang Wang, J.-S. Pan, C-M. Chen, *Security analysis and improvement on an image encryption algorithm using Chebyshev generator, Journal of Internet Technology*, Vol. 20, No. 1, pp. 13-23, 2019

[16] T. Y. Wu, K.-H. Wang, C.-M. Chen, Jimmy M.-T. Wu, J.-S. Pan, A Simple Image Encryption Algorithm Based on Logistic Map, Advances in Intelligent Systems and Computing, volume 891, pp 241-247, 20181225. *The Fifth Euro-China Conference on Intelligent Data Analysis and Applications* (ECC 2018), October 12-14, 2018, Xi'an, China

[17] Q. Y. Li, Zhaolin Lu and Leida Li, Perceptual Evaluation of Scanning Electron Microscopy Imaging, Journal of Network Intelligence, vol. 4, no. 1, pp. 10-16, Feb 2019.

[18] Chate Harold and Niquice Nelta, Blind Images Quality Assessment of Distorted Screen Content Images, *Journal of Network Intelligence*, vol. 3, no. 2, pp. 91-101, May 2018.

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

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