Spatiotemporal Data Fusion Using Temporal High-Pass Modulation and Edge Primitives

J. Malleswara Rao, C. V. Rao, A. Senthil Kumar, B. Lakshmi, and V. K. Dadhwal

Abstract—Most spaceborne sensors have a tradeoff between high spatial and high temporal resolutions. This tradeoff limits the use of remote sensing data in various applications that require images in both the high spatial and high temporal resolutions. In this paper, we propose a novel technique to create a fine spatial and high temporal resolution images at a ground-based data processing system. Resourcesat-2 is one of the Indian Space Research Organization missions, and it carries the Linear Imaging Self-Scanning sensors (LISS III and LISS IV) and an Advanced Wide-Field Sensor (AWiFS) onboard. The spatial resolution of LISS III is 23.5 m, and that of AWiFS is 56 m. The temporal resolution of LISS III is 24 days, and that of AWiFS is five days. The proposed method creates a synthetic LISS-III image at 23.5-m spatial and five-day temporal resolutions. It is based on the subpixel relationship between a single AWiFS–LISS-III image pair, which is acquired before or after the prediction date. In temporal data composition, spurious spatial discontinuities are inevitable for land-cover type changes. These discontinuities were identified with temporal edge primitives and were smoothed with a spatial-profile-averaging method. A synthetic LISS-III image for time $t_k$ is predicted from an AWiFS image at time $t_k$, and a single AWiFS–LISS-III image pair at time $t_j$, where $t_j \neq t_k$. Experimental results demonstrated that the proposed method is superior in terms of the computational efficiency and prediction accuracy with the other existing methods.

Index Terms—Advanced Wide-Field Sensor (AWiFS) and Linear Imaging Self-Scanning (LISS-III) sensor data, Resourcesat-2, spatiotemporal data fusion.

I. INTRODUCTION

ACCURATE quantification of vegetation dynamics (phenology) is required in regional to global scale to improve ecosystem models and in the understanding of interannual variability in terrestrial carbon exchange and climate–biosphere interactions [1]. The vegetation dynamics and the structural information can be obtained through the satellite images, which are acquired in both high spatial and high temporal resolutions. However, there is a tradeoff between high spatial and high temporal resolutions when designing the spaceborne sensors. For example, high spatial resolution (HSR) sensors usually result in a smaller swath width, so that the revisit time is increased to view the same location on Earth [2]. Conversely, high temporal resolution sensors have a more frequent revisit rate and produce a larger swath width with a lower spatial resolution [3], [4]. These factors limit applications that require the data in both the high spatial and high temporal resolutions. Another possible way to obtain images at high spatial and temporal resolutions is to merge the data from sensors with different spatial and temporal characteristics.

The Indian Space Research Organization Resourcesat-2 mission aims to map and monitor the natural resources of the Earth’s surface [5]. The Resourcesat-2 Linear Imaging Self-Scanning (LISS-III) sensor’s data have proven its usefulness in mapping and quantifying the land-cover type changes [6]. Its 24-day temporal resolution is a longer period to monitor the biodynamics of the Earth’s surface and may create difficulty in mapping the vegetation dynamics in a timely manner. The spatial resolution of the LISS-III sensor is 23.5 m, and the temporal resolution of the Advanced Wide-Field Sensor (AWiFS) is five days. This paper reports a novel fusion method to merge the LISS-III sensor data with AWiFS sensor data to create a synthetic LISS-III image at 23.5-m spatial and five-day temporal resolutions.

In recent decades, many studies have concentrated on the fusion of a high-resolution panchromatic band and low-resolution multispectral bands from one or more sensors [7]. Intensity–hue–saturation, principal component analysis, and Grams–Schmidt fusion techniques are component-substitution fusion techniques [8]–[10]. Brovey, high-pass filtering, smoothing filter-based intensity modulation, and synthetic variable ratio fusion techniques are modulation-based fusion techniques [11]–[14]. The generalized Laplacian pyramid and wavelet- and curvelet-transform-based fusion techniques are multiscale analysis fusion techniques [15]–[17]. All these techniques are spatial and spectral fusion techniques. In contrast to this spatial and spectral fusion, spatiotemporal data fusion techniques predict the unknown HSR image from its low-spatial-resolution (LSR) observation and a prior LSR–HSR image pair.

In recent times, spatiotemporal data fusion methods have gained a significant importance because of its ability to exploit both spatial and temporal resolutions of different sensors. As per the literature, spatiotemporal data fusion methods are developed mainly based on mathematical modeling and single-image super-resolution concepts. The spatial and temporal adaptive reflectance fusion model (STARFM) [18], enhanced STARFM [19], and spatial and temporal adaptive algorithm for mapping reflectance change (STAARCH) methods [20] are based on the...
The sparse-representation-based spatial and temporal reflectance fusion model [21] and spatiotemporal data fusion proposed by Song and Huang [22] are recently developed methods. These methods are based on the single-image super resolution using dictionary learning (DL). These methods establish a correspondence between the patches of the Landsat image and their corresponding patches of the MODIS image using sparse representation for time \( t_0 \). With this correspondence, a synthetic Landsat image is predicted for time \( t_k \) from the MODIS image at time \( t_k \). The DL-based methods have a better performance in comparison with the STARFM and the ESTARFM methods. However, the methods proposed by Huang and Song [21], [22] have more computational complexity due to dictionary training and learning. The size of the dictionary needs to be increased with [23] the size of the images for DL-based methods. Learning the large dictionaries raises the computational time exponentially so that these two methods are tested on smaller size images. However, remote sensing applications demand computationally efficient methods to generate the high spatial and high temporal resolution images for the larger study areas.

Until now, the spatiotemporal data fusion methods are primarily defined and implemented for MODIS and Landsat images only. There is a large spatial resolution difference between these images. One pixel of the MODIS image is equivalent to around the 16 × 16 pixel block of the Landsat image. The MODIS images were interpolated to the size of a Landsat image in the prediction process. If the upsampling factor is large, the interpolation errors will be high. Consequently, the pixel-to-pixel correspondence suffers in geometric accuracy and the bias occurred in the interpolation of reflectance values. The spatial resolution ratio across AWiFS and LISS-III images is smaller compared with that of MODIS and Landsat images; i.e., one pixel of AWiFS is equivalent to approximately 2 × 2 pixels of LISS III so that the geometric registration and interpolation errors are very less. This equivalence relation is called a subpixel relationship between the pixel of AWiFS and its corresponding 2 × 2 pixels of LISS III.

Images acquired in different spatial resolutions have the problem of mixed pixels [24]. These mixed pixels occur at LSR, and the same pixel response can be distinguishable at fine spatial resolution (FSR). If the spatial resolution ratio is small, we can predict the individual pixels within the mixed-pixel response of the LSR image by establishing a subpixel relationship between the LSR pixel and its corresponding pixel block in the FSR image. This prediction requires prior information about the spatial weights of the individual pixels within the mixed pixel. In this paper, these spatial weights were derived by establishing a subpixel relationship between the previously acquired single AWiFS–LISS-III image pair, and therefore, a synthetic LISS-III image can be created at 23.5-m spatial and five-day temporal resolutions.

This paper is organized in four sections. In Section I, we described the motivation of the current research. Section II provides the theoretical basis of the proposed method. Section III illustrates the experimental results and comparisons with the state-of-the-art techniques. The conclusion of the current study is reported in Section IV.

II. METHODOLOGY

In this paper, the satellite images that are acquired at LSR with high temporal resolution (such as AWiFS) were called as low spatial and high temporal (LSHT) images. The images that are acquired at high spatial resolution with low temporal resolution (such as LISS III) were called as fine spatial and low temporal (FSLT) images. The aim of the current research is to create a synthetic fine spatial and high temporal (FSHT) image for time \( t_k \) with an LSHT image at time \( t_k \) and prior single LSHT–FSLT image pair at time \( t_0 \), where \( t_0 \neq t_k \).

The proposed method involves three stages. In the first stage, high-frequency details were injected into the LSHT image at time \( t_k \) by deriving the high-resolution spatial weights with a subpixel relationship between the single LSHT–FSLT image pair at time \( t_0 \) and the LSHT image at time \( t_k \). The spurious spatial discontinuities are inevitable in temporal composition of different data sets [15]. In the second stage, these spurious spatial discontinuities were detected with the temporal edge primitives. In the third stage, these spurious spatial discontinuities were smoothed with the spatial-profile-averaging method.

A. Stage 1: Temporal High-Pass Modulation

The LSHT and FSLT images were corrected for geometric accuracy and radiometric normalization and to nullify the atmospheric effects. The scale of the LSHT image was made proportionate to the FSLT image, i.e., one pixel of the LSHT image is equivalent to the corresponding 2 × 2 pixels of the FSLT image. If the 2 × 2 pixels of the FSLT image are homogeneous, then the corresponding pixel in the LSHT image will have the same reflectance, and it can be obtained by simple average of the four pixels’ reflectance values. If the 2 × 2 pixels of the FSLT image are heterogeneous, then the surface reflectance of the corresponding pixel in the LSHT image will have a mixed response. This mixed pixels’ reflectance can be obtained by the weighted average of the corresponding 2 × 2 pixels of the FSLT image at time \( t_k \), which is described by the following equation:

\[
\begin{align*}
    a(i,j,t_k) &= \sum_{p=2i-1}^{2i} \sum_{q=2j-1}^{2j} w(p,q,t_k) \cdot l(p,q,t_k) + \varepsilon_k \quad \text{(1)}
\end{align*}
\]

where \( a(i,j,t_k) \) is the LSHT image pixel value at location \((i,j)\) at time \( t_k \), \( l(p,q,t_k) \) is the FSLT image pixel value at location \((p,q)\) at time \( t_k \), \( w(p,q,t_k) \) is the weighted coefficients to aggregate the FSLT image pixels to the LSHT image pixel such that \( \sum_p \sum_q w(p,q,t_k) = 1 \), and \( \varepsilon_k \) is the difference in reflectance values between the aggregation of 2 × 2 pixels of the FSLT image and their corresponding LSHT image pixel. This difference in reflectance arises due to the spectral or spatial mixing of the pixels between two sensors for a given area. The mathematical determination of the 2 × 2 pixels of the FSLT image for a given LSHT image pixel is an ill-posed problem because the unique solution cannot be determined without additional information.
Suppose there are no changes in the land-cover type during the period of time \(t_0\) to \(t_k\), then \(\varepsilon_a = \varepsilon_k\). From (1), we have
\[
\sum_{p=2i-1}^{2i} \sum_{q=2j-1}^{2j} w(p, q, t_k) \cdot l(p, q, t_k) = a(i, j, t_k) - a(i, j, t_0) + \sum_{p=2i-1}^{2i} \sum_{q=2j-1}^{2j} w(p, q, t_0) \cdot l(p, q, t_0).
\]
(2)

This is the ideal situation that cannot be frequently fulfilled by the LSHT and FSLT images. The subpixel relationship between the pixels of LSHT and FSLT images is complex with factors as follows: 1) surface reflectance of the LSHT image pixels is not homogenous, and it can contain mixed-pixel response; 2) the Earth surface features may change from one class to another; 3) changes due to the crop growth; and 4) changes in the sun elevation angle and relative satellite azimuth angle with respect to the pixel. All these factors play an important role in altering the reflectance values from time \(t_0\) to \(t_k\).

In the prediction process, the temporal changes from time \(t_0\) to \(t_k\) were measured from the LSHT images, and spatial information was considered from the FSLT image at time \(t_0\). A synthetic FSHT image for time \(t_k\) is computed from (3) by using the LSHT image at time \(t_k\) and a single LSHT–FSLT image pair at time \(t_0\), i.e.,
\[
\begin{pmatrix}
(l(2i - 1, 2j - 1, t_k) & l(2i - 1, 2j, t_k) \\
(l(2i, 2j - 1, t_k) & l(2i, 2j, t_k))
\end{pmatrix}
= \begin{pmatrix}
(l(2i - 1, 2j - 1, t_0) & l(2i - 1, 2j, t_0) \\
(l(2i, 2j - 1, t_0) & l(2i, 2j, t_0))
\end{pmatrix}
+ [4 \cdot (a(i, j, t_k) - a(i, j, t_0))] \cdot W.
\]
(3)

The size of the LSHT image is \(m \times n\), and that of the FSLT image is \(2m \times 2n\). Equation (3) gives the \(2 \times 2\) pixel blocks of the FSHT image corresponding to a pixel \((i, j)\) in the LSHT image at time \(t_k\). In the second term, the temporal difference from time \(t_0\) to \(t_k\) is multiplied by a scalar value of 4. It contributes the temporal changes to 4 pixels in the \(2 \times 2\) pixel block of the FSLT image at time \(t_0\) with respect to the weight matrix \(W\). For example, if it is a homogeneous region, the \(2 \times 2\) weight matrix \(W\) contains equal weights. Hence, each pixel gets equal temporal difference in the \(2 \times 2\) pixel block of the FSLT image at time \(t_0\). For heterogeneous regions, the weight matrix \(W\) contains unequal weights. To manage spatial heterogeneity in 4 pixels of the \(2 \times 2\) pixel block, the temporal difference is multiplied by a scalar value of 4 and the weight matrix \(W\). The weight matrix \(W\) contains the proportional spatial weights for time \(t_k\). These weights were derived using the spatial information from the FSLT image at time \(t_0\) and the resampled LSHT images at time \(t_0\) and \(t_k\).

The small spatial resolution ratio between AWiFS and LISS-III sensors, the same time of acquisition, and the same spectral bandwidths are the ideal conditions to fuse the data from different sensors. These ideal conditions and spectral normalization of AWiFS with LISS III make the resampled AWiFS pixel values spectrally almost similar to the corresponding LISS-III pixel values, but they have spatial resolution difference. Therefore, there is no need to search for the spectrally similar pixels as in the STARFM method to derive the weight matrix \(W\).

The weight matrix \(W\) is represented as
\[
W = \begin{pmatrix}
w(2i - 1, 2j - 1, t_k) & w(2i - 1, 2j, t_k) \\
w(2i, 2j - 1, t_k) & w(2i, 2j, t_k)
\end{pmatrix}.
\]
It is defined in the following section.

1) Weight Matrix \(W\) Formulation: To determine the weight matrix, the LSHT images were extended to the size of the FSLT image by cubic convolution interpolation. Then, the pixels of FSLT and LSHT images are in one-to-one correspondence. The weight matrix \(W\) formulation is illustrated as follows.

Let \(\bar{a}\) be the resampled LSHT image to the size of the FSLT image. The temporal changes were obtained as
\[
T = |\bar{a}_{t_k} - \bar{a}_{t_0}|
\]
where \(\bar{a}_{t_k}\) is the \(2 \times 2\) pixel block in the interpolated LSHT image at time \(t_k\) corresponding to a pixel \((i, j)\) in an original LSHT image \(a\), similarly \(\bar{a}_{t_0}\). This temporal difference \(T\) measures the changes that are occurred in the \(2 \times 2\) pixel blocks from time \(t_0\) to \(t_k\). Any temporal changes that are detected by the LSHT image at time \(t_k\) are predicted in the FSHT image by using \(T\). The changes in spatial weights were computed by considering spatial information from the FSLT image at time \(t_0\) and the temporal changes from \(T\). Thus
\[
C = L_0 \cdot T
\]
where \(\cdot^*\) denotes element-wise multiplication, and
\[
L_0 = \begin{pmatrix}
l(2i - 1, 2j - 1, t_0) & l(2i - 1, 2j, t_0) \\
l(2i, 2j - 1, t_0) & l(2i, 2j, t_0)
\end{pmatrix}.
\]
The weight matrix \(W\) is given by
\[
W = C \cdot \sum_{p=2i-1}^{2i} \sum_{q=2j-1}^{2j} C(p, q)
\]
where \(\cdot^*\) denotes element-wise division.

The weight matrix \(W\) determined the fraction of temporal changes and the proportional weights for the spatial details to predict the FSHT image for time \(t_k\).

B. Stage 2: Temporal Edge Primitives

The human visual system is more sensitive to sharp contrast changes than the smooth regions of an image [25]. Sharp contrast changes form image primitives such as contours, edges, or ridges. In temporal data composition, spurious spatial discontinuities are inevitable [26] for land-cover type changes. The reason for these spurious spatial discontinuities is that the pixels around the changed feature are originated from images with different dates with different sun-pixel-sensor viewing geometries. We identified these spurious spatial discontinuities using temporal edge primitives, which were extracted with the following steps.

1) High-frequency details were extracted from the LSHT images at time \(t_0\) and \(t_k\) using a \(3 \times 3\) Laplacian high-pass...
5856 IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, VOL. 53, NO. 11, NOVEMBER 2015

Fig. 1. Identifying the spurious spatial discontinuities using temporal edge primitives. (a) Temporal edge primitives. (b) Spurious edges overlaid on the predicted image of stage 1.

\[
\begin{pmatrix}
-1 & -1 & -1 \\
-1 & 8 & -1 \\
-1 & -1 & -1
\end{pmatrix}
\]  It highlights the points, lines, and edges in the image and suppresses the uniform and smoothly varying regions.

2) From these high-frequency details, edges were detected with the Canny edge detection method. These edges were called as temporal edge primitives for time \( t_0 \) and \( t_k \).

3) Subtract the edges of time \( t_0 \) from the edges of time \( t_k \).

If the edges are in the same location at time \( t_0 \) and \( t_k \), both edges are cancelled. If the edges are in different locations at time \( t_0 \) and \( t_k \), both edges are retained. Black edges represent the edge pixels at time \( t_0 \) in Fig. 1(a). White edges represent the edge pixels at time \( t_k \) in Fig. 1(a). Due to subpixel-level geometric bias between the two LSHT images at time \( t_0 \) and \( t_k \), the same edge may represent in both black and white as connected edge pixels side by side, as shown in Fig. 1(a).

While the edge pixels are cancelled, the white edge pixels and the connected edge pixels are important to retain the spatial discontinuities for time \( t_k \). After ignoring all these edge pixels, only black edge pixels are identified as spurious spatial discontinuities, and they are overlaid on the predicted image of stage 1, as shown in Fig. 1(b). The surrounding pixels of these black edge pixels will have spurious pixel values, which were occurred due to land-cover type changes from time \( t_0 \) to \( t_k \). These spurious spatial discontinuities were smoothed by spatial profile averaging between the homogeneous pixels of the LSHT image at time \( t_k \) and the pixels around the black edges in the predicted image of stage 1.

C. Stage 3: Spatial Profile Averaging

Consider a \( 7 \times 7 \) window around the black edge pixels as 2-D spatial profiles in the predicted image of stage 1. At the same location, consider the pixel values in a \( 7 \times 7 \) window of the LSHT image at time \( t_k \) as 2-D spatial profiles. These 2-D profiles represent the shape of the surface. The shape-based averaging preserves the contiguous regions of the images [27]. The average of these two 2-D spatial profiles smoothed the spurious spatial discontinuities occurred at stage 1. The predicted image in stage 3 is shown in Fig. 2(b), which is almost similar to the original in Fig. 2(c). The yellow boxes in Fig. 2(b) show the smoothed spurious discontinuities. If the size of the window increases, spatial profile averaging will oversmooth the image. Fig. 3 shows the flow chart of the proposed method.

III. EXPERIMENTAL RESULTS AND COMPARISONS

The aim of the current study is to develop a computationally efficient spatiotemporal data fusion technique to predict the synthetic LISS-III image for time \( t_k \) from an AWiFS image at time \( t_k \) and a single AWiFS–LISS-III image pair at time \( t_0 \), where \( t_0 \neq t_k \). The prediction of land-cover type changes is more difficult than the crop phenology in spatiotemporal data fusion [10]. The STARFM and ESTARFM methods face difficulties in predicting the land-cover type changes [10]. The method recently developed by Song and Huang (2013) (SH method) [11] improves the prediction accuracy of land-cover type changes when compared with the STARFM and ESTARFM methods. However, the SH method [11] has high computational complexity.

The STARFM and ESTARFM methods are specifically defined for the MODIS and Landsat sensors. The SH method is also defined for the MODIS and Landsat sensors. However, it is based on the single-image super-resolution technique, and it uses a single MODIS–Landsat image pair as prior knowledge. The proposed method also uses a single AWiFS–LISS image pair as prior knowledge. The single-image super-resolution technique can be applied for any data set. Hence, the proposed method is compared with the SH method [11].
Since the MODIS and Landsat images have large spatial resolution difference, the SH method is implemented in two layers. Each layer contains two stages: a super-resolution stage and a high-pass modulation stage. The first layer output is an input to the second layer to improve the spatial resolution. Since the spatial resolution ratio of AWiFS and LISS-III images are very small when compared with that of MODIS and Landsat images, the proposed method is compared only with one layer framework of the SH method with a dictionary size of 800, and the number of iterations for the DL is 30.

In preprocessing, an AWiFS and the LISS-III image pixel values were converted into the reflectance values with the COST method. Furthermore, the reflectance values of the AWiFS image were normalized to the reflectance values of LISS III by the linear relationship between the homogeneous regions of the AWiFS image and their corresponding regions of the LISS-III image [28].

We evaluated the effect of the proposed method on seven different data sets. Two of them are shown in Figs. 4 and 5 for visual comparison, and their quantitative results are shown in Tables I and II. The quantitative results for the remaining five experimental data sets are given in Table III. The first two experimental data sets are 12 km × 12 km areas, and they are selected from the Pantnagar, Uttarakhand state of India. A single AWiFS–LISS-III image pair acquired on 24-Oct-2012 was used as prior knowledge to predict the synthetic LISS-III image for 17-Nov-2012 by using an AWiFS image acquired on 17-Nov-2012. These data sets have both crop phenology and land-cover type changes such as bare soil on 24-Oct-2012 converted into wetland on 17-Nov-2012, as shown in yellow boxes in Fig. 4(a) and (c). The remaining five data sets are also 12 km × 12 km areas, and they are selected from the Surathgarh and Hisar regions of India.

The comparison of the four bands (Green B2–Red B3–NIR B4–SWIR B5) between the proposed method and the SH method in terms of total computational time in seconds, root mean square error (RMSE), structural similarity index (SSIM) [29], and spectral angle mapper (SAM) is described in Table I. The total computational time for our method is 36 s, and that for the SH method is 2520 s; this indicates that the proposed method is computationally efficient, and our method is 70 times faster than the SH method. All the experiments were executed on an Intel Core 2 Duo CPU at 2.66 GHz with 2-GB RAM. The SH method is based on the DL. The DL-based methods involve in training the LR-HR image patches to identify the best dictionary to predict the desired high-resolution pixels. The size of the dictionary is directly proportional to the size of the image. The training and learning of the dictionary increase the computational time exponentially for large-size images. However, the proposed method is based on primitive matrix addition, scalar multiplication, and the detection of edges. Hence, the proposed method is computationally efficient.

The average RMSE of the four bands is 0.0152 and 0.0149 for the SH method and 0.0141 and 0.0122 for our method, for the experimental data sets 1 and 2, respectively. It indicates that the proposed method shows better image quality when compared with the SH method. In our method, the spurious discontinuities were smoothed, and hence, spurious values were minimized. These spurious values are occurred due to the data composition of multiresolution data sets, which are acquired in different times. For example, in Fig. 4(a), bare soil is converted into wetland in Fig. 4(c). These conversions are shown in yellow boxes in Fig. 4(a) and (c). Because of these rapid changes, spurious spatial discontinuities were occurred in Fig. 4(e). However, our method proposes a novel approach to identify these spurious edges at stage 2, and then, these edges were smoothed in stage 3. Hence, the predicted images, shown in Figs. 4(f) and 5(f), are more similar to the original image than the SH method. The yellow boxes in Figs. 4(f) and 5(f) show the better prediction patches of our method compared with the SH method in land-cover type changes. In these boxes, Figs. 4(e) and 5(e) contain spurious spatial discontinuities, whereas in Figs. 4(f) and 5(f), these spurious spatial discontinuities were smoothed with spatial profile averaging. Hence, the proposed method obtained less RMSE compared with the SH method.

The average SSIM of the four bands is 0.8664 and 0.8537 for the SH method and 0.8825 and 0.8782 for our method, for the experimental data sets 1 and 2, respectively. This shows that the proposed method retains better structural information than the SH method. The SSIM measures the spatial quality of the

Fig. 4. Comparison between the actual and the predicted reflectance values for the experimental data set 1. (a) AWIFS acquired on 24-oct-2012. (b) LISS III acquired on 24-Oct-2012. (c) AWIFS acquired on 17-Nov-2012. (d) Original LISS III acquired on 17-Nov-2012. (e) Synthetic LISS III for 17-Nov-2012 predicted with the SH method. (f) Synthetic LISS III for 17-Nov-2012 predicted with the proposed method.

Fig. 5. Comparison between the actual and the predicted reflectance values for the experimental data set 2. (a) AWIFS acquired on 24-oct-2012. (b) LISS III acquired on 24-Oct-2012. (c) AWIFS acquired on 17-Nov-2012. (d) Original LISS III acquired on 17-Nov-2012. (e) Synthetic LISS III for 17-Nov-2012 predicted with the SH method. (f) Synthetic LISS III for 17-Nov-2012 predicted with the proposed method.
structures locally in an image [18]. The spurious discontinuities were smoothed locally, and hence, the local patches of the synthetic image were almost similar to the original image in our method. Therefore, the proposed method obtained high SSIM values. The proposed method has a less spectral distortion than the SH method with reference to the SAM values shown in Table I. Table II shows the prediction accuracy $R^2$ comparison of our method with the SH method between the predicted reflectance and the actual reflectance. The average $R^2$ of the four bands is 0.77 and 0.73 for the SH method and 0.79 and 0.76 for our method, for the experimental data sets 1 and 2, respectively. This indicates that the proposed method predicts both crop phenology and land-cover type changes more accurately than the SH method.

With two experimental data sets, we cannot conclude the advantage of the proposed method. Hence, the proposed method was tested with five more different data sets containing both crop phenology and land-cover type changes. All these experiments demonstrated that the proposed method is superior to the SH method in all quantitative parameters, as shown in Table III. The total computational time for our method is 36 s, and that for the SH method is 2520 s for each experiment. Each value in Table III is the average value of the four bands’ quality parameter.

The prediction accuracy of the proposed method primarily depends on the subpixel correspondence between the $2 \times 2$ pixel block of LISS III and a corresponding pixel of AWiFS. In all the experiments, we ensured such a subpixel relationship. Originally, one pixel of AWiFS is equivalent to the $2.38 \times 2.38$ pixels of LISS III with respect to their spatial resolutions. The pixel values under the fractional indexes were interpolated to create

<table>
<thead>
<tr>
<th>Experimental Dataset</th>
<th>Method</th>
<th>Computational time (in sec)</th>
<th>SSIM</th>
<th>RMSE</th>
<th>SAM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>B2</td>
<td>B3</td>
<td>B4</td>
</tr>
<tr>
<td>1</td>
<td>Proposed Method</td>
<td>36</td>
<td>0.9107</td>
<td>0.8999</td>
<td>0.8642</td>
</tr>
<tr>
<td></td>
<td>SH Method</td>
<td>2520</td>
<td>0.9034</td>
<td>0.8828</td>
<td>0.8466</td>
</tr>
<tr>
<td>2</td>
<td>Proposed Method</td>
<td>36</td>
<td>0.8997</td>
<td>0.8912</td>
<td>0.8683</td>
</tr>
<tr>
<td></td>
<td>SH Method</td>
<td>2520</td>
<td>0.8710</td>
<td>0.8814</td>
<td>0.8402</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Experimental Dataset No.</th>
<th>$R^2$</th>
<th>SSIM</th>
<th>RMSE</th>
<th>SAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.87</td>
<td>0.85</td>
<td>0.9224</td>
<td>0.9057</td>
</tr>
<tr>
<td>2</td>
<td>0.87</td>
<td>0.86</td>
<td>0.8732</td>
<td>0.8615</td>
</tr>
<tr>
<td>3</td>
<td>0.91</td>
<td>0.89</td>
<td>0.9058</td>
<td>0.8955</td>
</tr>
<tr>
<td>4</td>
<td>0.80</td>
<td>0.79</td>
<td>0.8055</td>
<td>0.7902</td>
</tr>
<tr>
<td>5</td>
<td>0.88</td>
<td>0.87</td>
<td>0.9063</td>
<td>0.8930</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Experiments with a sub-pixel shift under a coarse resolution pixel in 8-directions</th>
<th>$R^2$</th>
<th>SSIM</th>
<th>RMSE</th>
<th>SAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>No shift (proposed approach)</td>
<td>0.8588</td>
<td>0.8789</td>
<td>0.0120</td>
<td>2.4606</td>
</tr>
<tr>
<td>south</td>
<td>0.8166</td>
<td>0.8291</td>
<td>0.0140</td>
<td>2.7838</td>
</tr>
<tr>
<td>south-west</td>
<td>0.8303</td>
<td>0.8429</td>
<td>0.0138</td>
<td>2.7681</td>
</tr>
<tr>
<td>west</td>
<td>0.8573</td>
<td>0.8701</td>
<td>0.0123</td>
<td>2.5513</td>
</tr>
<tr>
<td>north-west</td>
<td>0.8500</td>
<td>0.8688</td>
<td>0.0165</td>
<td>2.4983</td>
</tr>
<tr>
<td>north</td>
<td>0.8449</td>
<td>0.8652</td>
<td>0.0127</td>
<td>2.4701</td>
</tr>
<tr>
<td>north-east</td>
<td>0.8149</td>
<td>0.8199</td>
<td>0.0142</td>
<td>2.7920</td>
</tr>
<tr>
<td>east</td>
<td>0.8065</td>
<td>0.8183</td>
<td>0.0144</td>
<td>2.8255</td>
</tr>
<tr>
<td>south-east</td>
<td>0.7965</td>
<td>0.7811</td>
<td>0.0155</td>
<td>3.0873</td>
</tr>
</tbody>
</table>
a subpixel relationship. While doing so, we ensure the correspondence of the $2 \times 2$ pixels under a coarse-resolution pixel. Since every mixed pixel intrinsically contains the neighborhood information from the surrounding pixels, the inaccuracies due to mixed-pixel response were kept under control while assuming the resolution ratio at $1:2$. The geometric distortions were intentionally introduced by shifting the subpixels under a coarse-resolution pixel in all eight directions, which is an extreme possibility of mixed response due to all geometric inaccuracies. The proposed spatiotemporal data fusion is also tested on these data sets. From Table IV, we have observed that the approach presented in this paper performs better compared with the experiments with data sets having a subpixel shift under a coarse-resolution pixel. Therefore, it is considerable evidence for our assumption, i.e., the $2 \times 2$ pixel block of LISS III corresponds to the most relevant AWiFS mixed pixel in the proposed approach.

IV. Conclusion

A novel spatiotemporal data fusion has been developed to predict the synthetic LISS-III image for time $t_k$ from an AWiFS image at time $t_k$ and a single AWiFS–LISS-III image pair at time $t_0$, where $t_0 \neq t_k$. With this method, we can create a synthetic LISS-III image at 23.5-m spatial and five-day temporal resolutions, whereas the temporal resolution of LISS III is 24 days. The proposed method includes three stages. In the first stage, high-frequency details are injected into the AWiFS image at time $t_k$ by deriving the high-frequency details for time $t_k$ from a prior knowledge of a single AWiFS–LISS-III image pair at time $t_0$ and an AWiFS image at time $t_k$. Spurious spatial discontinuities are inevitable in temporal composition of multiresolution data sets, which are acquired in different times. In the second stage, these spurious spatial discontinuities are detected with the temporal edge primitives. In the third stage, these spurious spatial discontinuities are smoothed with spatial profile averaging. The proposed method is also applicable to other sensors that have similar characteristics as the AWiFS and LISS-III sensors. The SH method is applicable to two different sensors whose spatial ratio is up to 16, but the proposed method works well for 2. However, the computational complexity of the SH method is very high even for small spatial resolution ratios. This is due to training and learning of the low- and high spatial resolution dictionaries. Hence, the SH method is tested on the small-size images. However, our method is most useful for the spaceborne sensors whose spatial resolution ratio is 2 and to synthesize the images for larger areas at high spatial and high temporal resolutions. Moreover, our method creates the synthetic images of size 6000 × 6000 in 30 min on an Intel Core 2 Duo CPU at 2.66 GHz with 2-GB RAM system. All experimental results demonstrated that the proposed method is computationally efficient and predicts crop phenology and land-cover type changes more accurately than the SH method. Remote sensing applications that require LISS-III images at 23.5-m spatial and five-day temporal resolutions can benefit with the proposed technique.

REFERENCES


J. Malleswara Rao received the M.Sc. degree in mathematics from the Acharya Nagarjuna University, Guntur, India, in 2004 and the M.Tech. degree in spatial information technology from the Jawaharlal Nehru Technological University, Hyderabad, India, in 2009, where he is currently working toward the Ph.D. degree in digital image processing.

He is currently a Senior Research Fellow with National Remote Sensing Center, Hyderabad. His research interests include image super-resolution, spatiotemporal data fusion, and pattern recognition.

C. V. Rao received the M.E. degree in digital systems and engineering from Osmania University, Hyderabad, India, in 1985 and the Ph.D. degree in computer science and engineering from the Jawaharlal Nehru Technological University, Hyderabad, in 2009. He is currently the Head of the Special Products Division, National Remote Sensing Center, Indian Space Research Organization, Hyderabad. His research interests include image fusion, spatiotemporal data fusion, and image classification.

A. Senthil Kumar received the M.Sc. (Engg.) and Ph.D. degrees in image processing from the Indian Institute of Science, Bangalore, India, in 1985 and 1990, respectively.

He is currently the Group Director of the Geophysical and Special Products Group with the National Remote Sensing Center, Hyderabad, India. His research interests include sensor characterization, radiometric data processing, and geophysical product development from Indian Remote Sensing satellite images.

B. Lakshmi received the M.Sc. degree in physics and electronics from Sri Venkateswara University, Tirupathi, India, in 1975.

She is currently a Deputy Director with the Satellite Data Acquisition and Products Services Area, National Remote Sensing Center, Indian Space Research Organization, Hyderabad, India. Her research interests include data security, distributed architecture for remote sensing data services, and real-time satellite data processing.

J. Malleswara Rao received the M.Sc. degree in mathematics from the Acharya Nagarjuna University, Guntur, India, in 2004 and the M.Tech. degree in spatial information technology from the Jawaharlal Nehru Technological University, Hyderabad, India, in 2009, where he is currently working toward the Ph.D. degree in digital image processing.

He is currently a Senior Research Fellow with National Remote Sensing Center, Hyderabad. His research interests include image super-resolution, spatiotemporal data fusion, and pattern recognition.

C. V. Rao received the M.E. degree in digital systems and engineering from Osmania University, Hyderabad, India, in 1985 and the Ph.D. degree in computer science and engineering from the Jawaharlal Nehru Technological University, Hyderabad, in 2009. He is currently the Head of the Special Products Division, National Remote Sensing Center, Indian Space Research Organization, Hyderabad. His research interests include image fusion, spatiotemporal data fusion, and image classification.

A. Senthil Kumar received the M.Sc. (Engg.) and Ph.D. degrees in image processing from the Indian Institute of Science, Bangalore, India, in 1985 and 1990, respectively.

He is currently the Group Director of the Geophysical and Special Products Group with the National Remote Sensing Center, Hyderabad, India. His research interests include sensor characterization, radiometric data processing, and geophysical product development from Indian Remote Sensing satellite images.

B. Lakshmi received the M.Sc. degree in physics and electronics from Sri Venkateswara University, Tirupathi, India, in 1975.

She is currently a Deputy Director with the Satellite Data Acquisition and Products Services Area, National Remote Sensing Center, Indian Space Research Organization, Hyderabad, India. Her research interests include data security, distributed architecture for remote sensing data services, and real-time satellite data processing.

V. K. Dadhwal received the M.Sc. and Ph.D. degrees in plant physiology from the Indian Agricultural Research Institute, New Delhi, India, in 1978 and 1983, respectively.

He is currently a Director with the National Remote Sensing Center, Indian Space Research Organization, Hyderabad, India. His research interests include crop modeling, remote sensing applications in agriculture, terrestrial carbon cycle, land-cover–land-use change modeling, and land surface processes.