I.J. Image, Graphics and Signal Processing, **2015, 10, 47-59**
Published Online September 2015 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijigsp.2015.10.06

Spatiotemporal Data Fusion using Dictionary Learning and Temporal Edge Primitives

J. Malleswara Rao, C. V. Rao, A. Senthil Kumar, B. Gopala Krishna, V. K. Dadhwal
National Remote Sensing Centre, Indian Space Research Organization, Hyderabad, India.
Email: rao_cv@nrsc.gov.in

Abstract—Technological limitations restrict to acquire an image at high spatial and high temporal resolutions with space borne global sensors. In this paper, we propose a novel technique to create such images at ground-based data processing system. The Resourcesat-2 is one of the Indian Space Research Organization (ISRO) global missions and it carries Linear Imaging and Self-Scanning Sensors (LISS III and LISS IV) and an Advanced Wide-Field Sensor (AWiFS). 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 5 days. Objective of the paper is to create a synthetic LISS III image at 23.5 m spatial and 5-day temporal resolutions. A synthetic LISS III image at time t_k is created from an AWiFS image at time t_k and a single AWiFS–LISS III image pair at time t_0 which is acquired before or after the prediction time t_k, here $t_0 \neq t_k$. The proposed method involves three phases. The first is super resolution phase. In this phase, two transition images are obtained for the time t_0 and t_k by improving AWiFS spatial resolution. The second is high pass modulation phase. In this phase, the high frequency details which are obtained in the difference of LISS III image and the transition image of time t_0 are proportionally injected into the transition image at time t_k. In composition of multi-temporal images of different spatial resolutions, spurious spatial discontinuities are inevitable. In the third phase, these spurious discontinuities are identified and smoothed with the spatial-profile-averaging method. The proposed method achieves better prediction accuracy when compared to the state-of-the art techniques.

Index Terms—Spatiotemporal data fusion, AWiFS and LISS III sensor data, Dictionary learning, Sparse coding.

I. INTRODUCTION

Space borne global sensors observe synoptically the dynamic changes on the Earth's surface at predefined intervals. This observed data plays a key role to understand the global water, carbon, and nitrogen cycles [10]. The structural information of the Earth's surface can be obtained from the high spatial resolution sensor and its temporal variability can be obtained from the high temporal resolution sensor. But, there is a trade-off between high spatial and high temporal resolutions in designing space borne global sensors. For example, high spatial resolution sensors usually result in a smaller swath width, so that the revisit time is increased to view the same location on the Earth [5]. Conversely, high temporal resolution sensors have a more frequent revisit rate and produce larger swath width with a lower spatial resolution ([11]; [17]). These factors limit the applications that require the data in both high spatial and high temporal resolutions. Another possible way to meet the requirement is to merge the data from sensors with different spatial and temporal resolutions.

The ISRO Resourcesat-2 mission aims for the natural resources mapping and monitoring of the Earth's surface [17]. The Resourcesat-2 LISS III sensor's data have proven utility for land cover type changes [19]. Its 24-day temporal resolution is a long period for monitoring the bio dynamics of the Earth's surface features and may create difficulties in mapping vegetation conditions in a timely manner. The spatial resolution of LISS III sensor is 23.5 m and the temporal resolution of AWiFS sensor is 5 days. This paper presents 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 5-day temporal resolutions.

Recently, the spatiotemporal data fusion methods are developed through single-image super resolution techniques. These techniques are implemented with the dictionary learning and the sparse representation methods (Hunag and Song.2012; [21]). The sparse representation methods use low spatial resolution (LR) and high spatial resolution (HR) image patches to derive low and high resolution dictionaries. Each patch in an image can be represented as a linear combination of atoms in a dictionary and the corresponding sparse coefficients [26]. Assume that the dictionary which describes the image patches as a periodic table of the fundamental elements (atoms) in the Chemistry. The combination of atoms forms a molecule, in a similar manner the linear combination of atoms in a dictionary and the sparse coefficients forms an image patch [7].

Yang et al (2010) [23] proposed a single image super resolution based on the sparse representation. This method jointly trains two dictionaries for high frequency details of LR and HR images, and these dictionaries are used to enforce the similarity of sparse representations between the low and high resolution patch pair. Compared to the Freeman et.al (2002) [9] approach, the learned dictionary pair is more compact representation of image patch pairs and it produces more competitive results.

The “Sparse representation-based Spatio-Temporal reflectance Fusion Model” (SPSTFM) developed by Huang and Song (2012) [12]. It is based on the single-image-super resolution though dictionary learning. The HR and LR dictionaries are jointly trained with the image patches of HR and LR images. With these two dictionaries, it establishes a relationship between the structures within the HR images and their corresponding LR images. This model requires two MODIS (500 m)-Landsat (30 m) image pairs at time t_1 and t_3 to predict a synthetic Landsat image for the MODIS image at time t_2 $(t_1 < t_2 < t_3)$. Whereas single-image-super resolution uses numerous LR-HR image pairs for a prior knowledge for the desired HR image.

Due to cloud coverage and low revisit rate, the two Landsat images may not be available. Moreover, it is very time consuming to identify these two Landsat images for a particular observation period $(t_1 < t_2 < t_3)$. To overcome this difficulty, the Song and Huang (2013) [21] developed a spatiotemporal data fusion method with only one MODIS-Landsat image pair. This fusion is achieved in two stages. In the first stage, transition images are obtained for the time t_1 and t_2. The transition images are derived from the LR and HR dictionaries. The LR dictionary is trained with the gradient feature image patches of MODIS image and the HR dictionary is obtained by forming a minimization problem with high frequency image patches of Landsat image and the same sparse coefficients which are obtained in the LR dictionary training. In second stage, the high frequency details which are obtained from the difference image of Landsat image and the transition image of time t_1 are proportionally injected into the transition image of time t_2. Therefore the synthetic Landsat image for the time t_2 is obtained with the MODIS image at time t_2 and a single MODIS-Landsat image pair at time t_1.

Spatiotemporal data fusion makes it possible to generate synthetic images at high spatial and high temporal resolutions. In synthesis of multi-sensor and multi-temporal data, spurious spatial discontinuities are inevitable [13]. The aforementioned methods have not dealt with these spurious spatial discontinuities. We propose a novel approach to minimize these spurious spatial discontinuities in the spatiotemporal data fusion.

Objective of the paper is to create a synthetic LISS III image for the 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 LR dictionary was trained with the gradient feature image patches of an AWiFS image for time t_0 using K-SVD algorithm [1], and the corresponding HR dictionary was derived by formulation of minimization problem with the high frequency image patches of LISS III image at time t_0 and the same sparse coefficients which were derived in the LR dictionary training. The two transition images were derived using these LR and HR dictionaries and an orthogonal-matching-pursuit (OMP) sparse coding technique [6]. Then, the high frequency details which were obtained in the difference of LISS III image and the transition image at time t_0 were proportionally injected into the transition image at the time t_k. The resultant image contains spurious spatial discontinuities. These discontinuities were identified with the temporal edge primitives and smoothed with the spatial-profile-averaging method.

II. Related study

A. Sparse representation and dictionary learning

If values of a vector $\boldsymbol{x}$ are most of them close to zero, that vector is said to be a sparse. A signal $\boldsymbol{x}$ is not sparse, we can make it sparse by transforming into another domain, for example, if $\boldsymbol{x}$ is sine, clearly it is not sparse. But Fourier transform of $\boldsymbol{x}$ is sparse [22].

A signal $\boldsymbol{x}$ can be written as a linear combination of T elementary waveforms (signal atoms), such that

$$\boldsymbol{x} = \boldsymbol{D\alpha}) \tag{1}$$

Where $\boldsymbol{\alpha} = [x_1, x_2, \dots, x_M]^{\mathrm{T}}$ and $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \dots, \alpha_N]^{\mathrm{T}}$, α_i are called the sparse representation coefficients of $\boldsymbol{x}$ in the dictionary $\boldsymbol{D} = [\boldsymbol{d_1}, \boldsymbol{d_2}, \dots, \boldsymbol{d_M}]$. Suppose the size of the signal $\boldsymbol{x}$ is $N \times 1$, size of $\boldsymbol{\alpha}$ is $M \times 1$ and the size of the dictionary $\boldsymbol{D}$ is $N \times M$ and whose columns $\boldsymbol{d_j}$ are the atoms, in general normalized to a unit l_2 norm, i.e. for all $j = 1,2, \dots, M$, $\|\boldsymbol{d_j}\|^2 = \sum_{i=1}^{N} |d_{ij}|^2$=1.

The dictionary $\boldsymbol{D} = [\boldsymbol{d_1}, \boldsymbol{d_2}, \dots, \boldsymbol{d_T}]$ is an indexed collection of atoms $\boldsymbol{d_j}$**.** The interpretation of $\boldsymbol{d_j}$ depends upon the dictionary. For example, frequency for the Fourier dictionary (i.e. sinusoids), Kronecker basis (standard unit vectors) for the Dirac dictionary, position-scale for the wavelet dictionary, translation-duration-frequency for cosine packets, and position-scale-orientation for the curvelet dictionary in two dimensions [22]. All these dictionaries are fixed dictionaries to their corresponding mathematical transforms. The sparse representation provides further sparsity by learning the over complete dictionaries directly on the data. An over complete dictionaries are occurred if the number of coefficients are larger than the number of signal samples (i.e. $M > N$). This redundancy is avoided in data compression applications. But in image restoration applications this redundancy had proven the advantage to have fast analysis and synthesis [22]. The aforementioned sparse representation formulation in Equation (1) is for a single vector $\boldsymbol{x}$. Now, the following paragraphs describe the dictionary learning procedure for the image patches.

Mathematical transforms may not always represent all the contained features in an image, because they are based on the fixed dictionaries. Sparse representation methods facilitate to learn the dictionary adaptively from the local structures within the image [20]. Each image patch $\boldsymbol{x_i}$ of size $\sqrt{n} \times \sqrt{n}$ is lexicographically ordered and stacked into a matrix $\boldsymbol{X} = [\boldsymbol{x_1}, \boldsymbol{x_2}, \dots, \boldsymbol{x_N}]$, where each $\boldsymbol{x_i}$ is a column vector of size $n \times 1$ and hence the size of the matrix $\boldsymbol{X}$ is $n \times N$. To find the best dictionary to represent a given training samples X, we have to solve the following optimization problem.

$$\min_{\boldsymbol{D},\boldsymbol{S}} \|\boldsymbol{X}-\boldsymbol{DS}\|_2^2 \ subject\ to\ \|\boldsymbol{s_i}\|_0 \le k_0\ for\ all\ i. \quad (2)$$

Where $\boldsymbol{D} = [\boldsymbol{d_1}, \boldsymbol{d_2}, \ldots, \boldsymbol{d_M}]$ and $\boldsymbol{S} = [\boldsymbol{s_1}, \boldsymbol{s_2}, \ldots, \boldsymbol{s_N}]$, each $\boldsymbol{d_i}$ is of size $n \times 1$ and hence the size of $\boldsymbol{D}$ is $n \times M$, each $\boldsymbol{s_i}$ is of size $M \times 1$ and hence the size of $\boldsymbol{S}$ **is** $M \times N$. The term $\|\boldsymbol{s_i}\|_0$ means that the number of non zero terms in the sparse column vector $\boldsymbol{s_i}$ is fixed and predetermined number of coefficients k_0. The above optimization problem can be solved with K-Singular Value Decomposition (K-SVD) algorithm [1]. The K-SVD algorithm is an iterative process to identify the best dictionary and sparse coefficients matrix. It updates an atom in the dictionary and the corresponding sparse coefficients at the same time. This simultaneous updating is speed up the convergence of the training process [7]. More details about the K-SVD can be found in Aharon et al. (2006) [1].

We found the dictionary $\boldsymbol{D}$ and the corresponding sparse coefficients matrix $\boldsymbol{S}$ for a given training samples $\boldsymbol{X}$ using the K-SVD algorithm. Now, we can find the sparse coefficients for a specific patch $\boldsymbol{x}$ (lexicographically ordered) with respect to the trained dictionary $\boldsymbol{D}$ in Equation (2) by using following approximation formula.

$$\hat{\boldsymbol{s}} = arg\ \min_{\boldsymbol{s}} \|\boldsymbol{s}\|_0 \ \ subject\ to\ \|\boldsymbol{x}-\boldsymbol{Ds}\|_2^2 \le \varepsilon \quad (3)$$

Here ε is the error tolerance. The above sparse approximation problem can be solved with the sparse coding technique such as Orthogonal-Matching-Pursuit (OMP) [6]. The OMP start initially with $\boldsymbol{s^k = 0}$, it iteratively constructs a k-term aproximant $\boldsymbol{s^k}$ by keeping set of active columns in the dictionary $\boldsymbol{D}$ at each stage. More details about the active columns selection is described in Elad. (2010) [7].The chosen active columns reduces the error $\|\boldsymbol{x}-\boldsymbol{Ds}\|_2^2$ at each stage, if it falls below the specified error threshold ε, the process terminates.

B. Single-image-super resolution through dictionary learning

The single-image-super resolution techniques enhance the spatial resolution of a single low spatial resolution image. The single-image-super resolution ([2], [9], [4], [7]), first proposed by and Kanade (2000) [2], attempts to capture prior correspondence between LR-HR image patches. This correspondence is derived from the numerous LR-HR image pairs and applies this correspondence to predict an HR image from a LR image. Jiji et al. (2004) [15] proposed the wavelet transform-based single-image super resolution. The wavelet coefficients are learned from a set of LR-HR training image pairs at finer scales for the desired high resolution image. The learned image further regularizes with smoothness prior while super-resolving a given low resolution image. Wu et al. (2011) [25] proposed contourlet transform based single-image super resolution. A low-resolution (LR) image is decomposed sub bands with the contourlet transform. The unknown high-frequency coefficients for the input low-resolution image are inferred by a belief propagation algorithm. These high-frequencies are obtained from the Markov Random Fields (MRF) model together with the contourlet coefficients of numerous LR-HR image pairs in the training data.

The wavelet and contourelet transforms are based on the fixed dictionaries. In recent times, single -image-super resolution methods use adaptive dictionaries which are learned from the local structures within an image. Yang et al (2010) [23] first proposed a super resolution method based on the dictionary learning and the sparse representation. The low and high resolution dictionaries are jointly trained from the low and high resolution image patches. It enforces the similarity of sparse representations between the low and high resolution image patch pairs with respect to their LR and HR dictionaries. Hence, the linear combination of the sparse coefficients of a given low resolution image patch and the atoms in high resolution dictionary produces a high resolution image patch.

Zeyde et al. (2012) [26] improves the Yang's algorithm in terms of the computational complexity and the algorithm architecture. Instead of training both low and high resolution dictionaries, Zyede's method trained only low resolution dictionary using K-SVD algorithm. The high resolution dictionary is derived by formulation of minimization problem with the high frequency image patches of high resolution image and the sparse coefficients matrix which is derived in the low resolution dictionary training. The sparse-land models on the image patches works as regularization term or prior information in both Yang et al (2010) [23] and Zeyde et al. (2012) [26] methods. The aforementioned methods use numerous LR-HR image pairs for the dictionary training. But our spatiotemporal fusion method used only one LR-HR image pair. The multi-temporal images of the same geographic area have the high structural similarity between two consecutive dates so that we have used only one LR-HR image pair. In this paper, we used K-SVD algorithm to train the dictionary for the low and high resolution image patches. The sparse representation of a given image patch was obtained with the sparse coding technique OMP.

III. Proposed Methodology

The aim of the study is to create a synthetic LISS III image at time t_k with an AWiFS image at time t_k and a single AWiFS–LISS III image pair at time t_0, here $t_0 \neq t_k$. Such synthetic LISS III image can be created through a spatiotemporal data fusion technique. Recently, Song and Hunag (2013) [21] developed a spatiotemporal data fusion based on the single-image-super resolution technique. This method uses a single MODIS-Landsat image pair as a prior knowledge. Originally, Song and Huang (2013) [21] method is two layer framework, each layer consists of two stages, the first is a super resolution stage and the second is a high pass modulation stage. Since the spatial resolution ratio of AWiFS and LISS III images are very small when compared to that of MODIS and Landsat images, the proposed method was adopted only one layer framework of the Song and Huang (2013) method [21]. In composition multi-sensor and multi-temporal images, spurious spatial discontinuities are inevitable. The Song and Huang (2013) [21] method have not deal with these spurious spatial discontinuities. Moreover, this method enhances the spatial resolution of low resolution image with super resolution technique and then high pass modulation technique. But both super-resolution and the high-pass-modulation techniques improve the spatial resolution. Their sequential implementation over enhance the spurious spatial discontinuities in the synthetic images. In this paper, we propose a novel approach to identify these spurious spatial discontinuities through temporal edge primitives. These discontinuities were smoothed with the spatial profile averaging method.

A. Super resolution of low spatial resolution image

The ideal parameters in consideration to fuse the data from different sensors are identical spectral bands, similar solar geometry, similar orbital parameters and the same time of acquisition [24]. The ISRO Resourcesat missions make it possible to acquire AWiFS, LISS III data simultaneously at identical spectral bandwidths, different spatial and radiometric resolutions at various swaths. Hence, AWIFS and LISS images met the most of the ideal conditions due to the acquisition from the same platform. Although AWiFS and LISS III sensors have similar spectral bandwidths, surface reflectance values of AWiFS were normalized to the reflectance of LISS III by the linear relationship between the homogeneous regions of AWiFS image and their corresponding regions of LISS III image [3].

In this section, the spatial resolution of AWiFS image was improved by the dictionary learning. The improved AWiFS images were called as the “transition images” as in Song and Huang (2013) method [21]. Denote AWiFS, LISS III and the transition images at time t_0 as $\boldsymbol{A_0}$, $\boldsymbol{L_0}$ and $\boldsymbol{T_0}$, respectively. The super resolution of low spatial AWiFS image contains two steps: The low and high resolution dictionary training and the transition image prediction. For low resolution dictionary training, extract the 5×5 patches from the gradient feature space of $\boldsymbol{A_0}$. These patches were rearranged into columns and these columns were stacked into a training sample matrix $\boldsymbol{X}$. For high resolution dictionary, extract the 5×5 patches from the difference image $\boldsymbol{L_0} - \boldsymbol{A_0}$. These patches were also rearranged into columns and these columns were stacked into a training sample matrix $\boldsymbol{Y}$. The columns of $\boldsymbol{X}$ and $\boldsymbol{Y}$ are in one-to-one correspondence. The low resolution dictionary $\boldsymbol{D_l}$ is derived with following optimization problem

$$\{\boldsymbol{D_l}^*, \boldsymbol{S}^*\} = \underset{\boldsymbol{D_l}, \boldsymbol{S}}{\mathbf{arg\,min}} \|\boldsymbol{X} - \boldsymbol{D_l S}\|_2^2 \quad subjec\ to \quad \|\boldsymbol{s_i}\|_{\mathbf{0}} \le k_0\ for\ all\ i. \tag{4}$$

Where $\boldsymbol{D_l}^*, \boldsymbol{S}^*$ was computed with the K-SVD algorithm. And $\boldsymbol{S}^*$ is the sparse coefficients matrix for the training samples $\boldsymbol{X}$ with respect to the dictionary $\boldsymbol{D_l}^*$. To establish a relationship between low and high resolution dictionary, the same sparse coefficients matrix $\boldsymbol{S}^*$ was used to derive the high resolution dictionary. The high resolution dictionary $\boldsymbol{D_h}$ was constructed by minimizing the error on $\boldsymbol{Y}$ with the same sparse coefficients matrix $\boldsymbol{S}^*$

$$\boldsymbol{D_h}^* = \mathbf{arg\,min} \quad _{\boldsymbol{D_h}} \|\boldsymbol{Y} - \boldsymbol{D_h S}^*\|_2^2\,) \tag{5}$$

The Equation (5) can be solved directly with pseudo inverse (here $\boldsymbol{S}^*$ is a full rank matrix)

$$\boldsymbol{D_h} = \boldsymbol{Y}(\boldsymbol{S}^*)^+ = \boldsymbol{Y}\boldsymbol{S}^{*\boldsymbol{T}}\left(\boldsymbol{S}^*\boldsymbol{S}^{*\boldsymbol{T}}\right)^{-\mathbf{1}} \tag{6}$$

Now, we have both low and high resolution dictionaries $\boldsymbol{D_l}$ and $\boldsymbol{D_h}$. The transition image $\boldsymbol{T_0}$ was derived from $\boldsymbol{A_0}$, extract the 5×5 patches from the gradient feature space of $\boldsymbol{A_0}$, rearranged into columns, and these columns were stacked into a matrix $\boldsymbol{X_0}$. Denote the columns of $\boldsymbol{X_0}$ as $\boldsymbol{x_{0i}}$ and its sparse coefficients column $\boldsymbol{s_i}$ with respect to the dictionary $\boldsymbol{D_l}$was obtained with the sparse coding technique OMP. This sparse coefficients column $\boldsymbol{s_i}$ was used to obtain the corresponding high resolution patch. Since the low and high resolution training samples are enforced to represent the same set of sparse coefficients columns, the corresponding high resolution patch column can be predicted with $\boldsymbol{y_{0i}} = \boldsymbol{D_h} * \boldsymbol{s_i}$. After rearranging all the columns $\boldsymbol{y_{0i}}$ into patches, we obtained a difference image $\boldsymbol{Y_0}$ between $\boldsymbol{T_0}$ and A_0, and hence $\boldsymbol{T_0} = \boldsymbol{Y_0} + \mathrm{A}_0$ .In a similar manner, the transition image $\boldsymbol{T_k}$ was predicted for the time t_k.

B. Spatiotemporal data fusion via high pass modulation

The training image pair AWiFS $\boldsymbol{A_0}$ – LISS III $\boldsymbol{L_0}$ at time t_0 has many similarities with the image pair AWiFS $\boldsymbol{A_k}$ – LISS III $\boldsymbol{L_k}$ at time t_k in regard to crop phenology and land-cover-type changes in most of the time except sudden land-cover-type changes. The predicted transition images $\boldsymbol{T_0}$ and $\boldsymbol{T_k}$ are almost similar to the LISS III images. As in Song and Huang (2013) [21] method, we assume that the pixel purity is same between $\boldsymbol{L_0}$ ($\boldsymbol{L_k}$) and $\boldsymbol{T_0}$ ($\boldsymbol{T_k}$) and there is a linear relationship between $\boldsymbol{T_0}$ and $\boldsymbol{T_k}$ in temporal changes. Then

$$T_k(i,j) = aT_0\,(i,j) + b \tag{7}$$

Where (i,j) is a given pixel location and a and b are linear regression coefficients for relative temporal change from the time t_0 to t_k. Since the pixel purity is same between $\boldsymbol{L_0}$ ($\boldsymbol{L_k}$) and $\boldsymbol{T_0}$ ($\boldsymbol{T_k}$), it is reasonable to assume that the linear relationship between $\boldsymbol{L_0}$ and $\boldsymbol{L_k}$.

$$L_k(i,j) = aL_0\,(i,j) + b \tag{8}$$

From Equation (7) and Equation (8), we obtain

$$L_k(i,j) = T_k\,(i,j) + a[L_0\,(i,j) - T_0\,(i,j)\,] \qquad (9)$$

The scalar value a can be approximated as $T_k\,(i,j)/T_0\,(i,j)$. Then Equation (9) is written as

$$\boldsymbol{L_k} = \boldsymbol{T_k} + \left(\frac{\boldsymbol{T_k}}{\boldsymbol{T_0}}\right)[\boldsymbol{L_0} - \boldsymbol{T_0}\,] \qquad (10)$$

Equation (10) shows that the high frequency information of time t_0 proportionally transfers to the time t_k by using high pass modulation, which is also widely used in pan sharpening. Since $\boldsymbol{L_0}$, $\boldsymbol{T_0}$ and $\boldsymbol{T_k}$ are already known, we can predict synthetic LISS III image $\boldsymbol{L_k}$ for the time t_k.

C. Smoothing the spurious spatial discontinuities

In this paper, section 3.1 and 3.2 describe the first and the second phase of the proposed method, which are similar to the one layer framework of the Song and Huang (2013) method [21]. In third phase, we propose a novel approach to identifying and smoothing the spurious spatial discontinuities which are occurred in the spatiotemporal data fusion of multi-sensor and multi-temporal images. In the third phase, there are two stages. In the first stage, spurious spatial discontinuities were identified with the temporal edge primitives. In the second stage, these spurious discontinuities were smoothed with spatial-profile-averaging method.

C.1. Stage 1: Temporal edge primitives

The human visual system is more sensitive to sharp contrast changes than the smooth regions in an image (Ji et al.1994). Sharp contrast changes form image primitives such as contours, edges or ridges. In temporal data composition, spurious spatial discontinuities are inevitable for land-cover-type changes [13]. The reason for these spurious spatial discontinuities is the pixels around the changed feature are originated from images with different days with different sun-pixel-sensor viewing geometries. We identified these spurious spatial discontinuities using temporal edge primitives which were extracted with the following steps.

(a) High frequency details were extracted from AWiFS images at the time t_0 and t_k using 3×3 Laplacian high-pass filter $\begin{pmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{pmatrix}$. It highlights points, lines and edges in the image and suppresses uniform and smoothly varying regions.

(b) From these high frequency details, edges are detected with Canny edge detection method. These edges were called as temporal edge primitives at the time t_0 and t_k.

(c) Subtract the edges at the time t_0 from the edges at the time t_k.

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

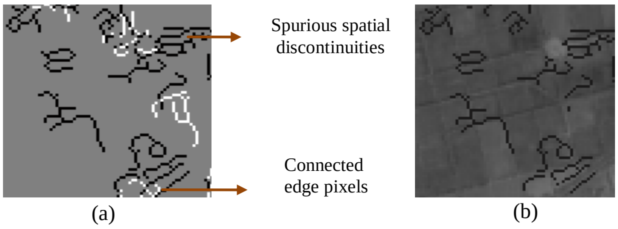


Fig.1. Indentifying the Spurious Spatial Discontinuities using Temporal Edge Primitives (a) Temporal Edge Primitives (b) Spurious Spatial Discontinuities in Stage 1

The edge pixels which are cancelled, white color edge pixels and connected edge pixels are important to retain the spatial discontinuities for the time t_k. After ignoring all these edge pixels, only black color edge pixels are identified as spurious spatial discontinuities and they are overlaid on the predicted image in section 3.2 as shown in Fig.1 (b). The surrounding pixels of these black color edge pixels will have spurious pixel values which were occurred due to land-cover -type changes from the time t_0 to t_k . These spurious spatial discontinuities are smoothed by the spatial profile averaging between the homogeneous pixels of AWiFS image at the time t_k and the pixels around the black colour edges in the predicted image in section 3.2.

C.2. Stage2: Spatial profile averaging

Consider a 7×7 window around the black colour edge pixels as 2-dimensional spatial profiles in the predicted image in section 3.2. At the same location, consider the pixel values in a 7×7 window of the AWiFS image at the time t_k as 2-dimensional spatial profiles.These 2-dimensional profiles represent the shape of the surface. Shape-based averaging preserves the contiguous regions of an image [18]. Average of these two 2-dimensional spatial profiles smoothed the spurious spatial discontinuities in the predicted images.

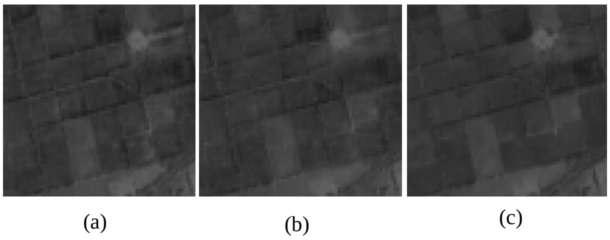


Fig.2. Smoothing Spurious Spatial Discontinuities Using Spatial Profile Averaging (a) Predicted Image after Super Resolution and High Pass Modulation Phases (b) Predicted Image after Super Resolution, High Pass Modulation and Third Phase (c) Original LISS III Image

The resultant image is shown in Fig.2 (b) which is almost similar to the original image in Fig.2(c). If the size of the window increased, spatial profile averaging over smoothed the image. Fig 3 shows the flowchart of the proposed method.

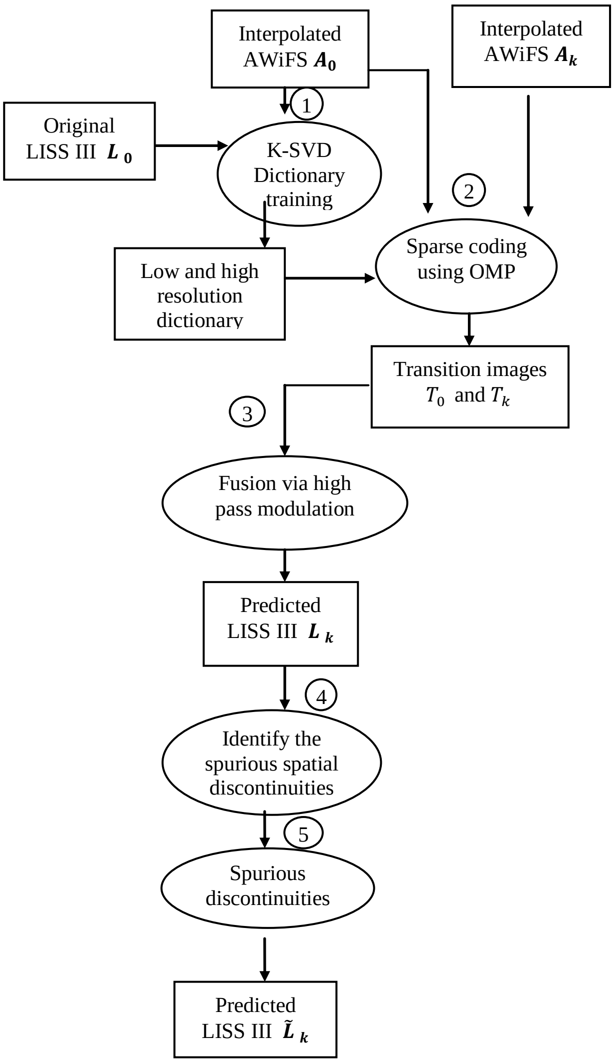


Fig.3. Flow Chart of the Proposed Spatiotemporal Data Fusion Using Dictionary Learning and Temporal Edge Primitives

IV. Experimental Results And Comparisons

In this section, the proposed method is compared with the recently developed Song and Huang (2013) method (SH method) [21]. Experiments are conducted with AWIFS and LISS III datasets. In pre-processing, AWiFS and LISS III image pixel values were converted into Top-Of-Atmosphere (TOA) reflectance. Further, reflectance values of AWiFS were normalized to the reflectance of LISS III by the linear regression between the homogeneous regions of AWiFS image and their corresponding regions of LISS III image with the procedure described by the [3].

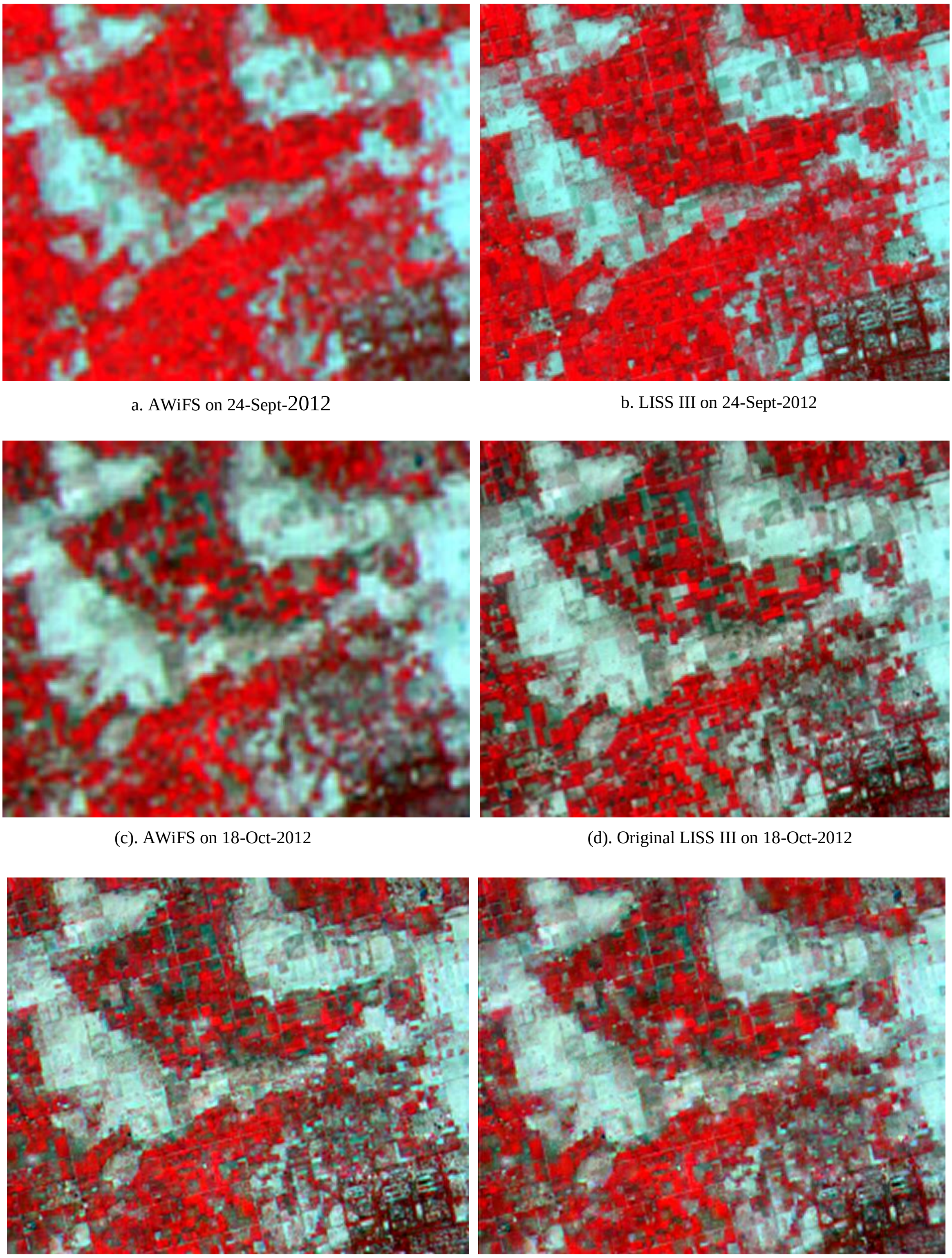
A. Experimental data set: 1 with focus on crop phenology changes

In this experiment, we have used previously acquired single AWiFS-LISS III image pair and a single AWiFS image to predict a synthetic LISS III image at AWiFS temporal resolution. For experimental demonstration of the proposed method, 12 km × 12 km area was selected in the Suratgarh area in Rajasthan state of India to predict the crop phenology at LISS III spatial and an AWiFS temporal resolutions. Single AWiFS–LISS III image pair acquired on 24-Sept-2012 was used as prior knowledge to predict a synthetic LISS III image for 18-Oct-2012 by using an AWiFS image acquired on 18-Oct-2012. These data sets show more crop phenology change from 24-sept-2012 to 18-Oct-2012 than the land-cover type changes as shown in Fig. 4.

The proposed method is compared with SH method with DL of dictionary size: 1000. Number of iterations is 30 for K-SVD and the size of the images is 512×512. Fig 4(e) shows the predicted synthetic LISS III image by using SH method and Fig 4(f) shows the predicted synthetic LISS III image by using the proposed method. Fig 4(f) more resembles than the Fig 4(e) to the original image in Fig 4(d).

The comparison of the four bands (Green-Red-NIR-SWIR) between SH method and the proposed method in terms root means squared error (RMSE), structural similarity index map (SSIM), and spectral angle mapper (SAM) is described in Table 1. The average RMSE of four bands for SH method is 0.0111 and for our method is 0.0100, it shows that the proposed method has better image quality when compared to the SH method. The average SSIM of four bands for SH method is 0.9074 and for our method is 0.9288, it shows that the proposed method shows high structural similarity with the original image than the SH method. Also, the proposed method has less spectral distortion than SH method with reference to the SAM values shown in Table 1.

The scatter plots with R^2 in Fig 5 and 6 (where the scale factor is 1000) show the reflectance distribution between the predicted and the actual surface reflectance of the four bands, Fig 5(a)-5(d) scatter plots of the predicted surface reflectance by using SH method and Fig. 6(a)-6(d) scatter plots of the predicted surface reflectance by using our method against the actual reflectance. The average R^2 of four bands for the SH method is 0.8575 and for the proposed method is 0.8850, it indicates that our method predicts reflectance values more accurately than SH method. Visually, it can be observed that the result of the proposed method in Fig 4(f) shows a similar perception as in original image Fig 4(d) when compared to the SH method.



a. AWiFS on 24-Sept-2012

b. LISS III on 24-Sept-2012

(c). AWiFS on 18-Oct-2012

(d). Original LISS III on 18-Oct-2012

(e). Predicted LISS III for 18-Oct-2012 using Song and Huang (2013) method

(f). Predicted LISS III for 18-Oct-2012 using proposed method

Fig.4. Predicted LISS III Images in Green-Red-NIR False Color Composition using (e) Song and Huang (2013) method and the (f) Proposed Method with the Prior Knowledge (a), (b) and (c).

Table 1. Quantitative Comparison of the Proposed Method against SH Method for the Dataset: 1

Method	SSIM				RMSE				SAM
	B2	B3	B4	B5	B2	B3	B4	B5	
Proposed method	**0.9434**	**0.9269**	**0.9175**	**0.9275**	**0.0071**	**0.0130**	**0.0100**	**0.0102**	**1.8523**
Song and Huang (2013)	0.9262	0.9011	0.8916	0.9108	0.0082	0.0143	0.0114	0.0107	2.1309

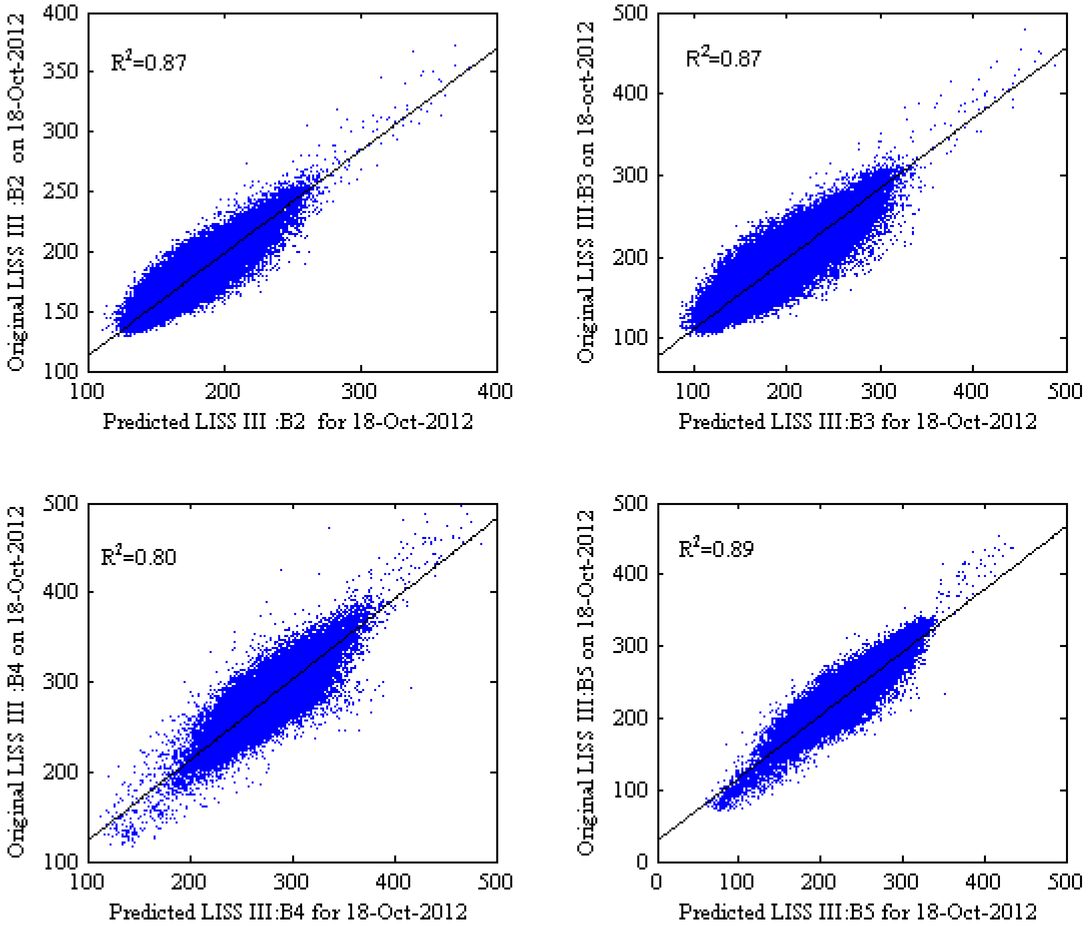
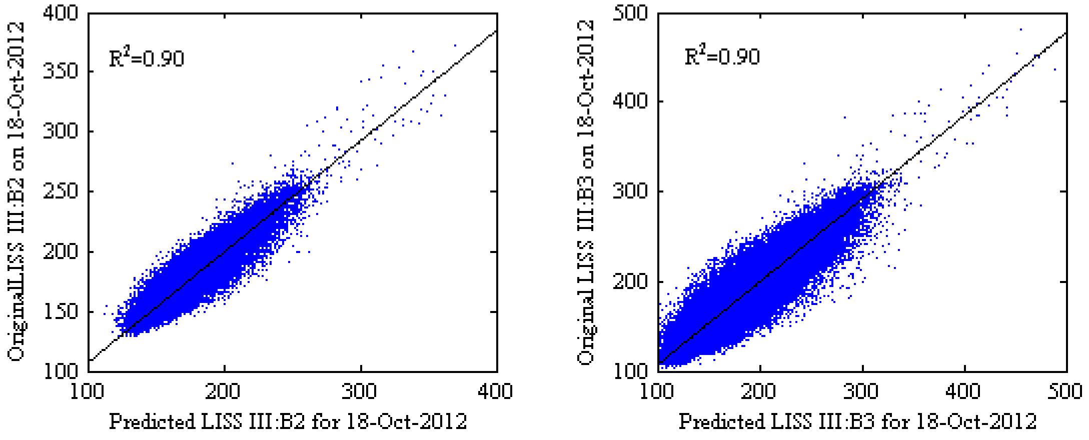


Fig.5. Scatterplots of the Predicted Against the Actual Surface Reflectance for the Song and Huang (2013) Method



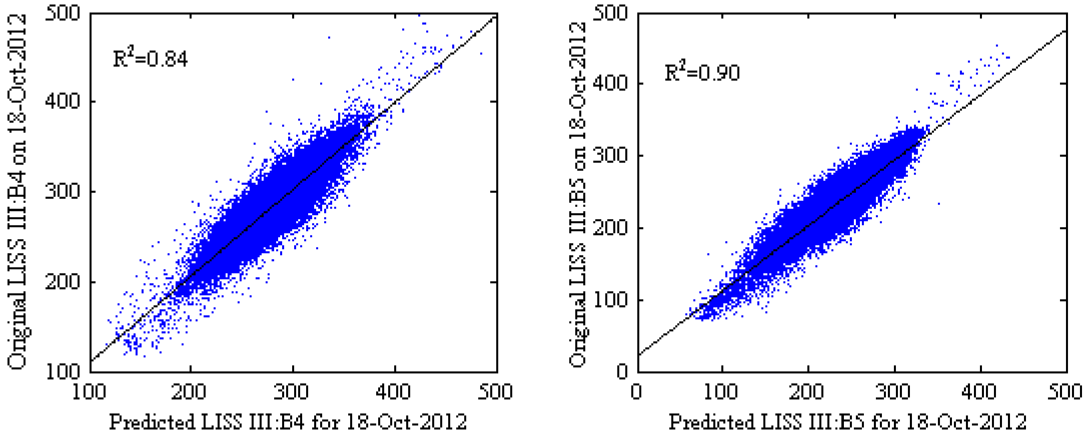


Fig.6. Scatter Plots of the Predicted Against the Actual Surface Reflectance for the Proposed Method

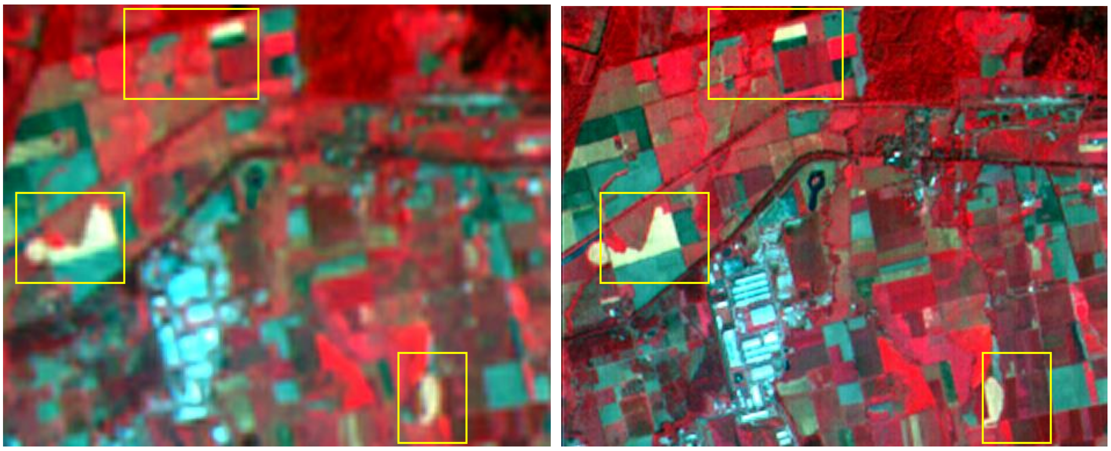
B. Experimental data set: 2 with focus on rapid land-cover-type changes

In this experiment, 12 km × 12 km area was selected in Pantnagar area in Uttaranchal state of India to predict the land-cover-type changes at LISS III spatial and AWiFS temporal resolution. Single AWiFS –LISS III image pair acquired on 24-Oct-2012 was used as prior knowledge to predict a synthetic LISS III image for 17-Nov-2012 by using the AWiFS image acquired on 17-Nov-2012. This data set has both crop phenology changes and land cover type changes such as bare soil on 24-Oct-2012 converted into wet land on 17-Nov-2012 as shown in Fig 7. These land cover type changes are shown in yellow colour boxes in Fig 7.

The proposed method was compared with SH method with DL of dictionary size: 1000. Number of iterations is 30 for K-SVD and the size of the images is 512×512. Fig 7(e) shows the predicted synthetic LISS III image by using SH method and Fig 7(f) shows the predicted synthetic LISS III image by using the proposed method. Fig 7(f) more resembles to the original image in Fig 7(d) than the Fig 7(e).

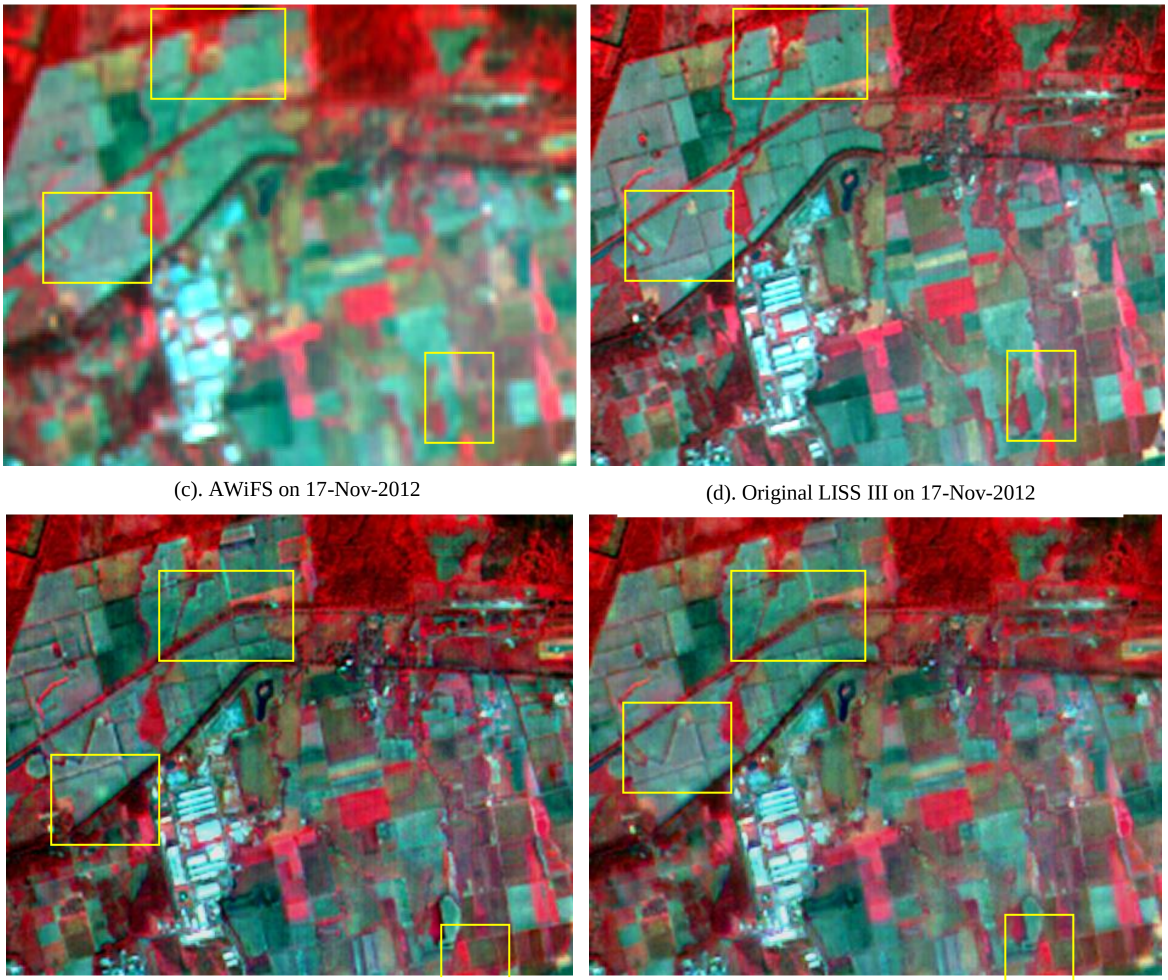
The comparison of the four bands (Green-Red-NIR-SWIR) between SH method and the proposed method in terms root means squared error (RMSE), structural similarity index map (SSIM), and spectral angle mapper (SAM) is described in Table 2. The average RMSE of four bands for SH method is 0.0130 and for our method is 0.0117, it shows that the proposed method has better image quality when compared to the SH method. The average SSIM of four bands for SH method is 0.8668 and for our method is 0.8966, it shows that the proposed method shows high structural similarity with the original image than the SH method. Also, the proposed method has less spectral distortion than SH method with reference to the SAM values shown in Table

The scatter plots with R^2 in Fig 8 and 9 (where the scale factor is 1000) show the reflectance distribution between the predicted and the actual surface reflectance of the four bands, Fig 8(a)-8(d) scatter plots of the predicted surface reflectance by using SH method and Fig 9(a)-9(d) scatter plots of the predicted surface reflectance by using our method against the actual reflectance. The average R^2 of four bands for the SH method is 0.7375 and for the proposed method is 0.7850, it indicates that our method predicts reflectance values more accurately than SH method. Visually, it can be observed that the result of the proposed method in Fig 7(f) shows a similar perception as in original image Fig 7(d) when compared to the SH method.



a. AWiFS on 24-Oct-2012

(b). LISS III on 24-OCT-2012



(c). AWiFS on 17-Nov-2012

(d). Original LISS III on 17-Nov-2012

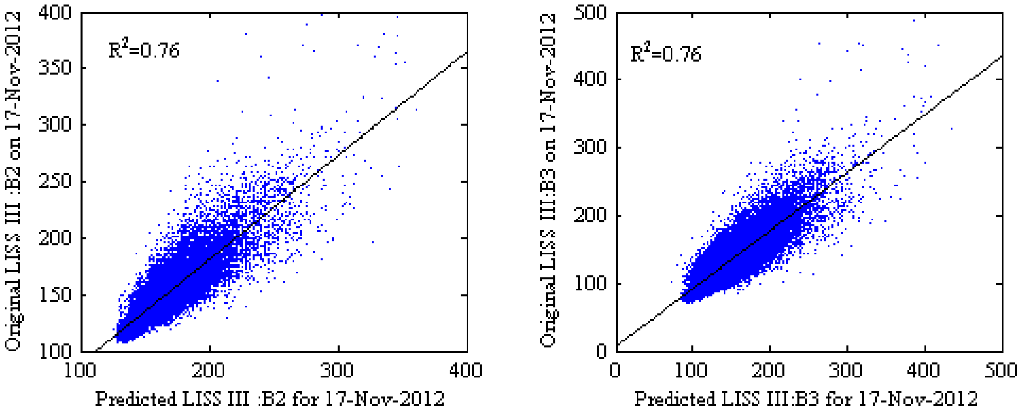
(e). Predicted LISS III for 17-Nov-2012 using Song and Huang (2013) method

(f). Predicted LISS III for 17-Nov-2012 using proposed method

Fig.7. Predicted LISS III Images in Green-Red-NIR False Color Composition using (e) Song and Huang (2013) Method and the (f) Proposed Method with the Prior Knowledge (a), (b) and (c).

Table 2. Quantitative Comparison of the Proposed Method against SH Method for the Dataset: 2

Method	SSIM				RMSE				SAM
	B2	B3	B4	B5	B2	B3	B4	B5	
Proposed method	**0.9251**	**0.9034**	**0.8844**	**0.8735**	**0.0076**	**0.0111**	**0.0153**	**0.0131**	**2.2450**
Song and Huang (2013)	0.9082	0.8736	0.8504	0.8350	0.0080	0.0124	0.0170	0.0147	2.5099



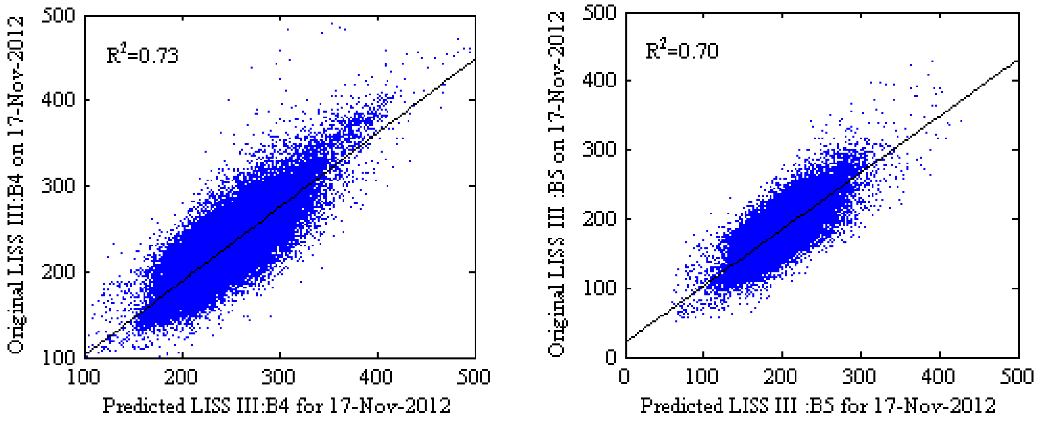


Fig.8. Scatter Plots of the Predicted Against the Actual Surface Reflectance of the Experimental Dataset: 2 for the Song and Huang (2013) Method

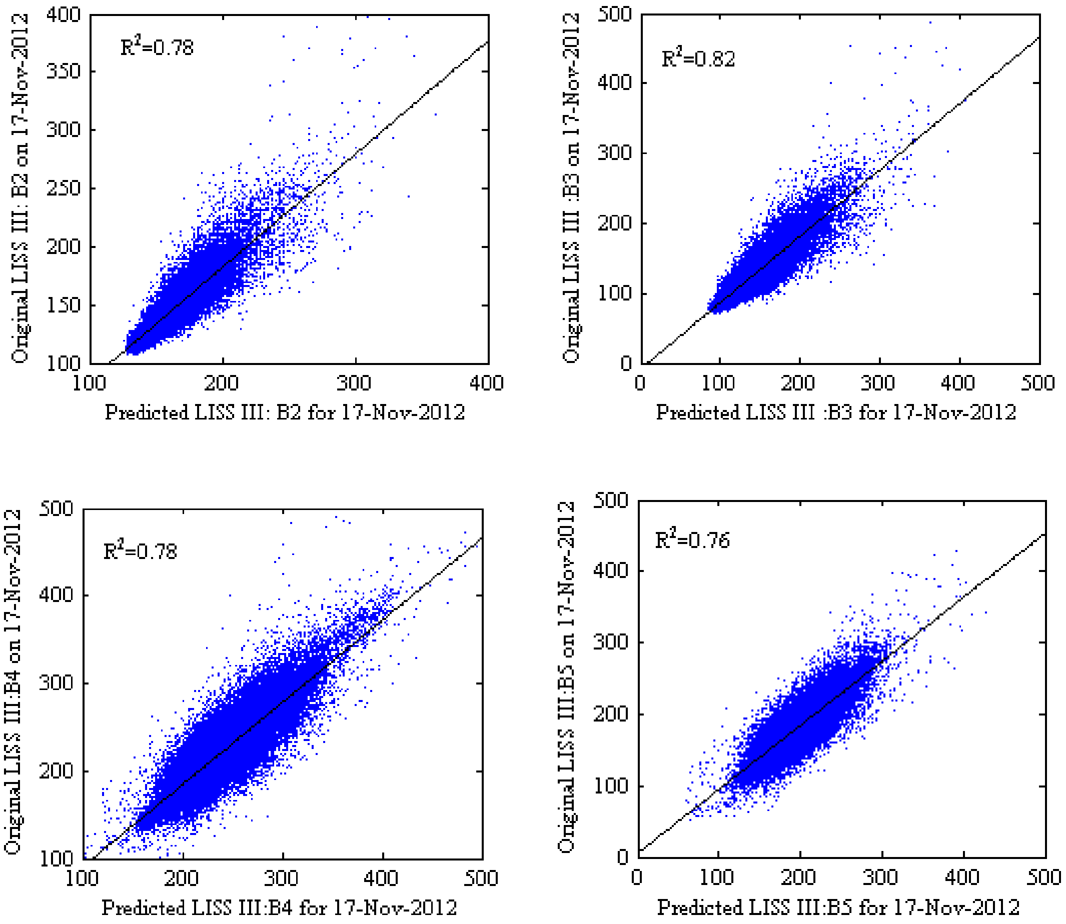


Fig.9. Scatter Plots of the Predicted Against the Actual Surface Reflectance of the Experimental Dataset: 2 for the Proposed Method

Although we have obtained a goodness of fit R^2 between predicted and the actual surface reflectance values better than the Song and Huang (2013) method [21], still there is a need to improve the prediction accuracy. In prediction of land-cover-type changes, both Song and Huang (2013) [21] method and the proposed method were suffered to retrieve the actual surface reflectance. The prediction accuracy R^2 for crop phenology changes is around 88% but for the land-cover-type changes is around 78%. By observing these quantitative and visual results, it is considerable evidence that as mentioned by Huang and Song (2012) [12] the accurate prediction of the land-cover-type changes is more difficult than the prediction of crop phenology changes.

V. Conclusion

We have proposed and demonstrated the spatiotemporal data fusion to create a synthetic LISS III image at time t_k with 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 proposed method involves three phases, the first is super resolution phase, the second is high pass modulation phase, and the third is smoothing the spurious spatial discontinuities. In the first phase, two transition images are obtained by enhancing the spatial resolution of AWiFS image at time t_0 and t_k through single-image-super resolution technique. It requires a prior knowledge about the desired high resolution image. This prior knowledge is obtained through a single AWiFS–LISS III image pair at time t_0. The image patches of AWiFS–LISS III image pair are used to train the low and high resolution dictionaries with K-SVD algorithm, and the corresponding sparse coefficients are approximated with OMP. For a given low resolution AWiFS image patch of time t_k, the sparse coefficients are obtained through the sparse coding technique OMP. The same sparse coefficients are used to derive the corresponding high resolution image patch for the time t_k by using the high resolution dictionary. The derived high resolution images in the super resolution phase are called as transition images for the time t_0 and t_k.

In high-pass-modulation phase, the high frequency details which are obtained in the difference of LISS III image and the transition image of time t_0, are proportionally injected into the transition image at time t_k. In synthesis of multi-sensor and multi-temporal images, spurious spatial discontinuities are inevitable. Until now, no method has dealt with these spurious spatial discontinuities in spatiotemporal data fusion. In third phase of our method, we identified these spurious discontinuities and smoothed with the spatial-profile-averaging method.

Space borne global sensors have a trade-off between high spatial and high temporal resolutions. The proposed method can be used as an alternative to generate high spatial and high temporal images. With this approach, we can create a synthetic LISS III image at 23.5 m spatial and 5-day temporal resolutions. Although the proposed method captures crop phenology changes more similar to the original, still further improvement is required to predict the land-cover-type changes. Future work of this paper is to improve the prediction accuracy of the both crop phenology and land-cover-type changes.

Acknowledgement

This research was supported under technology development project (TDP) at National Remote Sensing Centre, ISRO, Department of Space, Government of India, through the project [0606313AE901].

References

[1] Aharon, M., Elad, M., & Bruckstein, A. (2006). svd: An algorithm for designing overcomplete dictionaries for sparse representation. Signal Processing, IEEE Transactions on, 54(11), 4311-4322.

[2] Baker.S and T. Kanade, 2000 ,"Limits on super-resolution and how to break them," in Proc. IEEE Conf. Comput. Vision Pattern Recognit., vol. 2, pp. 372–379.

[3] Chander, G., M.J. Coan, andP.L. Scaramuzza.2008. "Evaluation and comparison of the IRS-P6 and the Landsat sensors." Geoscience and Remote Sensing, IEEE Transactions on, 46:209-221.

[4] Chang H., D.-Y. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in Proc. IEEE Conf. CVPR, 2004, vol. 1, pp. I-275–I-282.

[5] Coops.N.C, M.Johnson,M.A. Wulder, J.C.White, 2006. "Assessment of QuickBird high spatial resolution imagery to detect red attack damage due to mountain pine beetle infestation." Remote Sensing of Environment, vol.103, no.1, pp.67-80,July.2006.

[6] Davis, G., Mallat, S., & Avellaneda, M. (1997). Adaptive greedy approximations. Constructive approximation, 13(1), 57-98.

[7] Elad.M and D. Datsenko, 2009 "Example-based regularization deployed to super-resolution reconstruction of a single image," Computer J., vol. 52, no. 1, pp. 15–30, Jan..

[8] Elad, M. (2010). Sparse and redundant representations: from theory to applications in signal and image processing. Springer.

[9] Freeman, W. T., Jones, T. R., & Pasztor, E. C. (2002). Example-based super-resolution. Computer Graphics and Applications, IEEE, 22(2), 56-65.

[10] Gitelson, A. A. (2004). Wide dynamic range vegetation index for remote quantification of biophysical characteristics of vegetation. Journal of plant physiology, 161(2), 165-173.

[11] Holben B.N., "Characteristics of maximum-value composite images from temporal AVHRR data." International Journal of Remote Sensing, vol.7, no.11, pp.1417-1434, 1986.

[12] Huang.B. and H.Song. "Spatiotemporal reflectance fusion via sparse representation." IEEE Transactions on Geoscience and Remote Sensing, vol.50, no.10, pp.3707-3716, Oct.2012.

[13] Huete.A, C. Justice, W.Van Leeuwen,. "MODIS vegetation index (MOD13),"Algorithm theoretical basis document, 1999.

[14] *ISRO, Annual Report 2012.www.isro.org/pdf/AnnuaReport2012.pdf* (Accessed on 18-03-2014).

[15] Jiji, C. V., Joshi, M. V., & Chaudhuri, S. (2004). Single‐frame image super‐resolution using learned wavelet coefficients. International journal of Imaging systems and Technology, 14(3), 105-112.

[16] Ji T.L., M.K. Sundareshan, H. Roehrig, 1994. "Adaptive image contrast enhancement based on human visual properties." Medical Imaging, IEEE Transactions on, vol.13, no.4, pp. 573-586, April.

[17] Justice C.O., J.R.G.Townshend, B.N.Holben, E.C.Tucker, "Analysis of the phenology of global vegetation using meteorological satellite data." International Journal of Remote Sensing, vol.6, no.8, 1271-1318, 1985.

[18] Mitchell.H.B., "Data Fusion: Concepts and Ideas." Springer, 2012.

[19] Ranganath.R, R.Navalgund, R.PSingh, "The evolution of the earth observation system in India." Journal of the Indian Institute of Science, vol.90, no.4, pp. 471-488, April.2012.

[20] Rubinstein, R., Bruckstein, A. M., & Elad, M. (2010). Dictionaries for sparse representation modeling. Proceedings of the IEEE, 98(6), 1045-1057.

[21] Song.H and B.Huan,"Spatiotemporal satellite image fusion through one-pair image learning." IEEE Transactions on Geoscience and Remote Sensing, vol. 51,

no.4, pp.1883-1896, April.2013.

[22] Starck, J. L., Murtagh, F., & Fadili, J. M. (2010). Sparse image and signal processing: wavelets, curvelets, morphological diversity, pp.2-3,Cambridge University Press.

[23] Yang, J., Wright, J., Huang, T. S., & Ma, Y. (2010). Image super-resolution via sparse representation. Image Processing, IEEE Transactions on, 19(11), 2861-2873.

[24] Walker, J.J., K. M.De Beurs, R. H.Wynne, F.Gao. 2012. "Evaluation of Landsat and MODIS data fusion products for analysis of dryland forest phenology." Remote Sensing of Environment, 117:381-393.

[25] Wu, W., Z. Liu, X. He, W. Gueaieb. 2011. "Single-image super-resolution based on Markov random field and contourlet transform." Journal of Electronic Imaging, 20(2): 023005-023005.

[26] Zeyde, R., M. Elad, M. Protter. 2012. "On single image scale-up using sparse-representations." In Curves and Surfaces, Springer Berlin Heidelberg.

Authors' Profiles

J.Malleswara Rao received M.Sc degree in Mathematics from Acharya Nagarjuna University, and M.Tech in spatial information technology in 2009 from JNTUH, Hyderabad, India. He is currently working as a SRF in NRSC and pursuing Ph.D degree in spatial information technology at JNTUH, Hyderabad, India.

His research interests include image super resolution, spatio-temporal data fusion, and pattern recognition.

C.V.Rao received M.E. degree in Digital systems Engg in 1985 from Osmania University, India and Ph.D degree in Computer science and Engg in 2009 from JNTUH, Hyderabad, India. He is currently working as a Head of Data and Products Quality Evaluation Division, NRSC, ISRO, Hyderabad, India.

His research interests include image fusion, spatio-temporal data fusion, and Image classification

A.Senthil Kumar received M.Sc(Engg.) and Ph.D from Indian Institute of Science, Bangalore in the field of Image processing in 1985 and 1990 respectively. He is currently working as a Group director for Geophysical and Special products group in NRSC, Hyderabad, India.

His research interests include sensor characterization, radiometric data processing and Geophysical products development from IRS satellite images

B. Gopala Krishna currently working as a Deputy Director, DPPA&WAA, NRSC, ISRO, Hyderabad, India.

His research interests include image fusion, spatio-temporal data fusion, and Image classification

V.K.Dadhwal received M.Sc and Ph.D in Plant Physiology from Indian Agricultural Research Institute, New Delhi in year 1978 and 1983 respectively. He is currently working as a Direcor, NRSC, ISRO, Hyderabad, India.

His research interests include Crop modeling, Remote Sensing Applications in Agriculture, Terrestrial Carbon Cycle, Land Cover Land Use Change Modelling, and Land Surface Processes.

How to cite this paper: J. Malleswara Rao, C. V. Rao, A. Senthil Kumar, B. Gopala Krishna, V. K. Dadhwal,"Spatiotemporal Data Fusion using Dictionary Learning and Temporal Edge Primitives", IJIGSP, vol.7, no.10, pp.47-59, 2015.DOI: 10.5815/ijigsp.2015.10.06