Chapter 4

Development of Morphological Component Analysis Features-based Sparse Representation Technique for MS Image SR

4.1 Introduction

Many remote sensing satellites, like, IKONOS, QuickBird, Landsat, etc., acquires several low-resolution (LR) images of the same area on different wavelengths, forming a LR multispectral (LR-MS) image, along with a high-resolution panchromatic (HR-PAN) image [108]. Indian remote sensing satellites, ResourceSat-2 provides

(a) (b) (c)

Figure 4.1: Sample LR-MS images: (a) ResourceSat-2 LISS-III (b) ResourceSat-2 LISS-IV (c) QuickBird

only LR-MS images using LISS-III and LISS-IV sensors, which do not have an HR PAN band, like in the QuickBird. Fig. 4.1 shows sample LR-MS images captured by LISS-III, LISS-IV and QuickBird sensors. Underlying LR-MS band images are very useful in many remote sensing image analysis procedures, but due to the low spatial resolution, they have shortcomings in visualizing on latest high-definition (HD) displays.

In order to reconstruct a MS image with high spatial and spectral resolution, either both LR-MS and PAN images are fused (i.e in pansharpening) or sparse re-

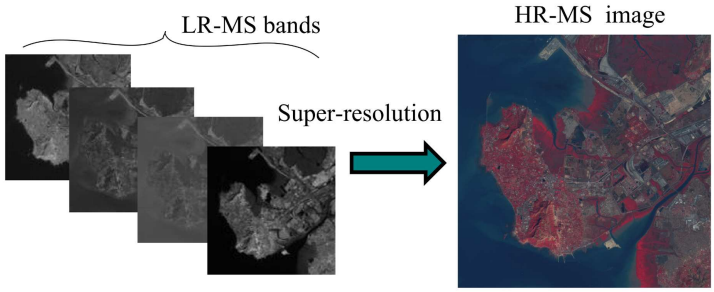


Figure 4.2: Example of MS image super-resolution

construction is done using the bands of the single LR-MS input image [96] only. Fig. 4.2 shows an example of SR of a LISS-III MS image taken by ISRO's ResourceSat-2 satellite, where each LR-MS band corresponds to a spatial resolution of about 23.5 m. The SR output is reconstructed for an upscale factor of 4, which provides enhanced spatial resolution of about 5.8 m. In the pioneering sparse representation-based SR work by Yang *et al.* [115] for natural images, they apply SR to the luminance component (Y) after the given RGB image is converted to the $YCbCr$ format, while the two color channels are bicubic interpolated. Their method has provisions for modifying the regularization problem by designing either new effective *a priori* (regularization constraints) terms or developing better schemes for feature extraction to enhance the accuracy of sparse representation.

In this chapter, we have investigated a new feature extraction strategy for enhancement of the performance of SISR. We explore the morphological component analysis (MCA) for identifying texture and cartoon images and developed a multicore parallel algorithm for sparse representation-based MS image SR. To perform dictionary learning from MS images, we have performed the principle component analysis (PCA) for the selection of significant bands. The algorithm is implemented using the multicore parallel processing strategy to reduce the computational time. The details are explained in the following sections.

4.2 MCA-based image decomposition

Real images have morphological diversities in terms of geometrical and texture features. MCA decomposes the images into piecewise-smooth (cartoon) and texture components for different applications, like image restoration, image inpainting, etc. The high-frequency texture features are highly useful for an optimal sparse representation compared to its low-frequency counterpart i.e. the smooth-structure layer [116].

An $\sqrt{N} \times \sqrt{N}$ image $\mathbf{X}$ is assumed to have a texture component $\mathbf{X}_t$ along with a structure component $\mathbf{X}_s$. Fig. 4.3 shows examples of smooth and texture components present in a MS image. The overcomplete dictionaries: $\{\mathbf{D}_t, \mathbf{D}_s\} \in \mathbb{R}^{N \times L}$ establish sparse representations corresponding to $\mathbf{X}_t$ and $\mathbf{X}_s$, separately [39] i.e.,

$$\mathbf{X}_t = \mathbf{D}_t \boldsymbol{\alpha}_t, \tag{4.1}$$

$$\mathbf{X}_s = \mathbf{D}_s \boldsymbol{\alpha}_s. \tag{4.2}$$

We can write,

$$\mathbf{X} = \mathbf{X}_t + \mathbf{X}_s. \tag{4.3}$$

Therefore,

$$\mathbf{X} = \mathbf{D}_t \boldsymbol{\alpha}_t + \mathbf{D}_s \boldsymbol{\alpha}_s, \tag{4.4}$$

where $\boldsymbol{\alpha}_t$ and $\boldsymbol{\alpha}_s$ are sparse coefficient vectors for texture and structure components, respectively. To initialize dictionaries for the training different transforms, like the wavelet, the overcomplete DCT are used for the texture part, while the wavelet, the curvelets, the contourlets, etc. are used for the overall structure part.

MCA seeks an optimal sparse representation of $\mathbf{X}$ over a coupled dictionary containing both $\mathbf{D}_t$ and $\mathbf{D}_s$ as follows:

$$\begin{aligned} \left\{\boldsymbol{\alpha}_t^{opt}, \boldsymbol{\alpha}_s^{opt}\right\} = \arg \min_{\{\boldsymbol{\alpha}_t, \boldsymbol{\alpha}_s\}} \{\|\boldsymbol{\alpha}_t\|_1 + \|\boldsymbol{\alpha}_s\|_1\} \\ \text{subject to} \quad \|\mathbf{X} - \mathbf{D}_t \boldsymbol{\alpha}_t - \mathbf{D}_s \boldsymbol{\alpha}_s\|_2^2 \leqslant \varepsilon \end{aligned}. \tag{4.5}$$

(a) (b) (c)

Figure 4.3: MCA decomposition results: (a) the example MS band image (b) the structure part and (c) the texture part

A total variation (TV) regularization prior may also be added in Eq. 4.5 for better recovery of piecewise smooth objects and enhanced edges. The equivalent unconstraint version of Eq. 4.5 is written as follows:

$$\begin{aligned}\left\{\boldsymbol{\alpha}_t^{opt}, \boldsymbol{\alpha}_s^{opt}\right\} = \underset{\{\boldsymbol{\alpha}_t, \boldsymbol{\alpha}_s\}}{\arg\min} \left\{\|\boldsymbol{\alpha}_t\|_1 + \|\boldsymbol{\alpha}_s\|_1\right\} + \\ \lambda_1 \left\|\mathbf{X} - \mathbf{D}_t\boldsymbol{\alpha}_t - \mathbf{D}_s\boldsymbol{\alpha}_s\right\|_2^2 + \gamma TV\left\{\mathbf{D}_s\boldsymbol{\alpha}_s\right\}\end{aligned} . \tag{4.6}$$

Eq. 4.6 is an ℓ_1- minimization problem, which is solved iteratively via simultaneous sparse approximation and dictionary updates using the block-coordinate-relaxation algorithm [11]. Here, ε is the error, which signifies the quality up to which the decomposed layers are being approximated.

4.3 PCA-based band selection

A MS image consists of several bands, representing images of the same spatial region acquired at different wavelengths. Therefore, such band images have spectral redundancy among them, which can be explored to obtain a band reduced MS image or its corresponding RGB image for dictionary learning. PCA is a linear transformation technique that finds the projection of the image vectors corresponding to different bands onto a subspace spanned by the principal eigenvectors arranged in the order of decreasing eigenvalues such that the principal components contains almost the 99% of the information of the input MS image. The algorithmic steps for the PCA-

based band reduction technique for efficient dictionary learning from MS image are as follows:

▷ *Convert the image into a set of vectors and subtract mean:*

First, an image is assumed as a set of vectors of equal length, where each vector represents a column in the image. An image $\mathbf{X}^{m\times n}$ can be written as $\mathbf{x} = [\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, ..., \mathbf{x}_n]^T$, where n is the number of columns and each vector $\mathbf{x}_k$ is of length m. The same idea is applicable for MS images, where number of columns will be represented by the number of spectral bands and length of each column vector $\mathbf{x}_k$ will become mn (assuming that band images are of size $m \times n$). The set of PCA transformed vectors $\mathbf{y}$ can be obtained from $\mathbf{x}$ as follows:

$$\mathbf{y} = A\left(\mathbf{x} - \mathbf{m}_x\right), \tag{4.7}$$

where $\mathbf{m}_x$ is the vector consisting of mean values from all the input vectors in $\mathbf{x}$. Subtracting the mean is also known as centering and it reduces the chances of influence by any other components on the principal components.

▷ *Find transformation matrix A from the covariance matrix:*

The matrix A in Eq. 4.7 is obtained from the covariance matrix $\mathbf{C}_x$ such that rows of A is formed by the eigenvectors of $\mathbf{C}_x$ arranged from top to bottom according to the descending order of eigenvalues. The formula for finding covariance matrix is given as,

$$\mathbf{C}_x = E\left[\left(\mathbf{x} - \mathbf{m}_x\right)\left(\mathbf{x} - \mathbf{m}_x\right)^T\right] = \frac{1}{K}\sum_{k=1}^{K}\mathbf{x}_k\mathbf{x}_k^T - \mathbf{m}_x\mathbf{m}_x^T. \tag{4.8}$$

Since each vector $\mathbf{x}_k$ has size $mn \times 1$, therefore size of $\mathbf{C}_x$ becomes $mn \times mn$. Here, elements on the main diagonal of $\mathbf{C}_x$ are the variances of $\mathbf{x}$ and all other elements in $\mathbf{C}_x$ represents the covariance between vectors $\mathbf{x}_i$ and $\mathbf{x}_j$. Mathematically,

$$\mathbf{C}_x\left(i,i\right) = E\left[\left(\mathbf{x}_i - \mathbf{m}_i\right)^2\right] \tag{4.9}$$

$$\mathbf{C}_x\left(i,j\right) = E\left[\left(\mathbf{x}_i - \mathbf{m}_i\right)\left(\mathbf{x}_j - \mathbf{m}_j\right)\right] \tag{4.10}$$

After obtaining the covariance matrix $\mathbf{C}_x$, its eigenvalues and corresponding eigenvectors are calculated to determine the matrix A.

▷ *Transform the original dataset:*

Since the rows of the transformation matrix $\mathbf{A}$ are orthonormal, inverse transform can be performed to transform the original data using the PCA coefficient as follows:

$$\mathbf{x} = \mathbf{A}^T\mathbf{y} + \mathbf{m}_x \tag{4.11}$$

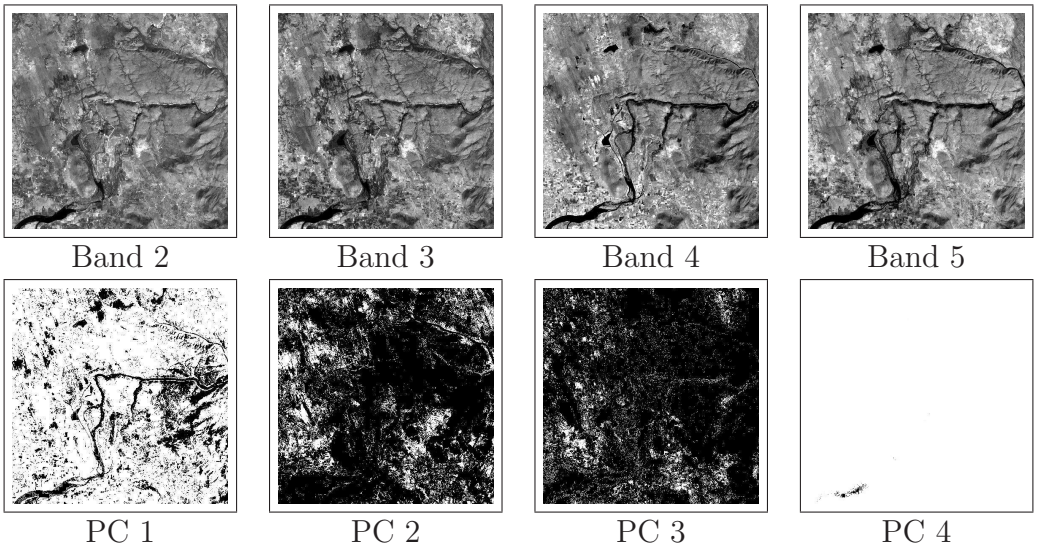


Figure 4.4: Example of PCA-based significant band selection from LISS-III data

Fig. 4.4 presents a MS image from LISS-III dataset containing four spectral bands. PCA is applied to find the significant bands and the transformed images are shown in terms of the principal components: PC1, PC2, PC3 and PC4, where they have information of about 74.01%, 20.69%, 4.14% and 1.14%, respectively of the given MS image. It can be found that the first three principal components are able to provide approximately 99% information of the original image.

4.4 Proposed method

The proposed sparse representation-based MS SR work consists of the following three stages:

i. Dictionary learning

ii. Sparse reconstruction

iii. Multicore parallel implementation.

4.4.1 Dictionary learning using PCA and MCA

A schematic of the proposed MS image-based dictionary learning is presented in Fig. 4.5. An image database, consisting of cropped MS images of sizes 512×512 and 1024×1024, is formed by selecting RoIs having useful land cover features from orthorectified GeoTIFF images. We apply 2D PCA on each MS image and bands corresponding to the three highest principal components (PCs) are selected and converted to an RGB image. It is ensured that the three selected PCs constitute at least 99% information of the actual MS image.

Now, to prepare two sets of training patch datasets for dictionary training: $\mathbf{X}^h$ and $\mathbf{X}^\ell$, first, the Y-channel of the YCbCr image obtained from the RGB is downsampled and blurred to produce an LR image. The downsampling factor d is equal to the desired upscale ratio for the final SR output of the scheme. Now, the HR patch vectors $\mathbf{x}^h$ of size $n^2 \times 1$ are extracted from the MS band at its original scale. On the other hand, LR patch vectors $\mathbf{x}^\ell$ are generated as follows: first, the LR image is upscaled by 2 to obtain an interpolated HR image using bicubic interpolation. In order to extract high frequency features from the bicubic interpolated image, first, MCA is carried out, which decomposes it into its structure and texture components. Then, on the texture component, we apply first- and second-order gradient filters in horizontal and vertical directions. This results in the four gradient maps from where

four LR feature patches are extracted. Finally, they are concatenated to obtain the LR feature vector for training. Thus, the size of a LR feature patch $\mathbf{x}^{\ell}$ becomes $((2n/d \times 2n/d) \times 4) \times 1$. We follow the same feature extraction strategy for different zoom in factors.

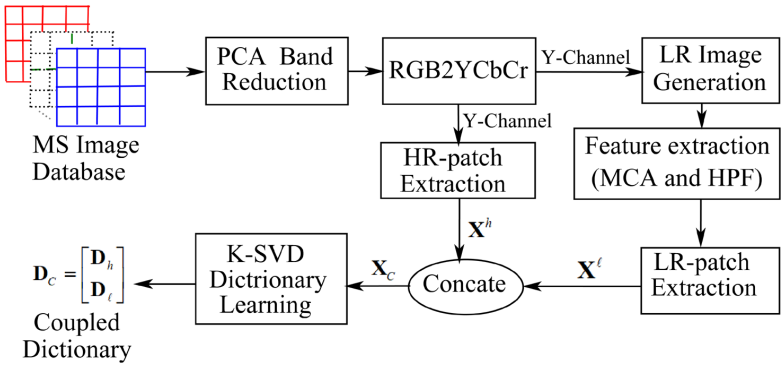


Figure 4.5: Proposed MS image dictionary learning scheme using PCA and MCA

In [115], authors learnt HR and LR dictionaries, namely, $\mathbf{D}_h$ and $\mathbf{D}_\ell$ from datasets $\mathbf{X}^h$ and $\mathbf{X}^\ell$, respectively. In the sparse representation framework, it is assumed that both $\mathbf{X}^h$ and $\mathbf{X}^\ell$ share a common sparse representation vector $\boldsymbol{\alpha}$ with their individual dictionaries as shown in Eq. 4.12.

$$\{\mathbf{D}_h, \mathbf{D}_\ell, \boldsymbol{\alpha}\} = \arg \min_{\{\mathbf{D}_h, \mathbf{D}_\ell, \boldsymbol{\alpha}\}} \frac{1}{p} \left\|\mathbf{X}^h - \mathbf{D}_h\boldsymbol{\alpha}\right\|_2^2 + \frac{1}{q} \left\|\mathbf{X}^\ell - \mathbf{D}_\ell\boldsymbol{\alpha}\right\|_2^2 + \lambda_2 \left(\frac{1}{p} + \frac{1}{q}\right) \|\boldsymbol{\alpha}\|_1 , \tag{4.12}$$

where p and q represent the size of HR and LR patch vectors, respectively and λ_2 is the regularization parameter. The above equation may be simplified by rearranging $\mathbf{D}_h$ and $\mathbf{D}_\ell$ into a joint dictionary, $\mathbf{D}_C$. Similarly, $\mathbf{X}^h$ and $\mathbf{X}^\ell$ may also be combined. Thus, Eq. 4.12 may be rewritten as follows:

$$\{\mathbf{D}_C, \boldsymbol{\alpha}\} = \arg \min_{\{\mathbf{D}_C, \boldsymbol{\alpha}\}} \|\mathbf{X}_C - \mathbf{D}_C\boldsymbol{\alpha}\|_2^2 + \lambda_2 \|\boldsymbol{\alpha}\|_1 , \tag{4.13}$$

where $\mathbf{X}_C = \begin{bmatrix} \frac{1}{\sqrt{p}}\mathbf{X}^h \\ \frac{1}{\sqrt{q}}\mathbf{X}^\ell \end{bmatrix}$ and $\mathbf{D}_C = \begin{bmatrix} \frac{1}{\sqrt{p}}\mathbf{D}_h \\ \frac{1}{\sqrt{q}}\mathbf{D}_\ell \end{bmatrix}$. Eq. 4.13 is solved based on K-SVD dictionary learning method [38], which in essence solves the following two steps, alternately over individual patch vectors.

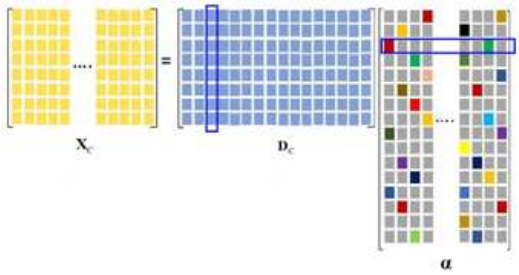


Figure 4.6: Schematic of sparse coding and dictionary update steps

▷ *Patch-based sparse coding step:*

First, an overcomplete dictionary $\mathbf{D}_C$ of given size is randomly initialized and then using this dictionary, it finds the sparse representation vector $\boldsymbol{\alpha}_i$ for each patch vector $\mathbf{x}_i$ from the combined dataset $\mathbf{X}_C$.

▷ *Dictionary update step:*

Given the sparse representations calculated above, next task is to update the dictionary. Here, each column $\mathbf{d}_i$ of the dictionary $\mathbf{D}_C$ is updated for the above obtained sparse vector $\boldsymbol{\alpha}_i$ by minimizing the error term $\|\mathbf{x}_i - \mathbf{d}_i\boldsymbol{\alpha}_i\|$ iteratively.

A schematic representation of coupled dictionary learning is shown in Fig. 4.6. K-SVD method has advantages since it is faster and can be trained for large number of sample patches. In line with the training applied to MS images, a coupled K-SVD dictionary training for PAN image database may also be carried out. We may apply the proposed dictionary learning method starting from the patch extraction stage as shown in the Fig. 4.5 for the PAN image. However, in this case, PCA-based band reduction and YCbCr transformations from MS images are not required.

4.4.2 SR reconstruction based on MCA and sparse representation

Given the trained coupled dictionary, the proposed method solves two regularization problems; one is seeking a patch-wise sparse solution and the other having the image formation model as a global prior. Fig. 4.7 pictorially depicts the proposed MS image SR reconstruction method.

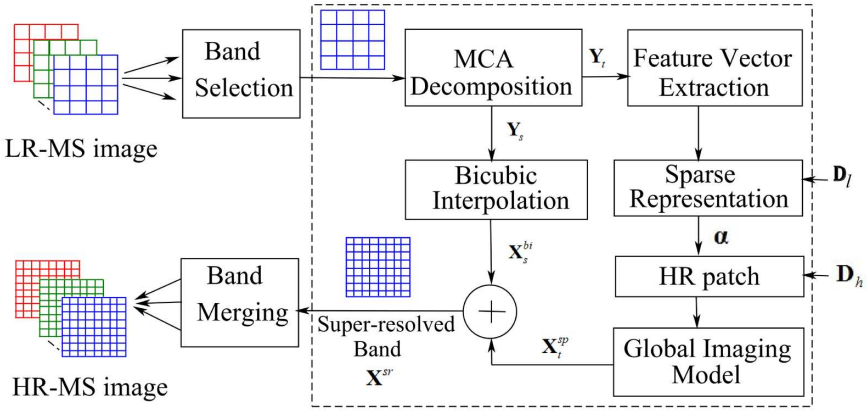


Figure 4.7: Proposed MS SR method

Here, MCA is performed on each LR-MS band separately to extract texture and cartoon components from the bands. The texture image '$\mathbf{Y}_t$' has diversified high-frequency features and thus favors for a better sparse representation. On the other hand, the structure or cartoon image '$\mathbf{Y}_s$' contains the low-frequency structural information, which is upscaled to a size equal to the size of the proposed super-resolved output by using bicubic interpolation, which is relatively simpler and reduces the overall computational cost of the method.

After MCA decomposition, '$\mathbf{Y}_t$' is first upscaled by 2 by bicubic interpolation. Upsampling before feature extraction works better as establishing a correspondence between HR and upsampled LR image patches is relatively easier [115]. Now, upsampled LR image is passed through 1D feature extraction filters of first- and second-orders, respectively as follows:

$$f = [1, \quad 0, \quad -1] \tag{4.14}$$

$$s = [1, \quad 0, \quad -2, \quad 0, \quad 1] \tag{4.15}$$

Linear convolution on the $\mathbf{Y}_t$ with the four filters f, f^T, s and s^T produces four filtered outputs; each of size equal to the size of $\mathbf{Y}_t$. Next, feature patches with single pixel overlapping are extracted from each of the above outputs. Four feature patches, corresponding to each pixel location of the filtered images, are concatenated to get a single feature vector, ${\mathbf{y}_t}^{\ell}$. Therefore, the size of each feature vector will be four times than that of the upsampled LR patch vector. Finally, the feature vectors containing high frequency information will be used for sparse representation.

A sparse representation problem is then formulated to represent ${\mathbf{y}_t}^{\ell}$ with the pre-trained overcomplete dictionary which may be written as follows:

$$\begin{aligned} \min \|\boldsymbol{\alpha}\|_1 \ \text{subject to} \ \left\|\mathbf{D}_\ell \boldsymbol{\alpha} - \mathbf{y}_t^\ell\right\|_2^2 &\leqslant \varepsilon_1 \\ \|E_p \mathbf{D}_h \boldsymbol{\alpha} - \mathbf{w}\|_2^2 &\leqslant \varepsilon_2 \end{aligned}. \tag{4.16}$$

Eq. 4.16 can also be rewritten as an unconstrained minimization problem as follows:

$$\hat{\boldsymbol{\alpha}} = \min_{\boldsymbol{\alpha}} \left\|\tilde{\mathbf{D}}\boldsymbol{\alpha} - \tilde{\mathbf{y}}\right\|_2^2 + \lambda_2 \|\boldsymbol{\alpha}\|_1 , \tag{4.17}$$

where $\tilde{\mathbf{D}} = \begin{bmatrix} \mathbf{D}_\ell \\ E_p \mathbf{D}_h \end{bmatrix}$, $\tilde{\mathbf{y}} = \begin{bmatrix} \mathbf{y}_t^\ell \\ \mathbf{w} \end{bmatrix}$, and λ_2 is the regularization parameter. The operator E_p in the above equation extracts the region of pixel overlap between the target patch and previously reconstructed adjacent HR patches; $\mathbf{w}$ represents the values of already reconstructed HR image patches on the region of overlap. The solution to Eq. 4.17, $\hat{\boldsymbol{\alpha}}$ will be used to generate the corresponding HR patch vector ${\mathbf{x}_\mathbf{t}}^h$ by multiplying $\hat{\boldsymbol{\alpha}}$ with the HR dictionary $\mathbf{D}_h$ as follows:

$$\mathbf{x}_t^h = \mathbf{D}_h \hat{\boldsymbol{\alpha}}. \tag{4.18}$$

Next, reconstructed HR patches $\mathbf{x}_t^h$ are positioned onto the target HR image grid to produce an intermediate reconstructed image $\mathbf{X}_0$. Since, $\mathbf{X}_0$ may not satisfy

exactly the image acquisition model due to noise or any other reconstruction error, the target HR image for the texture part $\mathbf{X}_t^{\text{sp}}$ is obtained by applying the global image constraint based regularization as follows:

$$\mathbf{X}_t^{\text{sp}} = \arg\min_{\mathbf{X}_t^{\text{sp}}} \|SH\mathbf{X}_t^{\text{sp}} - \mathbf{X}_t\|_2^2 + c\,\|\mathbf{X}_t^{\text{sp}} - \mathbf{X}_0\|_2^2\,, \tag{4.19}$$

where c is regularization parameter. As shown in Fig. 4.7 to get the target super-resolved image for a MS band $\mathbf{X}^{\text{sr}}$, the solution of Eq. 4.19 i.e. $\mathbf{X}_t^{\text{sp}}$ is combined with the bicubic interpolated output of the structural component, denoted by $\mathbf{X}_s{}^{\text{bi}}$. Finally, the individual super-resolved bands obtained as above are merged to form the target HR-MS image.

4.5 Parallel implementation

Modern high-performance computing (HPC) system features shared memory based multicore processors which make them suitable to execute single instruction, multiple data (SIMD) operations efficiently. OpenMP parallel programming interface associated with C, C++ or FORTAN language can be used for converting a sequential code to a parallel one.

4.6 Results and discussion

4.6.1 Performance evaluation and comparisons

4.6.1.1 Database preparation

We have carried out our simulations using one PAN and a few LR-MS image datasets as shown in Table 4.1. PAN images are acquired by the CartoSat-2 series satellites (sensor: single panchromatic camera) of Indian Space Research Organization

(ISRO), while the two MS datasets are acquired from ResourceSat-2 satellite (sensors: LISS-III and LISS-IV) of ISRO. These datasets are procured from ISRO's NRSC data centre (NDC)[1]. Earth observation satellites, namely, ResourceSat-2 provides low and medium resolution MS images, and CartoSat-2 provides sub-meter resolution PAN images. Another MS dataset (2.8 m resolution) is obtained from the Global Land Cover Facility (GLCF)[2], which are captured by the QuickBird satellite for public access. Some other freely available MS images are also collected from ISRO's Bhuvan portal[3]. Mostly, these images are of land cover and earth observation types, selected from different geographical locations within India. More details about the selection of training and test images are described in the later part of this section.

Table 4.1: Details of collected PAN and MS image datasets

Data Set	Satellites	Band Information	Source
Panchromatic -I Resolution: 0.65 m	CartoSat-2	1 band: monochrome	NRSC
Multispectral -I Resolution: 23.5 m	ResourceSat-2 LISS-III	B2 (VIS): 0.52 -0.59 nm B3 (VIS): 0.62 -0.68 nm B4 (NIR): 0.77 -0.86 nm B5 (SWIR): 1.55 -1.75 nm	NRSC & Bhuvan
Multispectral -II Resolution: 5.8 m	ResourceSat-2 LISS-IV	B2 (G): 0.52 -0.59 nm B3 (R): 0.62 -0.68 nm B4 (NIR): 0.76 -0.86 nm	NRSC
Multispectral -III Resolution: 2.8 m	QuickBird-2	B1 (B): 0.45 -0.53 nm B2 (G): 0.52 -0.60 nm B3 (R): 0.63 -0.69 nm B4 (NIR): 0.76 -0.90 nm	GLCF

4.6.1.2 Simulation environment

Simulations of the proposed work is carried out on a remote server with following specifications- Model: Dell PowerEdge R730 server; Processors: 2 x Intel Xeon

[1]NRSC Data Center. Available at `http://uops.nrsc.gov.in:33333/ImgeosUops/FinalImgeosUops/FinalImgeosUops.html#` (last seen on 20 April 2019)

[2]GLCF portal for satellite data. Available at `http://glcf.umiacs.umd.edu`(last seen on 20 March 2019)

[3]Bhuvan Open Data Archive. Available at `https://bhuvan-app3.nrsc.gov.in/data/download/index.php` (accessed on 20 April 2019)

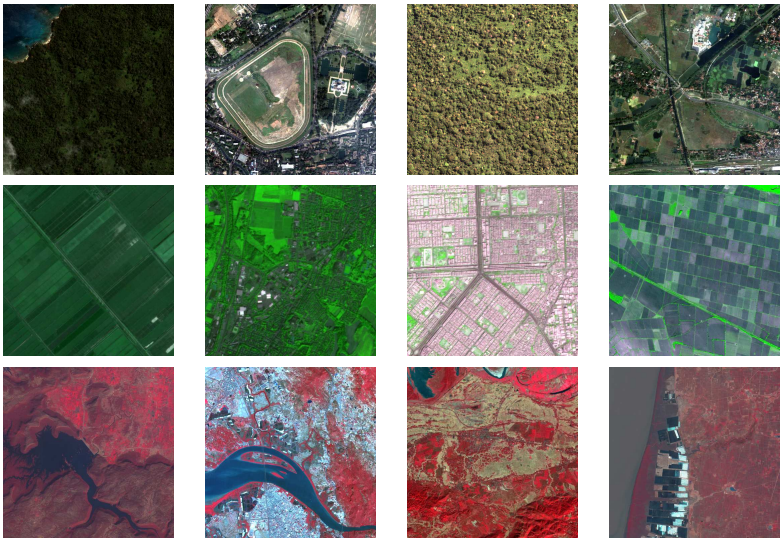


Figure 4.8: Example of training MS datasets: first row: QuickBird; second row: LISS-IV and third row: LISS-III

as.2620v4 running at 2.1 GHz, 20M cache, HT, 8C116T (8SW); RAM: 128 GB, 2400 MHz DDR4; HDD: 2 x 1.2 TB; Cores: 16 (32 virtually). The server is running on Ubuntu 16.04.3 LTS, 64 bit Linux operating system installed with open source computer vision (OpenCV) library packages (version 3.2.0) and OpenMP parallel programming tool for multicore programming.

For validation of the results several quantitative metrics are used for reference-based evaluation e.g. PSNR, MSSIM, sCC, UIQI, ERGAS, SAM, etc., along with a non-reference based parameter known as the NIQE. Also, spectral quality of super-resolved images are evaluated based on the spectral signature and end-member identification via spectral unmixing. In this work, we perform band wise reconstruction for all the test MS images and the visual results are shown using false RGB images of the results.

4.6.1.3 Comparison of SR results with different methods

In this section, we present results of MS image SR on different datasets as mentioned earlier for various upscales.

A) **SR of QuickBird MS images**

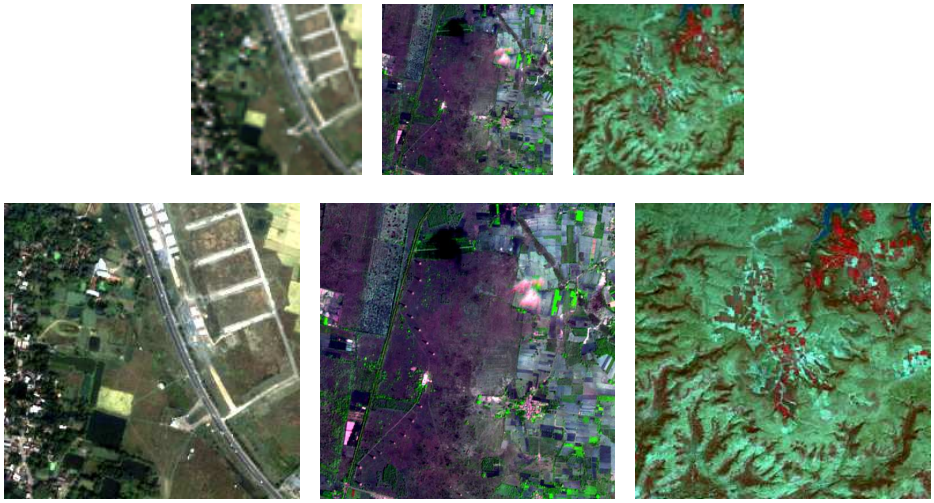


Figure 4.9: Test images acquired by QuickBird, LISS-IV and LISS-III sensors: first row: LR-MS input; second row: ground-truth

QuickBird sensor captures MS images which are very useful for land cover assessment. The dataset of this experiment is Multispectral[3] referred in Table 4.1. It consists of three visible bands, namely, blue, green, and red along with one NIR band of 2.8 m resolution. Training dataset includes 25 different MS images over the areas, like, the Chilika Lake, the Sundarbans, Ujong Kulon, Yala, etc. Few sample images of this dataset are displayed in Fig. 4.8. We extract LR and HR-patches of size 2×2 and 8×8, respectively from the dataset and trained HR and LR dictionaries of size 64×256 as detailed in subsection 4.4.1.

For reconstruction, the test image (ground-truth) is a 256×256 MS band image of resolution 2.8 m (capturing a scene over the India-Sundarban area passed on 2 November, 2002 by the satellite). Since, we do not have a HR-MS image, we consider the given MS image as the ground-truth. The LR-MS test image is obtained by applying the Gaussian low-pass filter (LPF) of size 5 and standard deviation $\sigma = 0.5$ on it followed by downsampling by a factor 4 as shown in Fig. 4.9. Thus, the size

Table 4.2: Quantitative measures for super-resolution reconstruction of the QuickBird test image (i.e. Multispectral-III) corresponding to upscale ratio 4.

Image	Methd	PSNR	MSSIM	ERGAS	SAM	UIQI	sCC	NIQE
	Bicubic	25.32	0.8015	7.663	4.438	0.7766	0.9690	17.92
	SparseFI	26.46	0.8402	6.714	4.296	0.8274	0.9761	13.10
Test3	Yang *et al.*	26.47	0.8499	6.762	4.280	0.8305	0.9767	14.04
	Moustafa *et al.*	26.52	0.8496	6.694	4.284	0.8298	0.9764	13.56
	Chen *et al.*	26.39	0.8290	7.149	4.328	0.8080	0.9740	13.70
	Lucas *et al.*	26.18	0.8070	7.477	4.423	0.7830	0.9700	13.41
	Proposed	26.60	0.8505	6.611	4.275	0.8312	0.9772	12.96

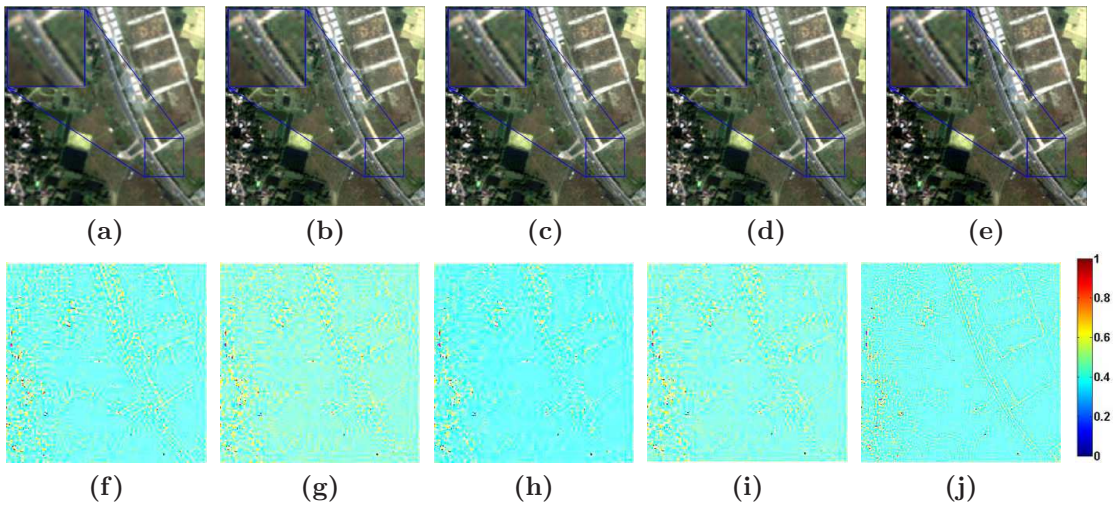


Figure 4.10: Visual outputs of different methods for 4× upscaling of the QuickBird test image. Row-wise: (a–e) SR images and (f–j) corresponding error images with ground-truth; left to right: Bicubic, SparseFI, Yang *et al.*, Moustafa *et al.*, and the proposed method.

of LR MS test image is 64×64 with a resolution of 11.2 m.

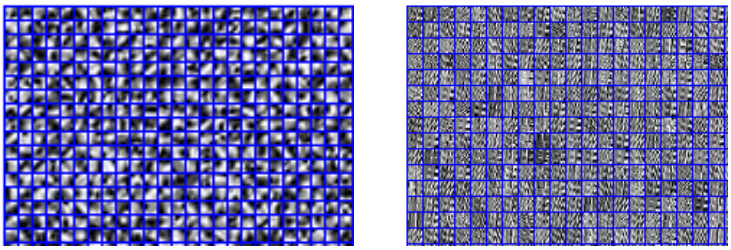
SR results are compared with different methods. A visual representation of SR outputs are shown in Fig. 4.10. Results show that the proposed method produces the least visible error compared to others. In order to validate this claim, we also carry out quantitative evaluation in terms of PSNR, MSSIM, ERGAS, SAM, UIQI, sCC and NIQE and results are presented in Table 4.2. It is observed that for upscale ratio 4, the proposed algorithm shows on an average improvements in PSNR by 1.28 dB over bicubic and 0.2-0.4 dB over other MS SR methods. While Moustafa *et al.* gives comparable results to that of the proposed method.

We have also carried out experiments for other zooming factors. For 2× upscaling, PSNR achieved for the proposed method, bicubic and Moustafa *et al.* are 28.86 dB, 26.43 dB, and 28.35 dB, respectively. Again, for 3× upscaling, these values are 27.48 dB, 25.89 dB, and 27.21 dB for the proposed, bicubic and Moustafa *et al.*, respectively. So, compared to the bicubic method, our method shows improvements of 2.43 dB and 1.59 dB for 2× and 3× upscales, respectively. Similarly, compared to Moustafa *et al.*, improvements of 0.51 dB and 0.27 dB are observed in case of the proposed method for 2× and 3× upscaling, respectively. Superior performances are

also achieved for other parameters as well for the proposed method. Closer results in case of Moustafa *et al.* may be because it is a self-learning method; the dictionary is learned from the given LR image itself. On the other hand, the proposed method utilizes a learned dictionary from noise-free MS images from the same satellite.

B) **SR of LISS-IV MS images**

The dataset of this experiment is Multispectral[2] as referred in Table 4.1. LISS-IV MS data consist of two visible bands (B2, B4) and one NIR band (B3) with a resolution of 5.8 m. The swath coverage is 70 km and bands are quantized using 10 bits.



(a) **(b)**

Figure 4.11: Trained dictionary pair using the proposed method for LISS-IV data: $\mathbf{D}_h$ (left) and $\mathbf{D}_\ell$ (right)

The training dataset consists of 30 images of different sizes: 256×256, 512×512 and 1024×1024. A set of representative images from the training dataset are displayed in Fig. 4.8. For selection of training images, RoIs are to be selected such that they contain homogeneous regions in each band. This ensures high feature contents in each of the training image and thereby supports better dictionary learning. For this experiment, 50,000 sample patches are selected from the training dataset to learn HR and LR dictionaries. Fig. 4.11 shows a visual representation of trained HR and LR dictionaries from the above dataset. We consider extracting patches of size 5×5 from both HR and upscaled LR MS images during dictionary training detailed in subsection 4.4.1 that produces LR and HR dictionaries of size 25×256 and 100×256, respectively.

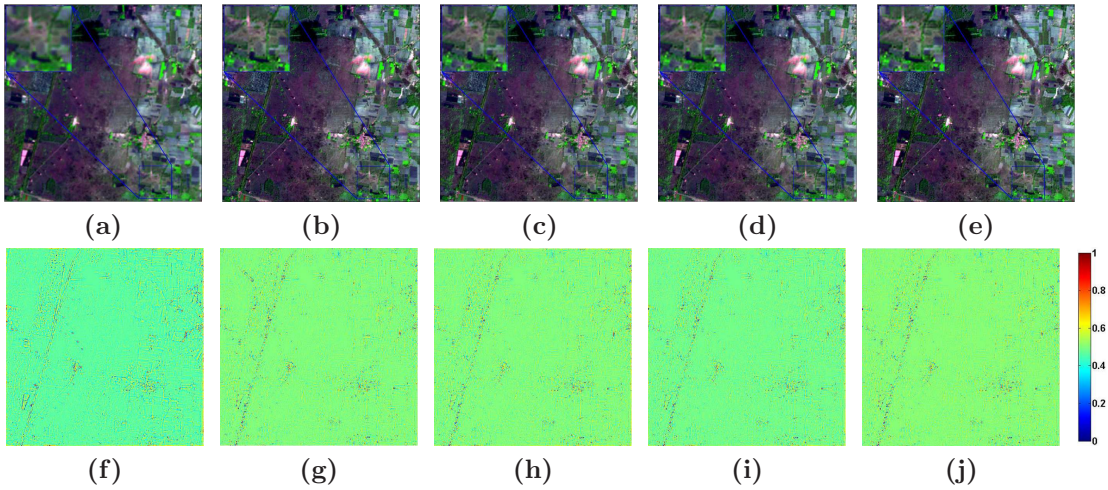


Figure 4.12: Visual outputs of different methods for 2× upscaling of LISS-IV test image. Row-wise SR images (a–e) and their corresponding error images (f–j) with ground-truth; left to right: Bicubic, SparseFI, Yang *et al.*, Moustafa *et al.*, and the proposed method

For reconstruction, we create test LR MS images by performing blurring and downsampling operations on each band of the input LISS-IV MS image. After SR reconstruction of test LR MS images, results are compared with the ground-truth. In this experiment, as shown in Fig. 4.9, a 512×512 size ground-truth band image is obtained from the LISS-IV with 5.8 m resolution passing over a region of India on 18 March, 2017. We apply LPF of size 5 with a standard deviation $\sigma = 0.5$ on the ground-truth and downsampled it by 2. So, the size of the LR MS test image for our experiment is 256×256 having resolution of 11.6 m.

Table 4.3: Quantitative measures for super-resolution reconstruction of the LISS-IV test image (i.e. Multispectral-II) corresponding to upscale ratio 2.

Image	Methd	PSNR	MSSIM	ERGAS	SAM	UIQI	sCC	NIQE
	Bicubic	30.11	0.9694	5.492	3.316	0.8454	0.9818	15.79
	SparseFI	32.54	0.9968	4.205	2.726	0.9135	0.9892	14.11
Test1	Yang *et al.*	31.37	0.9835	4.748	3.310	0.8783	0.9862	14.80
	Moustafa *et al.*	32.48	0.9937	4.233	2.769	0.9073	0.9892	14.25
	Chen *et al.*	32.35	0.9910	4.649	2.925	0.8900	0.9870	12.90
	Lucas *et al.*	32.45	0.9920	4.827	3.027	0.8870	0.9860	12.44
	Proposed	32.76	0.9974	4.106	2.663	0.9181	0.9897	12.91

SR reconstruction is carried out for different upscale ratios. A comparison of output images by different methods for upscale ratio 2 is shown in Fig. 4.12. Error images corresponding to each method are also shown. It is observed that the proposed method gives the least visible error compared to others.

Results of different objective parameters for LISS-IV test image are presented in Table 4.3. It is clearly observed that the performance of the proposed method is superior than other methods. On an average the increase in PSNR value of the proposed method for 2× zooming is about 2.7 dB over the bicubic, while it is approximately 0.3-0.4 dB better than Moustafa *et al.* and Lucas *et al.*. Similarly, improvement of values are also noticed for MSSIM and UIQI. Here, the NIQE value is also significantly reduced, which indicates a better reconstructed image. We have also conducted SR of the same image for 3× and 4× upscaling, where the proposed method is able to show better performances compared to others.

C) **SR of LISS-III MS images**

The dataset of this experiment is Multispectral[1] as referred in Table 4.1. LISS-III images are of medium resolution (i.e. 23.5 m) and consist of four spectral bands: three visible (B2, B3, B4) and one NIR (B5). The swath coverage is 140 km and different bands are quantized using 10 bits. We form a training dataset consisting of 25 MS images, which are selected in a manner similar to that explained for LISS-IV. A set of representative images for the training dataset are shown in Fig. 4.8. These images are considered as HR data from which LR images are generated for learning HR and LR dictionaries for our experiments.

We have considered extracting HR-patches of size 9×9 with 3 pixels overlapping and LR-patches of size 3×3 with 1 pixel overlapping to get HR and LR dictionaries of sizes 81×256 and 144×256, respectively.

Table 4.4: Quantitative measures for super-resolution reconstruction of the LISS-III test image (i.e. Multispectral-I) corresponding to upscale ratio 3.

Image	Methd	PSNR	MSSIM	ERGAS	SAM	UIQI	sCC	NIQE
	Bicubic	25.95	0.8108	7.724	4.023	0.7915	0.9454	18.06
	SparseFI	26.54	0.8414	7.082	3.917	0.8273	0.9545	15.23
Test2	Yang *et al.*	26.88	0.8607	6.787	3.859	0.8402	0.9570	14.96
	Moustafa *et al.*	27.10	0.8636	6.764	3.614	0.8491	0.9586	14.17
	Chen *et al.*	27.07	0.8510	7.014	3.683	0.8360	0.9550	13.89
	Lucas *et al.*	26.96	0.8500	7.091	3.757	0.8350	0.9540	12.01
	Proposed	27.24	0.8719	6.663	3.567	0.8601	0.9590	12.41

For reconstruction, a MS image band of size 510×510 with 23.5 m is obtained from a LISS-III image that contains scene over the location of Giri Forest, India and

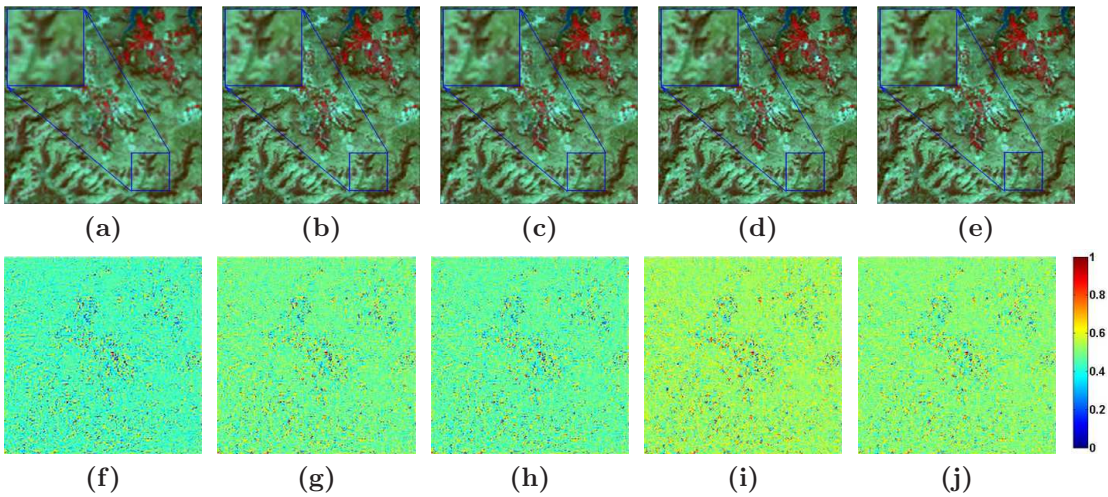


Figure 4.13: Visual outputs of different methods for 3× upscaling of the LISS-III test image. Row-wise: SR images (a–e) and their corresponding error images (f–j) with ground-truth; left to right: Bicubic, SparseFI, Yang *et al.*, Moustafa *et al.*, and the proposed method

having date of pass on 8 January, 2017. This image is considered as the ground-truth. The corresponding LR image for testing is obtained by applying LPF of size 5 and standard deviation $\sigma = 0.5$ followed by downsampling by a factor of 3. So, the size of the LR test image is 170×170 and resolution 70.5 m. These images are shown in Fig. 4.9.

Here, SR is performed for upscale ratio 3. Results are evaluated and compared with other MS image SR algorithms both visually and quantitatively and shown in Fig. 4.13 and Table 4.4, respectively. It is observed that the proposed method performs comparatively better than others even for a 3× upscaling of the test LR image. From visual inspection, it is clear that the proposed method produces less visible errors than others.

On an average the increase in PSNR value of the proposed method for 3× upscaling is about 1.29 dB over the bicubic output, while it is approximately 0.15 -0.3 dB more than Moustafa *et al.*, Chen *et al.* and Lucas *et al.*. Similarly, improvement of values are also noticed for MSSIM, UIQI and NIQE. We have also conducted SR of the same image for 2× and 4× zooming where the proposed method is able to show better performance compared to others.

Since all the above SR reconstructions are done band wise, for final visualization of the super-resolved MS image, a false color RGB image is shown using a combination of three band images for each dataset.

D) **SR of panchromatic images**

The data set of this experiment is Panchromatic[1] as referred in Table 4.1. A monochrome CartoSat-2 PAN image with a swath coverage of 9.6 km and spatial resolution around 0.65 m is considered. As mentioned above in MS image SR, we select 20 HR-PAN RoIs with homogeneous details for learning the dictionary. LR versions for each of these images are generated using the same process used for generating LR-MS test data. Examples of few sample images used in dictionary learning are shown in Fig. 4.14.

Figure 4.14: Example of training data set for CartoSat-2 PAN images of the regions of Patna, Chilka Lake and Amrtitsar, respectively

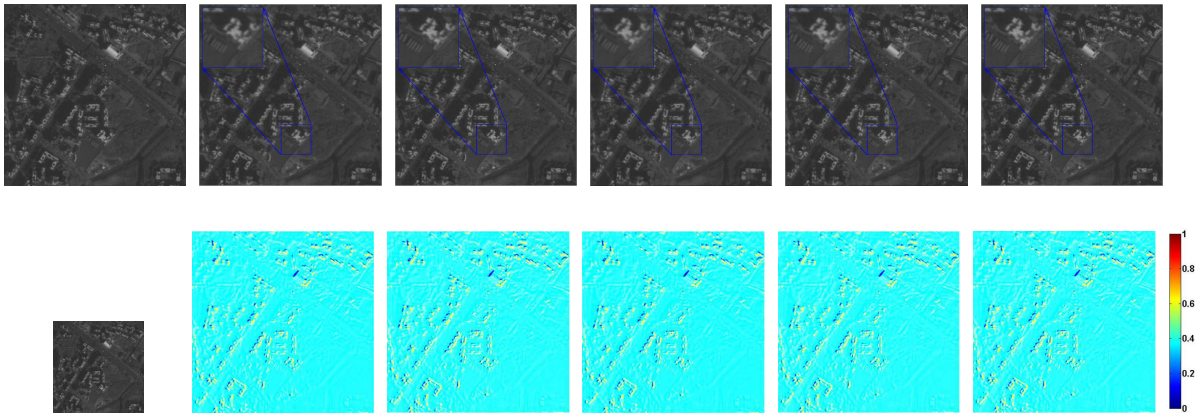


Figure 4.15: SR results of CartoSat-2 PAN test image for 2 times zooming. Column-wise from left to right: ground-truth and input images, resulted HR and error images by bicubic, Yang's method, Chen's method, Lucas's method and the proposed method

PAN image SR simply applies SISR algorithm which is applied to each band of MS image during SR. A ground-truth image of size 512×512 is obtained from

Table 4.5: Quantitative measures for super-resolution reconstruction of CartoSat-2 PAN image corresponding to upscale ratio 2.

Image	Methds	PSNR	MSSIM	UIQI	NIQE
Test3	Bicubic	40.13	0.9921	0.8586	24.29
	Yang *et al.*	41.43	0.9965	0.8825	17.39
	Chen *et al.*	41.32	0.9950	0.8720	18.16
	Lucas *et al.*	40.97	0.9950	0.8730	20.94
	Proposed	41.52	0.9966	0.8862	17.07

CartoSat-2B satellite of ISRO passing over Mumbai, India on 10 December 2015. An LR test image of size 256×256 is created by blurring and down-sampling. SR results are shown in Fig. 4.15. On an average the increase in PSNR value of the proposed method for PAN image SR is about 1.4 dB over the bicubic output, while it is approximately 0.1-2 dB than Yang's and Chen's methods. Similarly, improvement of values are also noticed for MSSIM, UIQI. The NIQE value is also significantly reduced, which indicates a better reconstruction in case of the proposed method.

E) **SR of real MS images**

In this work, experiments are carried out to check the performance of the proposed method, when input LR images itself are fed, without any preprocessing, like blurring and downsampling, to the proposed method for SR at different upscaling ratios. We evaluate the quality of super-resolved outputs by computing the no-reference evaluation metric, i.e. NIQE as mentioned above besides visual analysis. If we consider the QuickBird MS image as the input LR image directly and apply SR reconstruction to it for 2× upscaling, the NIQE values for both the input and output images obtained are 14.86 and 11.51, respectively. Similarly, we calculate NIQE values of reconstructed images with inputs taken from both LISS-IV and LISS-III sensors. NIQE measures for input and reconstructed LISS-IV images are 21.70 and 19.37, respectively, while those of LISS-III image are 24.25 and 21.49, respectively. In addition, the obtained visual results for different real test MS images are also better and at par with those obtained for downsampled LR-MS test images already explained above. We have shown the SR results in Fig. 4.16 for the three images discussed above.

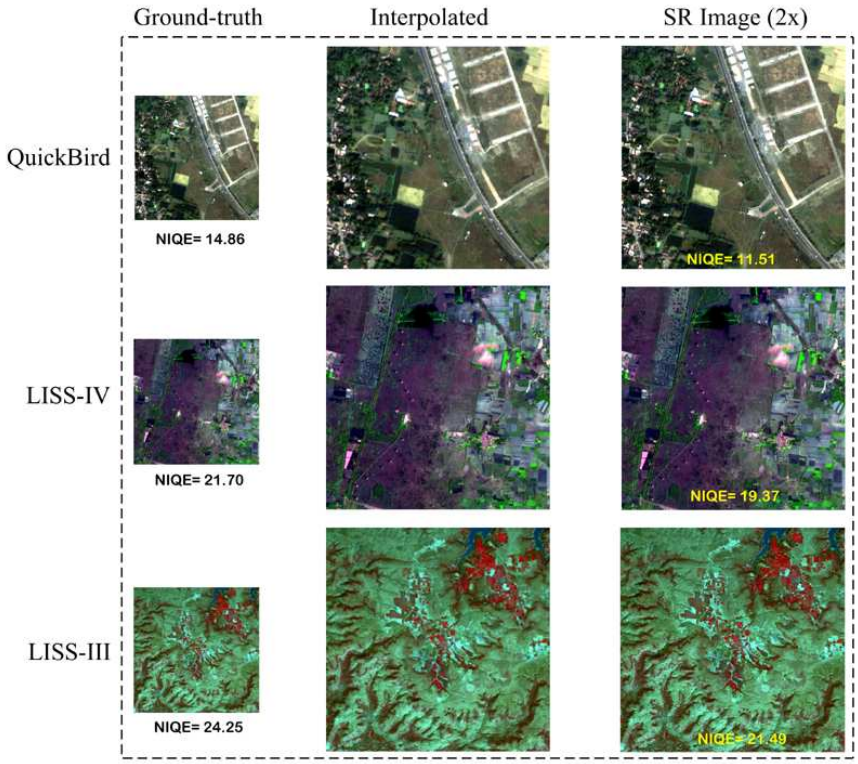


Figure 4.16: Visual outputs of real MS image SR from different sensors

4.6.2 Spectral signature comparisons

Different surfaces reflect or absorb electromagnetic radiation in varying amounts depending on the composition of their constituent materials, surface orientations, and so on. The reflectance of a material is also affected by the wavelength of electromagnetic waves, and by examining the reflactance difference in remote sensing images, we may gain a good notion about different land cover classes. A spectral signature, also known as a spectral graph, is a plot of the reflectance values of various band images with respect to the wavelength.

An RoI of homogeneous region corresponding to identical locations in each band is selected. Now, average intensities of pixels in the RoI corresponding to each band are plotted against the band numbers. Spectral profile plots of the QuickBird test MS image and corresponding HR-MS image are shown in Fig. 4.17, where it can be observed that spectral orientations of ground-truth and test images are not disturbed even after SR reconstruction.

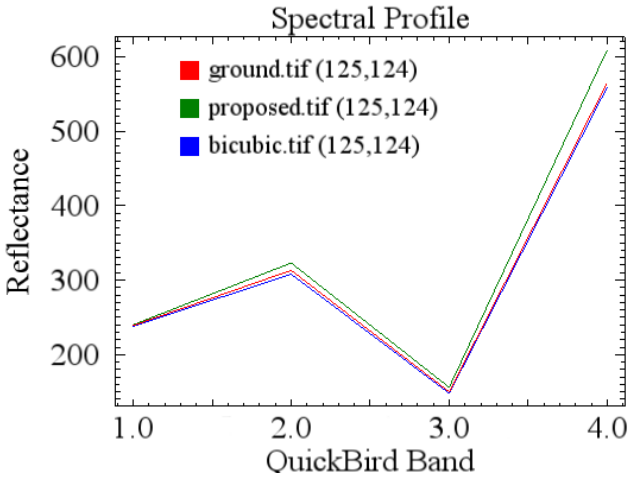


Figure 4.17: Comparisons of profiles: Spectral signatures for QuickBird ground-truth image, bicubic interploated image and super-resolved image by the proposed method

4.6.3 End-members identification via spectral unmixing

Decomposition of the spectral signature of a mixed pixel into a set of end-members and their corresponding abundances is known as spectral unmixing. A linear mixture model considers a mixed image; pixel is being a linear mixture of the end-members with fractional abundances which are positive and always add up to one. Spectral unmixing results are highly dependent on the input end-members; changing the end-members changes the results.

There are different methods of performing linear spectral unmixing, like, maximum likelihood based unmixing, spectral angle mapper, etc. In this work, we use ENVI software to perform spectral unmixing of LR and SR MS images using the maximum likelihood method and compare the effect of SR in end-member identifications. This is done by first, selecting different end-members from a MS image as an RoI in their abundance and then obtaining a mixed image with these end-member constituents.

In this example, three different types of end-members are selected from the given LISS-IV MS image that includes water body, vegetation and barren soil. As shown in the Fig. 4.18 the RoIs are selected for three end-member types, namely, 'water body' consisting of total 457 number of pixels, 'vegetation' consisting of 722 pixels

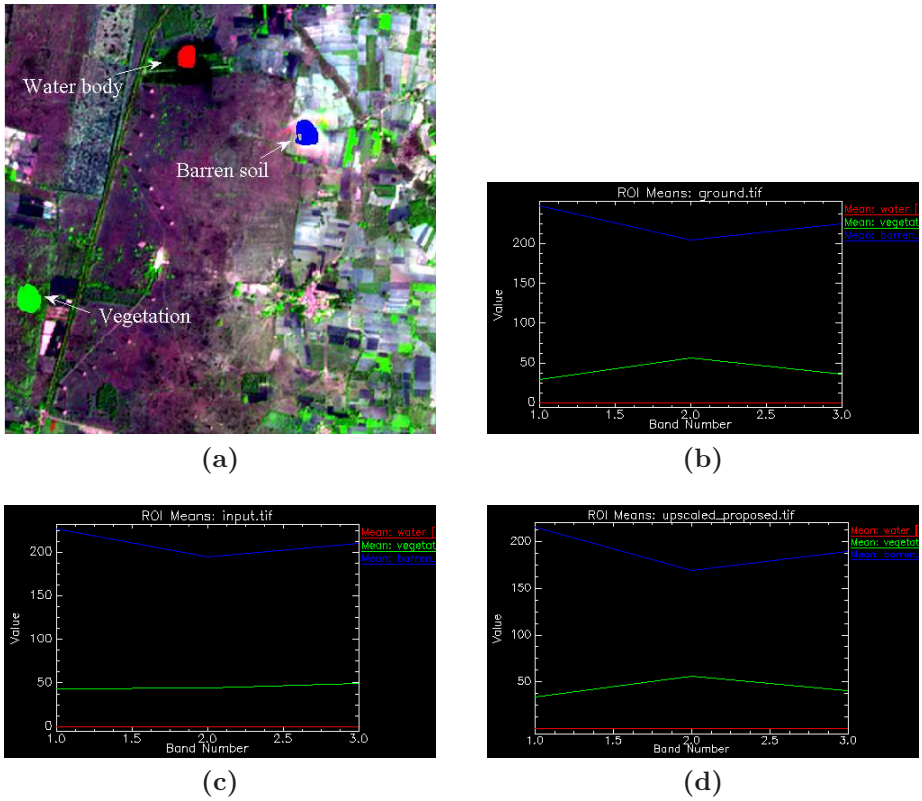


Figure 4.18: End-member identification from sepctral unmixing on LISS-IV MS image: (a) RoI image (b) spectral profiles for three end-members from ground truth image (c) spectral profiles for three end-members from LR-MS test image (d) spectral profiles for three end-members from resulted SR image

and 'barren soil' consisting of 606 pixels, respectively. The unmixed end-member profile distributions are displayed for the ground truth MS image, LR test image and SR MS image. It can be observed that, the SR reconstructed image by the proposed algorithm is better able to identify an end-member type of the true MS image compared to the LR test image.

4.6.4 Speed-up calculation

Fig. 4.19(a) shows computational time required with varying number of cores for the proposed dictionary learning algorithm. It can be noticed that the CPU time required for the multicore implementation of a dictionary of size 256 is reduced to 1 min. 49 secs with 16 cores from its sequential time of 52 min. 33 secs. The speed-up is about 28 times. In a similar way, as shown in Fig. 4.19(b), the highest speed-up

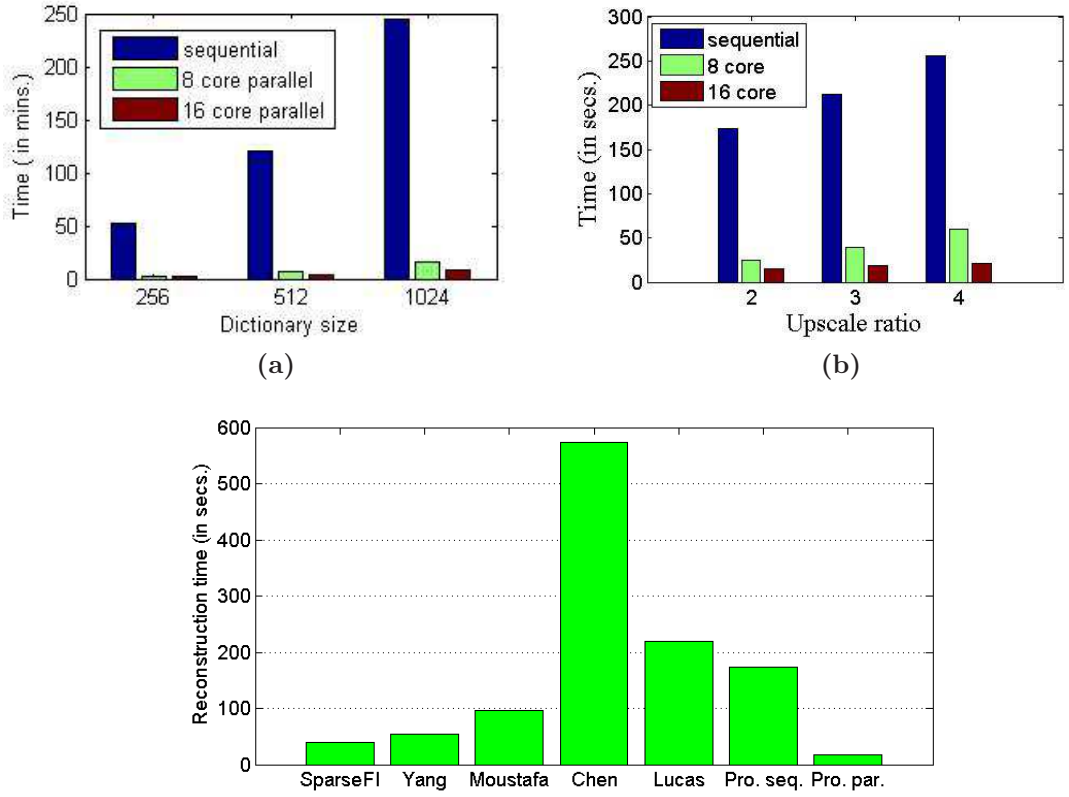


Figure 4.19: Plot of computational time versus number of cores (a) dictionary training time for different sizes (b) reconstruction time for different upscale ratio; and (c) Comparison of reconstruction time of different methods for zoom factor 2

for the SR reconstruction with a 256 size dictionary is achieved corresponding to $4\times$ upscaling, which is given as, 255.23 (secs.)/21.26 (secs.) $\approx$ 12.

Here, with a target to show a near real-time execution time, more focus is given on achieving a better speed-up of the proposed algorithm through parallel processing. Although the sequential execution time has less significance in achieving more speed-up, for general idea of the method, a comparison is also shown with all other methods in Fig. 4.19(c). It is observed that the proposed method with parallel implementation is able to provide fast results compared to others.

4.6.5 Effect of dictionary size

Fig. 4.20 shows plot of PSNR value versus dictionary size for the proposed method with zoom factor 4 for different bands of QuickBird Test image. It is observed that changing of dictionary size from 256 to 512 or 512 to 1024 contributes only a small change (about 0.05 – 0.10 dB) in PSNR while the parallel execution time of the algorithm is increased by approximately 5 times. Thus, a dictionary size of 256 is considered in the proposed algorithm to obtain a faster output.

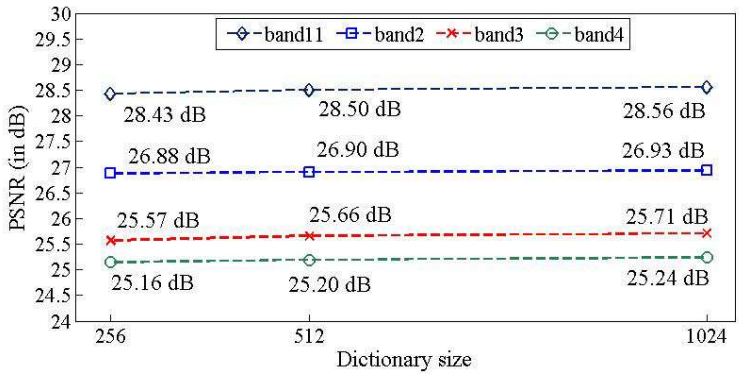


Figure 4.20: Comparison of band image's PSNR values with respect to different dictionary size in the proposed algorithm

4.6.6 Regularization parameter dependency

In sparse regularization problems, a balance between the sparsity of the solution and fidelity of the approximation is maintained by the regularization parameter λ. It controls the noise level of the input image and helps in optimal reconstruction of the target image by adjusting its value. Generally, for a more distorted data, the corresponding λ value should be also high for robust reconstruction. So, an empirical evaluation is conducted by setting different regularization parameter values and speculating the outputs for a set of test images to finalize its value.

In our experiment, we solve three regularization problems, namely, MCA decomposition, dictionary training and SR reconstruction. The regularization parameters controlling these problems are set as follows: For MCA decomposition in Eq. 4.6, we consider same parameter values as given in [39] and set the sparsity regularization

parameter λ_1 to 1 while the TV regularization parameter γ is set to [0, 2.5]. Again, for sparse representation during dictionary learning and SR reconstruction, the regularization parameter λ_2 in Eq. 4.12 and Eq. 4.13 is set to a common value. Here, the value of λ_2 depends upon the level of noise in the input image and we empirically set its value as 0.15 in all our experiments, which generally yields satisfactory results.

4.7 Conclusion

In this chapter, we demonstrate a sparse representation and morphological feature extraction-based MS SISR technique. Overcomplete dictionaries are learnt for both PAN and LR-MS datasets collected from QuickBird and ResourceSat-2, respectively. We consider both objective and visual analysis to evaluate reconstructed HR MS images. Results are compared with other recent methods and observe that the proposed method outperforms others both qualitatively and quantitatively. Speed-up is also achieved for the proposed algorithm using OpenMP-based parallel processing. As a future research, this work can be considered for GP-GPU based hardware implementations to achieve speed-ups suitable for real-time remote sensing applications.