INTERNATIONAL JOURNAL OF IMAGE AND DATA FUSION
https://doi.org/10.1080/19479832.2021.2019132

Taylor & Francis
Taylor & Francis Group

A region based remote sensing image fusion using anisotropic diffusion process

Bikash Meher[a], Sanjay Agrawal[a], Rutuparna Panda[a] and Ajith Abraham[b]

[a]Department of Electronics & Telecomm. Engg, VSS Univ. Of Tech, Burla, India; [b]Machine Intelligence Research Labs, Auburn, Washington, USA

ABSTRACT
The aim of remote sensing image fusion is to merge the high spectral resolution multispectral (MS) image with high spatial resolution panchromatic (PAN) image to get a high spatial resolution MS image with less spectral distortion. The conventional pixel level fusion techniques suffer from the halo effect and gradient reversal. To solve this problem, a new region-based method using anisotropic diffusion (AD) for remote sensing image fusion is investigated. The basic idea is to fuse the 'Y' component only (of YCbCr colour space) of the MS image with the PAN image. The base layers and detail layers of the input images obtained using the AD process are segmented using the fuzzy c-means (FCM) algorithm and combined based on their spatial frequency. The fusion experiment uses three data sets. The contributions of this paper are as follows: i) it solves the chromaticity loss problem at the time of fusion, ii) the AD filter with the region-based fusion approach is brought into the context of remote sensing application for the first time, and iii) the edge info in the input images is retained. A qualitative and quantitative comparison is made with classic and recent state-of-the-art methods. The experimental results reveal that the proposed method produces promising fusion results.

ARTICLE HISTORY
Received 18 July 2021
Accepted 6 December 2021

KEYWORDS
Image fusion; remote sensing; pansharpening; anisotropic diffusion; spatial frequency

1. Introduction

Remote sensing applications such as change identification, land cover classification and hazard monitoring require high spatial and spectral resolution images. However, due to the technological constraints in existing remote sensors, most of the remote sensing satellites like WorldView, SPOT and IKONOS give image data with different spectral and spatial resolutions. The high resolution and low spectral band is referred to as the PAN image, whilst the low resolution and high spectral band is referred to as the MS image. Because of the physical restraints, the MS image obtained from remote sensors usually has low spatial resolution. Therefore, it is needed to improve the spatial resolution of the MS image combined with the PAN image to obtain better spatial and spectral information. To achieve this, fusion of PAN and MS images is required, which can give complete information of the scene. This is called a pan-sharpening image (Zhang 2010, Xu *et al*. 2014, Pohl and van Genderen 2015, Li *et al*. 2018, Pandit and Bhiwani 2019).

CONTACT Rutuparna Panda r_ppanda@yahoo.co.in Department of Electronics & Telecomm. Engg, Vss Univ. Of Tech, Burla, India

Image fusion is performed at three levels, pixel, feature and decision. The pixel-based fusion method is the lowest level amongst the three levels of fusion (Ghassemian 2016). It takes into account the features of the remote sensing images acquired by diverse sensors and focuses on the statistical analysis of the pixel information to get the fused image. In the feature-level fusion method, a group of image pixels form a continuous region (using segmentation). It is also named region-based fusion. From this region, different features such as edges and texture can be extracted or classified into a variety of images from the same geographical location. The features are extracted employing different techniques. Then, the feature info is merged to enhance the spatial information (Mirzapour and Ghassemian 2015, Bai *et al*. 2014). The decision-level image fusion is the highest level of fusion. Here, the features are extracted from the regions and based on the classification and the decision is taken for the fusion (Luo *et al*. 2013, Mahmoudi *et al*. 2015, Tuia *et al*. 2018). Many remote sensing image fusion methods are reported in the last two decades. In a broad sense, they can be classified into different groups, i.e. component substitution (CS), multiresolution approach (MRA), model-based, hybrid, etc. In addition, deep learning-based fusion techniques have been suggested in recent years.

In CS-based techniques, the MS image is transformed into another colour space and replaced with the components obtained from the PAN image. Then, inverse transform is used to get the improved MS image. Many CS-based methods for remote sensing image fusion are reported in the literature. The Brovey transform (BT) is also known as colour normalisation transform as it involves the red-green-blue (RGB) colour transform procedure. However, the output image found using this technique experiences spectral distortion. The IHS-based technique utilises the colour transformation technique (Tu *et al*. 2001). These types of techniques are generally simple algorithms and have fast calculation speed. However, the contrast of the fused image reduces and produces poor fusion results. A nonlinear IHS scheme is proposed for the fusion of MS and PAN images in Ghahremani and Ghassemian (2016). The authors mainly focused on the estimation of the intensity value. The approximate intensity values are calculated using the local and global synthesis approach. However, the computational efficiency of the method is less as compared to other methods due to the patch-by-patch synthesis.

The MRA-based methods incorporate high frequency details extracted from the PAN image into the up-sampled MS image. Many methods are reported for the problem on hand. The discrete wavelet transform (DWT)-based method (Pajares and de La Cruz 2004) first extracts the spatial information from the PAN image. The extracted information is injected into the MS image to enhance the spatial resolution and decrease the colour distortion. But it does not extract the geometric structure of the image efficiently. Also, it suffers from the blurring effect. The shift invariant characteristics and directionality property of non-subsampled contourlet transform (NSCT) (Liu *et al*. 2015) overcome the shortcomings of the DWT method. However, the NSCT method is computationally less efficient. Shahdoosti and Ghassemian (2015) used the optimal filtering method, which is able to extract relevant and non-redundant information from the PAN images. The optimum filter coefficients are obtained utilising the statistical properties of the images, which are more consistent with the type and texture of remote sensing images as compared to other types of kernel-like wavelets. Restaino *et al*. (2016) fused the MS and PAN images using the morphological half gradient operator. The authors found an alternative MRA approach by investigating a non-linear MRA method implemented with

morphological pyramids. The process uses nonlinear decomposition using the morphological operator. Nowadays, edge-preserving filters such as guided filter and cross-bilateral filter are very popular and used in image fusion applications. These techniques preserve the significant edge information of the source images. However, the disadvantage of the cross-bilateral filter-based method is that it generates gradient reversal in the output image. Similarly, the fused image obtained with guided filter produces a halo effect. Tan *et al*. (2020) proposed remote sensing image fusion based on co-occurrence filtering (CoF) and the multi-scale morphological gradient domain dual-channel pulse-coupled neural network (PCNN). The CoF has the biggest advantage that the sharp details in the local region can be preserved when smoothing the fine textures. Wang *et al*. (2021b) proposed a MS and PAN image fusion method based on adaptive textural feature extraction using an adaptive guided filter and grey-level co-occurrence matrix (GLCM). The disadvantage of MRA-based methods is the selection of the decomposition level. Furthermore, these methods produce spatial distortion, i.e. blurring effect.

In recent years, mode- based fusion methods such as sparse representation (SR)-based (Li *et al*. 2013), compressed sensing-based (Ghahremani *et al*. 2019) and hierarchical Bayesian model (Golipour *et al*. 2015) are reported. Li *et al*. (2013) proposed remote sensing image fusion with SR over learned dictionary. The dictionary for PAN and low-resolution MS images is learned from the source images adaptively. Furthermore, a new approach is suggested to construct the dictionary for unknown high-resolution MS images without the training set. The sparse coefficients of the PAN image and the low-resolution MS image are found using the orthogonal matching pursuit (OMP) algorithm. The advantage of the SR technique is that it does not need a specific basis kernel-like wavelet, curvelet, etc. The limitations are as follows: (a) it is difficult to design a universal dictionary and (b) it is computationally intensive.

The hybrid methods improve the fusion performance by combining many different methods. Some popular hybrid methods reported are wavelet and SR (Cheng *et al*. 2015), SR and guided filter (Ma *et al*. 2019), à trous wavelet transform and IHS transform (Xin and Feng 2019), etc. Tambe *et al*. (2021) proposed the fusion of PAN and MS images by merging the principal component analysis (PCA) and rotated wavelet transform. They used this technique to eliminate the colour distortion, shifting effect and shift distortion in the fused image. This technique generally maps the MS image to a feature space by utilising the PAN image substituting its highest correlation component to enhance the spatial resolution of the output image. However, the computational complexity is high as compared to the individual PCA and non-subsampled rotated wavelet transform.

Recently, the deep learning approaches were proposed for the remote sensing image fusion problem. Scarpa *et al*. (2018) improved the architecture and training data of their previous convolutional neural network-based fusion work (Masi *et al*. 2016). They included a target-adaptive tuning phase to solve the problem of insufficient training data and allowed users to apply the proposed architecture to their own data. Azarang and Kehtarnavaz (2020) presented a multi-objective deep learning method for the pan sharpening. Liu *et al*. (2020) proposed a two-stream fusion network for the remote sensing image fusion and reconstructed the fused image from the fused features. Wang *et al*. (2021a) proposed a dual path fusion network to reduce the spectral distortion and enhance the spatial textural information. They have used two networks, i.e. the global subnetwork (GSN) and the local subnetwork (LSN) for the fusion. It gives good fusion results. However,

it has certain limitations such as the computational complexity is high and it relies on paired data and requires a large number of training samples. These methods show good fusion results as compared to the state-of-the-art methods. However, they require a large number of training samples and the computational time is also very high.

From the literature, it is known that the pixel level fusion is simple and easy to implement. However, it has certain limitations such as blurring effect and high sensitivity to noise. The decision-level fusion is complex as compared to pixel- and feature-level fusion. To overcome the problems of pixel-based fusion, the region-based fusion, which belongs to feature-level fusion, is preferred, which possesses certain advantages like less sensitivity to noise and use of semantic fusion rules (Meher *et al*. 2019).

Despite the availability of so many methods for remote sensing image fusion, the fusion accuracy and the fusion effect still need to be improved. Many problems still need to be solved such as selection of features, preserving loss of information and simplification of fusion rules. In this paper, the MS and PAN image fusion algorithm based on AD is proposed. Few papers are published on the successful use of AD filters for the fusion of infrared and visible images (Bavirisetti and Dhuli 2016, Yin and Zhang 2019), multi-exposure image fusion (Bhateja *et al*. 2019), etc.

As far as our knowledge is concerned, the AD filter is not used for remote sensing image fusion. This has motivated us to carry out the region-based remote sensing image fusion employing AD filters. For the first time, the AD filter-based approach is brought into the context of remote sensing image fusion. The AD filter is an edge-preserving and edge-smoothing filter. The AD procedure decomposes the input images into base layers and detail layers. These layers are combined using the region-based fusion rules. The FCM is used to segment the layers. The spatial frequency (SF) of each segmented region is compared. The fused image of base layers and the fused image of detail layers is obtained based on this comparison. The final output image is obtained by superposition of the fused detail and base layers.

The prime objectives of the suggested fusion technique are as follows: (i) integrating the information from the input MS and PAN images to produce the fused image that includes both the spectral and spatial information. (ii) Preserving the original colour information with less degradation. To achieve this, the colour and luminance information of the input images need to be separated. Therefore, the coloured input image has to be transformed from RGB colour space to another suitable colour space. Hence, choosing the suitable colour space for the fusion is another objective of the work. It is observed that some colour spaces such as Hue-Saturation-Value (HSV) colour space produce the ghosting effect in the output images. As found in the literature, the YCbCr colour space does not produce ghosting artefacts in the output images. The proposed fusion technique uses only the Y component of the MS image, whilst the chromaticity components (Cb and Cr) of the input image are preserved without any changes. (iii) Preserving edges, capturing fine details and smoothing the input images using the AD filter. (iv) Classification of the PAN image and the Y component of the MS image into various regions is an important step in region-based fusion. Hence, FCM is used to cluster the AD coefficients of the input images.

The remaining part of this article is organised in the following way: Section 2 explains the theory of AD and brief idea about the SF; Section 3 explains the suggested fusion scheme; Section 4 analyses the findings; finally, the conclusion in Section 5 gives a brief summary and critique of the findings.

2. Theoretical background

This section presents the different concepts used to model our proposed method. A brief idea of AD and SF is presented in the following sub-sections.

2.1. Anisotropic diffusion

The scale space technique (Babaud *et al.* 1986) produces coarser resolution images by convolving the original image with a Gaussian kernel. However, this method has major disadvantage: it is difficult to get precisely the locations of the meaningful edges at the coarser scale. Perona and Malik (1990) proposed a new definition of the scale space technique using a diffusion process called the AD process. It has transformed the application of the heat equation to digital images. AD is used to pre-process the image to efficiently retain the image texture details. The image is viewed as a heat field, with each pixel acting as a heat flow. It is determined whether to diffuse to the surroundings based on the relationship between the current pixel and the surrounding pixels. When the distance between the current pixel and the surrounding pixels is large, the surrounding pixels may form a boundary. The current pixel will not diffuse to the boundary as a result and the boundary will be preserved. Furthermore, it is known that the isotropic diffusion uses inter-region smoothing where the edges are not detected properly. The AD process overcomes this problem. It employs intra-region smoothing. The benefit of this approach is that, at every coarser resolution, the edges are sharp. The concept of AD has been applied in image processing to divide the image. It separates the pixels of the source image into two regions: homogeneous (base layer) and non-homogeneous (detail layer). The base layers are obtained by processing the input images using the AD technique. The detail layers are obtained by subtracting the base layers from the source images.

The AD equation for an input image I of dimension $M \times N$ (Perona and Malik 1990) is expressed as follows:

$$I^t = \mathrm{div}(\mathrm{c}(\mathrm{m}, n, t)\nabla I) = c(m, n, t)\Delta I + \nabla c \cdot \nabla I \tag{1}$$

where *div* is the divergence operator, $c(m, n, t)$ is the rate of diffusion or conduction coefficient, Δ is the Laplacian operator, ∇ is the gradient operator and I^t is the filtered image after t iterations. $c(m, n, t)$ is usually chosen as a function of image gradient to preserve the edges of the image. After solving Equation (1), it is expressed as follows:

$$I^{t+1} = I^t + \lambda[c_N \cdot \bar{\nabla}_N I^t + c_S \cdot \bar{\nabla}_S I^t + c_E \cdot \bar{\nabla}_E I^t + c_W \cdot \bar{\nabla}_W I^t] \tag{2}$$

where λ is a stability factor that is defined in the range $0 \leq \lambda \leq 0.25$, c_N, c_S, c_E *and* c_W represent the conduction coefficients in the four directions, i.e. north (N), south (S), east (E) and west (W), respectively. Here, $\bar{\nabla}$ represents the nearest neighbour difference and is defined as

$$\begin{aligned} \bar{\nabla}_N I_{m,n} &= I_{m-1,n} - I_{m,n} \\ \bar{\nabla}_S I_{m,n} &= I_{m+1,n} - I_{m,n} \\ \bar{\nabla}_E I_{m,n} &= I_{m,n+1} - I_{m,n} \\ \bar{\nabla}_W I_{m,n} &= I_{m,n-1} - I_{m,n} \end{aligned} \tag{3}$$

Similarly, the conduction coefficients c_N, c_S, c_E *and* c_W are represented as follows:

$$
\begin{aligned}
c^t_{N_{m,n}} &= f\left(\left\|(\nabla I)^t_{m+\left(1/2\right),n}\right\|\right) = f\left(\left|\overline{\nabla}_N I^t_{m,n}\right|\right),\\
c^t_{S_{m,n}} &= f\left(\left\|(\nabla I)^t_{m-\left(1/2\right),n}\right\|\right) = f\left(\left|\overline{\nabla}_S I^t_{m,n}\right|\right),\\
c^t_{E_{m,n}} &= f\left(\left\|(\nabla I)^t_{m,n+\left(1/2\right)}\right\|\right) = f\left(\left|\overline{\nabla}_E I^t_{m,n}\right|\right),\\
c^t_{W_{m,n}} &= f\left(\left\|(\nabla I)^t_{m,n-\left(1/2\right)}\right\|\right) = f\left(\left|\overline{\nabla}_W I^t_{m,n}\right|\right).
\end{aligned}
\tag{4}
$$

It is to be noted that the conduction coefficients can also be computed considering the eight directions (i.e. N, S, E, W, northeast (NE), northwest (NW), southeast (SE) and southwest (SW)). In fact, the results obtained in this work consider the eight directions. In Equation (4), $f(\cdot)$ is a non-negative monotonically decreasing function with $f(0) = 1$. Various functions can be employed for $f(\cdot)$. Here, two functions proposed in Perona and Malik 1990 are used as follows:

$$f(\nabla I) = e^{-(\|\nabla I\| k)^2} \tag{5}$$

$$f(\nabla I) = 1 + (\|\nabla I\| k)^2 \tag{6}$$

These two functions show a compromise between the image smoothing and retention of the edges. The function in Equation (5) is effective for images consisting of high contrast edges, whereas Equation (6) is effective for images consisting of wide regions. The constant k is used in both the functions. Its value is employed to select the permissible boundary region based on the edge strength. For a given image I, the AD is represented as AD(I).

2.2. Spatial frequency

The spatial frequency, which originated from the human visual system, indicates the overall active level in an image space. The human visual system is too complex to be fully understood with present physiological means, but the use of SF has led to an effective objective quality index for image fusion. The SF of an image block is defined as follows:

Consider an image I of size $M \times N$, where M equals the number of rows and N the number of columns. The row frequency (RF) and column frequency (CF) of the image block are given by

$$RF = \sqrt{MN \sum_{m=0}^{M-1} \sum_{n=1}^{N-1} \left[I(m,n) - I(m,n-1)\right]^2} \tag{7}$$

where $I(m,n)$ is the grey value of a pixel at position (m,n). Similarly,

$$CF = \sqrt{MN \sum_{n=0}^{N-1} \sum_{m=1}^{M-1} [I(m,n) - I(m-1,n)]^2} \tag{8}$$

The total SF of the image is then given as

$$SF = \sqrt{(RF)^2 + (CF)^2}. \tag{9}$$

The SF quantifies the amount of frequency content in an image. It contains the spatial details an image holds. It shows the clarity or sharpness of the image. The higher the value of SF, the richer the information the image contains. Moreover, the remote sensing images, which contain the forest area, building, road etc., are considered as high-frequency objects. Hence, in remote sensing image fusion, preserving the high frequency content is important in addition to enhancing the spectral quality (Li *et al*. 2001). Thus, we have used the SF as a feature for the problem on hand to select appropriate regions for fusion.

3. Proposed fusion scheme

The suggested fusion technique is based on the separation of the colour and the luminance components of the image. We have emphasised on retaining the original colour information. Therefore, we have separated the colour and the luminance components of the source images. Hence, the RGB image is converted into another colour space, which separates the colour and luminance components. It is observed that some colour spaces like IHS introduce the ghosting effect in the fused images (Umbaugh 2005). So to eliminate such an effect, we have chosen the YCbCr colour space. There may be chances of loss of colour information as the transform coefficients are generated after the transformation. To minimise this effect, we have applied transformation only to the 'Y' component of the source images. The other components 'Cb' and 'Cr' remain unchanged. The PAN image is fused only with the 'Y' component of the MS image. Here, we have used the FCM clustering technique for segmentation of the input images. The reason is that FCM is a soft clustering technique where the same pixel may belong to multiple clusters as per the membership value. The segmentation results are more accurate as compared to other clustering techniques such as k-means. Note that the accuracy of segmentation influences the fusion results also.

The block diagram of the suggested technique is depicted in Figure 1. The input images are assumed to be registered. Let the two input images PAN and 'Y' component of the MS image be denoted as $I_{PAN}(m,n)$ and $I_Y(m,n)$, respectively.

The RGB MS image is transformed into YCbCr colour space. The AD process is employed to the PAN image and only the 'Y' component of the MS image. The FCM clustering is utilised to segment the detail layers and the base layers obtained after applying the AD process. The SF of each of the segmented regions in both the detail layer and the base layer is computed. Then, the fused base layer and the fused detail layer are obtained by comparing the corresponding regions in the 'Y' component image and the PAN image. The new 'Y' component is obtained by linear combination of the fused

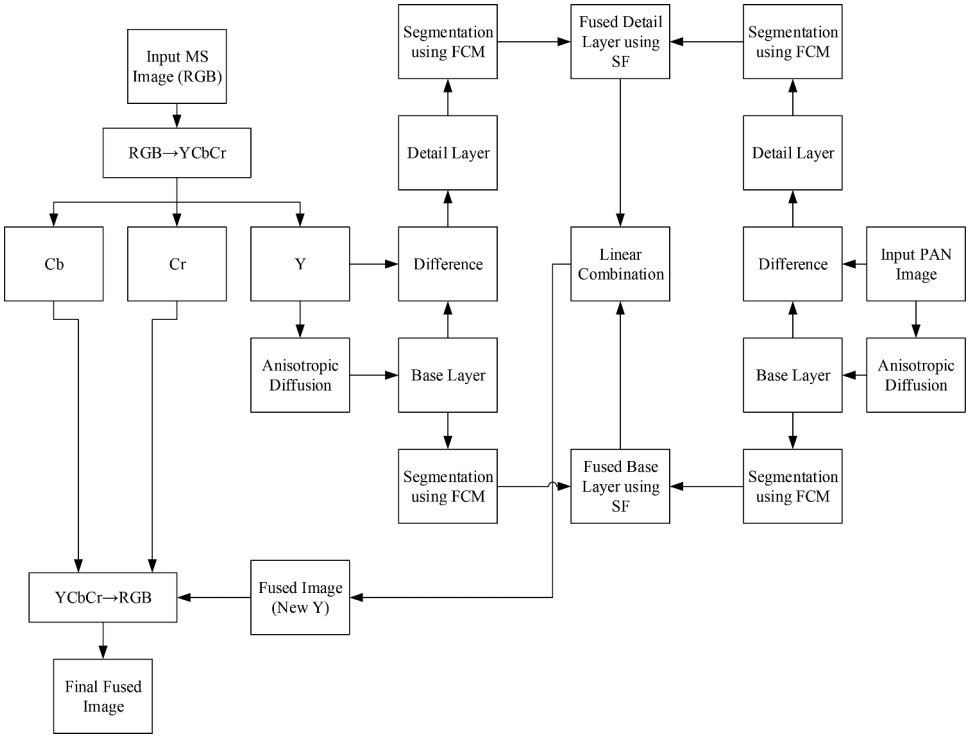

Figure 1. Block diagram of the suggested fusion technique.

base layer and the fused detail layer. Finally, the output image is obtained by merging the new 'Y' component with 'Cb' and 'Cr' components and converting back to RGB colour space.

The different steps of the suggested technique are presented as follows:

3.0.1. Step 1: extracting the Base Layers and the Detail Layers utilising the AD Method

Apply the AD process to acquire the base layers and detail layers from the input images using Equation (1). The base layers are represented as $B_{I_{PAN}}(m,n)$ and $B_{I_Y}(m,n)$ for the $I_{PAN}(m,n)$ and $I_Y(m,n)$ images, respectively. They are expressed as follows:

$$\begin{aligned}(B_{I_{PAN}}(m,n) &= \mathrm{AD}(I_{PAN}(m,n))\\ (B_{I_Y}(m,n) &= \mathrm{AD}(I_Y(m,n))\end{aligned} \tag{10}$$

The detail layers $D_{I_{PAN}}(m,n)$ and $D_{I_Y}(m,n)$ are similarly obtained by subtracting the base layers from the input images using the following equations:

$$\begin{aligned}(D_{I_{PAN}}(m,n) &= I_{PAN}(m,n) - B_{I_{PAN}}(m,n)\\ (D_{I_Y}(m,n) &= I_Y(m,n) - B_{I_Y}(m,n)\end{aligned} \tag{11}$$

3.0.2. Step 2: fusion of detail layers and base layers

The fusion of detail layers is carried out as follows:

(i) The FCM is applied to the detail layers of both I_Y and I_{PAN} images for segmentation. In this study, we have taken 10 clusters, as the remote sensing images contain a number of objects representing the main features such as urban area, agriculture land and deep water. The number of clusters is chosen after testing with several values. Hence, 10 segmented regions will be obtained for each of the images.
(ii) The SF of each of the segmented regions in the detail layer is determined using Equation (9) for both the images.
(iii) The SF of each region of the detail layer of the 'Y' component image is compared with the corresponding region of the detail layer of the PAN image. The final fused detail layer is obtained as

$$D_{Fus}(m,n) = \begin{cases} D_{I_{Y_i}}(m,n) & if \quad SF(D_{I_{Y_i}}) > (D_{I_{PAN_i}}) \\ D_{I_{PAN_i}}(m,n) & otherwise \end{cases} \tag{12}$$

where i= 1, . . .,10; $D_{I_{Y_i}}$ and $D_{I_{PAN_i}}$ is the ith region of the D_{I_Y} and $D_{I_{PAN}}$, respectively.

The same process is followed for generating the final fused base layer. The final fused base layer is denoted as $B_{Fus}(m,n)$.

3.0.3. Step 3: final fusion layer obtained by the superposition of $D_{Fus}(m,n)$ and $B_{Fus}(m,n)$

The new 'Y' component of the fused image is found by linear summation of the fused base layer and the fused detail layer. This process is followed because we got the detail layer by subtracting the base layer from the input image in Equation (11),

$$Y = D_{Fus}(m,n) + B_{Fus}(m,n) \tag{13}$$

3.0.4. Step 4: combine the new Y component with Cb and Cr components

The final output image is obtained by summation of the new 'Y' component with the 'Cb' and 'Cr' components of the MS input image and converted back to RGB colour space.

4. Results and discussion

This section presents the assessment of the proposed approach for remote sensing image fusion. A detailed description of the data sets used is presented in the following subsection followed by a brief description of the image fusion indices. Then, the compared methods are outlined. Finally, the results are presented in the form of figures and tables with discussions.

4.1. Data sets

The suggested technique is evaluated using three groups of images from the Worldview-2 satellite, two groups of images from the IKONOS satellite and two groups of images from the QuickBird satellite. The PAN and the MS images captured by the WorldView-2 satellite are of 0.5 m and 1.8 m spatial resolution, respectively. The PAN and the MS images captured by the QuickBird satellite are of 0.7 m and 2.8 m spatial resolution, respectively. Similarly, the PAN and the MS images captured by the IKONOS satellite are of 1 m and 4 m

spatial resolution, respectively. We have considered the fusion case with three spectral bands: Red (R), Green (G) and Blue (B) only. The MS and PAN images to be fused are geometrically registered and the size of the images is 256×256. As the reference images are rarely available for validation, in this paper, we have used the original MS image as the reference image (ground truth) as per Wald's protocol (Wald *et al.* 1997). In the experiments, the original MS image is resampled and then used as the input test data. Such a kind of test data is called as simulated data. Hence, as per Wald's protocol, the test data in this paper belong to simulation data.

4.2. Image fusion indices

For comparison, several image fusion indices are used. The indices are standard deviation (SD), average gradient (AG), correlation coefficient (CC), peak signal to noise ratio (PSNR), structural similarity index (SSIM), the root mean square error (RMSE), the relative global synthetic error (ERGAS) and universal quality index (Q). Note that the best in class results are displayed in bold. The metrics used are described in detail in the respective literature. However, they are presented in brief as follows:

SD: In general, the standard deviation represents the contrast of an image. An image having a high SD will indicate a high contrast. It is described by the degree of deviation between pixel's intensity levels of the image. It is expressed as

$$SD = \sqrt{M \times N \sum_{m=1}^{M} \sum_{n=1}^{N} (F(m,n) - \overline{F})^2} \tag{14}$$

where $F(m,n)$ is the output fused image F at pixel location (m,n) and $\overline{F}$ represent the average value of the output image (Wang and Chang 2011). The SD value of the fused image should be high.

AG: The average gradient computes the image clarity. It is expressed as follows:

$$AG = (M-1)(N-1) \times \sum_{m=1}^{M-1} \sum_{n=1}^{N-1} \sqrt{(F(m,n) - F(m+1,n))^2 + (F(m,n)) - F(m,n+1))^2 2} \tag{15}$$

where $M \times N$ is the size of the image and $F(m,n)$ is the pixel of the output image (Dong *et al.* 2015). The large value of AG indicates that the output image is of larger resolution.

CC: It is expressed as follows:

$$CC(F,Z) = \frac{\sum_{m=1}^{M} \sum_{n=1}^{N} [F(m,n) - \bar{F}][Z(m,n) - \bar{Z}]}{\sqrt{\sum_{m=1}^{M} \sum_{n=1}^{N} [F(m,n) - \bar{F}]^2 \times \sum_{m=1}^{M} \sum_{n=1}^{N} [Z(m,n) - \bar{Z}]^2}} \tag{16}$$

where Z is the reference image. $\bar{F}$ and $\overline{Z}$ represent the mean value of the output image and the reference image, respectively. *CC* is utilised to determine how the spectrum information is conserved (Yakhdani and Azizi 2010). The value should be close to 1.

PSNR: It is stated as follows:

$$PSNR = 10\log|F_{\max}^2 M \times N\sum_{m=1}^{M}\sum_{n=1}^{N}[F(m,n) - Z(m,n)]^2| \tag{17}$$

where $F_{\max}$ is the maximum grey value of the output image F (Naidu 2010). The value of PSNR should be high.

SSIM: It is expressed as follows:

$$SSIM(F,Z) = (2\mu_F\mu_Z + b_1)(2\sigma_{FZ} + b_2)(\mu_F^2 + \mu_Z^2 + b_1)(\sigma_F^2 + \sigma_Z^2 + b_2) \tag{18}$$

where μ_F is the average value of the output image F, μ_Z is the average value of the reference image Z, ${\sigma_F}^2$ denotes the variance of the output image, ${\sigma_Z}^2$ denotes the variance of the reference image, σ_{FZ} represent the covariance of the output image F and the reference image Z and b is a constant (Wang *et al*. 2004). The high value of SSIM shows more similarity between the reference and the output image.

RMSE: It calculates the quality of the output image by relating the reference and the output image. It is stated as (Zoran 2009)

$$RMSE(F,Z) = \sqrt{M \times N\sum_{m=1}^{M}\sum_{n=1}^{N}[F(m,n) - Z(m,n)]^2} \tag{19}$$

The RMSE value should be smaller.

ERGAS: It provides a universal quality measure of the output image F and is expressed as follows:

$$ERGAS = 100kl\sqrt{p\sum_{i=1}^{p}[RMSE(i)Mean(i)]^2} \tag{20}$$

where k denotes the PAN image spatial resolution, l represents the MS image spatial resolution, p denotes the number of bands of the output image F, $Mean(i)$ denotes the average value of the ith band of the reference image Z and $RMSE(i)$ calculates the RMSE between the ith band of Z and the output image F (Wald 2002). The value of ERGAS should be low.

Universal quality index (Q): The metric Q is mathematically expressed as

$$Q = \left(\frac{\sigma_{FZ}}{\sigma_F\sigma_Z}\right)\cdot\left(\frac{2\bar{F}\bar{Z}}{(\bar{F})^2 + (\bar{Z})^2}\right)\cdot\left(\frac{2\sigma_F\sigma_Z}{\sigma_F^2 + \sigma_Z^2}\right) \tag{21}$$

where $\bar{F}$ and $\bar{Z}$ denote the average values of the fused image F and the reference image Z respectively, σ_F and σ_Z represent the standard deviation of F and Z, respectively and $\sigma_F\sigma_Z$ is the covariance between F and Z (Wang and Bovik 2002). The value of this metric ranges between −1 and +1. The highest value should be preferred and it is close to +1.

4.3. Compared Methods

The proposed technique has been compared with many classic and recent state-of-the-art methods: IHS (Tu *et al*. 2001), BT (Earth Resource Mapping Pty Ltd 1990), DWT (Pajares and de La Cruz 2004), NSCT (Liu *et al*. 2015), SR (Yang and Li 2010), DWT-SR (Liu *et al*. 2015),

Table 1. Performance comparison for group1 images from the WorldView-2 data set.

Method	SD	AG	CC	PSNR	SSIM	RMSE	ERGAS	Q
IHS	60.9322	14.4218	0.8233	16.5029	0.9991	38.1415	8.4061	0.8160
BT	82.8489	13.4938	0.8143	11.2804	0.9971	69.5892	34.6119	0.6172
DWT	63.2286	15.0694	0.8570	17.5443	0.9993	33.8326	7.8565	0.8520
NSCT	64.1907	14.5119	0.8716	17.8664	0.9994	32.6017	7.5601	0.8654
SR	59.4783	13.4729	0.8515	17.3695	0.9993	34.5286	7.6948	0.8455
DWT-SR	63.1557	15.0874	0.8371	16.9473	0.9992	36.2402	8.4453	0.8344
PSQ-PS	67.2538	12.8828	0.5084	12.0841	0.9980	63.4358	14.6486	0.5020
MF-HG	73.8446	19.4777	0.9029	17.8274	0.9994	32.7469	8.1561	0.8794
NIHS	57.8227	10.9166	0.6483	14.3286	0.9987	48.9904	11.8328	0.6478
ADF	61.9711	12.7602	0.8955	18.9373	0.9995	28.8189	6.7038	0.8916
Proposed	62.6920	13.3344	0.9121	19.6715	0.9996	26.4831	6.2487	0.9084

preserving spectral quality for pan sharpening (PSQ-PS) (Shahdoosti and Ghassemian 2015), morphological half gradient (MF-HG) (Restaino *et al*. 2016), non-linear intensity-hue-saturation (NIHS) (Ghahremani and Ghassemian 2016) and our implementations of the pixel-based anisotropic diffusion based fusion (ADF). A method on the deep learning target-adaptive convolutional neural network (TACNN) (Scarpa *et al*. 2018) is also added for a comparison.

4.4. Fusion results and discussion

The proposed method is compared with classic and recent state-of-the-art methods. A comparison of the proposed method with other methods for group1 images from the WorldView-2 data set is shown in Table 1.

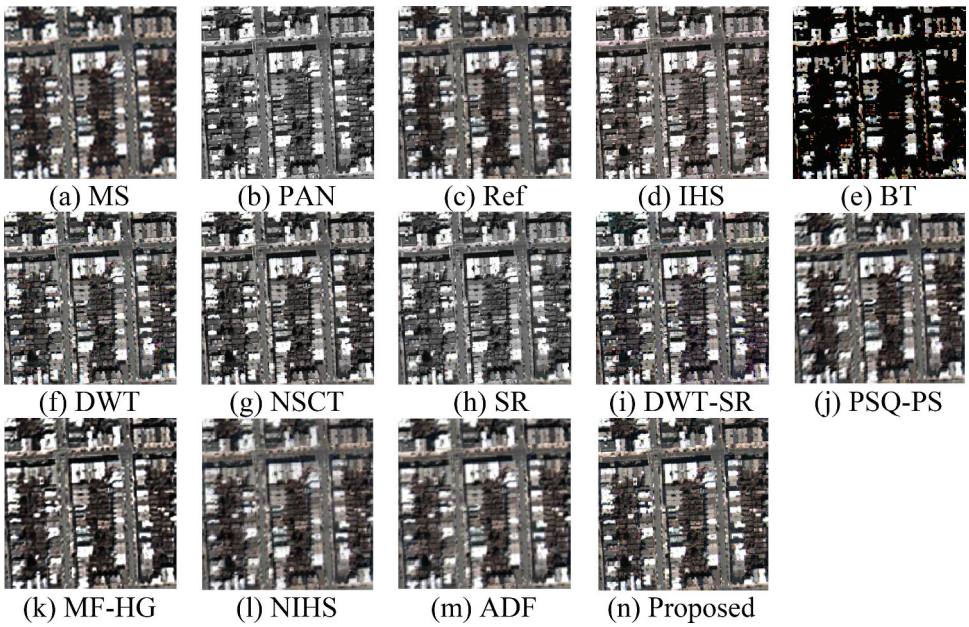

Figure 2. Fusion results of group1 images from the WorldView-2 data set: (a) MS input image, (b) PAN input image, (c) Reference image and (d)-(n) fused images.

Table 2. Performance comparison for group2 images from the WorldView-2 data set.

Method	SD	AG	CC	PSNR	SSIM	RMSE	ERGAS	Q
IHS	70.2432	10.9272	0.8424	16.5120	0.9991	38.1024	10.3916	0.8374
BT	83.6887	9.6771	0.8248	11.9389	0.9974	64.5158	34.6725	0.6471
DWT	68.6583	11.0220	0.8742	17.6131	0.9993	33.5681	9.1256	0.8708
NSCT	68.9706	10.9593	0.8748	17.6000	0.9993	33.6180	9.1371	0.8710
SR	68.9703	11.0061	0.8423	16.6002	0.9991	37.7407	10.1288	0.8380
DWT-SR	68.5656	11.0363	0.8468	16.7328	0.9992	37.1730	9.8926	0.8438
PSQ-PS	71.6730	10.8184	0.7113	13.7283	0.9988	52.4963	13.5328	0.7046
MF-HG	74.3095	12.2163	0.7029	13.4082	0.9987	54.4668	13.9956	0.6914
NIHS	68.1856	8.7353	0.7398	14.4923	0.9991	48.0765	12.4248	0.7356
ADF	71.3982	9.8709	0.8863	17.7422	0.9993	33.0702	9.2423	0.8796
Proposed	68.4204	8.9104	0.9428	20.9363	0.9997	26.4966	6.4620	0.9396

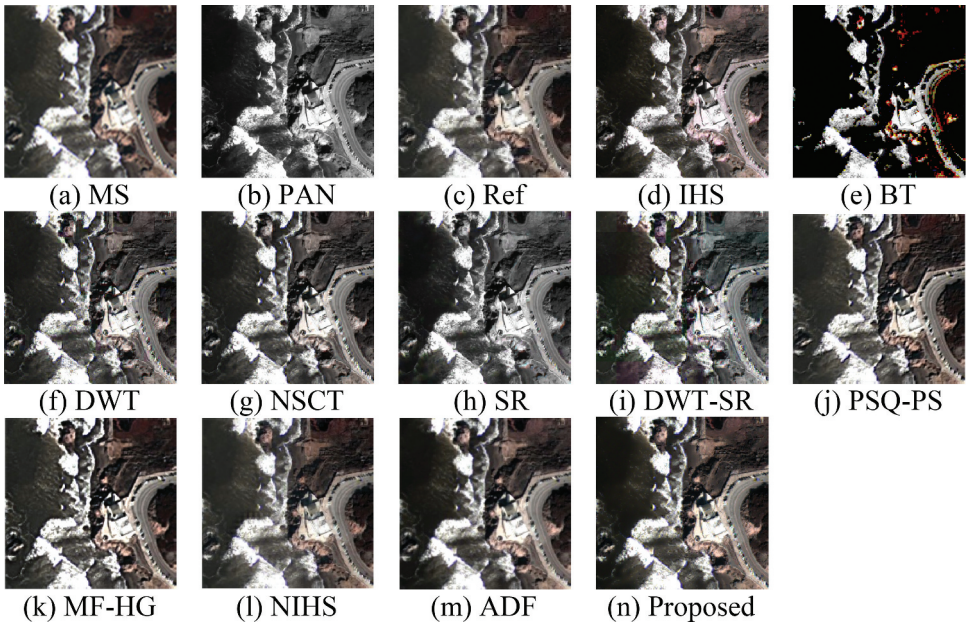

Figure 3. Fusion results of group2 images from the WorldView-2 data set: (a) MS input image, (b) PAN input image, (c) Reference image and (d)-(n) fused images.

It is seen from Table 1 that the suggested technique gives best in class values in terms of CC, PSNR, SSIM, RMSE, ERGAS and Q indices. It is observed that the BT method outperforms in terms of SD and MF-HG methods in terms of AG. The reason may be that it directly gets the most original spectral content from the MS image with zero truncation error.

A visual comparison of the fusion result is shown in Figure 2. It is observed that the visual clarity of the output images in (e) is not good as compared to the other techniques. Moreover, the spectral information is insignificant in (e). On the other hand, the edges are more clearly visible in Figure (h), (j), (l), (m) and (n). The fused image obtained with the

Table 3. Performance comparison for group3 images from the WorldView-2 data set.

Method	SD	AG	CC	PSNR	SSIM	RMSE	ERGAS	Q
IHS	49.0694	14.0350	0.7956	16.0261	0.9990	40.2982	12.3366	0.7310
BT	45.3006	7.9086	0.6506	13.5305	0.9982	53.7226	85.8940	0.3491
DWT	50.1273	14.2050	0.8585	18.5719	0.9994	30.1015	10.6729	0.8230
NSCT	50.8479	14.1297	0.8701	18.7341	0.9995	29.5346	10.4821	0.8326
SR	49.0651	14.0354	0.7929	16.0078	0.9990	40.5203	12.3702	0.7271
DWT-SR	50.8866	14.1373	0.8579	18.5652	0.9994	30.1017	10.7941	0.8230
PSQ-PS	50.5424	11.5088	0.6407	15.9121	0.9994	40.8266	16.7691	0.6301
MF-HG	61.8218	21.1533	0.7972	16.3035	0.9991	39.0275	15.0744	0.7403
NIHS	48.6393	9.6334	0.5726	15.3719	0.9992	43.4458	17.5145	0.5647
ADF	51.1551	10.3656	0.5243	14.3381	0.9989	48.9369	17.5713	0.5031
Proposed	49.1815	11.8883	0.9018	20.2875	0.9996	24.6774	9.1697	0.8763

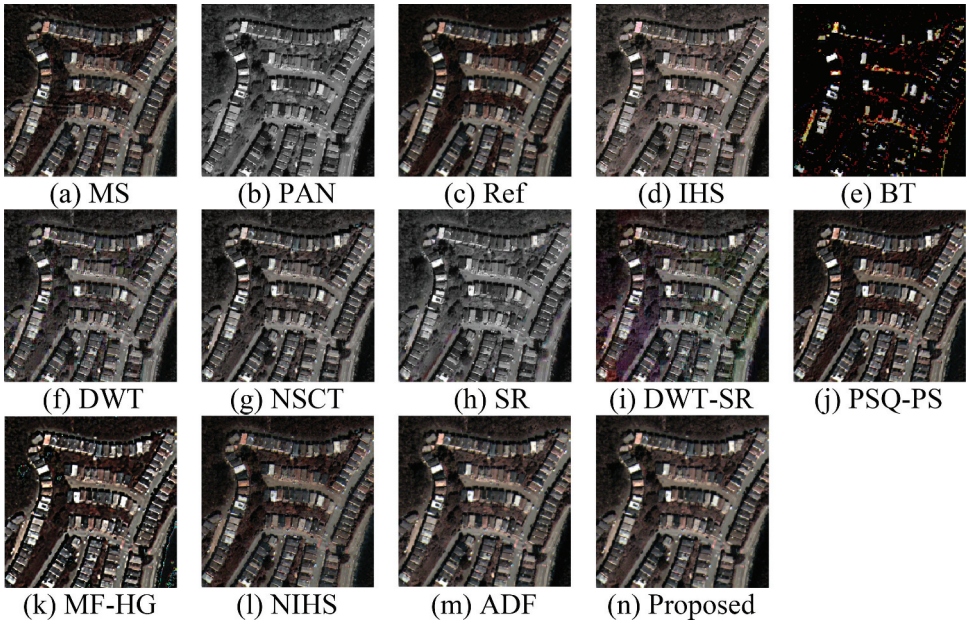

Figure 4. Fusion results of group3 images from the WorldView-2 data set: (a) MS input image, (b) PAN input image, (c) reference image and (d)-(n) fused images.

Table 4. Performance comparison for group1 images fromthe QuickBird data set.

Method	SD	AG	CC	PSNR	SSIM	RMSE	ERGAS	Q
IHS	40.8217	19.2845	0.5392	14.5382	0.9985	47.8272	9.4763	0.5240
BT	61.8211	17.6025	0.7035	10.4716	0.9961	76.3794	54.3992	0.4274
DWT	41.2172	19.7907	0.6595	16.6631	0.9991	37.4809	8.2091	0.6722
NSCT	41.3113	19.6859	0.7278	17.5148	0.9991	33.9910	7.4742	0.7205
SR	38.1460	18.7130	0.5346	14.7758	0.9986	46.7233	9.3396	0.5169
DWT-SR	40.7730	17.9048	0.6492	17.1941	0.9993	35.2598	8.4080	0.6474
PSQ-PS	45.2733	16.7167	0.6728	17.0161	0.9994	35.9532	8.5595	0.6705
MF-HG	41.6592	16.8123	0.6837	17.6446	0.9994	33.4435	8.4449	0.6835
NIHS	34.1539	12.0636	0.7298	18.8189	0.9996	29.2146	7.2377	0.7120
ADF	33.1352	12.8147	0.7090	17.9523	0.9992	32.2795	7.2283	0.6825
Proposed	38.7961	9.0120	0.7853	18.8196	0.9995	29.2119	6.5191	0.7777

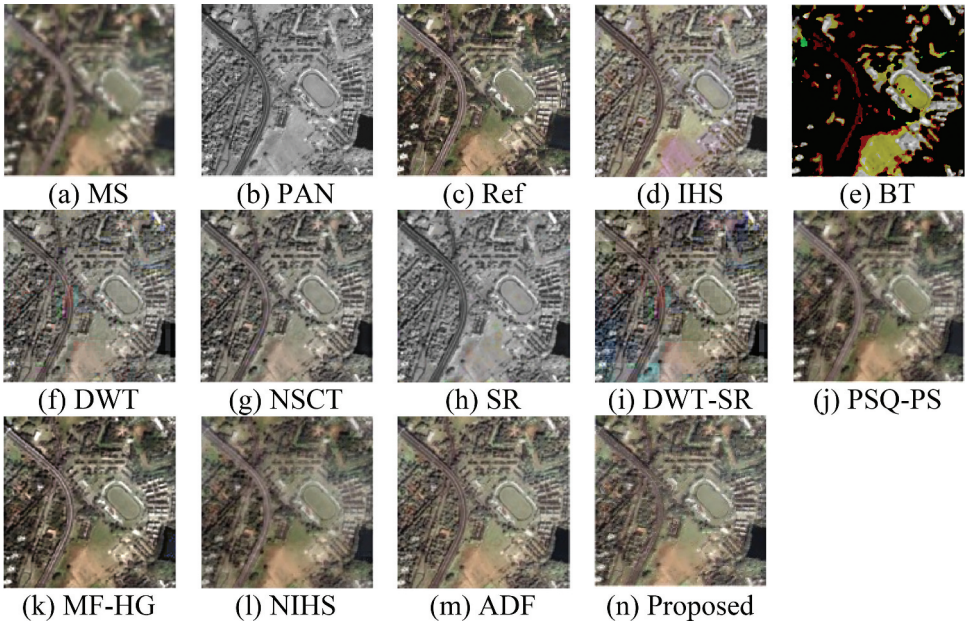

Figure 5. Fusion results of group1 images from the QuickBird data set: (a) MS input image, (b) PAN input image, (c) reference image and (d)-(n) fused images.

proposed procedure shows better spectral and spatial information as compared to the other techniques. Furthermore, the region-based approach in (n) outclasses the pixel-based approach in (m). The contrast is improved and the edges are finer.

From Table 2, it is seen that the suggested technique gives best values in terms of CC, PSNR, SSIM, RMSE, ERGAS and Q indices. The MF-HG gives best value in terms of AG. The reason may be the preservation of more edge information of the HF components. The BT method shows high value in terms of SD.

It is seen from Figure 3(e) that the output image found using the BT technique produces spectral distortion. The visible clarity is not so good. Most of the images retain the edge information but lacks spectral content. The figure (n) retains most of the spectral information and it looks visually more prominent as compared to the other methods. The retention of the PAN image conserves most of the detail information and the spectral information is conserved by the MS image.

Table 5. Performance comparison for group2 images from the QuickBird data set.

Method	SD	AG	CC	PSNR	SSIM	RMSE	ERGAS	Q
IHS	75.5290	20.5402	0.4837	10.5931	0.9975	75.3701	12.1750	0.4183
BT	93.3044	32.6907	0.6593	10.5834	0.9972	75.4169	19.4059	0.5146
DWT	64.7050	22.3902	0.6255	13.4735	0.9989	54.0848	9.7627	0.5869
NSCT	65.0379	22.2941	0.5847	13.0172	0.9990	56.9976	10.1847	0.5470
SR	70.4629	21.6175	0.5710	11.6196	0.9980	67.0249	10.9216	0.5096
DWT-SR	74.2990	21.4711	0.5380	11.2185	0.9971	70.2727	11.6103	0.4714
PSQ-PS	51.5608	16.9034	0.7320	16.9310	0.9994	36.3073	7.3885	0.7281
MF-HG	53.4754	20.4813	0.7860	17.6560	0.9992	33.3997	6.8578	0.7785
NIHS	46.5953	15.3429	0.8304	19.4649	0.9996	27.1206	5.6313	0.8304
ADF	48.0100	14.8781	0.8695	20.4266	0.9990	24.2784	4.9638	0.8690
Proposed	46.6570	14.8720	0.9225	22.8509	0.9991	18.3653	3.8198	0.9225

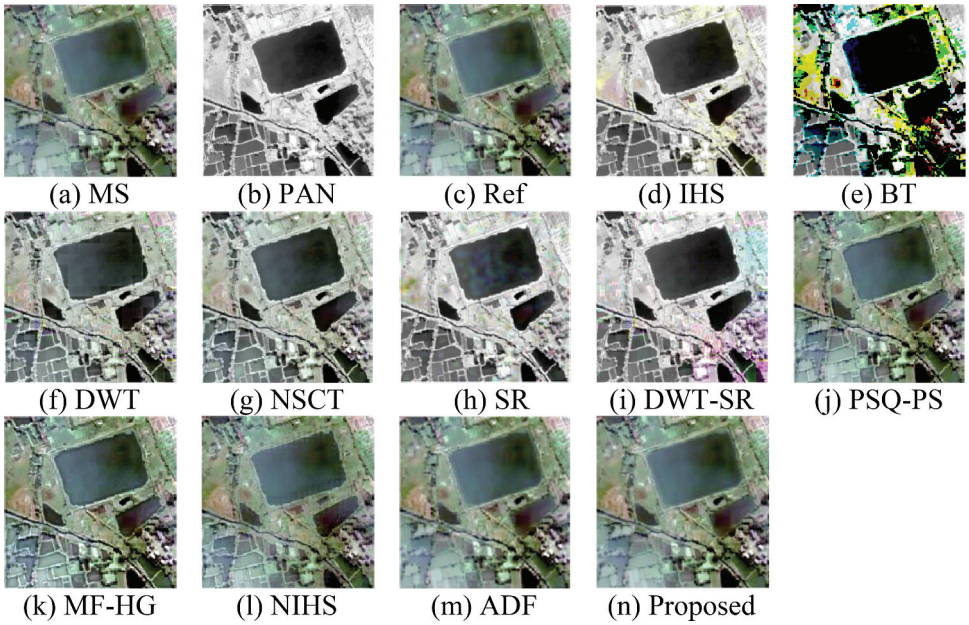

Figure 6. Fusion results of group2 images from the QuickBird data set: (a) MS input image, (b) PAN input image, (c) reference image and (d)-(n) fused images.

Table 6. Performance comparison for group1 images from the IKONOS data set.

Method	SD	AG	CC	PSNR	SSIM	RMSE	ERGAS	Q
IHS	57.1765	19.6420	0.5593	13.8491	0.9988	51.8888	8.9762	0.5515
BT	74.4610	21.3132	0.7194	12.6174	0.9982	59.8701	14.7415	0.6398
DWT	54.5365	19.1299	0.7084	15.9188	0.9994	40.9265	7.2528	0.7033
NSCT	53.9534	18.6471	0.7011	15.8647	0.9995	41.1545	7.2879	0.6967
SR	51.3681	16.6222	0.7432	16.6359	0.9996	37.9623	6.5526	0.7394
DWT-SR	54.6506	18.9394	0.6100	14.7149	0.9987	47.3977	8.3726	0.6042
PSQ-PS	55.8405	16.4490	0.7857	17.1698	0.9993	35.3405	6.2652	0.7798
MF-HG	58.0360	20.6530	0.7603	16.5103	0.9992	38.1587	6.9110	0.7510
NIHS	49.3971	14.3334	0.8192	18.6552	0.9995	29.7702	5.4295	0.8191
ADF	51.7233	14.8112	0.7148	16.4593	0.9993	38.3637	6.8696	0.7136
Proposed	52.6227	15.3086	0.8580	19.7602	0.9996	26.2137	4.8451	0.8561

The metrics CC, PSNR, SSIM, RMSE, ERGAS and Q have best values in the case of the proposed technique as shown in Table 3. However, the SD and AG values are best for the MF-HG method.

It is seen from Figure 4(e) that the output image obtained using the BT method looks blurred. Furthermore, the edges are also not clear. The output in (h) and (i) lacks the spectral information. Figure 4(j), (k) and (n) look visually more prominent. The output image found using the suggested technique retains most of the spectral infornation of the input images. Furthermore, it retains most edge information.

From Table 4, it is found that the CC, PSNR, RMSE, ERGAS and Q values are high for the proposed method as compared to the other methods. However, the SSIM value of the proposed method is close to the best in class value. In fact, the proposed method is the second contestant in the case of the SSIM value.

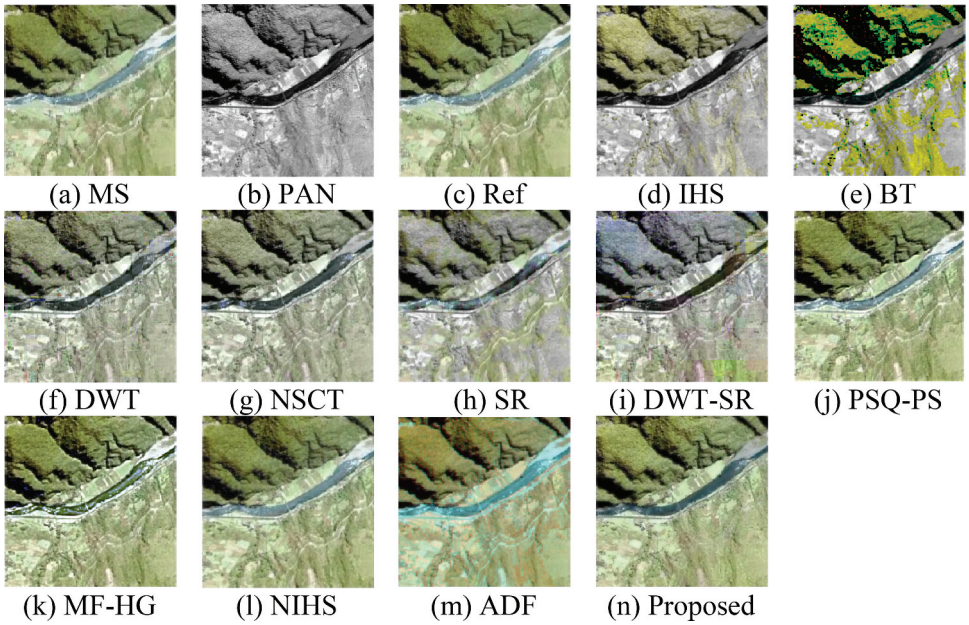

Figure 7. Fusion results of group1 images from the IKONOS data set: (a) MS input image, (b) PAN input image, (c) reference image and (d)-(n) fused images.

Table 7. Performance comparison for group2 images from the IKONOS data set.

Method	SD	AG	CC	PSNR	SSIM	RMSE	ERGAS	Q
IHS	29.1859	10.8485	0.4709	10.1262	0.9963	79.5847	13.3616	0.3733
BT	39.5533	10.0040	0.3925	10.3633	0.9965	77.3508	16.9379	0.1241
DWT	33.9254	12.3921	0.6142	14.5673	0.9989	47.6934	10.5309	0.5688
NSCT	32.8545	12.4359	0.5992	14.6790	0.9989	47.0829	10.4875	0.5550
SR	28.4986	10.6482	0.3597	11.2031	0.9970	70.2948	12.8661	0.2979
DWT-SR	38.4416	11.7791	0.4893	12.5605	0.9980	60.1988	12.1731	0.4388
PSQ-PS	35.4997	10.7722	0.7010	19.2375	0.9998	27.8402	9.0466	0.7006
MF-HG	33.0679	6.5930	0.6799	19.2180	0.9997	27.9040	9.4163	0.6766
NIHS	38.7031	11.4761	0.6229	17.8239	0.9998	32.7605	10.3677	0.6218
ADF	29.8549	9.9323	0.6994	14.7812	0.9988	46.5030	10.0841	0.6323
Proposed	31.8205	4.9963	0.7967	15.0607	0.9989	45.0304	9.6851	0.7272

It is observed from Figure 5(e) that it produces high colour distortion as compared to the other output images. By visually comparing the output images with the reference image, it is seen that most of the methods produce pan-sharpened images with some degree of blurriness as in Figure 5 (j) and (l). In this context, the output obtained using the proposed method is better than the other methods.

It is seen from Table 5 that the proposed technique outperforms other methods in terms of CC, PSNR, RMSE, ERGAS and Q values. The BT method shows high SD and AG values, whereas the NIHS method shows the best value in terms of SSIM. It is interesting to note that the proposed method is close to the best in class value.

From Figure 6(e), it is seen that the BT method generates the spectral distortions. The fused images Figure 6(d), (f), (g), (h) and (i) contain less colour information, whereas the fused images Figure 6(j), (k), (l), (m) and (n) preserve more spectral and spatial resolutions.

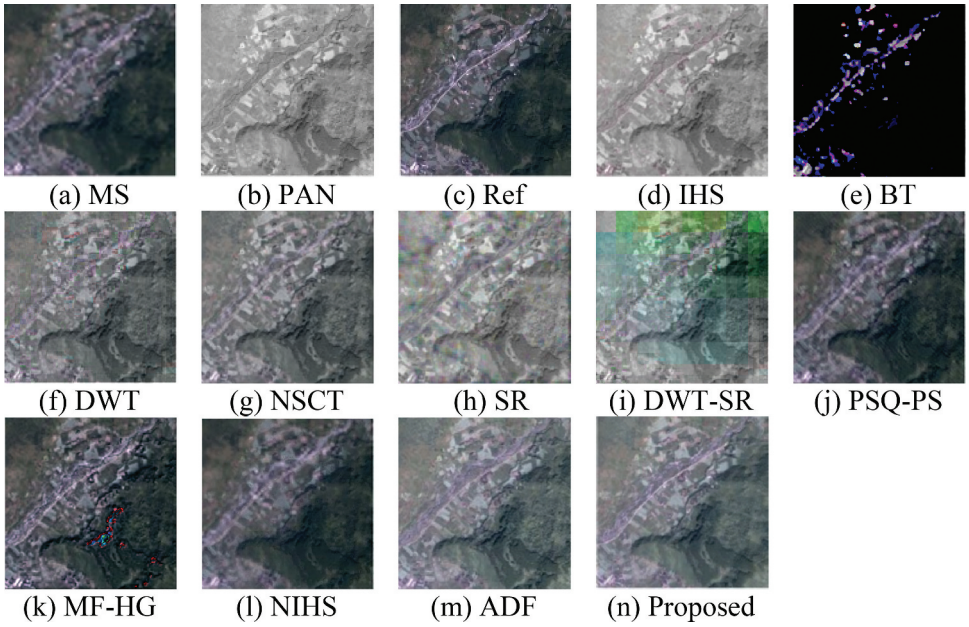

Figure 8. Fusion results of group2 images from the IKONOS data set: (a) MS input image, (b) PAN input image, (c) reference image and (d)-(n) fused images.

From Table 6, it is found that the CC, PSNR, SSIM, RMSE, ERGAS and Q values obtained using the proposed technique that outperforms other methods. The BT method shows high value in terms of SD and the MF-HG method shows high value in terms of AG.

The colour of the fused image in Figure 7(e) is unnatural, which is produced by the BT method. The output images Figure 7(h) and (i) contain less spectral information. The river part is visible more prominent in Figure 7(n). Also, the fused image Figure 7(n) contains more spectral and spatial details.

It is observed from Table 7 that the proposed technique outperforms other methods in terms of CC and Q values. The SD value is high for the BT method. The NSCT method gives the best AG value. The PSQ-PS method gives best output in terms of PSNR, SSIM, RMSE and ERGAS. Nonetheless, the proposed method is close to the winners. A distinct feature about this case is that the proposed method gives the best value for CC and Q. The possible reason may be the region-based approach where the pixel correlation is taken into consideration. Furthermore, the AD approach retains the necessary information required for a better quality output, resulting in a higher Q.

From Figure 8, it is seen that the spectral as well as spatial information is degraded in Figure 8(e). Some colour distortion is seen in Figure 8(i) and (k). Figure 8(j) and (l) looks blurry. However, the fused image in Figure 8(n) retains most of the spectral and spatial contents as compared to the other methods.

From the outputs shown above, we can infer that the contrast of the fused image is high with more spectral distortion in the case of the BT-based fusion method. The reason may be that the colour transform is obtained by multiplicative action with the PAN components. In DWT-based methods, the max coefficients of the low frequency and the

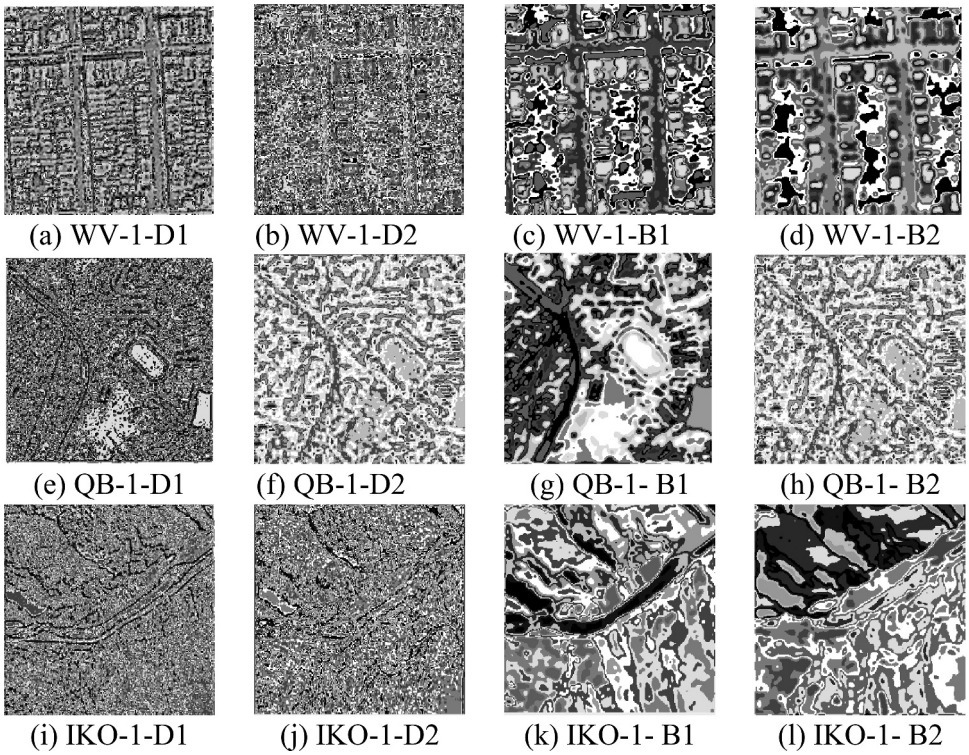

Figure 9. Segmentation results of group1 image pairs from different data sets: (a and b) detail layers for WorldView-2, (c and d) base layers for WorldView-2, (e and f) detail layers for QuickBird, (g and h) base layers for QuickBird, (i and j) detail layers for IKONOS and (k and l) base layers for IKONOS.

high frequency are fused, which is not able to retain all the spectral contents of the MS image. Similarly, due to the directional properties of NSCT, the spectral information from high-frequency components may not be preserved well in the fused image. The 'choose-max' rule may cause spatial inconsistency in the fused image in the SR-based method. In the case of the DWT-SR-based method, the high-frequency coefficients are fused using the maximum selection fusion rule and only the sparse representation-based fusion rule is used for low-frequency components. The more spectral components are contained in high-frequency components. So less spectral information may be preserved in the fused image.

To further strengthen our claim, a recently proposed pan-sharpening method based on MS and PAN image fusion in Wang *et al*. 2021b is considered for comparison. The average fusion metrics RMSE, ERGAS and SSIM of two data sets i.e. Worldview-2 and Quickbird, are given. The values obtained for Worldview-2 are RMSE: 20.73, ERGAS: 15.5 and SSIM: 0.76. The values obtained for Quickbird are RMSE: 24.49, ERGAS: 17.48 and SSIM: 0.64. It is observed that most of the metrics obtained are better in the case of our proposed method.

The segmentation results of group1 image pairs from different data sets is shown in Figure 9. The segmented detail layers for the 'Y' component of the MS image and the PAN image from WorldView-2 data set are shown in Figre 9(a), (b). Similarly, the segmented

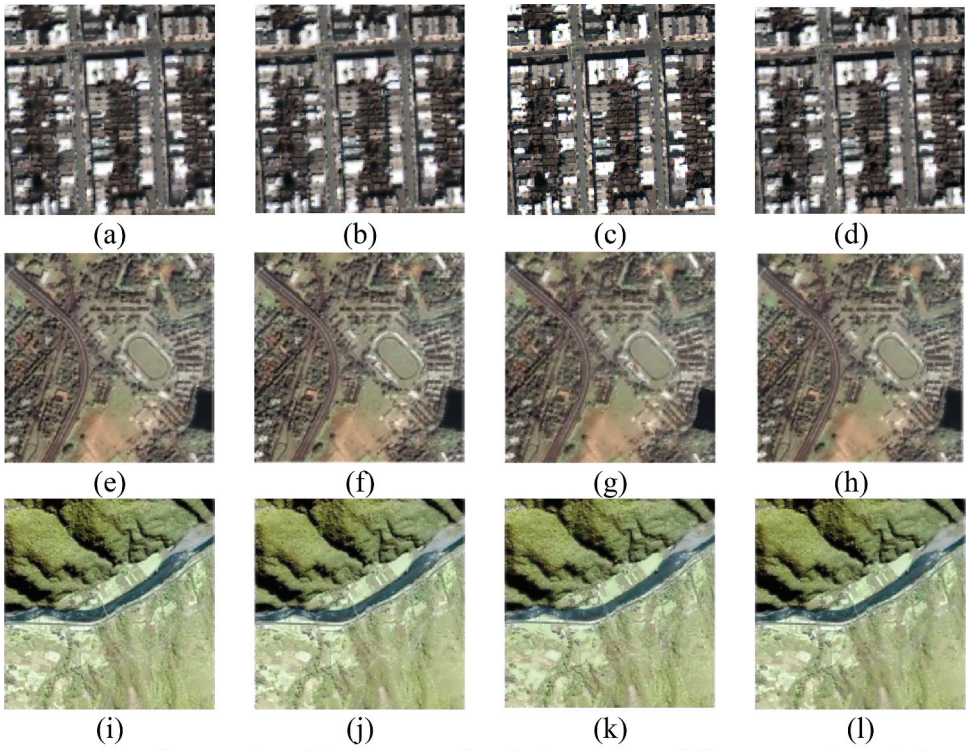

Figure 10. Fusion results of the proposed technique using different numbers of clusters (5, 7, 10 and 12) for segmentation using group1 images from different data sets: (a-d) WorldView-2 data set, (e-h) QuickBird data set and (i-l) IKONOS data set.

Table 8. Performance metric showing the effect of the number of clusters in segmentation using group1 images.

Data set	Clusters	SD	AG	CC	PSNR	SSIM	RMSE	ERGAS	Q
World View-2	5	57.9816	11.6794	0.8090	16.8468	0.9992	36.6608	8.6120	0.8073
	7	58.8887	12.0929	0.8508	17.7705	0.9994	32.9620	7.6858	0.8486
	10	62.6920	13.3344	0.9121	19.6715	0.9996	26.4831	6.2487	0.9084
	12	60.3724	11.6502	0.8706	18.2683	0.9995	31.1262	7.2887	0.8682
QuickBird	5	36.9958	8.9000	0.7283	17.9889	0.9994	32.1442	7.1120	0.7174
	7	37.8322	9.0893	0.7493	18.1884	0.9994	31.4139	6.9330	0.7397
	10	38.7961	9.0120	0.7853	18.8196	0.9995	29.2119	6.5191	0.7777
	12	39.2884	8.7354	0.5780	16.0121	0.9991	40.3596	8.7451	0.5708
IKONOS	5	51.5492	8.6046	0.8293	18.7092	0.9995	29.5874	5.4315	0.8275
	7	51.3770	8.2494	0.8402	18.9669	0.9995	28.7226	5.1905	0.8383
	10	52.6227	15.3086	0.8580	19.7602	0.9996	26.2137	4.8451	0.8561
	12	51.2024	8.0699	0.8360	18.9170	0.9995	28.8882	5.3774	0.8344

base layers for the 'Y' component of the MS image and the PAN image from WorldView-2 data set are shown in Figure 9(c) and (d). Similarly, the segmented results of the other two data sets are also shown. It is interesting to note that the base layers preserve the local spatial structure. The high-quality spectral contents of the fused image are provided by

Table 9. Performance metric showing comparison with the deep learning-based method.

Image	Method	SD	AG	CC	PSNR	SSIM	RMSE	ERGAS	Q
Stockholm-WV-2	TACNN	63.0545	7.0713	0.9077	19.4147	0.9996	27.2872	2.7930	0.9290
	Proposed	66.9697	7.8602	0.9159	19.5122	0.9996	26.9822	4.1893	0.9148
Adelaide-WV-3	TACNN	54.2805	5.6283	0.8455	18.5362	0.9996	30.1867	3.2560	0.9380
	Proposed	59.9008	7.6574	0.9430	20.9781	0.9998	22.7847	3.3231	0.9372

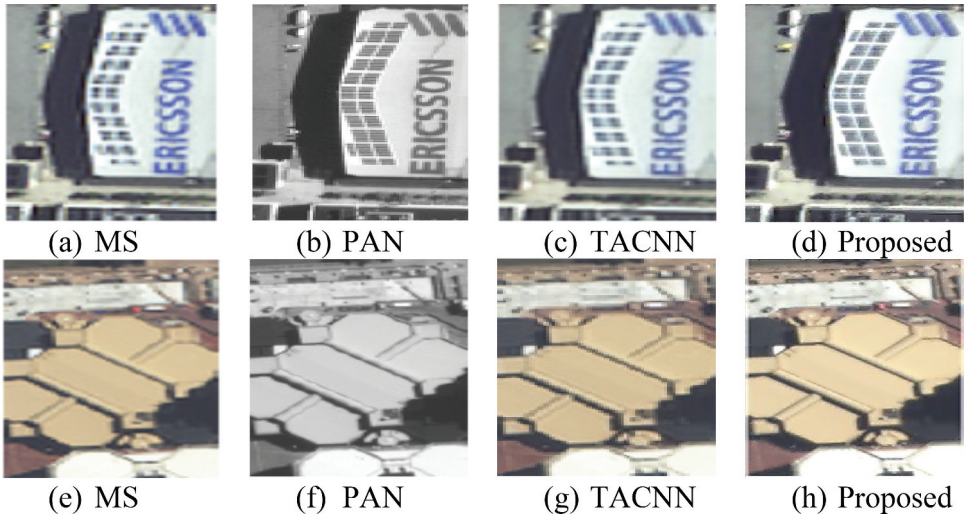

Figure 11. Fusion results: (a-d) Stockholm-WV-2 image and (e-h) Adelaide-WV-3 image.

Table 10. Performance comparison for considering four directions and eight directions.

Data set	Images	Directions	SD	AG	CC	PSNR	SSIM	RMSE	ERGAS	Q
Worldview-2	Group1	4	61.8784	12.2795	0.9005	19.1443	0.9995	28.1403	6.5461	0.8968
		8	62.6920	13.3344	0.9121	19.6715	0.9996	26.4831	6.2487	0.9084
	Group2	4	71.4230	9.6384	0.8867	17.7541	0.9993	33.0248	9.2324	0.8800
		8	68.4204	8.9104	0.9428	20.9363	0.9997	26.4966	6.4620	0.9396
	Group3	4	49.1814	11.8681	0.9017	20.2868	0.9996	24.6776	9.1698	0.8762
		8	49.1815	11.8883	0.9018	20.2875	0.9996	24.6774	9.1697	0.8763
Quickbird	Group1	4	43.4525	10.7681	0.8345	15.7644	0.9990	41.5259	7.2965	0.7886
		8	38.7961	9.0120	0.7853	18.8196	0.9995	29.2119	6.5191	0.7777
	Group2	4	37.6012	8.8452	0.7062	17.2251	0.9993	35.0986	7.6046	0.6916
		8	46.6570	14.8720	0.9225	22.8509	0.9991	18.3653	3.8198	0.9225
IKONOS	Group1	4	53.0719	9.2455	0.8631	19.2897	0.9996	27.6748	4.9426	0.8598
		8	52.6227	15.3086	0.8580	19.7602	0.9996	26.2137	4.8451	0.8561
	Group2	4	30.2548	4.8433	0.7225	14.4326	0.9987	48.4074	10.2733	0.6474
		8	31.8205	4.9963	0.7967	15.0607	0.9989	45.0304	9.6851	0.7272

these structures. Furthermore, it can be seen from the figure that the coarse details are provided by the base layers eliminating the texture. On the other hand, the detail layers retain the texture information and the finer details.

The fusion results of the proposed method using group1 images from WorldView-2, QuickBird and IKONOS using different clusters are shown in Figure 10. It is observed that the fused images obtained by segmenting the images into 10 clusters look visibly more

clear and retain more spectral information as compared to the fused images obtained by segmenting the images into 5,7 and 12 clusters. This shows that the accuracy in segmentation influences the fusion results.

It is seen from Table 8 that almost all performance metrics obtained with 10 clusters provide best values. Based on these observations, we have selected ten clusters for segmentation of the source images.

A comparison of the proposed method with a deep learning approach is presented in Table 9. The visual results are shown in Figure 11. It is observed from the figure that the proposed method gave better result with the Stockholm-WV-2 image. However, the output with the Adelaide-WV-3 image is identical.

It is to note that the results presented above are based on considering eight directions for computing the conduction coefficients in equation (3) and (4). A comparison with results obtained by considering four directions for all the test images is shown in Table 10.

It is seen from Table 10 that most of the metric values obtained using eight directions are better than those obtained using four directions. Furthermore, the diagonal edge information is also preserved using eight directions. As a result, the fused image retains more spatial information and spectral content.

5. Conclusion

Application of the region-based technique for the fusion of remote sensing images is limited. In this sense, our suggested method shows a new research direction to the researchers working in this area. Unlike earlier studies, which are either based on YCbCr colour space or IHS colour space or directly on RGB colour space, this study is based on a combination of YCbCr colour space and AD process. The aim is to preserve the spectral contents together with the spatial resolution. The AD filter could also retain the edge information as it is a class of edge-preserving filters. The chromaticity loss is less because the fusion is carried out only on the 'Y' component of the MS image along with the PAN image. Moreover, the Cr and Cb components remain unchanged that preserve the colour information, which is desirable. A limitation of the proposed technique is the selection of an appropriate segmentation technique to find the regions. The AD filter technique has good smoothing performance; however, the traditional diffusion filtering technique blurs the edges and details. To improve the fusion performance by reducing the hazy edges, some diffusion weighting coefficients can be used. It is believed that the suggested method may provide a pleasing MS image with high spatial and spectral resolution. The idea may be extended for fusion of biomedical images for smart healthcare services.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The author(s) reported that there is no funding associated with the work featured in this article.

References

Azarang, A. and Kehtarnavaz, N., 2020. Image fusion in remote sensing by multi-objective deep learning. *International Journal of Remote Sensing*, 41 (24), 9507–9524. doi:10.1080/01431161.2020.1800126

Babaud, J., *et al*., 1986. Uniqueness of the gaussian kernel for scale-space filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (1), 26–33. doi:10.1109/TPAMI.1986.4767749

Bai, X., *et al*., 2014. Object classification via feature fusion based marginalized kernels. *IEEE Geoscience and Remote Sensing Letters*, 12 (1), 8–12.

Bavirisetti, D.P. and Dhuli, R., 2016. Fusion of infrared and visible sensor images based on anisotropic diffusion and karhunen-loeve transform. *IEEE Sensors Journal*, 16 (1), 203–209. doi:10.1109/JSEN.2015.2478655

Bhateja, V., Singhal, A., and Singh, A., *et al*., 2019. Multi-exposure image fusion method using anisotropic diffusion. *Advances in Intelligent Systems and Computing*, 797, 893–900.

Cheng, J., *et al*., 2015. Remote sensing image fusion via wavelet transform and sparse representation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 104, 158–173. doi:10.1016/j.isprsjprs.2015.02.015

Dong, L., *et al*., 2015. High quality multi-spectral and panchromatic image fusion technologies based on curvelet transform. *Neurocomputing*, 159, 268–274. doi:10.1016/j.neucom.2015.01.050

Earth Resource Mapping Pty Ltd, 1990. The brovey transform explained. *EMU Forum*, 2 (11). Available at: http://www.ermapper.com/forum_new/emuf2-11.htm#aiticle_5

Ghahremani, M., *et al*., 2019. Remote sensing image fusion via compressive sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 152, 34–48. doi:10.1016/j.isprsjprs.2019.04.001

Ghahremani, M. and Ghassemian, H., 2016. Nonlinear IHS: a promising method for pan-sharpening. *IEEE Geoscience and Remote Sensing Letters*, 13 (11), 1606–1610. doi:10.1109/LGRS.2016.2597271

Ghassemian, H., 2016. A review of remote sensing image fusion methods. *Information Fusion*, 32, 75–89. doi:10.1016/j.inffus.2016.03.003

Golipour, M., Ghassemian, H., and Mirzapour, F., *et al*., 2015. Integrating hierarchical segmentation maps with MRF prior for classification of hyperspectral images in a bayesian framework. *IEEE Transactions on Geoscience and Remote Sensing*, 54 (2), 805–816. doi:10.1109/TGRS.2015.2466657

Li, H., *et al*., 2018. An image fusion method based on image segmentation for high-resolution remotely-sensed imagery. *Remote Sens*, 10 (5), 790. doi:10.3390/rs10050790

Li, S., Kwok, J.T., and Wang, Y., *et al*., 2001. Combination of images with diverse focuses using the spatial frequency. *Information Fusion*, 2 (3), 169–176. doi:10.1016/S1566-2535(01)00038-0

Li, S., Yin, H., and Fang, L., *et al*., 2013. Remote sensing image fusion via sparse representations over learned dictionaries. *IEEE Transactions on Geoscience and Remote Sensing*, 51 (9), 4779–4789. doi:10.1109/TGRS.2012.2230332

Liu, X., Liu, Q., and Wang, Y., *et al*., 2020. Remote sensing image fusion based on two-stream fusion network. *Information Fusion*, 55, 1–15. doi:10.1016/j.inffus.2019.07.010

Liu, Y., Liu, S., and Wang, Z., *et al*., 2015. A general framework for image fusion based on multi-scale transform and sparse representation. *Information Fusion*, 24, 147–164. doi:10.1016/j.inffus.2014.09.004

Luo, B., *et al*., 2013. Decision-based fusion for pansharpening of remote sensing images. *IEEE Geoscience and Remote Sensing Letters*, 10 (1), 19–23. doi:10.1109/LGRS.2012.2189933

Ma, X., *et al*., 2019. Remote sensing image fusion based on sparse representation and guided filtering. *Electronics*, 8 (3), 303. doi:10.3390/electronics8030303

Mahmoudi, F.T., Samadzadegan, F., and Reinartz, P., *et al*., 2015. Object recognition based on the context aware decision-level fusion in multiviews imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8 (1), 12–22. doi:10.1109/JSTARS.2014.2362103

Masi, G., *et al*., 2016. Pansharpening by convolutional neural networks. *Remote Sensing*, 8 (7), 594. doi:10.3390/rs8070594

Meher, B., *et al*., 2019. A survey on region based image fusion methods. *Information Fusion*, 48, 119–132. doi:10.1016/j.inffus.2018.07.010

Mirzapour, F. and Ghassemian, H., 2015. Improving hyperspectral image classification by combining spectral, texture, and shape features. *International Journal of Remote Sensing*, 36 (4), 1070–1096. doi:10.1080/01431161.2015.1007251

Naidu, V.P.S., 2010. Discrete cosine transform-based image fusion. *Defence Science Journal*, 60 (1), 48–54. doi:10.14429/dsj.60.105

Pajares, G. and de La Cruz, J.M., 2004. A wavelet-based image fusion tutorial. *Pattern Recognition*, 37 (9), 1855–1872. doi:10.1016/j.patcog.2004.03.010

Pandit, V.R. and Bhiwani, R.J., 2019. Using image segmentation for fusion of multispectral to panchromatic imagery. In *2019 Fifth International Conference on Image Information Processing (ICIIP)*, IEEE, 23–28, Shimla, India.

Perona, P. and Malik, J., 1990. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12 (7), 629–639. doi:10.1109/34.56205

Pohl, C. and van Genderen, J., 2015. Structuring contemporary remote sensing image fusion. *International Journal of Image and Data Fusion*, 6 (1), 3–21. doi:10.1080/19479832.2014.998727

Restaino, R., *et al*., 2016. Fusion of multispectral and panchromatic images based on morphological operators. *IEEE Transactions on Image Processing*, 25 (6), 2882–2895. doi:10.1109/TIP.2016.2556944

Scarpa, G., Vitale, S., and Cozzolino, D., *et al*., 2018. Target-adaptive CNN-based pansharpening. *IEEE Transactions on Geoscience and Remote Sensing*, 56 (9), 5443–5457. doi:10.1109/TGRS.2018.2817393

Shahdoosti, H.R. and Ghassemian, H., 2015. Fusion of MS and PAN images preserving spectral quality. *IEEE Geoscience and Remote Sensing Letters*, 12 (3), 611–615. doi:10.1109/LGRS.2014.2353135

Tambe, R.G., Talbar, S.N., and Chavan, S.S., *et al*., 2021. Fusion of multispectral and panchromatic images by integrating standard PCA with rotated wavelet transform. *Journal of the Indian Society of Remote Sensing*, 49, 2033-2055.

Tan, W., *et al*., 2020. Remote sensing image fusion via boundary measured dual-channel pcnn in multi-scale morphological gradient domain. *IEEE Access*, 8, 42540–42549. doi:10.1109/ACCESS.2020.2977299

Tu, T.M., *et al*., 2001. A new look at IHS-like image fusion methods. *Information Fusion*, 2 (3), 177–186. doi:10.1016/S1566-2535(01)00036-7

Tuia, D., Volpi, M., and Moser, G., *et al*., 2018. Decision fusion with multiple spatial supports by conditional random fields. *IEEE Transactions on Geoscience and Remote Sensing*, 56 (6), 3277–3289. doi:10.1109/TGRS.2018.2797316

Umbaugh, S.E., 2005. *Computer imaging: digital image analysis and processing*. Boca Raton, FL: CRC press.

Wald, L., 2002. *Data Fusion: definitions and Architectures: fusion of Images of Different Spatial Resolutions*. Paris: Presses des MINES.

Wald, L., Ranchin, T., and Mangolini, M., *et al*., 1997. Fusion of satellite images of different spatial resolutions: assessing the quality of resulting images. *Photogrammetric Engineering and Remote Sensing*, 63 (6), 691–699.

Wang, J., *et al*., 2021a. A dual-path fusion network for pan-sharpening. *IEEE Transactions on Geoscience and Remote Sensing*, 1-14.

Wang, W. and Chang, F., 2011. A multi-focus image fusion method based on Laplacian pyramid. *Journal of Computers*, 6 (12), 2559–2566. doi:10.4304/jcp.6.12.2559-2566

Wang, X., *et al*., 2021b. The PAN and MS image fusion algorithm based on adaptive guided filtering and gradient information regulation. *Information Sciences*, 545, 381–402. doi:10.1016/j.ins.2020.09.006

Wang, Z., *et al*., 2004. Image quality assessment_from error visibility to structural similarity. *IEEE Trans on Image Processing*, 13 (4), 1–14. doi:10.1109/TIP.2003.819861

Wang, Z. and Bovik, A.C., 2002. A universal image quality index. *IEEE Signal Processing Letters*, 9 (3), 81–84. doi:10.1109/97.995823

Xin, H. and Feng, L., 2019. Remote sensing image fusion algorithm based on à trous wavelet transform and HIS transform. *In*: *Proceedings of Eleventh International Conference on Digital Image Processing (ICDIP)*, International Society for Optics and Photonics, 11179, 1117908, Guangzhou, China.

Xu, Q., Zhang, Y., and Li, B., *et al*., 2014. Recent advances in pansharpening and key problems in applications. *International Journal of Image and Data Fusion*, 5 (3), 175–195. doi:10.1080/19479832.2014.889227

Yakhdani, M.F. and Azizi, A., 2010. Quality assessment of image fusion techniques for multisensor high resolution satellite images (Case study: IRS-P5 AND IRS-P6 satellite images). *In*: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*. In W. W., & B. Székely (Eds.), Isprs TC VII symposium-100 years ISPRS (pp. Part7B). Vienna, Austria: IAPRS.

Yang, B. and Li, S., 2010. Multifocus image fusion and restoration with sparse representation. *IEEE Transactions on Instrumentation and Measurement*, 59 (4), 884–892. doi:10.1109/TIM.2009.2026612

Yin, S. and Zhang, Y., 2019. Singular value decomposition-based anisotropic diffusion for fusion of infrared and visible images. *International Journal of Image and Data Fusion*, 10 (2), 146–163. doi:10.1080/19479832.2018.1487886

Zhang, J., 2010. Multi-source remote sensing data fusion: status and trends. *International Journal of Image and Data Fusion*, 1 (1), 5–24. doi:10.1080/19479830903561035

Zoran, L.F., 2009. Quality evaluation of multiresolution remote sensing images fusion. *UPB Scientific Bulletin, Series C: Electrical Engineering*, 71, 38–52.