CHAPTER 2

ROLE OF DISCRETE WAVELET TRANSFORM AND NEURAL NETWORK IN IMAGE FUSION

2.1 BRIEF OUTLINE

Based on the previous research work, image fusion can be broadly categorized into three stages - simple image fusion, pyramid decomposition based fusion and discrete wavelet transform based fusion [41, 69, 80, 99, 101]. Simple fusion algorithms are averaging, maximum/minimum and principal component analysis (PCA) methods. These primitive fusion schemes perform the fusion right on the source images. Few drawbacks of simple image fusion techniques are they reduce the contrast of the fused image and produce spatial distortions in the fused image. But these methods do prove good for certain particular cases wherein the input images have an overall high brightness and high contrast.

Generally, pyramid transform consists of three major phases: decomposition, formation of the initial image for recomposition, and recomposition. The basic idea is to construct the pyramid transform of the fused image from the pyramid transforms of the source images and then the fused image is obtained by taking inverse pyramid transform. The pyramid transform fails to introduce spatial orientation selectively in the decomposition process which leads to the blocking effects in the fusion results [47].
After an in-depth study, the present thesis found that, the methods based on wavelet transforms have compactness, directional selectivity and orthogonality. Due to these, discrete wavelet transform based fusion methods are used in literature when compared to pyramid decomposition based fusion methods.

A wavelet is a wave-like oscillation with amplitude that starts out at zero, increases, and then decreases back to zero. A wavelet is a waveform that is limited in duration and has an average value of zero. Generally, wavelets are irregular and asymmetric. Wavelets are purposefully designed to have specific properties that make them useful for signal processing. A wavelet is a mathematical function which divides a given function into different scale components. The advantage of wavelet transforms over traditional Fourier transforms is in representing functions that have discontinuities and sharp peaks, and for accurately deconstructing and reconstructing finite, non-periodic and/or non-stationary signals. Wavelets can be used to extract information from audio signals and images. Sets of wavelets are generally needed to analyze data fully. A set of "complementary" wavelets will deconstruct data without gaps or overlap so that the deconstruction process is mathematically reversible.

The pyramid based transforms and wavelet based transforms are multi-resolution techniques. The function can be decomposed into different scale levels and each level can further be decomposed into different scale levels. The ability to decompose a function into different scale levels is called multi-resolution analysis (MRA). It is a
mathematical tool that is often used in wavelet-based image compression schemes. It requires two functions, that is a scaling function $\phi(t)$ and a mother wavelet $\psi(t)$.

Discrete Wavelet Transforms (DWT) can provide better spatial and spectral localization of image information as compared to other multi-resolution representations \[61\]. Generally, wavelet-based schemes perform better than standard schemes, especially in terms of minimizing color distortions. Hence, DWT is widely used for remote sensing, medical and multi-focus image fusion \[14, 94, 100\]. Few of the Discrete Wavelet Transforms (DWT) is Haar wavelets, Daubechies wavelets, Mathieu wavelets, Legendre wavelets, Symlet wavelets, Coiflet wavelets, etc.

Wavelets are a mathematical tool for hierarchical decomposing functions. After performing many successful applications in the area of signal processing, DWT has been accepted as a powerful image processing technique for image fusion. DWT provides efficient localization in both space and frequency domains. The DWT is a spatial frequency decomposition that provides a flexible multi-resolution analysis of an image. Wavelets give a better signal representation using multi-resolution analysis with balanced resolution at any time and frequency.

One of the objectives of image fusion is to achieve high-quality digital camera images from several degraded images \[97\]. According to \[66\], the goal of fusion is to achieve high spatial resolution together with a high-quality spectral content from two kinds of remote sensing
images - images with high quality in spectral content but low quality in spatial resolution and images with high spatial resolution but with a unique spectral band.

2.2 ADVANTAGES OF DWT FOR IMAGE FUSION

The wavelets-based approach is appropriate for performing fusion tasks for the following reasons:

- DWT is a multi-resolution approach. Therefore, DWT can manage different image resolutions. During the recent years, few researchers [2, 9, 50, 54, 68] have studied multi-resolution representation of a signal and concluded that multi-resolution information can be useful in a number of image processing applications including image fusion.

- DWT allows image decomposition in different kinds of coefficients preserving the image information. These coefficients from multiple images can be combined in multiple ways to obtain new coefficients, so that the information in the original images is collected appropriately.

- The inverse wavelet transform is applied to get back the final fused image, where the essential relevant information from the source images is preserved in the fused image.

Most of the constructions of DWT make use of the multi-resolution analysis, which defines it by a scaling function. An auxiliary or scaling function avoids the numerical complexity of evaluating an integral of each wavelet coefficient in DWT. Hence, wavelets are defined
by the wavelet function $\psi \ (t)$ (also called mother wavelet) and scaling function $\phi \ (t)$ (also called father wavelet) in the time domain. If a discrete signal is represented by $f \ (t)$, its wavelet decomposition is defined by using the Equation (2.1).

$$f(t) = \sum_{m,n} c_{m,n} \psi_{m,n}(t) \quad (2.1)$$

where $\psi_{m,n}(t)$ is the dilated and/or translated version of the mother wavelet which is defined by using the Equation (2.2), where $m$ and $n$ are integers.

$$\psi_{m,n}(t) = 2^{-m/2} \psi[2^{-m}t - n] \quad (2.2)$$

For an iterated wavelet transform, additional coefficients $a_{m,n}$ are required at each scale. At each scale $a_{m,n}$ and $a_{m-1,n}$ describe the approximations of the function $f$ at resolution $2^m$ and at the coarser resolution $2^{m-1}$ respectively, while the coefficients $c_{m,n}$ describe the difference between one approximation and the other. In order to obtain the coefficients $c_{m,n}$ and $a_{m,n}$ at each scale and position, a scaling function is needed which is defined similarly to Equation (2.2). The convolution of scaling function with the signal is implemented at each scale through iterative filtering of the signal with a low pass FIR filter ‘$h_n$’. The approximation coefficients $a_{m,n}$ at each scale can be obtained by using the recursive relation defined by using Equation (2.3).

$$a_{m,n} = \sum_k h_{2n-k} \ a_{m-1,k} \quad (2.3)$$

where top level $a_{m,n}$ is the sampled signal itself. In addition, by using a related high pass FIR filter ‘$g_n$’, the wavelet coefficients can be obtained by using Equation (2.4).
\[ c_{m,n} = \sum_k g_{2n-k} a_{m-1,k} \]

To reconstruct original signal, analysis filters can be selected from a biorthogonal set which have a related set of synthesis filters. These synthesis filters \( h \) and \( g \) can be used to perfectly reconstruct the signal using the reconstruction formula given in Equation (2.5).

\[ a_{m-1,l}(f) = \sum_n [h_{2n-l} a_{m,n}(f) + g_{2n-l} c_{m,n}(f)] \]

Equations (2.3) and (2.4) are implemented by filtering and subsequent downsampling. Conversely, Equation (2.5) is implemented by an initial upsampling and a subsequent filtering.

Two dimensional wavelet transform decomposes an image into four frequency bands: low–low (LL), low–high (LH), high–low (HL) and high–high (HH). Among these four, it decomposes the image into LH, HL and HH spatial frequency bands at different scales and the LL band at the coarsest scale. Figure 2.1 shows the normal layout of such decompositions. LL band contains average image information whereas other bands contain directional information due to spatial orientation. Higher absolute values of wavelet coefficients in the high bands correspond to salient features such as edges, lines, etc. Image fusion can be performed using selection-based rule of wavelet transform [47]. The selection criteria can be performed by using maximum absolute value i.e., the wavelet coefficient which has maximum value is selected, since it represents the presence of a dominance feature. Another selection criteria depends upon the consistency verification stage in which the selection of a hyper pixel in image A is changed to image B if a majority of the surrounding pixels are from image B. This scheme is
better than Laplacian Pyramid based fusion due to the compactness, directional selectivity and orthogonality of the Wavelet Transform. In [86], perceptual based weighting using wavelet transform is suggested. The wavelet coefficients from each image are combined using a weighted average method. These fused images are visually better than the fusion techniques based on the gradient pyramid.

2.3 DWT BASED IMAGE FUSION TECHNIQUE

For fusing low-resolution Multispectral (MS) image with a high-resolution Panchromatic (PAN) image using wavelet fusion technique, the PAN image is first decomposed into a set of low-resolution PAN images with corresponding wavelet coefficients (spatial details) for each level. An individual band of MS image replaces the low-resolution PAN at the resolution level of the original MS image. The high resolution spatial detail is injected into each MS band by performing inverse wavelet transform on each MS band together with the corresponding wavelet coefficients. Generally, in wavelet-based fusion schemes, detailed information is extracted from the PAN image using wavelet transforms and injected into MS image. This process is represented in Figure 2.2.

![Figure 2.1: DWT at second level decomposition.](image)
2.4 ARTIFICIAL NEURAL NETWORKS IN IMAGE FUSION

After an in-depth survey, the present thesis identifies that the principle of soft computing is to exploit the tolerance for vagueness, uncertainty, partial truth, robustness and low solution cost. According to Zadeh (1996 & 2005), Soft computing is considered in various application fields. Neural network (NN), genetic algorithm (GA), analytical hierarchy process (AHP) and quality function deployment (QFD) are the basic methods of soft computing and decision making tools. Neural network is a decision making tool that helps to interact with design and decision making systems under uncertainty conditions. Artificial neural network (ANN) or commonly just neural network is an interconnected group of artificial neurons that uses a computational model for information processing.
Hence, NN is one of the powerful tools which helps in solving the problems such as pattern classification, function approximation, etc. The fusion process considered in the present study can be considered as a classification problem. Classification is the important task of remote sensing applications because accuracy will improve when multiple source image data are introduced to the processing [62]. NN is one of the possible approaches to handle high dimension nature of hyper-spectral satellite sensor data [39]. In the present study, NN is trained using the block features of different pair of multimodal images. Once the classifier is obtained, it can be used for the fusion purpose. In the present study, feed forward neural network is explored. In this network, the information moves in only one direction i.e., forward direction and never goes backwards. The information passes from the input nodes through the hidden nodes and to the output nodes. There are no cycles or loops in the network.

The simplest kind of NN is a single-layer perceptron network which consists of a single layer of output nodes. The inputs are fed directly to the outputs via a series of weights. Multi-layer perceptron network consists of multiple layers of computational units interconnected in a feed-forward way. Each neuron in one layer has directed connections to the neurons of the subsequent layer. One of the famous learning techniques is back-propagation.

Back-propagation is a form of supervised training. The NN is presented with training data and the results are compared with the expected results. The difference between actual results and the
expected results causes error. The error is then fed back through the network. Back-propagation is a method whereby the weights and input threshold of the NN are altered in a way that causes this error to be reduced. Back-propagation is used to train a feed forward neural network.

NN consists of three layers – input layer, hidden layer and output layer. The input layer has several neurons which represent the feature factors extracted and normalized from the source images. The hidden layer has several neurons and the output layer can have one or more neurons. The schematic diagram of the NN-based image fusion method is depicted in Figure 2.3.

![Figure 2.3: NN-based Image fusion method.](image)

Generally, the $i^{th}$ neuron of the input layer connects with the $j^{th}$ neuron of the hidden layer by weight $W_{ij}$ and weight between the $j^{th}$ neuron of the hidden layer and the $t^{th}$ neuron of output layer is $V_{jt}$. The weighting function is used to simulate and recognize response
relationship between features of the fused image and the corresponding features from the source images.

**2.5 INTEGRATING DWT WITH NN**

To enhance the fusion results, DWT is integrated with NN, which is one of the feature extraction or detection machine learning applications. The first step of NN-based image fusion is to decompose the source images into several blocks with some defined size. Then, the features from these images of the corresponding blocks are extracted, and the normalized feature vector incident to neural network is constructed. The features under study can be spatial frequency, visibility, edge, etc. Next, select some vector samples to train the NN. After training, the NN model can remember a functional relationship and be used for further calculations. For these reasons, the NN concepts are adopted to develop strongly nonlinear models for multiple sensor data fusion.

Neural networks are successfully used in image fusion [7, 23, 48, 49, 76, 98]. Rong *et al.* presented a feature-level image fusion method based on segmentation region and neural networks. The results proved that, integrated fusion scheme is more efficient than that of traditional methods [85]. The present thesis concludes that the NN-based image fusion methods have more advantages than traditional statistical methods, especially when input multi-sensor data is incomplete or with much noise.
2.6 BLOCK BASED FEATURES

In feature-level image fusion, the selection of different features is an important task. The five different features are used to characterize the information level contained in a specific portion of the image are contrast visibility, spatial frequency, variance, Energy of Gradient (EOG), and edge information.

1. **Contrast Visibility:** It calculates the deviation of a block of pixels from the block’s mean value. It relates to the clearness level of the block. The visibility of the image block is calculated by using Equation (2.6).

   \[
   VI = \frac{1}{p \times q} \sum_{m,n \in B_k} \frac{|I(m,n) - \mu_k|}{\mu_k}
   \]  

   (2.6)

   where \(\mu_k\) and \(p \times q\) are the mean and size of the block \(B_k\) respectively.

2. **Spatial Frequency:** It measures the activity level in an image. It is used to calculate the frequency changes along the rows and columns of the image. It is calculated by using the Equations (2.7), (2.8) and (2.9).

   \[
   SF = \sqrt{(RF)^2 + (CF)^2}
   \]  

   (2.7)

   where

   \[
   RF = \frac{1}{p \times q} \sum_{m=1}^{p} \sum_{n=2}^{q} |I(m,n) - I(m,n-1)|^2
   \]  

   (2.8)

   \[
   CF = \frac{1}{p \times q} \sum_{n=1}^{q} \sum_{m=2}^{p} |I(m,n) - I(m-1,n)|^2
   \]  

   (2.9)

   where \(I\) is the image and \(p \times q\) is the image size. A large value of spatial frequency describes the large information level in the image and therefore it measures the clearness of the image.
3. **Variance**: It is used to measure the extent of focus in an image block. It is calculated by using Equation (2.10).

\[
Variance = \frac{1}{pq} \sum_{m=1}^{p} \sum_{n=1}^{q} (I(m, n) - \mu)^2
\]  

(2.10)

where \( \mu \) is the mean value of the block image and \( pq \) is the image size. A high value of variance shows the greater extent of focus in the image block.

4. **Energy of Gradient (EOG)**: It is used to measure the amount of focus in an image. It is calculated by using the Equations (2.11), (2.12) and (2.13).

\[
EOG = \sum_{m=1}^{p-1} \sum_{n=1}^{q-1} (a_m^2 + a_n^2)
\]  

(2.11)

where

\[
a_m = a(m + 1, n) - a(m, n)
\]  

(2.12)

and

\[
a_n = a(m, n + 1) - a(m, n)
\]  

(2.13)

where \( p \) and \( q \) represent the dimensions of the image block. A high value of energy of gradient shows greater amount of focus in the image block.

5. **Edge Information**: The Canny edge detector is used to identify the edge pixels in an image block. If the current pixel belongs to some edge in the image then it will return 1 otherwise 0. The edge feature is used to calculate the number of edge pixels contained within the image block.

### 2.7 QUALITY METRICS

Quality assessment is enviable to evaluate the possible benefits of fusion, to determine an optimal setting of parameters for a certain fusion scheme and to compare results obtained with different
algorithms. Goal of image quality assessment is to supply quality metrics that can predict perceived image quality automatically. While visual inspection has limitation due to human judgment, qualitative approach based on the evaluation of “distortion” in the resulting fused image is more desirable in mathematical modeling. The present thesis considered the following metric parameters.

1. **Standard Deviation (SD):** SD is used to measure the level of contrast in the fused image. It is calculated by using Equation (2.14).

\[
SD = \sqrt{\frac{\sum_{i=0}^{L-1} (i - \bar{i})^2 h_F(i)}{L}}
\]  

(2.14)

where \( \bar{i} = \frac{\sum_{i=0}^{L-1} ih_F}{L} \) and \( h_F \) is the normalized histogram of fused image and \( L \) is the number of gray levels. A well contrast image will have high standard deviation.

2. **Entropy:** Entropy is used to quantify the information contained in the fused image [70]. It also shows the probability distribution of pixels with the distinct grey values. A higher value shows good fusion results. It is calculated by using Equation (2.15).

\[
H = -\sum_{i=0}^{L-1} h_F(i) \log_2 h_F(i)
\]  

(2.15)

where \( h_F \) is the normalized histogram of the fused image and \( L \) is the number of gray levels.

3. **Correlation Coefficient (CC):** Pearson’s CC is denoted by ‘r’, and is used for comparing the source image and the fused image. If \( r = 1 \) then the two images are absolutely identical, if \( r = 0 \) then they
are completely uncorrelated and if \( r = -1 \) then they are completely anti-correlated. It is calculated by using Equation (2.16).

\[
    r = \frac{\sum_{i}(x_i - x_m)(y_i - y_m)}{\sqrt{\sum_{i}(x_i - x_m)^2} \sqrt{\sum_{i}(y_i - y_m)^2}}
\]

(2.16)

where \( x_i \) is the intensity of the \( i^{th} \) pixel in the source image, \( y_i \) is the intensity of the \( i^{th} \) pixel in the fused image, \( x_m \) is the mean intensity of the source image and \( y_m \) is the mean intensity of the fused image.

4. **Mean Squared Error (MSE):** MSE is used to measure the spectral distortion in the fused image. A smaller value shows good fusion results. It is calculated by using Equation (2.17).

\[
    MSE = \frac{\sum_{i=1}^{M} \sum_{j=1}^{N} (I_R(i,j) - I_F(i,j))^2}{MN}
\]

(2.17)

where \( I_R(i,j) \) denotes pixel \((i,j)\) of the image reference and \( I_F(i,j) \) denotes pixel \((i,j)\) of the fuse image, \( MN \) is the image size.

5. **Peak Signal to Noise Ratio (PSNR):** PSNR is used to reveal the radiometric distortion of the fused image compared to the original image. A higher value shows good fusion results. It is defined by using the Equation (2.18).

\[
    PSNR(dB) = 10 \log_{10} \left( \frac{255 \cdot 255}{MSE} \right)
\]

(2.18)

6. **Root Mean Squared Error (RMSE):** RMSE is used to measure the standard error in the fused image [25]. A smaller value shows good fusion results. It is calculated by using Equation (2.19).

\[
    RMSE = \sqrt{MSE}
\]

(2.19)

7. **Mean Absolute Error (MAE):** MAE is used to measure the average magnitude of errors in a set of forecasts, without considering
their direction. It measures accuracy for continuous variables. A smaller value shows good fusion results. It is calculated by using Equation (2.20).

$$MAE = \frac{1}{M*N} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} (I_R(x,y) - I_F(x,y))$$

(2.20)

where $I_R(x,y)$ denotes pixel $(x,y)$ of the image reference and $I_F(x,y)$ denotes pixel $(x,y)$ of the fused image and $M*N$ is the image size.

8. Mutual Information Measure (MIM): MIM measures the similarity of image intensity distribution of the corresponding image pair. The greatest value shows the better fusion result. The mutual information between the source image A and the fused image F is denoted by $I_{AF}$ and defined by using Equation (2.21).

$$I_{AF} = \sum_{a,f} P_{AF}(a,f) \log \frac{P_{AF}(a,f)}{P_A(a)P_F(f)}$$

(2.21)

where $P_{AF}$ is the jointly normalized histogram of A and F, $P_A$ and $P_F$ are the normalized histogram of A and F, and a and f represent the pixel value of the image A and F respectively. The mutual information between the source image B and the fused image F is denoted by $I_{BF}$ and is calculated similarly as $I_{AF}$. The mutual information between the source images A, B and the fused image F is given by Equation (2.22). Larger MIM value indicates better fusion result.

$$M_{AB}^F = I_{AF} + I_{BF}$$

(2.22)

9. Fusion Factor (FF): FF is used to measure the amount of information present in the fused image. A higher value of FF indicates that fused image contains moderately good amount of
information present in both the images. It is defined by using the Equation (2.23).

\[ FF = I_{AF} + I_{BF} \]  \hspace{1cm} (2.23)

where A and B are the source images and F is the fused image.

10. **The metric** $Q^{AB/F}$: $Q^{AB/F}$ evaluates the amount of edge information transferred from source images into fused image. The greatest value shows the better fusion result. It is defined by using Equation (2.24).

\[
Q^{AB/F} = \frac{\sum_{n=1}^{N} \sum_{m=1}^{M} (Q^{AF}(n,m)w^A(n,m)+Q^{BF}(n,m)w^B(n,m))}{\sum_{n=1}^{N} \sum_{m=1}^{M} (w^A(n,m)+w^B(n,m))} \]  \hspace{1cm} (2.24)

where $Q^{AF}(n,m) = Q^A_g(n,m)Q^A_\omega(n,m)$; $Q^A_g(n,m)$ and $Q^A_\omega(n,m)$ are the edge strength and orientation preservation values respectively; n and m represents the image location; and N, M are the size of images respectively. $Q^{BF}(n,m)$ is similar to $Q^{AF}(n,m)$. $w^A(n,m)$ and $w^B(n,m)$ reflect the importance of $Q^{AF}(n,m)$ and $Q^{BF}(n,m)$ respectively. The dynamic range of $Q^{AB/F}$ is [0, 1] and it should be as close to 1 as possible.

According to the survey, many researchers used the above discussed metrics. Wang used mutual information as a means of objective assessing image fusion performance [34]. The work of Yin et al. [12] focused on one popular mutual information based quality measure and weighted averaging image fusion. Qu et al, [63] proposed MIM metric. Xydes and Petrovic [87] proposed $Q^{AB/F}$ metric.