CHAPTER 3

MULTI-RESOLUTION ANALYSIS BASED FUSION
USING WAVELET TRANSFORM

3.1 INTRODUCTION

Representing the images as non-stationary signals with a time varying spectra allows one to analyse the various frequency components present in it and also identify the position where that component is present. These frequency and position values identified can be used to yield specific information pertaining to the image, thereby provide a resolution of the image in frequency and time (position). However, the time and frequency resolution method is subject to the constraint imposed by the Heisenberg’s uncertainty principle, which says that it is never possible to resolve both the frequency and the time of a signal simultaneously. To overcome this uncertainty problem, various signal representation methods have been proposed; beginning with the windowed Short Term Fourier Transform (STFT) and the Multi-Resolution Analysis (MRA) techniques involving the wavelet transform and the pyramidal decomposition methods.

The multi-scale image fusion algorithms used widely follow two stages. Firstly, they decompose the input image into multiple resolution levels using different decomposition methods. Then they combine the decomposition images of specific levels using fusion rules to obtain the fused output. The pyramid decomposition and wavelets are the commonly used multi-resolution image fusion schemes. The pyramidal decomposition has been implemented using various algorithms such as Toet’s ratio pyramid
(Toet 1989), the contrast pyramid based fusion (Toet et al. 1989), the Generalized Laplacian Pyramid (GLP) method put forward by Kim et al. (1993) and Burt’s Gaussian Pyramid (GP) and the Enhanced Laplacian Pyramid (ELP) (Burt 1984). However, in these methods the computation involved increase rapidly with an increase in the decomposition levels and hence a trade-off is sought between the accuracy and the computation in applications.

The Discrete Wavelet Transform (DWT) has been used as an efficient image decomposition method for fusion for the reasons that follow (Pajares and de La Cruz 2004). The DWT preserves the image information in the wavelet coefficients resulting from the image decomposition. Such coefficients can then be appropriately combined to obtain new coefficients, thereby collecting the information in the original images. Once the images are merged, the final fused image is obtained through the inverse discrete wavelet transform, which also preserve the information in the merged coefficients.

The different wavelet fusion methods involve decomposing the image into the high frequency (detail) and low frequency (approximation) sub-images and designing of rules to combine the corresponding sub-images to give the fused output. Recursive decomposition of the image takes place, with the approximation part of the previous level being decomposed further. Another method used in the literature is the Wavelet Package (WP) based fusion model as proposed by Jishuang and Chao (2001). This approach decomposes both the low frequency part and the high frequency part at each level recursively. The corresponding parts of the different images are then fused at the same level using fusion rules. While the WP method gives an equivalent performance, the computational complexity is high and so the wavelet transform based fusion is given preference.
This chapter discusses the two new DWT-based image fusion schemes put forth by the author. The section that follows gives an introduction to the Short Term Fourier Transform (STFT) and the concept of Multi-Resolution Analysis (MRA), explaining the need for the wavelet transform. Section 3.3 introduces the Continuous Wavelet Transform (CWT) along with a detailed mathematical analysis. Then the Discrete Wavelet Transform (DWT) is presented as the practical and efficient tool for MRA and the filter bank implementation of the same is described. Analysis of images using the Discrete Wavelet Transform is discussed in section 3.4. The proposed threshold based image fusion algorithm and the energy based fusion scheme using DWT are detailed in section 3.5. Section 3.6 presents results obtained using the two new algorithms. This chapter concludes with a summary of the proposed fusion schemes.

3.2 THE SHORT TERM FOURIER TRANSFORM AND MULTI-RESOLUTION ANALYSIS (“The Wavelet Tutorial” by Robi Polikar)

The STFT was the first transform developed to localize a non-stationary signal both in time and frequency. It is derived from the Fourier Transform (FT) and uses a local window to slide through the signal, thereby providing the necessary time localization. Within the region of this window the signal appears to be stationary and hence its Fourier Transform (FT) exists. The choice of the window function is critical as its width determines the segment of the signal where its stationarity is valid. This window function, of width ‘T’ is first located at the very beginning of the signal, say at t=0, such that it overlaps with the first T/2 seconds. The window function and the signal are then multiplied so as to choose only the first T/2 seconds of the signal. This product is considered to be a stationary signal whose FT is to be taken. The resultant spectrum will give the frequency representation of the small portion of the signal. The next step would be to shift this window to a
new location and repeating the process of multiplication and calculation of the FT until the end of the signal is reached. This process of windowing and Fourier transforming can be represented mathematically as:

$$\text{STFT}(t, f) = \int [x(t).w^*(t - t')].e^{-j2\pi ft} dt$$

(3.1)

where $x(t)$ is the 1-D signal under consideration, $w(t)$ is the window function and $*$ is the complex conjugate operator.

However, there is a resolution problem in the STFT representation as only the time intervals in which a band of frequencies that exist can be known because of the limitation imposed by Heisenberg’s uncertainty principle. Because of the finite length windows used to overcome the problem of stationarity, the frequency resolution gets poorer and hence identification of specific frequency components becomes difficult. Hence, with a narrow window, the time resolution gets better but the frequency resolution decreases. Wide windows give good frequency resolution, but poor time resolution; and they also cause violation of the condition of stationarity.

To overcome this resolution problem present in STFT, the technique called Multi-Resolution Analysis (MRA) is used; which analyzes the signal at different frequencies with different resolutions. Every spectral component is not resolved equally as was the case in the STFT. MRA is designed to give good time resolution and poor frequency resolution at high frequencies and good frequency resolution and poor time resolution at low frequencies. This approach makes sense especially when the signal at hand has high frequency components for short durations and low frequency components for long durations (Polikar 1994). The MRA approach using Continuous Wavelet Transform (CWT) and the Discrete Wavelet Transform (DWT) as adapted from Polikar (1994) is discussed in the next section.
3.3 THE WAVELET TRANSFORM

3.3.1 The Continuous Wavelet Transform (CWT)

The wavelet transform method is similar to the STFT approach in that the signal is multiplied by a wavelet function and the transform is computed separately for different time segments of the signal. However, the Fourier Transform of the product is not taken and also the width of the window is changed as the transform is computed for every single spectral component. The mathematical representation of the CWT is given by:

\[
\text{CWT}(\tau, s) = \frac{1}{\sqrt{s}} \int \psi(t) \psi^* \left( \frac{t - \tau}{s} \right) dt
\]

In the above equation, the transformed signal is a function of two variables, \( \tau \) and \( s \), which are the translation and scale parameters, respectively. \( \psi(t) \) is the transforming function, and it is called the mother wavelet as it is the main function from which the other window functions are generated.

The term translation corresponds to time information in the transform domain and refers to the shifting of the window location across the signal. The parameter scale in the wavelet analysis is defined in terms of frequency; low frequencies (high scales) correspond to a global information of a signal (that usually spans the entire signal), whereas high frequencies (low scales) correspond to a detailed information of a hidden pattern in the signal (that usually lasts a relatively short time). And scaling as a mathematical operation, either dilates or compresses a signal. Larger scales correspond to dilated (or stretched out) signals and small scales correspond to compressed signals. In terms of wavelet analysis, if \( \psi(t) \) is a given function, \( \psi(t/s) \) corresponds to a expanded (dilated) version of \( \psi(t) \) if \( s > 1 \) and to an contracted (compressed) version of \( \psi(t) \) if \( s < 1 \).
The CWT can be represented using the equation (3.1) as the inner product of the test signal with the basis functions \( \psi_{\tau,s}(t) \) as,

\[
\text{CWT}(\tau,s) = \Psi(\tau,s) \int_{-\infty}^{\infty} \pi(t) \psi^*_r(t) dt
\]

(3.3)

where,

\[
\psi_{\tau,s} = \frac{1}{\sqrt{s}} \psi \left( \frac{t - \tau}{s} \right)
\]

(3.4)

This definition of the CWT shows that the wavelet analysis is a measure of similarity in frequency content between the basis functions (wavelets) and the signal itself. If the signal has a major component of the frequency corresponding to the current scale, then the wavelet at the current scale will be similar or close to the signal at the particular location where this frequency component occurs. Therefore, the CWT coefficient computed at this point in the time-scale plane will be a relatively large number.

The Continuous Wavelet Transform is reversible if the admissibility condition of equation (3.4) is satisfied.

\[
c_{\psi} = \left\{ 2\pi \int_{-\infty}^{\infty} \frac{\hat{\psi}(\xi)}{\xi} d\xi \right\}^{1/2} < \infty
\]

(3.5)

where \( \hat{\psi}(\xi) \) is the FT of \( \psi(t) \). The above equation implies that \( \hat{\psi}(0) = 0 \), which is

\[
\int \psi(t) dt = 0
\]

(3.6)
The inverse wavelet transform is defined by:

\[
x(t) = \frac{1}{C_\psi^2} \int \int \Psi_x^{\psi}(\tau, s) \frac{1}{s^2} \psi \left( \frac{t - \tau}{s} \right) d\tau ds
\]  

(3.7)

where \( C_\psi \) is a constant that depends on the wavelet used and is called the admissibility constant that satisfies equation (3.4).

### 3.3.2 The Discrete Wavelet Transform (DWT)

Although the discretized continuous wavelet transform enables the computation of the continuous wavelet transform by computers, the wavelet series is simply a sampled version of the CWT, and the information it provides is highly redundant as far as the reconstruction of the signal is concerned. This redundancy, on the other hand, requires a significant amount of computation time and resources. The discrete wavelet transform (DWT), on the other hand, provides sufficient information both for analysis and synthesis of the original signal, with a significant reduction in the computation time.

The main idea is the same as it is in the CWT. A time-scale representation of a digital signal is obtained using digital filtering techniques, employing filters of different cut-off frequencies to analyse the signal at different scales. The signal is passed through a series of high pass filters to analyse the high frequencies, and it is passed through a series of low pass filters to analyse the low frequencies.

The resolution of the signal, which is a measure of the amount of detail information in the signal, is changed by the filtering operations, and the scale is changed by up-sampling and down-sampling (sub-sampling) operations.
3.3.2.1 Signal Decomposition by DWT

The DWT analyses the signal at different frequency bands with different resolutions by decomposing the signal into a coarse approximation and detail information. DWT employs two sets of functions, called scaling functions and wavelet functions, which are associated with low pass and high-pass filters, respectively. The decomposition of the signal into different frequency bands is simply obtained by successive high-pass and low-pass filtering of the time domain signal. The original signal $x[n]$ is first passed through a half-band high-pass filter $g[n]$ and a low-pass filter $h[n]$. After the filtering, half of the samples can be eliminated by subsampling by 2. This constitutes one level of decomposition and can mathematically be expressed as follows:

\[
y_{\text{high}}[k] = \sum_n x[n].g[2k - n]
\] (3.8)

\[
y_{\text{low}}[k] = \sum_n x[n].h[2k - n]
\] (3.9)

where $y_{\text{high}}[k]$ and $y_{\text{low}}[k]$ are the outputs of the high-pass and low-pass filters, respectively, after sub-sampling by 2.

This decomposition halves the time resolution since only half the number of samples now characterizes the entire signal. However, this operation doubles the frequency resolution, since the frequency band of the signal now spans only half the previous frequency band, effectively reducing the uncertainty in the frequency by half. The above procedure, which is also known as the sub-band coding, can be repeated for further decomposition. At every level, the filtering and sub-sampling will result in half the number of samples (and hence half the time resolution) and half the frequency band.
spanned (and hence doubles the frequency resolution). Figure 3.1 illustrates this procedure, where \( x[n] \) is the original signal to be decomposed, and \( h[n] \) and \( g[n] \) are low-pass and high-pass filters, respectively. The bandwidth of the signal at every level is marked on the figure as "f".

\[
X[n] \quad f = 0 - \pi
\]

\[
\begin{align*}
g[n] & \quad f = \pi/2 - \pi \\
\downarrow 2 & \\
\text{Level one DWT} & \\
\end{align*}
\]

\[
\begin{align*}
h[n] & \quad f = 0 - \pi/2 \\
\downarrow 2 & \\
\end{align*}
\]

\[
\begin{align*}
g[n] & \quad f = \pi/4 - \pi/2 \\
\downarrow 2 & \\
\text{Level two DWT} & \\
\end{align*}
\]

\[
\begin{align*}
h[n] & \quad f = 0 - \pi/4 \\
\downarrow 2 & \\
\end{align*}
\]

\[
\begin{align*}
g[n] & \quad f = \pi/8 - \pi/4 \\
\downarrow 2 & \\
\text{Level three DWT coefficients} & \\
\end{align*}
\]

\[
\begin{align*}
h[n] & \quad f = 0 - \pi/8 \\
\downarrow 2 & \\
\end{align*}
\]

\[
\ldots.
\]

**Figure 3.1 The sub-band coding algorithm**
The frequencies that are most prominent in the original signal will appear as high amplitudes in that region of the DWT signal that includes those particular frequencies. Also, the time localization of these frequencies will not be lost; however, the time localization will have a resolution that depends on which level they appear. This procedure in effect offers a good time resolution at high frequencies, and good frequency resolution at low frequencies. The frequency bands that are not very prominent in the original signal will have very low amplitudes, and that part of the DWT signal can be discarded without any major loss of information, allowing data reduction.

One important property of the discrete wavelet transform is the relationship between the impulse responses of the high-pass and low-pass filters. The high-pass and low-pass filters are not independent of each other, and they are related by

\[ g[L-1-n] = (-1)^n h[n] \]  

(3.10)

where \( g[n] \) is the high-pass, \( h[n] \) is the low-pass filter, and \( L \) is the filter length (in number of points). The two filters are odd index alternated reversed versions of each other and the low-pass to high-pass conversion is provided by the \((-1)^n\) term. Filters satisfying this condition are commonly used in signal processing, and they are known as the Quadrature Mirror Filters (QMF).

The reconstruction in this case is simple since half-band filters form orthonormal bases. The above procedure is followed in reverse order for the reconstruction. The signals at every level are up-sampled by two, passed through the synthesis filters \( g'[n] \), and \( h'[n] \) (high-pass and low-pass, respectively), and then added. The interesting point here is that the analysis and synthesis filters are identical to each other, except for a time reversal. Therefore, the reconstruction formula becomes (for each layer)
\[ x[n] = \sum_{k=-\infty}^{\infty} \left[ y_{\text{high}}[k] g[2k - n] + y_{\text{low}}[k] h[2k - n] \right] \quad (3.11) \]

If the filters are not ideal half-band, then perfect reconstruction cannot be achieved. However, under certain conditions, it is possible to find filters that provide perfect reconstruction such as that developed by Ingrid Daubechies, known as Daubechies' wavelets (Daubechies 1990).

### 3.4 DISCRETE WAVELET DECOMPOSITION OF IMAGES

Images represent two dimensional data and the principles of one dimensional DWT and the filter bank approach can be extended to images also. In the wavelet decomposition of an image, the decomposition is done row by row and then column by column. For instance, consider the case of an \( M \times N \) image. Each row is filtered and then down-sampled to obtain two \( M \times (N/2) \) images. Then each column is filtered and the output is sub-sampled to obtain four \( (M/2) \times (N/2) \) images. Of the four sub-images obtained, the one obtained by low-pass filtering the rows and columns is called the LL image. The one obtained by low-pass filtering the rows and high-pass filtering the columns is referred to as the LH image and that obtained by high-pass filtering the rows and low-pass filtering the columns is called the HL image. The sub-image obtained by high-pass filtering the rows and columns is called the HH image. Each of the sub-images obtained can be filtered and sub-sampled to produce four more sub-images. This process is continued until the desired sub-band structure is obtained. The LL sub-image represents the base image and is called the Approximation image. The other high frequency sub-images are called the details, with LH, HL and HH representing the Vertical, Horizontal and Diagonal Details.

Three of the most popular ways to decompose an image are: Pyramid, Spacl and Wavelet Package methods. In the structure of Pyramid decomposition, only the LL sub-image (approximation) is decomposed into
four more sub-images after each decomposition level. In the structure of Wavelet Package (WP) decomposition, each sub-image (LL, LH, HL, HH) is decomposed after each decomposition level. In the structure of Spatial, after the first level of decomposition, each sub-image is decomposed into smaller sub-images and then only the LL sub-image is further decomposed. These decomposition structures are illustrated in Figures 3.2 to 3.3.

<table>
<thead>
<tr>
<th>A_{LL}</th>
<th>A_{HL}</th>
<th>V_{HL}</th>
</tr>
</thead>
<tbody>
<tr>
<td>A_{LH}</td>
<td>A_{HH}</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>H_{LH}</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D_{HH}</td>
</tr>
</tbody>
</table>

Figure 3.2 Pyramid structure

<table>
<thead>
<tr>
<th>A_{LL}</th>
<th>A_{HL}</th>
<th>V_{LH}</th>
<th>V_{HL}</th>
</tr>
</thead>
<tbody>
<tr>
<td>A_{LH}</td>
<td>A_{HH}</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>H_{LH}</td>
<td>D_{LH}</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H_{HH}</td>
<td>D_{HH}</td>
</tr>
</tbody>
</table>

Figure 3.3 Package structure

<table>
<thead>
<tr>
<th>A_{LL}</th>
<th>A_{HL}</th>
<th>A_{HL}</th>
<th>V_{LH}</th>
<th>V_{HL}</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A_{LH}</td>
<td>A_{HH}</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>H_{LH}</td>
<td>D_{LH}</td>
<td>D_{HH}</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H_{HH}</td>
<td>D_{HH}</td>
<td></td>
</tr>
</tbody>
</table>

Figure 3.4 Spatial structure

This thesis considers the pyramidal wavelet decomposition structure for performing multi-level wavelet decomposition. If I_0 represents a grayscale image, then the first level decomposition using DWT will be,
\[ I_0 = I_{LL1} + I_{LH1} + I_{HL1} + I_{HH1} \] 

where, \( I_{LL1} \) represents the approximation, \( I_{LH1} \) the vertical detail, \( I_{HL1} \) the horizontal detail and \( I_{HH1} \) represents the diagonal details of the image. The approximation \( I_{LL1} \) will be further decomposed at the second level:

\[ I_{LL1} = I_{LL2} + I_{LH2} + I_{HL2} + I_{HH2} \] 

Recursively, the nth level decomposition will be:

\[ I_{LLn-1} = I_{LLn} + I_{LHn} + I_{HLn} + I_{HHn}, n=1,2,-- \] 

Thus the nth decomposition will comprise \( 3n+1 \) sub-image sequences.

The representation of a one-level wavelet based pyramid decomposition of the boat image is shown in Figure 3.5.
The reconstruction of the image from its decomposed sub-images is carried out using the synthesis filters $g'[n]$ and $h'[n]$, as explained in the previous section.

### 3.5 DISCRETE WAVELET TRANSFORM-BASED FUSION

The DWT based fusion techniques involve the initial multi-scale decomposition of the input images, combining the coefficients of the corresponding levels using fusion rules and finally synthesizing the fused image using Inverse Discrete Wavelet Transform (IDWT). This generic flow of fusion using the DWT is illustrated in Figure 3.6 (Pajares and de La Cruz 2004).

![Figure 3.6 Block diagram of generic fusion schemes](image)

There are different fusion rules that are adopted for fusing the approximation images and detailed images separately. Such a need arises because the approximation image represents the visible part of the decomposition level, corresponding to the low frequency component. Hence, while combining these images, the useful visible information must be transferred from the sources to the fused output image. The detail images represent the sharp, high frequency information. They are used to define the edges in the image and contribute to the overall shape of objects contained
within the image. The fusing of detail information requires preserving the edges in the source images and avoiding the introduction of artefacts in the fused image due to the high frequency components. Numerous fusion rules have been published in the literature as discussed in chapter 1. Most of these techniques used a weighting scheme to combine the coefficients. The work detailed in this thesis involves the fusion of two different source images; a visible spectrum image obtained using a Charge Coupled Device (CCD) camera and an infra-red (IR) image obtained using a Forward Looking Infra-Red (FLIR) image capture device. In the subsequent sections two weight based fusion rules are presented, one uses a threshold function to calculate the weights and the second method uses the energy of the wavelet coefficients to calculate the weights of the components to be fused.

![Figure 3.7 Flow of DWT-based data fusion method](image)

**Figure 3.7 Flow of DWT-based data fusion method**
3.5.1 Threshold and Weight-Based Image Fusion Algorithm

The proposed fusion algorithm uses a threshold value that is calculated from the intensity of the pixels in the approximation and detail sub-images. Let the input images from the CCD camera and the IR camera be $I_{vis}$ and $I_{ir}$ respectively. The two source images are subjected to DWT decomposition and let the low-frequency sub-images corresponding to the final decomposition level be $A_{vis}$ and $A_{ir}$. The detail sub-images are $H_{vis}$, $H_{ir}$ (horizontal details), $V_{vis}$, $V_{ir}$ (vertical details) and $D_{vis}$, $D_{ir}$ (diagonal details). The fusion algorithm uses two different rules, each for the approximation and the detail sub-images.

3.5.1.1 Approximation Sub-image Fusion: Threshold and Weight-Based Image Fusion Algorithm

Consider $A_{vis}(i, j)$ and $A_{ir}(i, j)$ to be a pair of pixels in the approximate sub-images. The threshold value $\text{Thresh1}$ is calculated as the mean of the maximum and minimum intensity of the pixels in the first image, $A_{vis}$:

$$\text{Thresh1} = \frac{\min(A_{vis}(i, j)) + \max(A_{vis}(i, j))}{2}$$

$g$ is a decision matrix that is defined by,

$$T_{next} = \frac{\text{mean}(A_{vis}(g)) + \text{mean}(A_{vis}(\sim g))}{2}$$

The new threshold value is calculated as the arithmetic mean of the mean intensity level greater than $\text{Thresh1}$ and the mean intensity level lesser than $\text{Thresh1}$ in the source $A_{vis}$. Mathematically,
\[
g(i, j) = \begin{cases} 
1, & \text{if } A_{\text{vis}}(i, j) \geq \text{Thresh1} \\
0, & \text{otherwise}
\end{cases}
\] (3.16)

where, \(\neg g\) refers to the pixel values for which \(g(i, j) = 0\).

The new threshold calculated is used to evaluate the condition as given in equation 3.15.

The Thresh1 is assigned the value of the new threshold and the process is repeated as long as the condition in equation 3.15 is not satisfied.

The second threshold, Thresh2, is calculated by applying the similar algorithm to the second image, \(A_{\text{ir}}\).

The flowchart representation of the fusion algorithm is shown in Figure 3.8.

![Flowchart of threshold and weight-based fusion algorithm](image)

**Figure 3.8 Flow of threshold and weight-based fusion algorithm**

The fusion weights \(w_1, w_2, w_3\) and \(w_4\) are chosen such that \(w_i\) and \((w_{i+1})\) satisfy \(w_i + (w_{i+1}) = 1\). Hence, after processing the fused approximation, coefficients are available in \(F_a(i, j)\) and can be used to calculate the Inverse DWT.
3.5.1.2 High Frequency Fusion: Decision-Based Image Fusion

Each of the detail components is fused separately by the weighted addition rule:

\[ H_{\text{fuse}}(i,j) = k_1(i,j)H_{\text{vis}}(i,j) + k_2(i,j)H_{\text{ir}}(i,j) \]  \hspace{1cm} (3.17)

for all values of i and j. Here, H represents the horizontal sub-image components obtained from decomposition.

The weighting coefficients k1 and k2 are given by the decision map;

\[
[k_1, k_2] = \begin{cases} 
[0.5, 0.5], & \text{if } D < k_1 \\
[1, 0], & \text{if } k_1 < D < k_2 \text{ AND } H_{\text{vis}} > H_{\text{ir}} \\
[0, 1], & \text{if } k_1 < D < k_2 \text{ AND } H_{\text{vis}} < H_{\text{ir}} 
\end{cases}
\]  \hspace{1cm} (3.18)

where

\[ D = \frac{H_{\text{vis}} - H_{\text{ir}}}{\max(H_{\text{vis}}, H_{\text{ir}})} \]

The above fusion is carried out for the vertical and diagonal components so as to obtain the fused detail components; \( H_{\text{fuse}}, V_{\text{fuse}} \) and \( D_{\text{fuse}} \).

Synthesizing the fused image

The fused low frequency and high frequency components are then used to reconstruct the fused image. This is done by applying the IDWT to the components, namely, \( A_{\text{fuse}}, H_{\text{fuse}}, V_{\text{fuse}} \) and \( D_{\text{fuse}} \). The reconstruction filter is used to process these components to give the fused image output \( I_{\text{fuse}} \).
3.5.2 DWT Coefficient Energy-Based Weighted Fusion Algorithm

This algorithm proposes a weighted fusion scheme for the approximation sub-images, where the weights are calculated based on the energy of the DWT coefficients. The detail sub-images are fused using the rule of choosing the maximum intensity pixels. This would effectively make use of the energy of the source images on finding out the weights of the corresponding source images at the fused output image. Here too we assume the images to be designated as $I_{vis}$ and $I_{ir}$ corresponding to the inputs from the CCD camera and the IR camera. The approximation sub-images are $A_{vis}$ and $A_{ir}$ and the detail components are $H_{vis}$, $H_{ir}$ (horizontal details), $V_{vis}$, $V_{ir}$ (vertical details) and $D_{vis}$, $D_{ir}$ (diagonal details). All the individual detail components can be combined and represented by $Det_{ir}$ and $Det_{vis}$ such that,

$$Det_{ir} = D_{ir} + H_{ir} + V_{ir}$$

$$Det_{vis} = D_{vis} + H_{vis} + V_{vis}$$

The image of a road has been considered for illustrating the fusion scheme. Both the IR image and the CCD camera image has been obtained and are decomposed using the DWT for a single level. From the decomposed images, the approximation sub-image and detail sub-image are shown in Figures 3.9 and 3.10, for each of the source images.
Figure 3.9 DWT decomposition of CCD image

Figure 3.10 DWT decomposition of IR image
3.5.2.1 Approximation Sub-image Fusion Rule

The basic fusion rule for the low frequency approximation is the weighted image fusion given by,

\[
\text{APP}_{\text{fuse}} = (w_1 * A_{\text{vis}}) + (w_2 * A_{\text{ir}})
\]

where,

\[w_1 \text{ and } w_2\] are the normalized weights computed from the energy of the coefficients \(A_{\text{vis}}\) and \(A_{\text{ir}}\).

The weights have been computed from the energy so that the image which has the highest energy would have more weightage at the fused output image. This implies that the fused output image would get more information from the source image which has higher energy and very less information would be taken from low energy content source image.

The energy of the approximation images, \(E_{a1}\) and \(E_{a2}\), are calculated using the general expression for energy,

\[
E_{a1} = \sum_{i=1}^{k} (A_{a1}(i,j))^2
\]

\[
E_{a2} = \sum_{i=1}^{k} (A_{a2}(i,j))^2
\]

The normalization factor \(E_t\) is then calculated as the maximum energy value of the two images.

\[
E_t = \text{Max}\{E_{a1}, E_{a2}\}
\]
Then, the weights $w_1$ and $w_2$ are given by,

$$ W_1 = \frac{E_{a1}}{E_t} \quad (3.25) $$

$$ W_2 = \frac{E_{a2}}{E_t} \quad (3.26) $$

where, $w_1$ and $w_2$ lie in the range of 0 to 1.

From experimentation it has been determined that using the normalization value $E_t$ for finding out the weights $w_1$ and $w_2$ gives better results than using the energy of the own source images directly.

The result of fusing the approximation components using the energy-based weighted fusion scheme is shown in Figure 3.11.

![Energy-Based Weighted Fusion](image)

**Figure 3.11** Energy-based weighted fusion of approximation sub-images
3.5.2.2 Detail Sub-image Fusion Rule

The detail components of the SWT decomposition are combined using the method of obtaining the fused output by selecting the maximum intensity pixels from each of the detail sub-image. This method ensures that the high frequency edge details that are present in each of the sub-images are appropriately retained in the fused image.

Mathematically, the decision map for this is given by,

\[
\text{Det}_{\text{fuse}}(i,j) = \begin{cases} 
\text{Det}_{\text{ir}}(i,j) & \text{if } |\text{Det}_{\text{ir}}(i,j)| > |\text{Det}_{\text{vis}}(i,j)| \\
\text{Det}_{\text{vis}}(i,j) & \text{if } |\text{Det}_{\text{ir}}(i,j)| < |\text{Det}_{\text{vis}}(i,j)|
\end{cases}
\]  

(3.27)

\[i = 1, 2, 3, \ldots S_i\]

\[j = 1, 2, 3, \ldots S_j\]

and \(S_i \times S_j\) are the dimensions of the detail sub-image

The resultant fused detail image is shown in Figure 3.12,
Synthesizing the fused image

The fused approximation image shown in Figure 3.11 and detail components using max rule shown in Figure 3.12 are then combined in order to synthesize to get the final fused output. The result in Figure 3.13 shows that information from both the sources is present in the output and the image also has an appreciably good subjective quality.

Figure 3.13 Result of fusing using the energy-based fusion algorithm

3.6 RESULTS

Experiments were carried out with the newly developed algorithms on different sets of images pertaining to surveillance and night vision applications. The performance metrics are given in Table 3.1 for gun and smoke screen images. DWT energy and threshold outputs are given in Figures 3.14 and 3.15 for the same set of images. A more detailed comparison of the results is given in Chapter 6. In addition, the average results obtained by conducting experiments on 40 sets of images are shown in Table 3.2. Here the results are compared with the existing DWT-Max algorithm (Huntsberger and Jawerth 1993, Zhang and Blum 1999) and the Laplacian-based fusion algorithm (Burt and Adelson 1983).
Table 3.1 Performance metrics for the newly designed MRA-based fusion algorithms

<table>
<thead>
<tr>
<th>Image</th>
<th>Fusion Scheme</th>
<th>En</th>
<th>SSIM</th>
<th>MI</th>
<th>SD</th>
<th>CE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gun DWT Energy</td>
<td>6.6157</td>
<td>0.517307</td>
<td>3.08789</td>
<td>59.3339</td>
<td>1.54435</td>
<td></td>
</tr>
<tr>
<td>Gun DWT Threshold</td>
<td>6.60156</td>
<td>0.451041</td>
<td>3.4833</td>
<td>85.6316</td>
<td>1.19518</td>
<td></td>
</tr>
<tr>
<td>Smoke Screen DWT Energy</td>
<td>7.09711</td>
<td>0.650225</td>
<td>3.08562</td>
<td>61.3231</td>
<td>-0.2233</td>
<td></td>
</tr>
<tr>
<td>Smoke Screen DWT Threshold</td>
<td>7.34082</td>
<td>0.559121</td>
<td>3.76699</td>
<td>90.2928</td>
<td>-0.2587</td>
<td></td>
</tr>
</tbody>
</table>

(a) DWT Energy                                   (b) DWT Threshold

Figure 3.14 Fusion output for Gun image

(a) DWT Energy                                   (b) DWT Threshold

Figure 3.15 Fusion output for smoke screen image
Table 3.2 Average value of different set of images

<table>
<thead>
<tr>
<th>Image Set of images</th>
<th>Fusion Scheme</th>
<th>DWT Energy</th>
<th>DWT Threshold</th>
<th>DWT Max</th>
<th>Laplacian</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>CE</td>
<td>1.00621</td>
<td>6.061468</td>
<td>6.653046</td>
<td>4.837856</td>
</tr>
<tr>
<td></td>
<td>EN</td>
<td>6.715652</td>
<td>6.833483</td>
<td>7.013706</td>
<td>6.329638</td>
</tr>
<tr>
<td></td>
<td>SSIM</td>
<td>0.581902</td>
<td>0.491262</td>
<td>0.701182</td>
<td>0.693228</td>
</tr>
<tr>
<td></td>
<td>MI</td>
<td>2.638773</td>
<td>3.176069</td>
<td>4.200587</td>
<td>2.181704</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>53.4335</td>
<td>70.55499</td>
<td>38.51432</td>
<td>35.54645</td>
</tr>
</tbody>
</table>

3.7 CONCLUDING REMARKS

In this chapter, the MRA-based fusion schemes developed by the author have been discussed. A new fusion algorithm that uses the energy of the DWT coefficients to determine the weights and a threshold-based weighted fusion has been developed. The energy-based fusion has been found to give a poor performance compared to the PCA-based algorithm put forward in this thesis in chapter 4. However, the proposed scheme is easier to implement in real time compared to the recently published DWT weighted fusion algorithm (Zheng et al. 2007). The fusion rule described by Zheng et al. 2007 involves the computation of the Principal Components of the images in order to determine the weights. This process is computationally intensive because of the calculation required to find the eigenvectors of the images, increasing in complexity for images of larger dimensions. From a complexity point of a view, the wavelet-based reduction method yields the order of $O(MN)$, where $N$ is the number of bands and $M$ is the number of pixels in the spatial domain. On the other hand, the total estimated complexity of PCA is $O(MN^2+N^3)$, which shows that the computation efficiency of a wavelet reduction technique is superior to the efficiency of the PCA method.
In this context, the proposed fusion rule gives a subjectively acceptable result and also satisfactory performance metrics. Also, in this approach the weights are chosen to be the energy content of the respective images. This ensures that the weights are not randomly selected and are proportionate to the intensity levels in the image.

The threshold based fusion scheme gives comparatively lesser performance than the energy technique as seen from Table 3.1.

In this chapter the average results obtained from a set of 40 different images shows the effectiveness of the two schemes designed by the author. The average value of Cross Entropy (CE) of the DWT energy algorithm is a little less than the existing DWT max (Huntsberger and Jawerth 1993, Zhang and Blum 1999) algorithm. However, the DWT threshold shows a better CE than the Laplacian fusion (Burt and Adelson 1983). The entropy (EN) of the new algorithms is approximately the same as the existing methods taken for comparison. The SSIM index of the new algorithms appears less compared to the existing methods however, the quality of the fused image based on human observation shows a perceptual improvement. Hence, the results validate the effectiveness of the two new DWT-based fusion schemes.