Chapter 1

Introduction

1.1 Image Fusion

Image Fusion refers to extraction of complementary information and removal of redundancy from multimodal images to obtain all relevant information in a single (hybrid) image [1]. The motivation is to obtain information of higher quality. Image fusion improves data interpretation and recognition by making use of multimodal image observations. Fused image has further applications of feature extraction , statistical calculation, object identification etc. Feature level image fusion requires identification and extraction of salient information of input images and its transfer into the output fused image. As most of signal information is carried by irregular structures and transient phenomena,fusion of edges is an important part of medical image processing. Various methods of image fusion are available using wavelet transform technique . Lifting wavelet transform domain, which is a multiresolution analysis enables to identify and fuse image features. It produces large coefficients near edges, thus revealing salient information. Image fusion can be performed at different levels- signal level, pixel level, feature level and decision level [1].

1.1.1 Types of Fusion

1. **Signal level fusion**: This is low level image fusion. At this level of fusion raw images obtained from multimodal imaging systems are fused. The greatest accuracy and information is achieved at this level of fusion. The disadvantage is that it requires transfer of all signals to a central processor which is difficult.

2. **Pixel level fusion**: In this level of fusion information contained associated with each pixel in an image is enhanced through multiple image combination. Fusion at this level can be performed either in spatial domain or in frequency domain.

3.**Feature level Fusion**: This is an intermediate level image fusion. This level can be used as a means of creating additional composite features. Features can be pixel intensities or edge and texture features. At first, relevant features are abstracted from input images and then fused. For this raw data is transformed and represented as feature vector sets. Various kind of features, such as signal amplitude or shape, length or image segments, are considered depending on the nature of images and the application of the fused image. The fused data can

also be used for classification or detection based on the fused feature set. Fusion at this level has an added advantage that standardized reconstruction procedure is not required[1].

4. **Decision level Fusion:** This is a high level fusion in which decisions coming from various fusion experts are fused. Different methods of decision fusion are voting methods, fuzzy logic based methods, statistical methods. There are various techniques of image fusion. They are roughly divided into two groups-multiscale decomposition based(MBD) fusion methods and non multiscale decomposition based (NMDB) fusion methods. Typical MDB fusion methods include pyramid based methods, discrete wavelet transform based methods. Typical NMDB fusion methods include adaptive weight averaging methods, neural network based methods, Markov random field based methods and estimation theory based methods. The proposed method is based on lifting wavelet transform based method which belongs to MDB category [1].

1.1.2 Applications

1. Image fusion is used in Medical Imaging Techniques ,for example to clear a tumor image.
2. Image fusion is used in Remote Sensing areas for making pictures more informative and Clear around boundaries and corners.

1.2 Wavelet transform

Multiresolution analysis has become one of the most promising methods in image processing. This makes wavelet transform a very useful tool for image fusion. It has been found that wavelet based fusion techniques out perform the standard fusion technique in spatial and spectral quality. The wavelet transform is of two types : Continuous Wavelet transform (CWT) and Discrete Wavelet transform (DWT)[2][3].

Continuous wavelet transform : X(t) is the input signal, then CWT of X(t) is defined as

$$X_w(a,b)=1/\sqrt{b}\int_{-\infty}^{+\infty} X(t)\ \psi(\frac{t-a}{b})dt \qquad (1.1)$$

where location factor a can be any real number, and scaling factor b can be positive real number. The mother wavelet $\psi(t)$ is a well-designed function so that the CWT has low computation complexity and is reversible. It is obvious that as b is larger, $\psi((t-a)/b)$ is more like a high-frequency signal, and thus output $Xw(a, b)$ would represent the high-frequency

component of $x(t)$ after inner product with $\psi((t-a)/b)$. Also, larger b implies the window size of $\psi((t-a)/b)$ is smaller; that is, the time resolution is smaller [2].

Discrete Coefficients Continuous wavelet transform

Although the CWT performs well in mathematics, it is hard to implement. Thus, it is not useful in practical. As we restrict the values of parameters a and b as $a = n2^{-m}$ and $b = 2^{-m}$, the CWT can be rewritten as

$$X_w(n,m)=2^{m/2}\int_{-\infty}^{+\infty} X(t)\ \psi\ (2^m t - n)dt \quad (1.2)$$

this special case is called CWT with discrete coefficients . The main reason of this setting is easy in implementation[2]. As the mother wavelet $\psi(t)$ satisfies

$$\psi(t)= 2\sum_k h_k\ \varphi(t)(2t-k)\ \ and\ \emptyset(t) = \ 2\sum_k h_k\ \varphi(t)(2t-k) \quad (1.3)$$

$X_w(n, m)$ can be computed from $X_w(n, m-1)$ by digital convolution,

$$X_w(n,m) = 2^{\frac{1}{2}}\ \sum_k g_k X_w(2n+k, m+1) \quad (1.4)$$

$$X_w(n,m) = \ 2^{\frac{1}{2}}\ \sum_k h_k X_w(2n+k, m+1) \quad (1.5)$$

The $\varphi(t)$, called scaling function, can be deemed as a low-pass filter compared to the high-pass filter $\psi(t)$. Although the setting of $a = n2^{-m}$ and $b = 2^{-m}$, we can only obtain some coefficients in the particular positions of the time-frequency distribution, However, these coefficients are enough for image processing[2].

Discrete Wavelet Transforms (DWT)

The DWT is similar to the DC-CWT except that the input signal is discrete. Therefore, the design rules for $\psi(t)$, $\varphi(t)$, $g[k]$ and $h[k]$ are similar as in the DC-CWT.2-D DWT is very useful for image processing because the image data are discrete and the spatial-spectral resolution is dependent on the frequency. The DWT has the property that the spatial resolution is small in low-frequency bands but large in high-frequency bands[2].

1.3 Lifting wavelet transform

Lifting wavelet transform is a multi resolution analysis used for the construction of the second generation wavelets. It is an efficient implementation of the wavelet transform algorithm. The discrete wavelet transform (DWT) can be viewed as a predictor-error decomposition. The scaling coefficients at a given scale (j) are "predictors" for the data at the next higher resolution or scale (j-1). The wavelet Coefficients are simply the "prediction errors" between the scaling coefficients and the higher resolution data that they are attempting to predict. This interpretation has led to a new framework for DWT known as the lifting scheme. It lifts the wavelet transform to a more sophisticated level and is implemented by factoring the wavelet transforms into lifting steps. The lifting scheme consists of the iteration of the following three steps[1][4].

(i)**Lazy wavelet transform**: This step divides the original data ($x[n]) \in R, n \in Z$) into its even and odd polyphase components $x_e[n]$ and $x_o[n]$ respectively where

$$x_e[n] = x[2n] \tag{1.6}$$

$$x_o[n] = x[2n+1] \tag{1.7}$$

(ii) **Predict**: This is also called as dual lifting. In this step the odd poly phase coefficients are predicted from the neighbouring even coefficients using a predictor P and the wavelet coefficients (high pass) or details are generated as the error in predicting the odd samples from the even using prediction operator[1].

$$d = x_o - P(x_e) \tag{1.8}$$

using these details one can recover the odd components as

$$x_o = P(x_e) + d \tag{1.9}$$

iii)**Update**:This is also termed as primal lifting.This step updates the even set using the wavelet coefficients to compute the scaling function coefficients (low pass).It applies an update operator U to detail coefficients obtained in previous step [1].

$$S = x_e + U(d) \tag{1.10}$$

This step is also invertible and reproduces x_e as

$$Xe = s - U(d) \tag{1.11}$$

Lifting scheme has the advantage of fast implementation of wavelet transform as it makes use of similarities between high pass and low pass filters. It provide perfect reconstruction of original signal which is not possible by standard implementation even though they are lossless in principle. It helps in saving auxiliary memory as the original signal is gradually replaced by its transform. Lifting scheme has further advantage of simplicity of inverse transform as the inverse transform is obtained by reverting the order of operations and inverting the signs in forward lifting steps. It also reduces the computational complexity by a factor of two as compared to non-lifting wavelet transform. It also provides flexibility as compared to classical wavelets[1][4].

Chapter 2

Wavelet Transform

2.1 Fourier Transform: *An Overview*

The Fourier transform is a frequency domain representation of a function. This transform contains exactly the same information as that of the original function; they differ only in the manner of presentation of the information[8].

Fourier transform is given by

Forward FT: $$X(f)=\int_{-\infty}^{\infty} x(t)e^{-j2\pi ft}\ dt \quad (2.1)$$

Inverse FT: $$X(t)= \int_{-\infty}^{\infty} X(f)e^{+j2\pi ft}\ df \quad (2.2)$$

X(t) is the continuous function in time and X(f) is its corresponding Fourier transform, which is a continuous function in frequency. This formula is mainly applied to the functions with bounded energy i.e X(t) should be an energy signal satisfying the following bound.

$$\int_{-\infty}^{\infty}|x(t)|^2\ dt < \infty \quad (2.3)$$

This Fourier transform is mainly used for the theoretical analysis and design of continuous signals and systems. In case of continuous periodic functions, the functions does not have a finite energy. If X(t) is periodic with a period of T and fundamental frequency of f_o =1/T, x(t) = x (t+T) for all t's, and if it has a finite power, the periodic function can then be expressed as a linear combination of harmonically related sinusoidal functions. The pair of equations, which defines the Fourier series(FS) of a periodic function, is stated by

Forward FS: $$C_k = 1/T \int_{-\frac{T}{2}}^{\frac{T}{2}} x(t)e^{-j2\pi kf_0}dt \quad (2.4)$$

Inverse FS: $$x(t) = \sum_{k=-\infty}^{\infty} c_k\ e^{+j2\pi kf_0}$$

where c_k 's are fourier coefficient of $x(t)$. The condition of having finite power for the periodic function $x(t)$ is stated by the following bound.

$$1/T \int_{-T/2}^{T/2}|x(t)|^2\ \mathrm{dt} < \infty \quad (2.5)$$

This transform converts a continuous periodic function to a sequence of complex numbers. In general, this sequence is infinite. In most practical cases, only finite number of c_k 's have significant values[8]. This transformation is also used in many areas of applied mathematics like solving partial differential equations. Advances in computers and digital technology resulted in design of discrete signals and systems and modifications in the Fourier transform. The Fourier transform that is applied to discrete sequences and referred to as discrete time Fourier transform (DTFT) is defined by the following pair of equations.

Forward DTFT: $$X\left(e^{j2\pi f}\right) = \sum_{n=-\infty}^{\infty} x[n]e^{-j2\pi nf} \quad (2.6)$$

Inverse DTFT: $$x[n] = \int_{-\pi}^{\pi} X(e^{j2\pi f})e^{+j2\pi nf}\,df \quad (2.7)$$

Where $x[n]$ is the discrete function and $X(e^{j2\pi f})$ is its corresponding Fourier transform. The transform function is continuous and periodic in the frequency domain, with the period of 2π . In this formulation, the frequency variable, F , is normalized by the sampling frequency F_s . In other words, if F_a is the actual frequency in Hz, $F = F_a / F_s$ is the normalized frequency . This transformation is commonly used for analysis of discrete signals and systems[8].

Calculation of DTFT by computer can only be carried out for finite sequences and for discrete samples of $X(e^{j2\pi f})$ in frequency domain. These requirements and constraints result in another formulation of the Fourier transform that is defined for periodic discrete functions. Let $x[n]$ be a periodic sequence with a period of N ; i.e., $x[n] = x[n+N]$ for all n 's, the pair of the Fourier transform relations, referred to as discrete Fourier transform (DFT), for $x[n]$, is defined by

Forward DFT: $$X[k] = \sum_{n=0}^{N-1} x[n]e^{-j2\pi nk/N} \quad \text{for } k = 0,1,2,\ldots,N-1 \quad (2.8)$$

Inverse DFT: $$x[n] = 1/N \sum_{n=0}^{N-1} X[k]e^{+j2\pi nk/N} \quad \text{for } n = 0,1,2,\ldots,N-1 \quad (2.9)$$

Where $x[n]$ and its DFT, $X[k]$, are periodic with the same period N . Although different formulations of the Fourier transform have real application in analyzing signals and systems

but only the DFT relations is practically used in real world computations. Some of the applications of DFT in signal processing are spectrum estimation, feature extraction, and frequency domain filtering. Due to advances in fast computation algorithms for DFT, known as Fast Fourier Transform (FFT), and high-speed hardware implementation, this approach is used for real-time digital signal processing (DSP). But DFT has performance limitations for various applications. Let $x[n]$ for n =0,1,2,....,N -1, be the sequence of real numbers obtained from sampling an analog temporal signal with sampling period of T seconds. The actual duration of this signal is therefore equal to T_0=NT seconds. When calculating the DFT of this sequence, the resultant sequence, $X[k]$, is in general a complex sequence in frequency domain. The actual distance between frequencies associated to the two consecutive samples of $X[k]$ is 1/ NT Hertz (Hz). Due to the symmetry properties of $X[k]$ and sampling constraints, center of $X[k]$ sequence corresponds to the maximum frequency of the signal. This frequency is $\mathrm{F_{max}} = \left(\frac{\mathrm{N}}{2}\right).\left(\frac{1}{\mathrm{NT}}\right) = 1/2\mathrm{T}$ Hz, which is determined by the sampling period T. Resolution of DFT is fixed at ΔF =1/NT =1/T_0 Hz and is depended on the duration of the original analog signal. Increasing number of samples by reducing the sampling period does not change the overall resolution.One main assumption in using DFT for calculation of the spectrum of a discrete signal is that the observed signal is stationary during the observation time T_0 . For most practical signals, this assumption is not valid. In this case and other similar cases, the Fourier transform is modified such that a two-dimensional time-frequency representation of the signal is obtained. The modified Fourier transform referred to as short-time or time-dependent Fourier transform, depends on a window function. For the discrete signals, this transformation, referred to as discrete short timeFourier transform (DSTFT) is

obtained by using a window function, $g[l]$, where

$$g[l] \neq 0 \ \ for \ \ 0 \leq l \leq L-1 \tag{2.10}$$

$$g[l] = 0 \ \ for \ \ l < 0 \ or \ \ l \geq L \tag{2.11}$$

The resultant forward Fourier transform in this case provides estimates of the instantaneous frequency spectrum of the signal at any desired time. The window $g[l]$ has a stationary origin, and as n changes, the signal slides past the window so that, at each value of n , a different portion of the signal is viewed[8].

The main purpose of the window in the time-dependent Fourier transform is to limit the extent of the transformed sequence so that the spectral characteristics are reasonably stationary over the duration of the window function. The more rapidly the signal characteristics change, the shorter the window should be. Resolution in frequency depends on the duration of the window function. In the discrete case and for the uniform window, the actual frequency resolution, in terms of the sampling period T , equals to $\Delta f = 1/LT$ which is the inverse of the actual size of the window. In general the resolution of the DSTFT can be related to the bandwidth of the window sequence[8] . Using RMS (Root-Mean Square) as a measure of bandwidth, the resolution is

$$\Delta f = \left(\sum_{k=0}^{N-1} k^2 |G[k]|^2 \,/\, \sum_{k=0}^{N-1} G|[k]|^2\right)^{1/2} \tag{2.12}$$

Where $G[k]$ is obtained by calculating DFT of the window sequence as follows.

$$G[k]=\sum_n g[n] e^{-\frac{j2\pi nk}{N}} ; \text{ for k=0,1,2,.....,N-1} \tag{2.13}$$

In this approach, two sinusoids will be discriminated only if they are more than Δf apart. Similarly, the spread in time is given by Δt as

$$\Delta t=\left(\sum_n n^2 |g[n]|^2 \,/\, \sum_n |g[n]|^2\right)^{1/2} \tag{2.14}$$

This parameter indicates resolution in time. In other words, two pulses in time can be discriminated only if they are more than Δt apart. As the window becomes shorter, frequency resolution decreases. On the other hand, as the window length decreases, the ability to resolve changes with time increases. Consequently, the choice of window length becomes a trade-off between frequency resolution and time resolution. Resolution in time and frequency cannot be arbitrarily small, because their product is lower bounded [8].

Time-Bandwidth Product , $\Delta t\, \Delta f \geq 1/4\pi$ (2.15)

This is referred to as the uncertainty principle, or Heisenberg inequality. In general, for DSTFT, after deciding about the window function, the frequency and time resolutions are fixed for all frequencies and all times respectively.This approach does not allow any variation in resolutions in terms of time or frequency[8].

2.1.1 Wavelet Transform

Wavelet transform can be defined for different class of functions. The intention in this transformation is to address some of the shortcomings of the STFT. Instead of fixing the time and the frequency resolutions Δt and Δf, one can let both resolutions vary in time-frequency plane in order to obtain a multiresolution analysis[7][8]. This variation can be carried out without violating the Heisenberg inequality . In this case, the time resolution must increase as frequency increases and the frequency resolution must increase as frequency decreases. This can be obtained by fixing the ratio of Δf over f to be equal to a constant c [8] *i.e* $\Delta f/f$ = c .

2.1.2 Merits of wavelet Transform over Fourier Transform

The Fourier transform has its limitations. For example, this transformation cannot be applied to non stationary signals. Although, the modified version of the Fourier transform, reffered to as short time (or variable time) Fourier transform can resolve some of the problems associated with the non stationary signals, but does not address all the issues of concern.The wavelet transform is applied to non stationary signals for the analysis and processing and provides an alternative to the short-time Fourier transform (STFT). In contrast to STFT, which uses a single analysis window, the wavelet transform uses short windows at high frequencies and long windows at low frequencies [8].

2.2 Continuous wavelet Transform(CWT)

The CWT or continuous-time wavelet transform of f(t) with respect to a wavelet ψ(t) is defined as

$$W(a,b)=\int_{-\infty}^{+\infty} f(t)\, \frac{1}{\sqrt{|a|}}\, \psi^{*}\left(\frac{t-b}{a}\right) dt \quad (2.16)$$

Where a and b are real and $*$ denotes complex conjugation.Thus ,the wavelet transform is a function of two variables. The a is called dilation variable while b is called translation variable .

The mother wavelet should posses following properties[3].

1. The function integrates to zero.

$$\int_{-\infty}^{+\infty} (t)dt = 0 \qquad (2.17)$$

2. It is square integrable or equivalently has finite energy.

$$\int_{-\infty}^{+\infty} | (t)|^2 dt < \infty \qquad (2.18)$$

2.3 Discrete Wavelet Transform (DWT)

This is a type of non redundant wavelet representation[3].

$$f(t) = \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} d(k,l) 2^{-k/2} \psi (2^{-k} t - l) \qquad (2.19)$$

Equation (2.19) uses discrete values for dilation and translation parameters. The dilation takes values of the form $a = 2^k$ where k is an integer. At any dilation 2^k, the translation parameter takes values of the form $2^k l$ where l is again an integer. The values $d(k, l)$ are related to values of the wavelet transform $W(a,b) = W[f(t)]$ at $a = 2^k$ and $b = 2^k l$ .This corresponding to sampling the coordinates (a,b) on a grid. The two-dimentional sequence $d(k, l)$ is referred to as the discrete wavelet transform (DWT) of f(t) [3].

The general block scheme of a wavelet or subband transform is shown below [3]. The forward transform uses two analysis filters $\tilde{h}$(low pass) and $\tilde{g}$(band pass) followed by subsampling,while the inverse transform first upsamples and then uses two synthesis filters $h(low\ pass)$ and g (high-pass) [5].

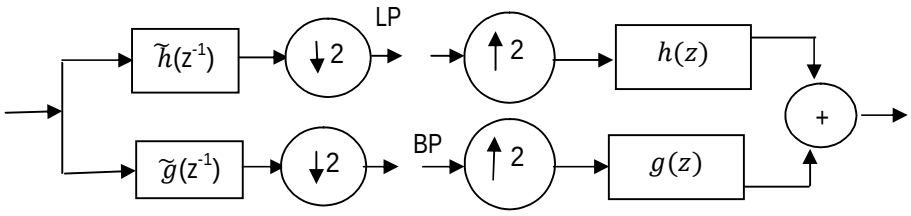

Fig 1: Discrete wavelet Transform

All these filters are considered as FIR filters. The condition for perfect reconstruction are given by

$$h(z)\tilde{h}(z^{-1}) + g(z)\tilde{g}(z^{-1}) = 2 \quad (2.20)$$

$$h(z)\tilde{h}(-z^{-1}) + g(z)\tilde{g}(-z^{-1}) = 0 \quad (2.21)$$

The modulation matrix M(z) is defined as below

$$\mathrm{M(z)} = \begin{bmatrix} h(z) & h(-z) \\ g(z) & g(-z) \end{bmatrix} \quad (2.22)$$

Similarly the dual modulation matrix $\widetilde{M}(z)$ can be defined.The perfect reconstruction condition can now be written as

$$\widetilde{M}(z^{-1})^{t}M(z)=2\mathrm{I} \quad (2.23)$$

Where I is the 2x2 identity matrix

The polyphase representation of a filter h is given by

$$h(z) = h_e(z^2) + z^{-1}h_o(z^2) \quad (2.24)$$

where h_e contains the even coefficients and h_o contains the odd coefficients

$$h_e(z) = \sum_k h_{2k}\, z^{-k} \text{ and } h_o(z) = \sum_k h_{2k+1} z^{-k} \quad (2.25)$$

$$h_e(z^2) = \frac{h(z)+h(-z)}{2} \text{ and } h_0(z^2) = \frac{h(z)-h(-z)}{2z^{-1}} \quad (2.26)$$

Assemble the polyphase matrix as

$$P(z)=\begin{bmatrix} h_e(z) & g_e(z) \\ h_0(z) & g_o(z) \end{bmatrix} \tag{2.27}$$

$$P(z^2)^t = 1/2\ M(z) \begin{bmatrix} 1 & z \\ 1 & -z \end{bmatrix} \tag{2.28}$$

The perfect reconstruction property is given by

$$P(z)\tilde{P}(z^{-1})^t = I \tag{2.29}$$

The wavelet transform does subsampling even and odd samples. This transform is called polyphase transform but in the context of lifting it is often reffered to as the lazy wavelet transform[5].

2.4 Wavelet Families

There are many members in the wavelet family, a few of them that are generally found to be more useful. Haar wavelet is one of the oldest and simplest wavelet. Therefore, any discussion of wavelets starts with the Haar wavelet. Daubechies wavelets are the most popular wavelets. They represent the foundations of wavelet signal processing and are used in numerous applications. The Haar, Daubechies, Symlets and Coiflets are compactly supported orthogonal wavelets. These wavelets along with Meyer wavelets are capable of perfect reconstruction. The Meyer, Morlet and Mexican Hat wavelets are symmetric in shape. The wavelets are chosen based on their shape and their ability to analyze the signal in a particular application. Haar wavelet is discontinuous, and resembles a step function[3][9].

2.4.1 Haar Wavelet

The Haar wavelet is a certain sequence of functions. It is now recognised as the first known wavelet. Haar used these functions to give an example of a countable orthonormal system for the space of square integrable functions on the real line. The study of wavelets, and even the term "wavelet", did not come until much later[3].

The Haar wavelet is also the simplest possible wavelet. The technical disadvantage of the Haar wavelet is that it is not continous , and therefore not differentiable[9].

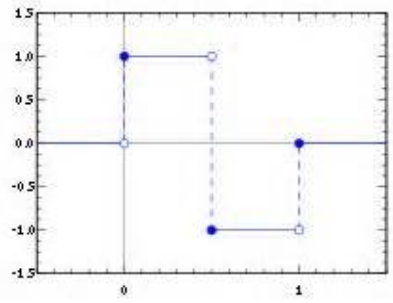

Fig 2. Haar wavelet

The Haar wavelet's mother wavelet function $\psi(t)$ can be described as

$$\psi(t) = \begin{cases} 1 & 0 \leq t < 1/2, \\ -1 & 1/2 \leq t < 1, \\ 0 & \text{otherwise.} \end{cases} \tag{2.30}$$

and its scaling function $\varphi(t)$ can be described as

$$\phi(t) = \begin{cases} 1 & 0 \leq t < 1, \\ 0 & \text{otherwise.} \end{cases} \tag{2.31}$$

wavelets are mathematical functions that were developed by scientists working in several different fields for the purpose of sorting data by frequency. Translated data can then be sorted at a resolution which matches its scale [10].

The Haar wavelet operates on data by calculating the sums and differences of adjacent elements. The Haar transform is computed using:

$$\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \tag{2.32}$$

2.4.2 Daubechies wavelets

Daubechies are compactly supported orthonormal wavelets and found application in DWT. Its family has got nine members in it [3][7].

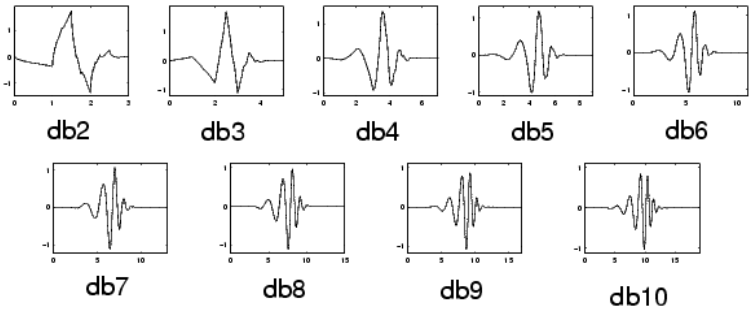

Fig 3: Daubechies wavelets

2.5 Multiresolution Analysis

A multiresolution analysis decomposes a signal into a smoothed version of the original signal and a set of detail information at different scales[3]. This type of decomposition is most easily Understand by thinking of a picture (which is a two dimensional signal). We remove from the picture information that distinguishes the sharpest edges, leaving a new picture that is slightly blurred. This blurred version of the original picture is a rendering at a slightly coarser scale. We then recursively repeat the procedure. Each time we obtain some detail information and a more and more blurred (or smoothed) version of the original image. Removal of the detail information corresponds to a bandpass filtering,and generation of the smoothed image corresponds to a lowpass filtering. Given the decomposition, we can reconstruct the original image. Once we have decomposed a signal this way, we may analyze the behavior of the detail information across the different scales. We can extract the regularity of a singularity,which characterizes the signal's behavior at that point. This provides an effective meansof edge detection. Furthermore, noise has a specific behavior across scales, and hence,in many cases we can separate the signal from the noise. Reconstruction then yields a relatively accurate noise free approximation of the original signal[9][10].

The wavelet transform specifies a multiresolution decomposition, with the wavelet defining the bandpass filter that determines the detail information. Associated with the wavelet is a smoothing function, which defines the complementary lowpass filter. Conditions to be described later ensure that the set consisting of the detail information at all scales and the smoothed version of the original signal contains no redundant information. In lieu of the wavelet transform's ability to localize in time and its ability to specify a multiresolution analysis, many potential application areas have been identified. These include edge

characterization, noise reduction, data compression, and sub-band coding. In order to analyze a nonstationary signal, we need to determine its behavior at any individual event. Multiresolution analysis provides one means to do this.

Multiresolution Analysis analyzes the signal at different frequencies with different resolutions. But every spectral component is not resolved equally as was the case in the STFT. MRA gives good time resolution and poor frequency resolution at high frequencies and good frequency resolution and poor time resolution at low frequencies. This approach is helpful when the signal at hand has high frequency components for short durations and low frequency components for long durations[9].

Decompose a given signal in a nested subspace structure, the question of relation between scaling and wavelet function with their filter coefficients remains to be resolved. To resolve this question, It need to solve refinement (dilation) equations as given by

$$\varphi(x) = \sum_{-\infty}^{\infty} h_k 2^{1/2} \varphi(2x-k) \tag{2.33}$$

$$\Psi(x) = \sum_{-\infty}^{\infty} g_k 2^{1/2} \varphi(2x-k) \tag{2.34}$$

It is able to use FIR filter bank structure to identify the relation between filter coefficients and scaling function (father wavelet) where Fourier Transform of the filter coefficients are used.

Both scaling and wavelet functions, can be constructed from Fourier transform of the low pass analysis filter in the filter bank using refinement (dilation) equations as given by

$$\varphi(x) = \sum_{-\infty}^{\infty} h_k 2^{1/2} \varphi(2x-k) \tag{2.35}$$

$$\Psi(x) = \sum_{-\infty}^{\infty} g_k 2^{1/2} \varphi(2x-k) \tag{2.36}$$

There are two approaches, both use successive approximation as follows.

1. The first approach (backward approach) is as follows. Assume that a scaling function is given, i.e. start with a scaling function $\varphi(x)$,
2. Use refinement equation in which the scaling function $\varphi(x)$ is written as a function of its translates at the resolution twice as fine i.e. $\varphi(2x)$ as given below.

$$\varphi(x) = \sum_{-\infty}^{\infty} h_k 2^{1/2} \varphi(2x-k) \tag{2.37}$$

$$\varphi(x) = \text{S}(\varphi(2x)) \tag{2.38}$$

Where S(.) is an operator.

3. Now provided there is a solution for refinement equation, scaling function $\varphi(x)$ may be determined from above by iteration of a linear transformation S($\varphi(x)$) and successive approximation.

No closed form solution exits for $\varphi(x)$, it can be solved by successive approximation. At each stage, a set of finite number of points may be obtained and used to derive additional points of the continuous function $\varphi(x)$ until sufficient number of points are obtained. Iterative process converges to a unique $\varphi(x)$. Often for $\varphi(x)$ of simple form such as unit pulse or triangular function, it is possible to solve for $\varphi(x)$ by identifying coefficients h(k) of the translates $\varphi(2x-k)$ and $\varphi(x)$ can be constructed[10].

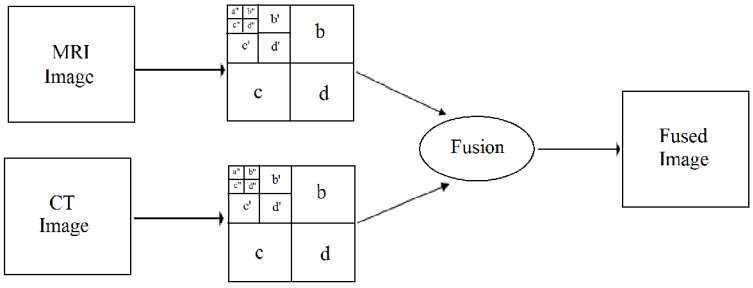

Fig 4: Image fusion using multiresolution analysis.

2.5.1 Signal Decomposition and Reconstruction

Decomposition Process

The image is high and low-pass filtered along the rows. Results of each filter are down-sampled by two. The two sub-signals correspond to the high and low frequency components along the rows, each having a size N by N/2. Each of the sub-signals is then again high and low-pass filtered, but now along the column data and the results are again down-sampled by two[9].

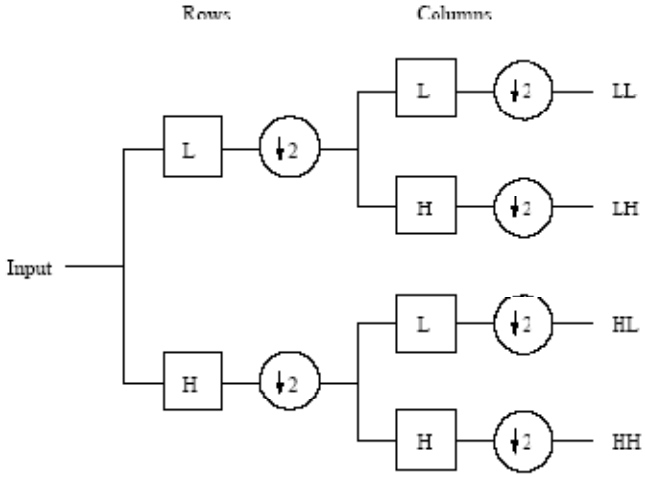

Fig.5 Signal Decomposition

Hence, the original data is split into four sub-images each of size N/2 by N/2 and contains information from different frequency components. Figure 3.15 shows the block wise representation of decomposition step.

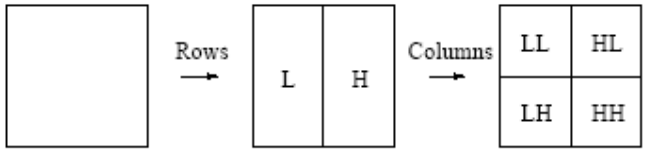

Fig.6: One DWT decomposition step

The LL subband obtained by low-pass filtering both the rows and columns, contains a rough description of the image and hence called the approximation subband. The HH Subband, high-pass filtered in both directions, contains the high-frequency components along the diagonals. The HL and LH images result from low-pass filtering in one direction and high-pass filtering in the other direction. LH contains mostly the vertical detail information, which corresponds to horizontal edges. HL represents the horizontal detail information from the vertical edges. The subbands HL, LH and HH are called the detail subbands since they add the high-frequency detail to the approximation image[7][9].

Reconstruction Process

The four sub-images are up-sampled and then filtered with the corresponding inverse filters along the columns.The result of the last step is added together and we have the original image again, with no information loss[7][9].

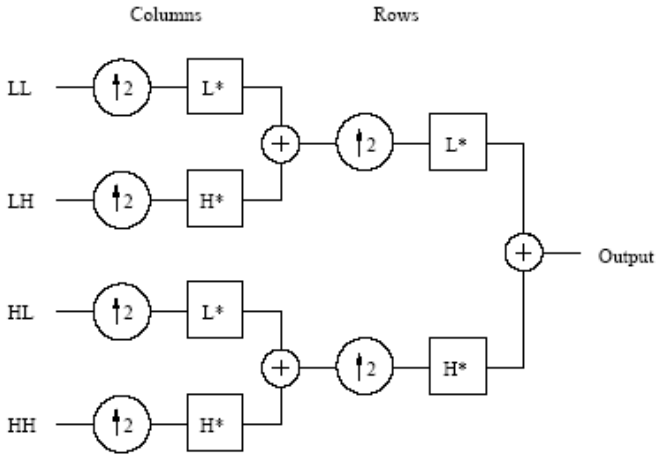

Fig.7 One Recostruction step of the four sub images

2.6 Wavelet Properties

Various properties of wavelet transforms is described below:

1) Regularity
2) The window for a function is the smallest space-set (or time-set) outside which function is identically zero.
3) The order of the polynomial that can be approximated is determined by number of vanishing moments of wavelets and is useful for compression purposes.
4) The symmetry of the filters is given by wavelet symmetry. It helps to avoid de phasing in image processing. The Haar wavelet is the only symmetric wavelet among orthogonals. For biorthogonal wavelets both wavelet functions and scaling functions that are either symmetric or antisymmetric can be synthesized.
5) Orthogonality: This property of wavelet transform implies that the inverse wavelet transform is the adjoint of forward wavelet transform. The necessary condition for the orthogonality of the wavelets is that the scaling sequence is orthogonal to any shifts of it by an even number of coefficients [7].

6) Filter length: Shorter synthesis basis functions are desired for minimizing distortion that affects the subjective quality of the image. Longer filters (that correspond to longer basis functions) are responsible for ringing noise in the reconstructed image at low bit rates[9].

7) Vanishing order is a measure of the compaction property of the wavelets. The synthesis wavelet, when orthogonal to the analysis scaling functions, is said to have p vanishing moments. In the case of orthogonal wavelets, the analysis wavelet function is same as the synthesis wavelet function. Thus, the synthesis as well as the analysis wavelets has the same vanishing moment. However, for biorthogonal wavelets, the analysis wavelet function $\psi(t)$ is different from the synthesis wavelet $\psi(t)$ [9].

A higher vanishing moment corresponds to better accuracy of approximation at a particular resolution. Thus, the lowest frequency subband captures the input signal more accurately by concentrating a larger percentage of the image's energy in the LL subband[9].

Chapter 3

Lifting wavelet Transform

Any discrete wavelet transform or two band subband filtering with finite filters can be decomposed into a finite sequence of simple filtering steps, which we call lifting steps but that are also known as ladder structures. This decomposition corresponds to a factorization of the polyphase matrix of the wavelet or subband filters into elementary matrices. Here, building the decomposition from basic wavelet filtering. This factorization provides an alternative for the lattice factorization, with the advantage that it can also be used in the biorthogonal,i.e, non-unitary case. Like the lattice factorization, the decomposition presented here asymptotically reduces the computational complexity of the transform by a factor two.It has other applications, such as the possibility of defining a wavelet –like transform that maps integers to integers. Various techniques to construct wavelet bases or to factor existing wavelet filters into basic building blocks are known.One of these is lifting. The original motivation for developing lifting was to build second generation wavelet i.e wavelets adapted to situations that do not allow translation and dilation like non-Euclidean spaces. First generation wavelets are all translates and dilates of one or a few basic shapes; the fourier transform is then the crucial tool for wavelet construction. A construction using lifting,on the contrary, is entirely spatial and therefore ideally suited for building second generation wavelets when fourier techniques are no longer available.

Consider a signal X with $x_k \in \mathbf{R}$.

$$X = (x_k)_{k \in z} \tag{3.1}$$

splited it into two disjoint sets which are called the polyphase components :the even indexed samples and odd indexed samples

$$\text{Even} = \quad X_e = (x_{2k})_{k \in z} \tag{3.2}$$

$$\text{Odd} = \quad X_o = (x_{2k+1})_{k \in z} \tag{3.3}$$

Typically these two sets are closely correlated.Thus it is possible one can build a good predictor **P** for the other set,e.g., the even

$$\mathbf{d} = X_0 - P(X_e) \tag{3.4}$$

Given the detail **d** and the odd,we can immediately recover the odd as

$$\mathbf{X_0} = P(X_e)+\mathbf{d} \tag{3.5}$$

If **P** is a good predictor, then **d** approximately will be a sparse set; in other words it is expected that the first order entropy to be smaller for **d** than for X_0. An easy predictor for an odd sample x_{2k+1} is simply the average of its two even neighbours; the detail coefficient then is

$$d_k=x_{2k+1}-(x_{2k}+x_{2k+2})/2 \tag{3.6}$$

From this it can be see that if the original signal is locally linear,the detail coefficient is zero.The operation of computing a prediction and recording the detail we will call a lifting step. This idea connects naturally with wavelets as follows. The prediction steps can take care of some of the spatial correlation ,but now we have a transform from (x_e,x_0) to (x_e,d). The frequency separation is poor since x_e is obtained by simply subsampling so that serious aliasing occurs.To correct this, it is being proposed a second lifting step,which replaces the evens with smoothed values **S** with the use of an update operator **U** apllied to the details :

$$\mathbf{S} = X_e + U(\mathbf{d}) \tag{3.7}$$

Again this step is trivially invertible : given (**S** ,**d**) , X_e can be recovered as

$$X_e = S - U(\mathbf{d}) \tag{3.8}$$

And then $\mathbf{X_0}$ can be recovered as explained earlier .This illustrates one of the built-in features of lifting .No matter how P and U are chosen ,the scheme is always invertible and thus leads to critically sampled perfect reconstruction filter banks.

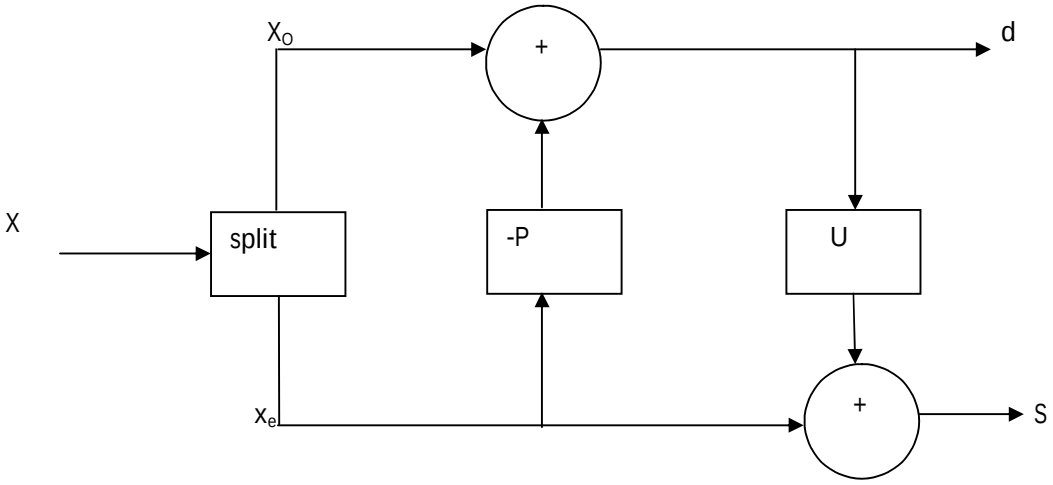

Fig.8 Block diagram of predict and update lifting steps

Every FIR wavelet or filter bank can be decomposed into lifting steps .This can be seen by writing the transform in the polyphase form .Statements concerning perfect reconstruction or lifting can then be made using matrices with polynomials or Laurent polynomial entries. A lifting step then becomes a so- called elementary matrix, that is, a triangular matrix (lower or upper)with all diagonal entries equal to one .Any matrix with polynomial entries and determinant one can be factored into such elementary matrices.

3.1 Filters and Laurent Polynomials

A filter h is a linear time invariant operator and is completely determined by its impulse response :$\{ h_k \epsilon R | k \epsilon z \}$. The z-transform of a FIR filter h is a Laurent polynomial h (z) given by[5]

$$h\,(z) = \sum_{k=kb}^{Ke} h_k z^{-k} \tag{3.9}$$

A 2x2 matrix of Laurent polynomials is

$$M(z) = \begin{bmatrix} a(z) & b(z) \\ c(z) & d(z) \end{bmatrix} \tag{3.10}$$

This matrices also form a ring,which is denoted by $M(2;R[z,z^{-1}])$.

3.2 The lifting scheme

The lifting scheme is an easy relationship between perfect reconstruction filter pair (h, g) that have the same low-pass or high-pass filter.One can then start from the Lazy wavelet and use lifting to gradually build one's way up to a multiresolution analysis with particular properties[5] .

A filter pair (h, g) is complementary in case the corresponding polyphase matrix P(z) has determinant 1.

If (h, g) is complementary, so is $(\tilde{h}, \tilde{g})$ .This states the lifting scheme .

3.2.1 Lifting Theorem

Let (h, g) be complementary . Then any other finite filter h^{new} complementary to h is of the form:

$$g^{new}(z) = g(z) + h(z)\ s(z^2) \tag{3.11}$$

where s(z) is a Laurent polynomial. Conversely any filter of this form is complementary to h.

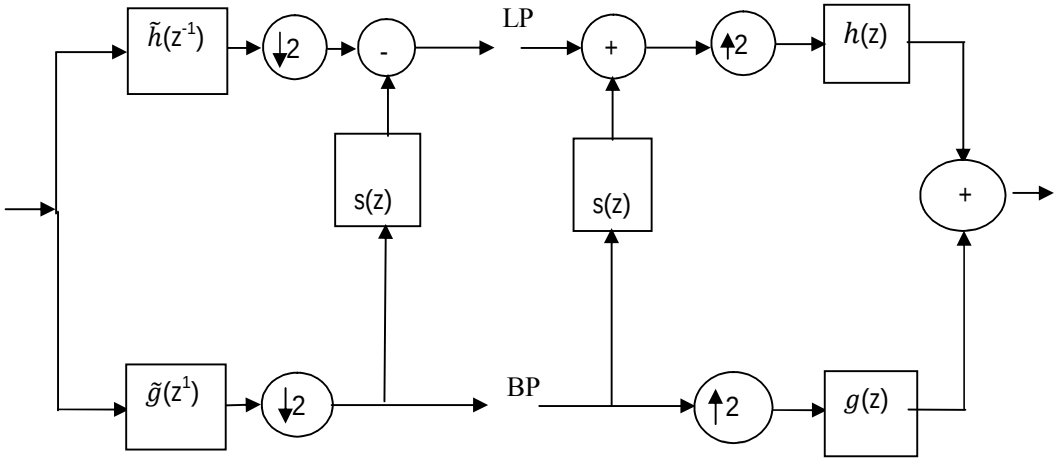

Fig.9: Lifting Scheme

3.2.1 Dual liftingTheorem

Let (h, g) be complementary. Then any other finite filter h^{new} complementary to g is of the form [5]:

$$h^{new}(z) = h(z) + g(z)\, t(z^2) \tag{3.12}$$

where t(z) is a Laurent polynomial. Conversely any filter of this form is complementary to g .

After dual lifting the new polyphase matrix is given by

$$P^{new}(z) = P(z) \begin{bmatrix} 1 & 0 \\ t(z) & 1 \end{bmatrix} \tag{3.13}$$

Dual lifting creates a new $\tilde{g}$ given by

$$\tilde{g}^{new}(z) = \tilde{g}(z) - \tilde{h}(z)\, t(z^{-2}) \tag{3.14}$$

Lifting and dual lifting are used to build Wavelet transforms starting from the lazy wavelet. There a whole family of wavelet is constructed from the lazy followed by one dual lifting and primal lifting step. All the filters h constructed this way are half band and the corresponding scaling functions are interpolating. Because of the many advantages of lifting,it is natural to try to build other wavelets as well,perhaps using multiple lifting steps.

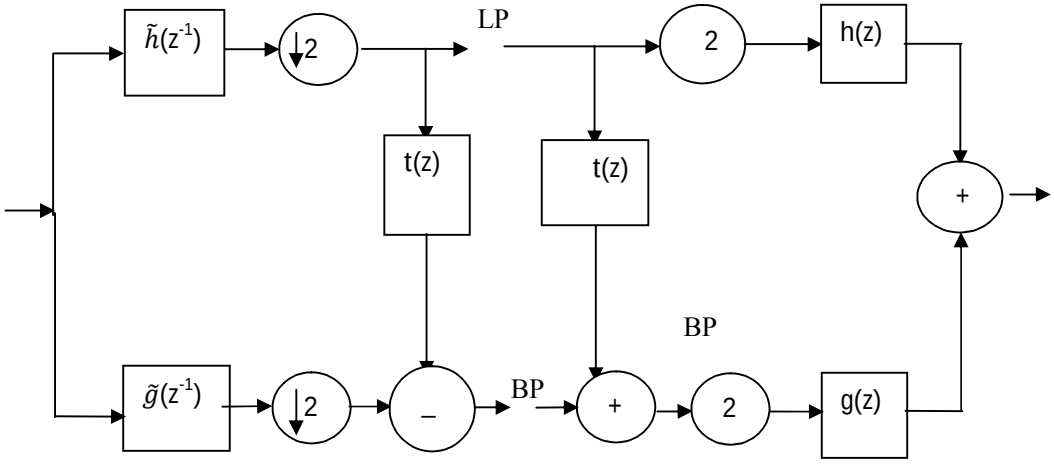

Fig.10: Dual Lifting

3.3 The Euclidean algorithm

The Euclidean algorithm was originally developed to find the greatest common divisor of two natural numbers, but it can be extended to find the greatest common divisor of to polynomials[5]. Here it need to find common factors of Laurent polynomials. The main difference with the polynomial case is again that the solution is not unique. Indeed the gcd of two Laurent polynomials is defined only up to a factor z^p . (This is similar to saying that the gcd of two polynomials is defined up to a constant.)Two Laurent polynomials are relatively prime in case their gcd has degree zero. Note that they can share roots at zero and infinity[5].

3.3.1 Laurent polynomial Theorem

Take two Laurent polynomials a(z) and b(z) ≠ 0 with |a(z)| ≥ |b(z)| . Let $a_0(z)$ = a(z) and $b_0(z)$ = b(z) and iterate the following steps starting from i = 0

$$a_{i+1}(z) = b_i(z) \tag{3.15}$$

$$b_{i+1}(z) = a_i(z) \,\%\, b_i(z) \tag{3.16}$$

Then $a_n(z)$ = gcd (a(z), b(z)), where n is the smallest number for which $b_n(z) = 0$

Given that $|b_{i+1}(z)| < |b_i(z)|$, There is an m so that $|b_m(z)| = 0$.The algorithm then finishes for n = m+1. The number of steps thus is bounded by n ≤ |b(z)| + 1. If we let $q_{i+1}(z)=a_i(z)/b_i(z)$,

We have that

$$\begin{bmatrix} a_n(z) \\ 0 \end{bmatrix} = \prod_{i=n}^{1} \begin{bmatrix} 0 & 1 \\ 1 & -qi(z) \end{bmatrix} \begin{bmatrix} a(z) \\ b(z) \end{bmatrix} \tag{3.17}$$

consequently

$$\begin{bmatrix} a(z) \\ b(z) \end{bmatrix} = \prod_{i=n}^{1} \begin{bmatrix} qi(z) & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} an(z) \\ 0 \end{bmatrix} \tag{3.18}$$

and thus $a_n(z)$ divides both a(z) and b(z). If $a_n(z)$ is a monomial,then a(z) and b(z) are relatively prime[5] .

3.4 THE FACTORING ALGORITHM

This section explains how any pair of complementary filters (h,g) can be factored into lifting steps. First note that $h_e(z)$ and $h_o(z)$ have to be relatively prime because any common factor would also divide det P(z) and we already know that det P(z) is 1. We can thus run the Euclidean algorithm starting from $h_e(z)$ and $h_o(z)$ and the gcd will be a monomial.

Given the non-uniqueness of the division we can always choose the quotients so that the gcd is a constant. Let constant be K. We thus have that

$$\begin{bmatrix}\mathrm{he(z)}\\ ho(z)\end{bmatrix}=\prod_{i=n}^{1}\begin{bmatrix}qi(z) & 1\\ 1 & 0\end{bmatrix}\begin{bmatrix}K\\ 0\end{bmatrix} \tag{3.19}$$

Note that in case $|h_o(z)| > h_e(z)$, the first quotient $q_1(z)$ is zero. We can always assume that *n* is even. Indeed if n is odd, we can multiply the h(z) filter with z and g(z) with $-z^{-1}$. This does not change the determinant of the polyphase matrix. It flips (up to a monomial) the polyphase components of h and thus makes n even again. Given a filter *h* we can always find a complementary filter g^0 by letting

$$\mathrm{P^o(z)} = \begin{bmatrix}h_e(z) & g_e^0(z)\\ h_o(z) & g_o^0(z)\end{bmatrix} = \prod_{i=1}^{n}\begin{bmatrix}q_i(z) & 1\\ 1 & 0\end{bmatrix}\begin{bmatrix}K & 0\\ 0 & 1/K\end{bmatrix} \tag{3.20}$$

Here the final diagonal matrix follows from the fact that the determinant of a polyphase matrix i one and n is even. Let us slightly rewrite the last equation. First observe that

$$\begin{bmatrix}q_i(z) & 1\\ 1 & 0\end{bmatrix} = \begin{bmatrix}1 & q_i(z)\\ 0 & 1\end{bmatrix}\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix} = \begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}\begin{bmatrix}1 & 0\\ q_i(z) & 1\end{bmatrix} \tag{3.21}$$

Using the first equation of (3.21) in case *i* is odd and the second in case *i* is even yields:

$$\mathrm{P^o(z)} = \prod_{i=1}^{\frac{n}{2}}\begin{bmatrix}1 & q_{2i-1}(z)\\ 0 & 1\end{bmatrix}\begin{bmatrix}1 & 0\\ q_{2i}(z) & 1\end{bmatrix}\begin{bmatrix}K & 0\\ 0 & 1/K\end{bmatrix} \tag{3.22}$$

finally, the original filter g can be recovered by applying theorem (3.15) now it is found that the filter g can always be obtained from g^0 with one lifting

$$\mathrm{P(z)} = \mathrm{P^0(z)}\begin{bmatrix}1 & s(z)\\ 0 & 1\end{bmatrix} \tag{3.23}$$

3.4.1 Factoring Theorem.

Given a complementary filter pair (h, g), then there always exist Laurent polynomials $s_i(z)$ and $t_i(z)$ for $1 \leq i \leq m$ and a non zero constant K so that

$$P(z) = \prod_{i=1}^{m} \begin{bmatrix} 1 & s_i(z) \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ t_i(z) & 1 \end{bmatrix} \begin{bmatrix} K & 0 \\ 0 & 1/K \end{bmatrix} \quad (3.24)$$

The proof follows from combining (3.22) and (3.23) ,setting m = n/2 + 1, $t_m(z) = 0$ and $s_m(z) = K^2 s(z)$. In other words every finite filter wavelet transform can be obtained by starting with the lazy wavelet followed by m lifting and dual lifting steps followed with a scaling.

The dual polyphase matrix is given by

$$\tilde{P}(z) = \prod_{i=1}^{m} \begin{bmatrix} 1 & 0 \\ -s_i(z^{-1}) & 1 \end{bmatrix} \begin{bmatrix} 1 & -t_i(z^{-1}) \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1/K & 0 \\ 0 & K \end{bmatrix} \quad (3.25)$$

from this we see that in the orthogonal case ($P(z) = \tilde{P}(z)$) we immediately have two different factorizations.

3.5 Haar Lifting wavelet Transform

For a unnormalized haar wavelet,

$$h(z) = 1 + z^{-1}, g(z) = -\frac{1}{2} + \frac{1}{2z^{-1}}, \tilde{h}(z) = \frac{1}{2} + \frac{1}{2z^{-1}} \text{ and } \tilde{g}(z) = -1 + 1z^{-1} \quad (3.26)$$

using the Euclidean algorithm polyphase matrix can be written as

$$P(z) = \begin{bmatrix} 1 & -\frac{1}{2} \\ 1 & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & -\frac{1}{2} \\ 0 & 1 \end{bmatrix} \quad (3.27)$$

Thus on the analysis side we have

$$P(z)^{-1} = \tilde{P}\left(\frac{1}{z}\right) = \begin{bmatrix} 1 & \frac{1}{2} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} \quad (3.28)$$

This corresponds to the following implementation of the forward transform

$$s_l^{(0)} = x_{2l} \quad (3.29)$$

$$d_l^{(0)} = x_{2l+1} \quad (3.30)$$

$$d_l = d_l^{(0)} - s_l^{(0)} \tag{3.31}$$

$$s_l = s_l^{(0)} + 1/2\ (d_l) \tag{3.32}$$

While the inverse transform is given by

$$s_l^{(0)} = s_l - 1/2(\ d_l) \tag{3.33}$$

$$d_l^{(0)} = \ d_l + s_l^{(0)} \tag{3.34}$$

$$x_{2l+1} = d_l^{(0)} \tag{3.35}$$

$$x_{2l} = s_l^{(0)} \tag{3.36}$$

The intermediate values computed during lifting are denoted with sequences $s^{(i)}\ and\ d^{(i)}$ where as x is the basic sequence of elements ,i.e $x = \{x_l |\ l \in Z\}$.

3.6 Computational Complexity of Lifting wavelet Transform

The cost of applying a filter h is $|h|+1$ multiplications and $|h|$ additions. The cost of the standard algorithm thus is $2(|h|+|g|)+2$. If the filter is symmetric and $|h|$ is even, the cost is $3|h|/2+1$. Let us consider a general case not involving symmetry.

Take $|h| =$ 2N, $|g|$ = 2M, and assume M$\geq N$. The cost of the standard algorithm now is 4(N+M)+2. Without loss of generality we can assume that $|h_e|$ = N, $|h_o|$ = N-1, $|g_e|$ = M and $|g_o|$ = M-1. In general the Euclidean algorithm started from the (h_e, h_o) pair now needs N steps with the degree of each quotient equal to one ($|q_i| = 1$ for $1 \leq i \leq N$).

To get the (g_e, g_o) pair, one extra lifting step is needed with |s| = M - N. The total cost of the lifting algorithm is:

Scaling	2
N lifting steps :	4N
Final lifting step:	2 (M- N +1)
Total	2 (N+ M+2)

3.5 Advantages of Lifting implementation

1. Lifting leads to a speedup when compared to the standard implementation[5].

2. Lifting allows for an in-place implementation of the fast wavelet transform,a feature similar to the fast Fourier transform. This means the wavelet transform can be calculated without allocating auxiliary memory[5].

3. All operations within one lifting step can be done entirely parallel while the only sequential part is the order of the lifting operations[5].

4. Using lifting it is particularly easy to build non linear wavelet transforms. A typical example are wavelet transforms that map integers to integers. Such transforms are important for hardware implementation and for lossless image coding[5].

5. Using lifting and integer to integer transforms, it is possible to combine biorthogonal wavelets with scalar quantization and still keep cubic quantization cells which are optimal like in the biorthogonal case.In a multiple description setting, it has been shown that this generalization to biorthogonality allows for substantial improvements[5].

6. Lifting allows for adaptive wavelet transforms. This means one can start the analysis of a function from the coarsest levels and then built the finer levels by refining only in the areas of interest[5].

Chapter4

Wavelet Transform Modulus Maxima

Points of sharp variations are the most important feature for analyzing the properties of transient signals or images. we concentrate on the canny edge detector which is equivalent to finding the local maxima of a wavelet transform modulus. There are many different types of sharp variation points in images. Edges created by occlusions, shadows, highlights, roofs, textures, etc. Have very different local intensity profiles. To label more precisely an image that has been detected, it is necessary to analyze its local properties. In mathematics singularities are generally characterized by their lipschitz exponents. The wavelet theory proves that these lipschitz exponents can be computed from the evolution across scales of the wavelet transform modulus maxima[6].

4.1 Lipschitz Regularity

Different types of singularities can be discriminated by measuring their local Lipschitz regularity[4].
Let $0 < \alpha < 1$. A function $f(x,y)$ is uniformly Lipschitz α over an open set Ω of R^2 if and only if there exists a constant K such that for all (x_0, y_0) and (x_0, y_0) in Ω.

$$|f(x_0, y_0) - f(x_1 - y_1)| \leq K|(x_0 - x_1)^2 + (y_0 - y_1)^2|^{\alpha/2} \quad (4.1)$$

4.2 Wavelet Transform Modulus Maxima Method

Suppose that wavelets Ψ^1 and Ψ^2 have exactly one vanishing moments and a compact support such that

$$\Psi^1 = -\partial\theta/\partial x, \quad \Psi^2 = -\partial\theta/\partial y \quad \text{and} \quad \int_{+\infty}^{-\infty} \theta(t)dt \neq 0 \quad (4.2)$$

θ is a smoothing function. Thus wavelet transform of f can be written as a multiscale differential operator

$$\begin{pmatrix} W^1 & f(x, y, 2^j) \\ W^2 & f(x, y, 2^j) \end{pmatrix} = \begin{pmatrix} f * & \bar{\psi}^1\, 2^j\,(x, y) \\ f * & \bar{\psi}^2 2^j (x, y) \end{pmatrix} \quad (4.3)$$

$$= 2^j \overline{\nabla} (f * \overline{\theta_{2^j}})(x, y) \quad (4.4)$$

where * denotes the convolution operator and

$$\bar{\psi}^1\ 2^j\ (x,y)\ =\ \bar{\psi}^2\ 2^j(x,y\ ) \qquad (4.5)$$

Practically this transform is computed by iterative filtering with a set of low pass filters h and high pass filters g associated with the wavelets ψ^{1} and ψ^{2} [1].W^1f and W^2f are detail images, since they contain horizontal and vertical details of f. The modulus of this gradient vector is proportional to the wavelet transform modulus

$$M f\ (x, y, 2^j) = \sqrt{|W^1 f(x,y,2^j) + |W^2 f(x,y,2^j)|} \qquad (4.6)$$

And its angle

$$A f\ (x, y, 2^j) = \alpha\ \ W^1\ (x, y, 2^j) \geq\ 0 \qquad (4.7)$$

$$\pi - \alpha\ \ W^1\ (\ x, y, 2^j\ ) < 0 \qquad (4.8)$$

Where

$$\alpha = \tan^{-1}\left(\frac{W^2 f(x,y,2^j)}{W^1 f(x,y,2^j)}\right) \text{ ,when } W^1 f\left(x,y,2^j\right) \neq 0 \qquad (4.9)$$

$$= \pm \frac{\pi}{2}\ ,\ \text{otherwise} \qquad (4.10)$$

Then the local maxima of the wavelet transform modulus

$M f\ (x,y,2^j)$ at a point (x_0, y_0) can be calculated by solving

$$\partial\ M\ f\left(x,y,2^j\ \right) = 0 \qquad (4.11)$$

4.2.1 Properties of Wavelet Modulus Maxima

Any point (s_0, x_0) such that $\left|Wf(s_0,x)\right| < \left|Wf(s_0,x_0)\right|$ when x belongs to either a right or the left neighbourhood of x_0 and $\left|Wf(s_0,x)\right| \leq \left|Wf(s_0,x_0)\right|$ when x belongs to the other side of the neighbourhood of x_0 .

A maxima line is any connected curve in the scale space (s, x) along which all points are modulus maxima .A modulus maxima (s_0,x_0) of the wavelet transform is a strict local maximum of the modulus either on the right or the left side of the x_0.If the wavelet transform has no modulus maximum at fine scales in a given interval,then the function is uniformly Lipschitz α for $\alpha < n$.

4.3 Detection and Measurement of Singularities

There is an approach to detect singularities, with a wavelet which is a Hardy function[4]. A Hardy function *g(x)* is a complex function whose Fourier transform satisfies

$$\hat{g}(w) = 0 \text{ , for } w < 0 \text{ .} \tag{4.12}$$

Let $f(x) \in L^2(R)$ and *Wf(s, x)* be the complex wavelet transform built with a Hardy wavelet.

A smoothing function is any real function $\theta(x)$ such that

$$\theta(x) = O(1/(1+x^2)) \tag{4.13}$$

and whose integral is nonzero. **A** smoothing function can be viewed as the impulse response of a low-pass filter[4].

$$\text{Let } \theta_s(x) = \left(\frac{1}{s}\right)\theta\left(\frac{x}{s}\right) \tag{4.14}$$

Let f(x) be a real function in $L^2(R)$. Edges at the scale s are defined as local sharp variation points of f(x) smoothed by $\theta_s(x)$.

4.4 Edge detection by wavelet transform

Let $\psi^1(x)$ *and* $\psi^2(x)$ *be* the two wavelets defined by

$$\psi^1(x) = \frac{d\theta(x)}{dx} \text{ and } \psi^2(x) = \frac{d^2\theta(x)}{dx^2} \tag{4.15}$$

The wavelet transforms defined with respect to each of these wavelets are given by

$$W^1 f(s,x) = f * \psi_s^1(x) \text{ and } W^2 f(s,x) = f * \psi_s^2(x) \tag{4.16}$$

$$W^1 f(s,x) = f * \left(s\frac{d\theta_s}{dx}\right)(x) = s\frac{d}{dx}(f * \theta_s)(x) \tag{4.17}$$

And

$$W^2 f(s,x) = f * \left(s^2\frac{d^2\theta_s}{dx^2}\right)(x) = s^2\frac{d^2}{dx^2}(f * \theta_s)(x) \tag{4.18}$$

For a fixed scale s , the local extrema of $W^1 f(s,x)$ along the x variable,correspond to the zero-crossings of $W^2 f(s,x)$ and to the inflection points of $f * \theta_s(x)$.

4.5 Multiscale Edge Detection

Multiscale edge detectors smooth the signal at various scales and detect sharp variation points from their first or second order derivative [6].The extrema of the first derivative correspond to the zero crossings of the second derivative and to the inflection points of the smoothed signal. A smoothing function is any function $\theta(x)$ whose integral is equal to 1 and that converges to 0 at infinity[6]. For example one can choose $\theta(x)$ equal to a Gaussian. Suppose that $\theta(x)$ is twice differential and define respectively $\psi^a(x)$ and $\psi^b(x)$ as the first and second order derivative of $\theta(x)$.

$$\psi^a(x) = \frac{d\theta(x)}{dx} \quad \text{and} \quad \psi^b(x) = \frac{d^2\theta(x)}{dx^2} \tag{4.19}$$

By definition, the functions $\psi^a(x)$ and $\psi^b(x)$ can be considered to be wavelets because their integral is equal to 0

$$\int_{-\infty}^{+\infty} \psi^a(x)dx = 0 \quad \text{and} \quad \int_{-\infty}^{+\infty} \psi^b(x)dx = 0 \tag{4.20}$$

Now dilation of any function $\xi(x)$ by a scaling factor s is denoted by

$$\xi_s(x) = 1/s\ \xi\left(\frac{x}{s}\right) \tag{4.21}$$

A wavelet transform is computed by convolving the signal with a dilated wavelet. The wavelet transform of $f(x)$ at the scale s and position x, computed with respect to the wavelet $\psi^a(x)$, is defined by

$$W_s^a f(x) = f * \psi_s^a(x) \tag{4.22}$$

The wavelet transform of $f(x)$ with respect to $\psi^b(x)$ is

$$W_s^b f(x) = f * \psi_s^b(x) \tag{4.23}$$

We derive that

$$W_s^a f(x) = f * \left(s\frac{d\theta_s}{dx}\right)(x) = s\frac{d}{dx}(f * \theta_s)(x) \text{ and} \tag{4.24}$$

$$W_s^b f(x) = f * \left(s^2 \frac{d^2\theta_s}{dx^2}\right)(x) = s^2 \frac{d^2}{dx^2}(f * \theta_s)(x) \tag{4.25}$$

The wavelet transforms $W_s^a f(x)$ and $W_s^b f(x)$ are respectively, the first and second derivative of the signal smoothed at the scale s . The local extrema of $W_s^a f(x)$ thus correspond to the zero crossings of $W_s^b f(x)$ and to the inflection points of $f * \theta_s(x)$.

In the particular case where $\theta(x)$ is a Gaussian ,the zero-crossing detection is equivalent to a Marr-Hildreth edge detection, whereas the extrema detection corresponds to a canny edge detection. When the scale s is large, the convolution with $\theta_s(x)$ removes small signal fluctuations. Therefore only the sharp variations of large structures can be detected. Detecting zero crossings or local extrema are similar procedures ,but the local extrema approach has some important advantages. An inflection point of $f * \theta_s(x)$ can either be a maximum or a minimum of the absolute value of its first derivative. The maxima of the absolute value of the first derivative are sharp variation points of $f * \theta_s(x)$, whereas the minima correspond to slow Variations. With a second derivative operator, it is difficult to distinguish these two types of zero crossings. On the contrary, with a first order derivative, we easily select the sharp variation points by detecting only the local maxima of $|W_s^a f(x)|$.In addition, zero crossings give position information but do not differentiate small amplitude fluctuations from important discontinuities.When detecting local maxima,we can also record the values of $W_s^a f(x)$ at the maxima locations, which measures the derivative at the inflection points.

The canny edge detector is easily extended in two dimensions .We denote by

$$xi_s(x, y)=1/s^2\, \xi(\frac{x}{s},\frac{y}{s}) \qquad (4.26)$$

the dilation by s of any 2-D function $\xi(x,y)$.we use the term 2-D smoothing function to describe any function $\theta(x,y)$ whose integral over x and y is equal to 1 and converges to 0 at infinity.The image f (x,y) is smoothed at different scales s by a convolution with $\theta_s(x,y)$.We then compute the gradient vector $\vec{\nabla}(f * \theta_s)(x,y)$.The direction of the gradient vector at a point (x_0,y_0) indicates the direction in the image plane (x,y) along which the directional derivative of $f(x,y)$ has the largest absolute value.Edges are defined as points (x_0,y_0) where the modulus of the gradient vector is maximum in the direction towards which the gradient vector points in the image plane.

Edge Points are inflection points of the surface $f * \theta_s(x,y)$.Let us relate this edge detection to a 2-D wavelet transform.We define two wavelet functions $\psi^1(x,y)$ and $\psi^2(x,y)$ such that

$$\psi^1(x,y) = \frac{\partial\theta(x,y)}{\partial x} \quad \text{and} \quad \psi^2(x,y) = \frac{\partial\theta(x,y)}{\partial y} \tag{4.27}$$

Let $\psi_s^1(x,y) = \frac{1}{s^2}\psi^1(\frac{x}{s},\frac{y}{s})$ and (4.28)

$$\psi_s^2(x,y) = \frac{1}{s^2}\psi^2(\frac{x}{s},\frac{y}{s}). \tag{4.29}$$

Let $f(x,y) \in L^2(R^2)$. The wavelet transform of $f(x,y)$ at the scale s has two components defined by

$$W_s^1 f(x,y) = f * \psi_s^1(x,y) \text{ and } W_s^2 f(x,y) = f * \psi_s^2(x,y) \tag{4.30}$$

Similarly to (4),one can easily prove that

$$\begin{pmatrix} W_s^1 f(x,y) \\ W_s^2 f(x,y) \end{pmatrix} = s\begin{pmatrix} \frac{\partial(f*\theta_s)(x,y)}{\partial x} \\ \frac{\partial(f*\theta_s)(x,y)}{\partial y} \end{pmatrix} = s\,\vec{\nabla}(f*\theta_s)(x,y) \tag{4.31}$$

Hence, edge points can be located from the two components $W_s^1 f(x,y)$ and $W_s^2 f(x,y)$ of the wavelet transform[6].

4.6 Wavelet Transform of Image

Multiscale sharp variation points can be obtained from a dyadic wavelet transform if

$$\psi^1(x,y) = \frac{\partial\theta(x,y)}{\partial x} \text{ and } \psi^2(x,y) = \frac{\partial\theta(x,y)}{\partial y} \tag{4.32}$$

Wavelet transform can be rewritten

$$\begin{pmatrix} W_s^1 f(x,y) \\ W_s^2 f(x,y) \end{pmatrix} = 2^j \begin{pmatrix} \frac{\partial}{\partial x}(f*\theta_{2^j})(x,y) \\ \frac{\partial}{\partial x}(f*\theta_{2^j})(x,y) \end{pmatrix} \tag{4.33}$$

$$= 2^j \vec{\nabla}(f*\theta_{2^j})(x,y) \tag{4.34}$$

The two components of the wavelet transform are proportional to the two components of the gradient vector $\vec{\nabla}(f*\theta_{2^j})(x,y)$.At each scale 2^j, the modulus of the gradient vector is proportional to

$$M_{2^j} f(x,y) = \sqrt{|W_{2^j}^1 f(x,y)|^2 + |W_{2^j}^2 f(x,y)|^2} \tag{4.35}$$

The angle of the gradient vector with the horizontal direction is given by

$$A_{2^j} f(x,y) = argument(W_{2^j}^1 f(x,y) + iW_{2^j}^2 f(x,y)) \tag{4.36}$$

Like in the canny algorithm, the sharp variation points of $f * \theta_{2^j}(x, y)$ are the points (x, y), where the modulus $M_{2^j} f(x, y)$ has a local maxima in the direction of the gradient given by $A_{2^j} f(x, y)$.The position of each of these modulus maxima as well the values of the modulus $M_{2^j} f(x, y)$ and the angle $A_{2^j} f(x, y)$ at the corresponding locations are recorded.

Chapter 5

Implementation

The term image refers to a two-dimensional light intensity $f(x, y)$, where x and y denote spatial coordinates and the value of f at any point (x, y) is proportional to the brightness (or gray level) of the image at that point [10] .

5.1 Digital Image

The term digital image refers to an image $f(x, y)$ that has been discretised both in spatial coordinates and brightness[10]. A digital image can be considered a matrix whose row and column indices identify a point in the image and the corresponding matrix element value identifies the gray level at that point. The elements of such a digital array are called picels (picture elements), or more commonly pixels. Images are built up of pixels that contain color information and are aligned with the Cartesian coordinate system.. The image's width is represented by the variable N the image's height with the variable M as shown below:

$$f(x, y) = \begin{pmatrix} f(0, 0) & f(0, :) & . & f(0, N-1) \\ f(:, 0) & f(1, :) & . & f(1, N-1) \\ . & . & . & . \\ f(M-1, 0) & f(M-:, 1) & . & f(M-1, N-1:) \end{pmatrix} \quad (5.1)$$

5.2 Implemented algorithm

The implemented algorithm consists of the following steps:

1. Decomposition of input images using lifting wavelet transform.
2. Calculation of modulus of gradient of lifting wavelet transform.
3. Fusion based on threholded value of gradient of wavelet coefficients.
4. Inverse lifting wavelet transform of fused image.
5. Evaluation of image fusion performance parameters

The two preregistered images A(x, y) and B(x, y) are decomposed into the subband images using the lifting scheme of wavelet transform. Suppose that the low frequency subimages of A and B are represented by $LA_J(x,y)$ and $LB_J(x,y)$ respectively and high frequency

subimages are represented by $HA_J^K(x,y)$ and $HB_J^K(x,y)$ respectively (J is the parameter of resolution). The modulus of the gradient of the wavelet coefficients generated from $HA_J^K(x,y)$ and $HB_J^K(x,y)$ are represented by $GA_J^K(x,y)$ and $GB_J^K(x,y)$ (for every J,K=1,2,3 corresponding to horizontal, vertical and diagonal directions respectively). Then exchange of the high frequency components takes place according to the following rule:

$$\text{If } GA_J^K(x,y) - GB_J^k(x,y) > \delta^K \sqrt{\frac{1}{N} \sum_{i=1}^{N} ((x_i - \bar{x})^2} \quad (5.2)$$

$$HF_J^K(x,y) = HA_J^K(x,y) \quad (5.3)$$

$$\text{if } GB_J^K(x,y) - GA_J^K(x,y) \geq \delta^K \quad (5.4)$$

$$HF_J^K(x,y) = HB_J^K(x,y) \quad (5.5)$$

Where δ^K is the threshold value and $HF_J^K(x,y)$ is the high frequency component of the fused image. Finally the resultant image is obtained using the inverse lifting wavelet transform.In this method the high frequency subimage fusion is performed using the modulus maxima criterion .As the modulus maxima criterion of the gradient of the wavelet coefficients in a multiresolution framework extract singularities and edges, this method improves the edge and boundary information in the fused image. The insignificant edges are eliminated by taking into account only the pixels with modulus superior to a threshold value.

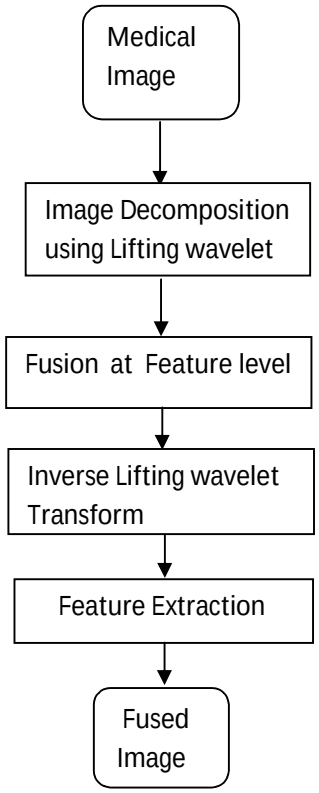

Fig.11. Diagram of Proposed Algorithm

5.3 **Image Fusion Performance Assessment**

The Parameters like standard deviation, entropy and average gradient are calculated for evaluation of image fusion performance[12].

Standard Deviation (SD)

The standard deviation (σ) which is the square root of variance reflects the distribution or spread in the data. The standard deviation measures the contrast in an image. Thus a high contrast image will have large variance and low contrast image will have a low variance. It indicates the closeness of the sharpened image to the original multispectral image at pixel level. The standard deviation measures the contrast in an image[12].

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x - \bar{x})^2} \qquad (5.6)$$

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \quad (5.7)$$

Where x_i is the data vector and x is the mean value

Standard deviation also reflects the detail information of an image. The bigger the standard deviation is, the richer the detail of the image is[12].

Entropy

Entropy measures information content in an image[12].

$$H = -\sum_{i=0}^{L-1} P_i log_2 P_i \quad (5.8)$$

Where L is the number of gray levels and

$$P_i = \frac{\text{Number of Pixels Di of each gray level i}}{\text{Number of Pixels D in the image}}$$

Gradient

Gradient has been used as a measure of image sharpness. The gradient at any pixel is the derivative of the difference values of neighboring pixels. Generally sharper image have higher gradient values. Thus any image sharpening method should result in increased gradient values because this process makes the images sharper compare to the low-resolution image. The gradient defines the contrast between details variation of pattern on the image and the clarity of the image. G is the index to reflect the expression ability of the little detail contrast and texture of the image[12]. It can be given by

$$\bar{G} = \frac{1}{(m-1)(n-1)} \sum_{i=1}^{m-1} \sum_{j=1}^{n-1} \sqrt{\frac{\Delta I_x^2 + \Delta I_y^2}{2}} \quad (5.9)$$

$$\Delta I_x = f(i+1, j) - f(i, j) \quad (5.10)$$

$$\Delta I_y = f(i, j+1) - f(i, j) \quad (5.11)$$

Chapter 6
Result and Discussion

This thesis presents feature level image fusion using Haar lifting wavelet transform.Feature fused is edge and boundary information, which is obtained using wavelet transform modulus maxima criteria. Simulation results show the superiority of the result as entropy, gradient, standard deviation are increased for fused image as compared to input images. The proposed methods have the advantages of simplicity of implementation, fast algorithm, perfect reconstruction and reduced computational complexity. Computational cost of Haar wavelet is very small as compared to other lifting wavelets.

We used T1 and T2 weighted MRI images as input and extracted the edges using modulus of wavelet transform function. Edges are fused using threshold value of modulus of wavelet transform function. Table I gives the values of entropy, gradient and standard deviation for input and output images. Then one of the input image is blurred and fusion is performed. The values show that these parameters are increased in case of fused image. Also, visually good image is obtained when one of the input image is blurred.

6.1 Figure of Results

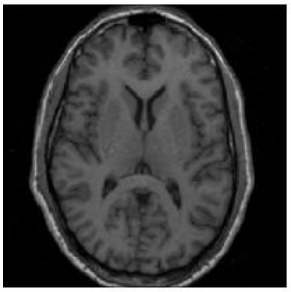

Fig12 : T1 weighted image

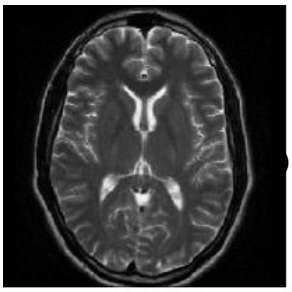

Fig13: T2 weighted image

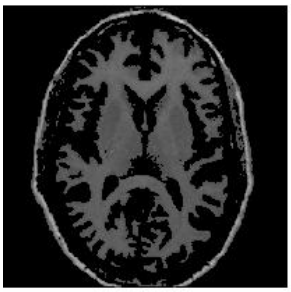

Fig14:T1 Blurred weighted image

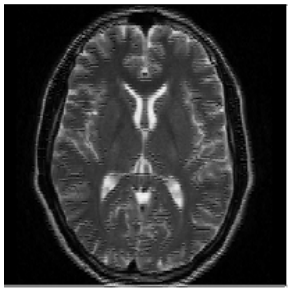

Fig15:Fusion result of T2 and blurred T1 weighted image

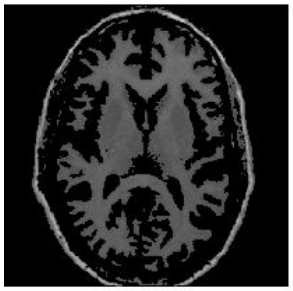

Fig16: Blurred T2 weighted image

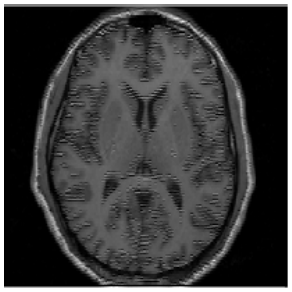

Fig17: Fusion result of T1 and blurred T2 weighted image

6.2 Fusion Performance Evaluation

S. No	Parameters	Source images	Fused Image
1	Average gradient	0.2891,0.3163	0.3221
2	Standard deviation	159.8302	161.2191
3	Entropy	3.6786	4.8183

Table 1: Performance Evaluation For T2 weighted and Blurred T1 weighted Images

S. No	Parameters	Source images	Fused Image
1	Average gradient	0.2887,0.3163	0.3212
2	Standard deviation	160.0630	161.2013
3	Entropy	4.6311	4.8191

Table 2 : Performance Evaluation For Unblurred T1 weighted and T2 weighted Images

6.3 Future work

In Future we can put our implementation on Colour Medical Images and can perform wavelet Decomposition at different Scales .

References

[1] Sudipta Kor & Umashanker Tiwary. "*Feature level Fusion Of Multimodal Medical Images In Lifting Wavelet Transform Domain"* In Proc.IEEE .San Francisco.Sept.2004

[2] Shih-Gu-Huang. *"Wavelet for Image Fusion"*.

[3] Raghuveer M.Rao and Ajit S.Bopardikar. *"Wavelet Transforms .Introduction To Theory and Application"*.Pearson, 2011

[4] Stephane Mallat and Wen Liang Hwang . *"Singularity Detection and Processing with wavelets".* IEEE Trans.Information Theory ,vol.14,no.2,March 1992

[5] Ingrid Daubechies And Wim Sweldens . *"Factoring Wavelet Transforms into Lifting Steps".*Sept1996

[6] Stephane Mallat and Sifen Zhong. "*Characterization of Signals from Multiscale Edges*". IEEE Trans.Pattern Analysis,Vol.14,no.7,July 1992

[7] Tinku Acharya and Ajoy K. Ray . *"Image Processing .Principles and Applications"*. Wiley Publication.1998

[8] Ali M.Reza. *"From Fourier Transform to wavelet Transform"* .White Paper, Oct 27, 1999

[9] L.Prasad and S.S. Iyengar "*Wavelet Analysis with application to ImageProcessing*". CRC Press,1997

[10] RC Gonzalez and RE Woods, *"Digital Image Processing"*. 2nd Ed., Englewood Cliffs, Prentice-Hall, Inc., 2002.

[11] G.P.Hegde, Dr. Nagaratna Hegde and Dr. I.V. Muralikrishna, "*Measurement of Quality Preservation of Pan-sharpened Image".* International Journal of Engineering Research and Development, *Volume 2, Issue 10 (August 2012), PP. 12-17*

[12] Xinman Zhang and Jiuqiang Han ."*Multiscale Contrast Image Fusion Scheme with Performance Measures*".Optica Applicata,vol.xxxiv,No 3, 2004

List of Figures & Tables

Figures **Page no**

Tables **Page no**