Chapter 2

Foundation of Denoising

It is difficult to get rid of noise while preserving detail features of images at the same time. To ensure an appropriate trade-off between noise suppression and detail preservation, suitable filters should be designed depending on the image type and noise models. In this chapter, common additive noise models, various image denoising methods, wavelet based algorithms, bilateral filters and performance evaluation methods are introduced.

2.1 Image and Noise Models

Many of the current denoising techniques are based on the assumption of noise models. In reality, assumptions may not always hold true due to the varied nature and sources of noise. An ideal denoising procedure requires a priori knowledge of the noise; while a practical procedure may not have the required information about the variance of the noise or the noise models. Thus, in practice, most of the algorithms estimate the variance of noise and assume the noise models to compare the performance with different denoising algorithms.

While there have been many assumed additive noise models, only those relevant to this thesis works, such as impulse noise models and Gaussian noise models, are discussed below.
2.1. Image and Noise Models

2.1.1 Impulse Noise Models

Impulsive noise can be caused by coding or decoding errors, transient noise, errors in analog-to-digital conversion, etc. The properties of spikes, including their statistical characteristics, also vary depending on the situation. The amplitudes of spikes can be different at different image regions. Sometimes, impulsive noise can corrupt separate lines, columns or fragments of images. However, a priori knowledge of spike properties is limited. In order to design image denoising algorithms under the presence of this type of noise, some general models of the spikes should be assumed. Impulse noise is a type of additive noise. The properties of impulsive noise are usually determined by the amplitude of spike $A$ and the probability of spike occurrence $p_r$. In many current impulse noise models for images, corrupted pixels are often replaced with values equal to or near the maximum or minimum intensity values of the allowable dynamic range. For example, salt-and-pepper impulsive noise typically corresponds to fixed values near 0 (minimum) or 255 (maximum), for an 8-bit gray scale image.

In this work, a more general noise model in which a noisy pixel is taken as an arbitrary value in the dynamic range according to some underlying probability distribution is introduced. Let $O(i, j)$ and $A(i, j)$ denote the intensity value of the original and the noisy image at position $(i, j)$, respectively. Then, for an impulse noise model with error probability $p_r$ is described as:

$$A(x) = \begin{cases} O(i, j), & \text{with probability } 1 - p_r \\ \eta(i, j), & \text{with probability } p_r \end{cases}$$

where $\eta(i, j)$ is an identically distributed, independent random process with an arbitrary underlying probability density function [10]. Recently, it has been demonstrated that a type of $\alpha$-stable distribution can approximate impulse noise more accurately than other models [9]. The parameter $\alpha$ controls the degree of impulsiveness, which increases as $\alpha$ decreases. The Gaussian ($\alpha = 2$) and the Cauchy ($\alpha = 1$) distributions are the only symmetric $\alpha$-stable distributions that have closed-form probability density functions. A symmetric $\alpha$-stable ($S\alpha S$) random variable is described by its characteristic function:

$$\Phi(t) = \exp(j\theta t - \gamma |t|^{\alpha})$$
where $j$ is the imaginary unit, $\theta$ is the location parameter (centrality), $\gamma$ is the dispersion of the distribution and $\alpha \in [0,2]$, which controls the heaviness of the tails, and it is the characteristic exponent.

### 2.1.2 Gaussian Noise Model

A type of noise, which occurs in all recorded images to a certain extent, is detector noise. This type of noise is due to the discrete nature of radiation, i.e., the fact that each imaging system records an image by counting photons. Allowing some assumptions (which are valid for many applications), this noise can be modeled with an independent, additive model, where the noise $n(i,j)$ has a $\mu$ mean Gaussian distribution described by its standard deviation ($\sigma$), or variance. This means that each pixel in the noisy image is the sum of the true pixel value and a random, Gaussian distributed noise value. The 1-D Gaussian distribution has the density function,

$$G(x) = \frac{1}{\sqrt{2\pi\sigma}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$  \hspace{1cm} (2.3)

where $\sigma$ is the standard deviation of the distribution, and $\mu$ is the expectation of the distribution.

### 2.1.3 Mixed Noise Model

Mixed noise can cause significant difficulties for filtering, image interpretation, and restoration. Different types of noise require the use of different filtering approaches. Therefore, the algorithms of mixed noise removal are often based on hybrid structures using fuzzy and adaptive methods of processing [12]-[13]. Let us consider the following typical model with additive noise and spikes,

$$g(i,j) = \begin{cases} 
  f(i,j) + n(i,j), & \text{with probability } 1 - p_r \\
  A_{imp}(i,j), & \text{with probability } p_r 
\end{cases}$$  \hspace{1cm} (2.4)

where $n(i,j)$ denotes the additive noise value for the $ij^{th}$ pixel, and $A_{imp}(i,j)$ is the amplitude of a spike, which may occur in the $ij^{th}$ pixel with probability $p_r$. 


2.2 Classification of Denoising Methods

There are two basic approaches to image denoising, spatial domain filtering methods and transform domain filtering methods [14].

2.2.1 Spatial Domain Denoising Methods

Spatial filtering is a traditional way to remove noise from image data. This filtering is of two types:

1. Linear Filtering:

   Most classical linear image processing techniques are based on the assumption that images are stationary. Even though linear filters are useful in a wide variety of applications, there are some situations in which they are not adequate. For instance, linear filters may involve image processing applications where both edge enhancement and noise reduction are desired. Usually edge enhancement can be considered as a high-pass filtering operation, while noise reduction is most often achieved using low-pass filtering operations. However, linear shift-invariant (LSI) filters do not take into account any structure in images. Therefore, the degree of smoothing is same over all parts of an image and will cause the loss of some detailed information of the image. However, it is well known that people are less sensitive to noise in more detailed regions of an image [15]. If the filter can smoothen less in these regions than in the less detailed regions, it will preserve the detailed information of the images while smoothing out noise. Conventionally, linear shift invariant filters do not adapt to image content.

2. Non-linear Filtering:

   Nonlinear filters modify the value of each pixel in an image based on the value returned by a nonlinear function that depends on the neighboring pixels. Nonlinear filters are mostly used for noise removal and edge detection. For example, the traditional nonlinear filter is the median filter. It can efficiently decrease additive noise, especially impulsive noise. There are also many improved median filters, such as the weighted median filter [16], center
2.2. Classification of Denoising Methods

weighted median filters [17], detail preserving median based filters [18], the multilevel hybrid median filter [19], etc. However, they do not get rid of all additive noise and blur edges to some degree.

2.2.2 Transform Domain Filtering

According to the choice of the "analysis function" [20], the transform domain filtering methods can be classified into the following two categories.

1. Spatial-Frequency Filtering:

Spatial-Frequency Filtering refers to low pass filters using Fast Fourier Transform (FFT). In frequency smoothing methods [15] the removal of the noise is achieved by designing a frequency domain filter and adapting a cut-off frequency to distinguish the noise components from the useful signal in the frequency domain. These methods are time consuming and depend on the cut-off frequency and the filter function behaviour. Furthermore, they may produce frequency artifacts in the processed image.

2. Wavelet domain:

As mentioned above, noise is usually concentrated in high frequency components of the signal, which correspond to small detail size when performing a wavelet analysis. Therefore, removing some high frequency (small detail components), which may be distorted by noise, is a denoising process in the wavelet domain. Many investigations have been made into additive noise suppression in images using wavelet transforms. Filtering operations in the wavelet domain can be categorized into wavelet thresholding, statistical wavelet coefficient model and undecimated wavelet transform based methods.

In this thesis works, the wavelet thresholding model has been used. In next section wavelets are discussed in detail.
2.3 Wavelets in Image Denoising

This section starts with an overview of some basic wavelet concepts. These can be found in many books and papers at many different levels of exposition. Some of the standard books are [5], [6], [21]-[23]. Introductory papers include [24]-[26], and more technical ones are [27]-[29]. The rest of the section presents ideas of various wavelet based image denoising methods and reviews the state of the art in this field.

2.3.1 The wavelet concept and its origins

The central idea to wavelets is to analyze a signal according to scale. Imagine a function that oscillates like a wave in a limited portion of time or space and vanishes outside of it. The wavelets are such functions: wave-like but localized. One chooses a particular wavelet, stretches it to meet a given scale and shifts it, while looking into its correlations with the to be analyzed signal. This analysis is similar to observing the displayed signal (e.g., printed or shown on the screen) from various distances. The signal correlations with wavelets stretched to large scales reveal gross features, while at small scales fine signal structures are discovered. It is therefore often said that the wavelet analysis is to see both the forest and the trees. In such a scanning through a signal, the scale and the position can vary continuously or in discrete steps. The latter case is of practical interest in this thesis works. From an engineering point of view, the discrete wavelet analysis is a two channel digital filter bank composed of the lowpass and the highpass filters, iterated on the lowpass output. The lowpass filtering yields an approximation of a signal (at a given scale), while the highpass filtering yields the details that constitute the difference between the two successive approximations.

The wavelet family is generated from a unique prototype function that is called a mother wavelet. Given a real variable \( x \), the function \( \Psi(x) \) is called a mother wavelet provided that it oscillates, averaging to zero \( \int_{-\infty}^{+\infty} \Psi(x), dx = 0 \) and that is well localized (i.e., rapidly decreases to zero when \( |x| \) tends to infinity). By convention it is centered around \( x = 0 \), and has a unit norm \( ||\Psi(x)|| \). In practice, applications impose additional requirements, among which, a given number of vanishing moments \( N_v \) is one.
2.3. Wavelets in Image Denoising

The mother wavelet \( \Psi(x) \), generates the other wavelets \( \Psi_{(a,b)}(x), a > 0, b > \mathbb{R}, \) of the family by change of scale \( a \) (i.e., by dilation) and by change of position \( b \) (i.e., by translation),

\[
\Psi_{(a,b)}(x) = \frac{1}{\sqrt{a}} \Psi \left( \frac{x - b}{a} \right), \text{ where } a > 0, b > \mathbb{R}
\]  

(2.6)

In fig. 2.1, several wavelets are shown that are obtained from the mother wavelet \( \Psi(x) = (1 - 2x^2)e^{-x^2} \), this wavelet is the second derivative of a Gaussian function and is called the Mexican hat. For the first time, it was used in computer vision, for multiscale edge detection [30]. The origins of the wavelet analysis can be traced to the 1909 as Haar wavelet and various "atomic decompositions" in the history of mathematics. A comprehensive review is available in [22]. The current use of the name "wavelet" is due to Grosman's and Morlet's work on geophysical signal processing, which led to the formalization of the continuous wavelet transform [31]. In the development of wavelets, the ideas from many different fields (including subband coding and computer vision) have merged. Excellent texts on this topic are [23] and Daubechies paper "Where do the wavelets come from" [32].

2.3.2 Continuous and dyadic wavelet transforms

The continuous wavelet transform (CWT) of a signal \( f(x) \) is defined as
2.3. Wavelets in Image Denoising

\[ Wf(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f(x)\Psi^* \left( \frac{x-b}{a} \right) dx = \langle f, \Psi_{(a,b)} \rangle \]  

(2.7)

where \( \Psi^*(x) \) denotes the complex conjugate of \( \Psi(x) \). The existence of the inverse transform is guaranteed if \( \int_{-\infty}^{+\infty} \frac{|\hat{\Psi}(W)|^2}{|W|} dW \triangleq C_\Psi < +\infty \) where \( \hat{\Psi}(W) \) is the Fourier transform of \( \Psi(x) \).

This is called the admissibility condition [6]. It implies \( \hat{\Psi}(0) = 0 \), and thus \( \Psi(x) \) can be viewed as an impulse response of a bandpass filter. Obviously, the CWT offers a great degree of freedom in the choice of a wavelet. The inverse transform is defined as

\[ f(x) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \left( Wf(a, b)\Psi_{(a,b)}(x)dadb \right)/C_\Psi \]  

(2.8)

The CWT is highly redundant, and is shift invariant. It is extensively used for the characterization of signals [33]. The evolution of the CWT magnitude across scales provides information about the local regularity of a signal. However, in most applications of the CWT, for the sake of memory savings, dyadic scales \( a = 2^j \) are commonly used. The corresponding transform \( Wf(2^j, b) \) is called the dyadic wavelet transform.

In case of images, one can use an arbitrary number \( N \geq 1 \) of (spatially oriented) wavelets \( \Psi_{2^j,u,v}^n(x,y) = \Psi^n(2^{-j}(x-u,y-v)) \), \( 1 \leq n \leq N \) which yields \( N \) component transform \( \{W^1f(2^j, u, v), ..., W^Nf(2^j, u, v)\} \), where

\[ W^n f(2^j, u, v) = \frac{1}{\sqrt{a}} \left( \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y)\Psi^{n*} \left( \frac{x-u}{2^j}, \frac{y-v}{2^j} \right) dudv \right) = \langle f, \Psi_{2^j,u,v}^n \rangle \]  

(2.9)

The conditions, which \( \Psi_{2^j,u,v}^n(x,y) \), \( 1 \leq n \leq N \) have to meet for a complete and a stable representation of an image, are presented in [6]. The dyadic wavelet transform with spatially oriented wavelets is extensively used in texture processing [15].

2.3.3 What makes wavelets useful in signal processing

In signal processing, the representation of signals plays a fundamental role. David Marr elaborated in [34] on this topic. For example, the Arabic numeral representation permits one to easily notice a power of 10, but more difficult to realize a
power of 2. With the binary representation, the situation is reversed. Meyer [22], wrote, “any particular representation makes certain information explicit at the expense of information that is pushed into the background and may be quite hard to recover”. The Fourier representation reveals the spectral content of a signal, but makes it impossible to recover the particular moment in time or the particular space coordinates in case of images where a certain change has occurred. This makes the Fourier representation inadequate when it comes to analyzing transient signals. In signal and image processing, concentrating on transients (like image discontinuities) is a strategy for selecting the most essential information from often an overwhelming amount of data. In order to facilitate the analysis of transient signals, i.e., to localize both the frequency and the time information in a signal, numerous transforms and bases have been proposed [6], [23]. Among those, the wavelet and the short time Fourier Transform (STFT) are quite standard. In the STFT (which is also called the window Fourier transform or the Gabor transform) the signal is multiplied by a smooth window function (typically Gaussian) and the Fourier integral is applied to the windowed signal. For a signal $f(x)$, the STFT [6]

$$S(\tau, w) = \int_{-\infty}^{+\infty} f(x)g(x - \tau)e^{-jwx} dx$$

(2.10)

where $g(x)$ is the window function. The basis functions of a STFT expansion are $g(x)$ modulated by a sinusoidal wave and shifted in time; the modulation frequency is changing while the window remains fixed. A few of these functions and the corresponding tilings [35] of the time-frequency plane are illustrated in fig. 2.2.

In wavelet analysis, the scale can be interpreted as the inverse of frequency. As opposed to STFT, which divides the time-frequency plane into equal blocks, the wavelet transform acts as a microscope [23] focusing on smaller time phenomena as the scale decreases. This behaviour permits a local characterization of signals, which the Fourier and the window Fourier transform do not. Other main advantageous properties of the wavelet transform are:

1. Multiresolution - a scale invariant representation;

2. Edge detection - large wavelet coefficients correspond to image edges;
2.4 Discrete Wavelet Transform

The discrete wavelet transform (DWT) is in literature commonly associated with signal expansion into biorthogonal wavelet bases. The same convention has been adopted in this work. Thus, as opposed to the highly redundant CWT, there is no redundancy in the DWT of a signal; the scale is sampled at dyadic steps $a \in \{2^j : j \in \mathbb{Z}\}$, and the position is sampled proportionally to the scale $b \in \{k2^j : (k, j) \in \mathbb{Z}^2\}$. By no means can a DWT be understood as a simple sampling from a CWT. In the first place, the choice of a wavelet is now far more restrictive: if we are dealing with finite-energy signals $f(x) \in L^2(\mathbb{R})$, the wavelet $\Psi(x)$ has to be chosen such that $\{\Psi(2^{-j}(x - 2^j k))\}_{(j,k) \in \mathbb{Z}^2}$ is a basis of $L^2(\mathbb{R})$. The first such basis was constructed by Alfred Haar in 1909, and the choice for better ones has culminated in Ingrid Daubechies's work [36]. The systematic framework for constructing wavelet bases, known as the multiresolution analysis, was mostly de-
2.4. Discrete Wavelet Transform

Developed by Stephane Mallat [37], and has merged the ideas of pyramidal algorithms [38] in computer vision and the filter banks for subband coding [39]. Books like [5], [6], [21]-[23] provide a comprehensive treatment of these topics. A particularly comprehensive filter bank point of view is presented in [40].

The orthogonal wavelets are rarely available as closed form expressions, but rather obtained through a computational procedure which uses discrete filters. The link between wavelets and these discrete filters is essential for understanding the Mallat's fast DWT algorithm and its extension to images. The rest of the sections address the non-decimated transforms and aspects that are especially important for image denoising.

2.4.1 Wavelet frames and bases

We can start with the notion of bases and frames. A series expansion of a signal \( f \) from some space \( S \) is represented in equation (2.11).

\[
f = \sum_i C_i \Phi_i \tag{2.11}
\]

where the elementary "atoms" or building blocks \( \Phi_i \) are simple waveforms. If the set \( \{ \Phi_i \}_{i \in \mathbb{Z}} \) is complete for the space \( S \) (meaning that all \( f \in S \) can be decomposed as in equation (2.11), then a dual set \( \{ \Phi_i \}_{i \in \mathbb{Z}} \) exists, such that the expansion coefficients \( C_i \) in equation (2.11) are given by inner products \( C_i = \langle \Phi_i, f \rangle \). A complete and linearly independent set \( \{ \Phi_i \}_{i \in \mathbb{Z}} \) is a basis of \( S \); its dual set is then also a basis \( \{ \Phi_i \}_{i \in \mathbb{Z}} \) of \( S \), and is biorthogonal to the primal one: \( \langle \Phi_i, \Phi_j \rangle = \delta(i-j) \), where \( \delta(i) \) is the Kronecker Delta. An important special case is when the set \( \Phi_i \) constitutes an orthonormal basis, where \( \langle \Phi_i, \Phi_j \rangle = \delta(i-j) \) (the dual basis is now equal to the primal one). If the set \( \{ \Phi_i \} \) is complete, but the functions \( \Phi_i \) are not linearly independent, the representation is redundant (overcomplete), and is not a basis but a frame.

A common wavelet basis of \( L^2(\mathbb{R}) \) is a family of functions

\[
\{ \Psi_{j,k}(x) = \frac{1}{\sqrt{2^j}} \psi \left( \frac{x - 2^j k}{2^j} \right) \}_{j,k \in \mathbb{Z}} \tag{2.12}
\]

for a suitably constructed \( \psi(x) \). Generalizations with other than dyadic scales, are discussed in [5] but these are beyond the scope of this work.
Any finite energy signal \( f(x) \) can be decomposed in a basis of the equation (2.12), as

\[
f(x) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} W_{j,k} \Psi_{j,k}(x)
\]  

(2.13)

where \( W_{j,k} \) are the wavelet coefficients, given by the inner products of \( f(x) \) with the dual basis functions \( \Psi_{j,k}(x) \)

\[
W_{j,k} = \langle f, \Psi_{j,k} \rangle = \int_{-\infty}^{\infty} f(x) \Psi_{j,k}^*(x) dx
\]  

(2.14)

The first example of a wavelet basis is the 1909 Haar system, where the wavelet is "blocky":

\[
\Psi(x) = \begin{cases} 
1, & \text{if } 0 \leq x < 1/2 \\
-1, & \text{if } 1/2 \leq x < 1 \\
0, & \text{otherwise}
\end{cases}
\]  

(2.15)

The Haar wavelet has a compact support (meaning \( \Psi(x) = 0 \) outside of a bounded interval). Moreover, the corresponding basis is orthonormal. However, these blocky wavelets are far from optimal for representing smooth functions. There are infinitely many other choices. The systematic way of their construction is indirect and starts from the scaling functions, which span the nested approximation spaces. The wavelets span the complementary spaces, which contain the differences between two successive approximations.

### 2.4.2 Multiresolution Analysis

Multiresolution analysis (MRA) results from a sequence of nested approximation spaces \( ... V_3 \subset V_2 \subset V_1 \subset V_0 ... \). By projecting a signal onto this sequence, a ladder of its approximations is obtained. We use the notation in which the index \( j \) refers to the resolution scale \( 2^j \) [37], the indexing is reversed.

For finite energy signals \( f(x) \), the \( V_j \) are subspaces of \( L^2(\mathbb{R}) \). By definition [27], the approximation spaces satisfy the following properties:

(i) The spaces are embedded as \( V_{j+1} \subset V_j \).
2.4. Discrete Wavelet Transform

(ii) The orthogonal projections \( P_{V_j} f(x) \) of \( f(x) \) onto \( V_j \) satisfy \( \lim_{j \to \infty} P_{V_j} f(x) = 0 \) and \( \lim_{j \to -\infty} P_{V_j} f(x) = f(x) \).

(iii) The \( V_j \) are generated by a scaling (father wavelet) function \( \varphi(x) \in L^2(\mathbb{R}) \), in the sense that, for each fixed \( j \), the family

\[
\{ \varphi_{j,k}(x) = \frac{1}{\sqrt{2^j}} \varphi\left( \frac{x - 2^j k}{2^j} \right) \}_{k \in \mathbb{Z}} \tag{2.16}
\]

is a stable basis (Riesz basis) of \( V_j \).

From this definition, it follows clearly that \( f(x) \in V_{j+1} \) is equivalent to \( f(2x) \in V_j \), and that \( V_j \) is invariant under translation of \( 2^j \).

The orthogonal projection of a signal \( f(x) \) onto \( V_j \), which is its best approximation \( f_j(x) \) at the scale \( 2^j \), is:

\[
P_{V_j} f(x) \triangleq f_j(x) = \sum_{k=-\infty}^{\infty} S_{j,k} \varphi_{j,k}(x) \tag{2.17}
\]

where \( S_{j,k} \) are the scaling coefficients. The details that constitute the difference between two successive approximations \( \Delta f_j(x) = f_{j-1}(x) - f_j(x) \) are contained in the detail space \( W_j \), which is a complement of \( V_j \) in \( V_{j+1} \):

\[
V_j \bigcap W_j = 0 \text{ and } V_{j+1} = V_j \bigoplus W_{j+1} \tag{2.18}
\]

The spaces \( W_j \) are differences between the \( V_j \), and the spaces \( V_j \) are sums of the \( W_j \). For some \( L < J \), \( V_L = (\bigoplus_{j=1}^{j=J} W_j) \bigoplus V_j \), i.e., for a function in these spaces \( fL(x) = \sum_{j=L+1}^{J} \Delta f_j(x) + f_j(x) \). This decomposition "telescopes" into the signal at the scale \( 2^L \).

By analogy with the approximation spaces, the detail spaces are built by dilating and shifting the mother wavelet \( \Psi(x) \), such that the family \( \Psi_{j,k}(x)_k \in \mathbb{Z} = 2^{-j/2}\Psi(2^{-j}(x) - k)_{k \in \mathbb{Z}} \) is a Riesz basis of \( W_j \).

The orthogonal projection of \( f(x) \) onto this space is

\[
P_{W_j} f(x) \triangleq f_j(x) = \sum_{k=-\infty}^{\infty} W_{j,k} \Psi_{j,k}(x) \tag{2.19}
\]

where wavelet coefficients \( W_{j,k} \) carry the necessary information to refine the signal approximation. Desired properties of approximation (e.g., the degree of smoothness) impose a particular choice of \( \varphi(x) \), from which the wavelet \( \varphi(x) \)
2.4. Discrete Wavelet Transform

Figure 2.3: Meyer wavelet $\psi(x)$ and scaling function $\phi(x)$, and amplitudes of their Fourier transforms $|\hat{\psi}(\omega)|$ and $|\hat{\phi}(\omega)|$.

directly follows. For example, for piece-wise constant approximations, $\varphi(x)$ is a box function: let $\varphi(x) = 1$ for $x \in [0, 1]$ and $\varphi(x) = 0$ elsewhere and $\varphi_{j,k}(x) = \varphi((x - 2^j k)/2^j)$; $f_j(x)$ is then piece-wise constant over intervals $[2^{-j} k, 2^{-j} (k + 1)]_{k \in \mathbb{Z}}$. The corresponding wavelet is the Haar wavelet. In this case, $\varphi_{j,k} k \in \mathbb{Z}$ and $\psi_{j,k}, k \in \mathbb{Z}$ are orthogonal bases of $V_j$ and $W_j$ respectively, and $V_j \perp W_j$. Another example of orthogonal wavelet and scaling functions together with their frequency responses is given in fig. 2.3.

A general MRA is not orthogonal; in a non orthogonal case, the dual functions $\tilde{\varphi}(x)$ and $\tilde{\psi}(x)$ are needed to express the coefficients:

$$S_{j,k} = \langle f, \tilde{\varphi}_{j,k} \rangle$$

(2.20)

$$W_{j,k} = \langle f, \tilde{\psi}_{j,k} \rangle$$

(2.21)

The dual families $\{\tilde{\varphi}_{j,k}\}_{k \in \mathbb{Z}}$ and $\{\tilde{\psi}_{j,k}\}_{k \in \mathbb{Z}}$ span the spaces $\tilde{V}_j$ and $\tilde{W}_j$ respectively, such that $\tilde{V}_j \perp V_j$ and $\tilde{W}_j \perp W_j$, for $i \neq j$, and $\tilde{V}_j \perp W_j$ and $\tilde{W}_j \perp V_j$ for all $j$. This biorthogonal setting gives more freedom in designing scaling and wavelet bases [41].

At this point it is to be noted that the wavelet series expansion in equation (2.13) can be rewritten as

$$f(x) = \sum_{j=-\infty}^{J} \sum_{k=-\infty}^{\infty} W_{j,k} \varphi_{j,k}(x) + \sum_{k=-\infty}^{\infty} S_{j,k} \varphi_{j,k}(x)$$

(2.22)

where the expansion of a signal into the scaling basis $\varphi_{j,k}$ replaces the aggregation of infinitely many details ($J + 1 < j < \infty$). According to the notation in
2.4. Discrete Wavelet Transform

2.4.3 Wavelets and discrete filters

A direct consequence of the property $V_{j+1} \subset V_j$ of approximation spaces is that the function $2^{-j/2} \varphi(2^{-j} x) \in W_j$ can be decomposed into the basis of $V_0$, which is $\varphi(x - k)_{k \in \mathbb{Z}}$. The same argument holds for the function $2^{-j/2} \Psi(2^{-j} x) \in W_j$, since $W_{j+1} \subset V_j$. Formally,

$$\frac{1}{\sqrt{2}} \varphi \left( \frac{x}{2} \right) = \sum_{k \in \mathbb{Z}} h_k \varphi(x - k) \quad (2.23)$$

$$\frac{1}{\sqrt{2}} \Psi \left( \frac{x}{2} \right) = \sum_{k \in \mathbb{Z}} g_k \varphi(x - k) \quad (2.24)$$

Equation (2.23) is called the dilation equation, two-scale equation or the scaling equation, while equation (2.24) is referred to as the wavelet equation. The sequences $h$ and $g$ can be interpreted as discrete filters.

In the biorthogonal case, similar relations are defined for the dual scaling and wavelet functions, via the dual $\tilde{h}$ and the dual $\tilde{g}$ filters. A case of great importance is when the impulse responses of these filters are finite (FIR); the corresponding wavelet and scaling functions are then of compact support. A necessary condition for the perfect reconstruction [40] (i.e., for the duality of $\varphi(x)$ and $\tilde{\varphi}(x)$ [27]) is $2 \sum_{n \in \mathbb{Z}} \tilde{h}_n h_{n+2k} = \delta_k$. Once $h$ and $\tilde{h}$ are specified, the wavelet filters follow as $g_n = (-1)^n \tilde{h}_{1-n}$ and $\tilde{g}_n = (-1)^n h_{1-n}$. The equations (2.23) and (2.24) are thus the core for the construction of wavelet bases [5], [6], [40], which is beyond the scope of this thesis. The construction of fast discrete wavelet transform algorithms is addressed in the next section.

2.4.4 DWT - A fast discrete wavelet transform algorithm

Mallat has introduced a fast, pyramidal filter bank algorithm [37] for computing the coefficients of the orthogonal wavelet representation; later it was generalized for the biorthogonal case. This algorithm, is in literature usually referred to as the discrete wavelet transform (DWT). The explanation of the algorithm is simple. One can show that the dilation equation (2.23), generalizes to $\varphi_{j+1,k} = \sum_{l \in \mathbb{Z}} h_{l-2k} \varphi_{j,l}$ and
2.4. Discrete Wavelet Transform

\[ S_{j+1,k} = \langle f, \varphi_{j+1,k} \rangle = \langle f, \sum_{l \in \mathbb{Z}} h_{l-2k} \varphi_{j,l} \rangle = \sum_{l \in \mathbb{Z}} \tilde{h}_{2k-l} S_{j,l} \tag{2.25} \]

where \( \tilde{h}_{k} = h_{-k} \) is the mirror filter. The scaling coefficients at the scale \( 2^{j+1} \) are thus computed by convolving the scaling coefficients from the previous, finer scale with the filter \( \tilde{h} \) and downsampling by 2. Similarly, one can show that

\[ W_{j+1,k} = \sum_{l \in \mathbb{Z}} \tilde{g}_{2k-l} S_{j,l} \tag{2.26} \]

One step of the above decomposition (the forward DWT) is depicted in fig. 2.4(a). At the reconstruction (the inverse DWT), one has

\[ S_{j,k} = \sum_{l \in \mathbb{Z}} h_{k-2l} S_{j+1,l} + \sum_{l \in \mathbb{Z}} g_{k-2l} W_{j+1,l} \tag{2.27} \]

![Diagram](image)

Figure 2.4: A fast orthogonal DWT: a decomposition (a), and a reconstruction (b) step. It is a filter bank algorithm: the lowpass and the highpass filters are followed by the down sampling by 2; at the reconstruction the up sampling by 2 precedes the filtering.

which can be interpreted as up-sampling (by introducing a zero between each two points) followed by filtering and summation of the filtered outputs in fig. 2.4(b). For an n-sample vector, the algorithm requires \( O(n) \) operations and is faster than the FFT, which has complexity \( O(n \log n) \). In the first step of the DWT decomposition the scaling coefficients are approximated by the input data samples. Specific wavelets, coiflets, [5], were designed to make the corresponding error negligible. In image processing, the error is usually neglected for other wavelets as well. The reason is that if the sampling interval is sufficiently small [42] then physical measurements are good approximations of wavelet scaling coefficients. Another possibility is to prefilter the samples [43] before computing the wavelet transform.
2.4.5 DWT in two dimensions

The MRA model from section 2.4.2 can be generalized to any positive dimension $n > 0$. Here we address the conventional separable two dimensional (2D) DWT [37]. Non-separable decompositions are described in [44]. In the separable 2D case, one can show [37], that the detail spaces of the (bi)-orthogonal MRA are spanned by the shifts and dilations of the tree “wavelets”: $\Psi^{LH}(x,y) = \varphi(x)\psi(y)$, $\Psi^{HL}(x,y) = \psi(x)\varphi(y)$ and $\Psi^{HH}(x,y) = \psi(x)\psi(y)$ The fast algorithm is a straightforward extension of the one in section 2.4, where the filter banks are applied successively to the rows and to the columns of an image. A decomposition step is shown in fig. 2.5(a), and a usual representation of the frequency subbands in fig. 2.5(b).

![Diagram of 2D DWT](image)

Figure 2.5: Two dimensional DWT. A decomposition step (a) and the usual organization of the subbands (b).

The DWT of an image yields fairly well decorrelated wavelet coefficients. However, these coefficients are not independent. One can see this in fig. 2.6; it shows that large-magnitude coefficients tend to occur near each other within subbands, and also at the same relative spatial locations in subbands at adjacent scales and orientations, as noted, e.g., in [45]. The positions of the large wavelet coefficients indicate image edges, i.e., the DWT has an edge detection property. (The use of wavelets for edge detection is mentioned in section 2.4.9). Fig. 2.6 also illustrates the sparsity of the DWT of images, which makes it in particular suitable for image coding and compression [46], [47]. The 2D DWT is extensively used for image denoising [26], [48]-[53] as well, but it suffers from limitation, as the next section explains.
2.4. Discrete Wavelet Transform

![Image](image.png)

Figure 2.6: An image (a) and its 2D DWT (b). Black pixels denote large magnitude wavelet coefficients.

2.4.6 Improvement over the limitations of DWT in denoising

A disadvantage of the DWT is that, in contrast to the CWT, this decimated representation is not invariant under translation. The lack of shift invariance makes it unsuitable for pattern recognition [28] and also limits the performance in denoising [54]. The latter is perhaps more clear from the viewpoint of the lack of redundancy: the redundancy of a representation, in general, helps to better estimate a signal from its noisy observation. In this respect, two approaches are common in wavelet based image denoising:

1. Cycle spining proposed in [54]: one averages denoising results of several cyclically shifted image versions

2. Denoising in a non-decimated wavelet representation.

There is a slight confusion in literature regarding the two: some authors [4], [55] refer to the first approach as using the redundant discrete wavelet transform (RDWT), while this notion is commonly and more naturally associated with the non-decimated transform. Even though the two approaches are sometimes, [56], regarded as “equivalent”, they should not be mixed up. The framework in which one works is quite different than the other. In the first (cycle-spining) case one removes noise from a decimated and thus decorrelated set of coefficients - the i.i.d. models are largely justified and thus the derivation of a MMSE estimator is
facilitated. In the second (non-decimated) case, the realistic statistical modelling of coefficients is far more difficult. There are however other advantages. In the first place, the interscale comparisons between wavelet coefficients yielding the detection of useful image features are largely facilitated.

2.4.7 Non-decimated discrete wavelet transform

In an undecimated wavelet transform, a signal is represented with the same number of wavelet coefficients at each scale. These coefficients are samples of the continuous wavelet transform at all integer locations at each dyadic scale $W_f(2^{-j}, k)_{j,k} \in \mathbb{Z}^2$. Such a redundant representation results from decomposing a signal into a family of wavelets $2^{j/2} \Psi(2^{-j}(x - k))$, also abbreviated by $\Psi_{j,k}(x)$ for notation simplicity. It can be represented by equation (2.28).

$$f(x) = \sum_{j=0}^{\infty} \sum_{k=-\infty}^{\infty} \langle f, \tilde{\Psi}_{j,k} \rangle \Psi_{j,k}(x)$$

but now the wavelets $\Psi_{j,k}(x)$ are not linearly independent. They do not constitute a basis but a frame. If the frame is tight, $\Psi(x) = \Psi(x)$. With respect to wavelet bases, the choice of a wavelet is less restrictive [27].

![Figure 2.7: An example of the redundant wavelet frame decomposition in three resolution levels. From left to right are represented lowpass images and detail images in LH, HL and HH subbands, respectively.](image)

A non-decimated wavelet transform approaches translation invariance and is therefore also called the Stationary Wavelet Transform. It is computed with the a
2.4. Discrete Wavelet Transform

trous algorithm [6], [57]. If we denote by \( h^t \) a filter where \( 2^t-1 \) zeroes are inserted between each two coefficients of the filter \( h \), then this algorithm is

\[
S_{j+1,k} = \sum_{l=-\infty}^{\infty} \tilde{h}^t_{k-l} S_{j,l}
\]

(2.29)

\[
W_{j+1,k} = \sum_{l=-\infty}^{\infty} \tilde{g}^t_{k-l} S_{j,l}
\]

(2.30)

\[
S_{j,k} = \sum_{l=-\infty}^{\infty} \tilde{h}^t_{k-l} S_{j+1,l} + \sum_{l=-\infty}^{\infty} \tilde{g}^t_{k-l} W_{j+1,l}
\]

(2.31)

To understand the inserting zeroes, equations (2.29) and (2.30) should be a consistent extension of the forward DWT in equations (2.25) and (2.26). All the DWT coefficients should reappear in this new transform. To get those coefficients among the redundant set, we have to skip the “extra” ones before applying convolutions. A pictorial explanation is in [43].

The non-decimated wavelet transform of an image is illustrated in fig. 2.7. It requires more calculations and calls for bigger memory than the decomposition into wavelet bases. However, it enables a better denoising quality. Also, the nonredundant representation is usually implemented for discrete signals or images whose size is a power of two, because the number of coefficients is halved in subsequent resolution levels. In contrast, the non-decimated transform is equally implemented for arbitrary input sizes.

2.4.8 Choosing a wavelet for image denoising

Important questions are which wavelet(s) to choose for image denoising and why? Firstly, denoising in general is facilitated in a sparse representation (i.e., one with relatively few non-negligible coefficients). Secondly, when choosing a wavelet for imaging applications, its influence on the visual quality should be taken into account as well. One goal is thus to produce as many as possible wavelet coefficients \( \langle f, \tilde{\Psi}_{j,k} \rangle \) that are close to zero. Apart from the regularity of the analyzed signal, this depends on the number of vanishing moments \( \tilde{N}_v \) as described in equation (2.5) and on the support size \( k \) of the analysis wavelet \( \tilde{\Psi}(x) \): \( \tilde{N}_v \) should be as large as possible and \( k \) as small as possible. \( \langle f, \tilde{\Psi}_{j,k} \rangle \) is large only if a signal discontinuity
2.4. Discrete Wavelet Transform

Figure 2.8: Examples of Daubechies $dbNv$ and Symlet $symNv$ scaling functions and wavelets.

is located within the support of $\Phi(x)_{j,k}$. Moreover, any polynomial component of $f(x)$ of a degree less than $\hat{N}_v$ lies in the complementary scaling space [58], yielding thus zero wavelet coefficients.

Figure 2.9: An example of spline biorthogonal wavelets and scaling functions of compact support of [41], with $N_v = 7$, $\hat{N}_v = 3$.

For the visual quality of images, the regularity and the symmetry of the synthesis wavelet $\Psi(x)$ are important. When reconstructing a signal from its (thresholded or quantized) wavelet coefficients, an error $\epsilon$ added to a coefficient $W_{j,k}$ will add the wavelet component $\epsilon \Psi_{j,k}(x)$ to the reconstructed signal. If $\Psi(x)$ is smooth then this error is smooth as well, and if we work with images it is less visible. The regularity of a wavelet usually increases with the number of its vanishing moments, even though this is not guaranteed in general [6]. The preference of symmetrical wavelets is due to the fact that our visual system is more tolerant of symmetric errors than asymmetric ones [5].
Having in mind the above requirements, now we address some of the wavelets at our disposal. In the orthogonal case, it is difficult to achieve a large number of vanishing moments and a small support size at the same time. The theoretical limit is \( K = 2N_y - 1 \) and is achieved in the Daubechies wavelets [5] usually denoted as \( dbN_y \). The shortest member of this group \( db1 \) is in fact the Haar wavelet. Two others, \( db2 \) and \( db8 \), are illustrated in fig. 2.8. It is noted that the lack of smoothness in the shorter one and the asymmetry of both. Except the Haar wavelet, compactly supported orthogonal wavelets cannot be symmetrical. This comes from the properties of filter banks as mentioned in [40]. The least asymmetrical compactly supported orthogonal wavelets, also constructed by Daubechies, are the symlets \( symN_y \) [5]; (an example, \( sym8 \), is illustrated in fig. 2.8). With biorthogonal wavelets [41], illustrated in fig. 2.9, the desired properties are easier to meet. Firstly, the compact support does not contradict the symmetry. \( \Psi(x) \) and \( \hat{\Psi}(x) \) have equal support size \( k \), but in general a different number of vanishing moments and a different regularity (one can "compromise" the properties of the analysis and the synthesis wavelet).

Figure 2.10: The cubic spline smoothing function \( \theta(x) \) (left) and its first derivative, the quadratic spline wavelet [59] \( \Psi(x) \) (right).

The non-decimated wavelet transform gives even more freedom in choosing a wavelet [6]. The number of vanishing moments is less important, since the sparsity is not the main argument now. In this framework, for image denoising, the quadratic spline wavelet [59] as in fig. 2.10, is often used ([60], [61]), because it is short and smooth.

This wavelet will be used extensively. Finally, as in most of the reported works in image denoising, the mother wavelets \( db8 \) and \( sym8 \) are used, because of their impressive performance, the same will be used in this thesis works for case
2.4. Multiscale edge detection

A specific type of the dyadic wavelet transform, introduced in [59] acts as a multiscale extension of the Canny [62] edge detector. The Canny algorithm detects points of sharp variation in an image \( f(x, y) \) by calculating the modulus of its gradient vector

\[
\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix}
\]  
\[ (2.32) \]

A point \((x_0, y_0)\) is defined as an edge if the modulus of \( \nabla f(x, y) \) is locally maximum at \((x_0, y_0)\) in the direction parallel to \( \nabla f(x_0, y_0) \). The multiscale version [59] of this edge detector uses two wavelets that are the partial derivatives of a smoothing function \( \theta(x, y) \)

\[
\Psi^1(x, y) = \left[ \frac{\partial \theta(x, y)}{\partial x} \right]
\]
\[ (2.33) \]

\[
\Psi^2(x, y) = \left[ \frac{\partial \theta(x, y)}{\partial x} \right]
\]
\[ (2.34) \]

The dyadic wavelet transform \( Wf(2^j, u, v) \) now consists of the two components \( (W^1f(2^j, u, v), W^2f(2^j, u, v)) \), where \( W^n f(2^j, u, v) = \langle f, \Psi^n_{2^j, u, v} \rangle \) as in equation (2.9). The local maxima of the modulus

\[
Mf(2^j, u, v) = \sqrt{(W^1f(2^j, u, v))^2 + (W^2f(2^j, u, v))^2}
\]
\[ (2.35) \]

in the direction of the angle \( Af(2^j, u, v) \), defined by

\[
\tan (Af(2^j, u, v)) = \frac{W^2f(2^j, u, v)}{W^1f(2^j, u, v)}
\]
\[ (2.36) \]

are exactly the Canny's edges of the smoothed image \( (f \ast \theta_{2^j})(x, y) \), where \( \theta_{2^j}(x, y) = \left( \frac{1}{2^j} \right) \theta \left( \frac{x}{2^j}, \frac{y}{2^j} \right) \). From these multiscale edges only, the algorithm of [59] computes an image approximation that is visually identical to the original one. An important application is in image denoising [33].

A particular choice, leading to a fast implementation, is when the smoothing function is separable. The two components of the wavelet transform are then
obtained by convolving only the rows and only the columns of the image with a 1D wavelet $\Psi(x)$, respectively. In [59], the quadratic spline wavelet from fig. 2.10 was proposed for this purpose. It has a short support and is continuously differentiable, and its dual is also a spline. This wavelet is the first derivative of the cubic spline $\theta(x)$ in fig. 2.10 which results from three convolutions of the box function defined on $[0, 1]$ with itself.

In practice, the above transform is implemented, using a 2D version of the a trous algorithm in section 2.4.7, yielding the discretized wavelet coefficients $W^j_{j,m,n} = W^j f(2^j, m, n)$ and the corresponding scaling coefficients $S^j_{j,m,n}$ as shown in fig. 2.11.

Figure 2.11: Decomposition of the non-decimated wavelet transform with two orientation subbands.

Figure 2.12: Reconstruction of the non-decimated wavelet transform with two orientation subbands.

It is thus a specific 2D non-decimated wavelet transform, with two instead of the classical three orientation subbands. The reconstruction process now involves two bandpass filters. An example illustrating the lowpass and the detail images is shown in fig. 2.12.
2.4.10 On some extensions of the classical wavelet scheme

Numerous extensions of the "classical wavelet scheme" exist, which will not be ad­
dressed in this thesis. A nice overview is given in [63]. The idea about the wavelet
packets is given in [64], [65], which yields more flexible and signal-adapted [66]
representations at the cost of slightly more complex algorithms; local trigonomet­
ric bases [67]; multiwavelets [29], [68], where instead of only one, several mother
wavelets are used in order to combine their useful properties; second generation
wavelets, using, e.g., the Sweldons' lifting scheme [69], [70], where the idea of
translation and dilation is abandoned, and the wavelet construction is adapted to
irregular samples, weights, and manifolds [71], [72].

Recent trends like ridgelets [73], curvelets [74] and bandelets [75] appear as
competitors to wavelets in image processing.

2.5 Image Denoising under Wavelet Domain

This section introduces the use of wavelets in image denoising. In denoising there
is always a trade-off between noise suppression and preserving actual image dis­
continuities. To remove noise without excessive smoothing of important details,
a denoising algorithm needs to be spatially adaptive. The wavelet representa­
tion, due to its sparsity, edge detection and multiresolution properties, naturally
facilitates such spatially adaptive noise filtering. A common procedure is:

1. Compute the DWT or non-decimated wavelet transform;

2. Remove noise from the wavelet coefficients and

3. Reconstruct the denoised image.

The scaling coefficients are usually kept unchanged, unless in certain cases of signal
dependent noise.

2.5.1 Noise in the wavelet domain

In the wavelet domain, the most essential information in a signal is compressed
into relatively few, large coefficients, which coincide with the areas of major spatial
activity (edges, corners, peaks, ...) in the image. On the other hand, noise is spread over all coefficients, and at typical noise levels (that are of practical importance) the important coefficients can be well recognized. For an additive model of a discrete image $f$ and noise $\varphi$

$$v = f + \varphi$$

(2.37)

The vector $v$ is the input image. The noise $\varphi$ is a vector of random variables, while the unknown $f$ is a deterministic signal. One usually assumes that the noise has zero mean ($E(\varphi) = 0$), so that the covariance matrix is

$$Q = E[(\varphi - E(\varphi))(\varphi - E(\varphi))^T] = E(\varphi\varphi^T)$$

(2.38)

On its diagonal are the variances $\sigma_l^2 = E(\varphi_l^2)$. If the covariance matrix is diagonal, i.e., if $E(\varphi_l, \varphi_k) = 0$ for $l \neq k$, the noise is uncorrelated and is called white. If all $\varphi_l$ follow the same distribution, they are said to be identically distributed. This implies $\sigma_l^2 = \sigma^2$, for all $l = 1, \ldots, n$. The Gaussian noise is with the probability density

$$p_\varphi(\varphi) = \frac{1}{(2\pi)^{n/2}\sqrt{\det(Q)}} e^{-1/2\varphi^T Q^{-1} \varphi}$$

(2.39)

If Gaussian noise variables are uncorrelated, they are also statistically independent $p_{\varphi_l}(\varphi_l) = \prod_l p_{\varphi_l}(\varphi_l)$. The reverse implication (independent variables are uncorrelated) holds for all densities. A common assumption is that the noise variables are independent, identically distributed (i.i.d.). Most of the methods in this thesis are specifically designed for the case of additive white Gaussian noise, which is often abbreviated as AWGN.

Now, the assumed noise model in the wavelet domain is described formally. Due to linearity of the wavelet transform, the additive model of equation (2.37) remains additive in the transform domain as well:

$$w = y + n$$

(2.40)

where $w = W_d v$ are the observed wavelet coefficients, $y = W_d f$ are the noise-free coefficients, $n = W_d \varphi$ is additive noise, and $W_d$ is an operator that yields the discretized wavelet coefficients. An orthogonal wavelet transform maps the white
noise in the input image into a white noise in the wavelet domain. Under such an orthogonal transform, i.i.d. noise with a variance \( \sigma^2 \) remains i.i.d. with the same variance \( \sigma^2 \).

The situation is slightly more complicated for bi-orthogonal and nondecimated transforms. In general, one can show that if the input noise \( \varphi_i \) is i.i.d., then in the wavelet domain the noise variance depends only on the resolution level and on the subband orientation. In other words, in each detail image \( w^d_j \) noise has a constant variance \( (\sigma^d_{n,j})^2 \), where

\[
(\sigma^d_{n,j})^2 = S^d_j \sigma^2
\]

and \( S^d_j \) is a function of the coefficients of the particular lowpass \( h \) and the highpass \( g \) filters used in the decomposition. Referring to fig. 2.5, to produce \( w^{LH}_j \) and \( w^{HL}_j \) subbands, the lowpass filter \( h \) is for both horizontal and vertical directions applied \( 2j - 1 \) times in total, and the highpass filter \( g \) only once (either horizontally or vertically). The subband \( w^{HH}_j \), results from in total \( 2(j - 1) \) lowpass filtering and 2 highpass filtering. Starting from this, and deriving the second-order cumulants of the wavelet coefficients, one can show [76] that

\[
S^{LH,HL}_j = \left( \sum_k g^2_k \right) \left( \sum_l h^2_l \right)^{2j-1}
\]

and

\[
S^{HH}_j = \left( \sum_k g^2_k \right) \left( \sum_l h^2_l \right)^{2(j-1)}
\]

which holds for both decimated and non-decimated case. In the same way, for the specific decomposition with two orientation subbands \( d \in x, y \) from section 2.4.2, one can derive

\[
S^x_j = S^y_j = \left( \sum_k g^2_k \right) \left( \sum_l h^2_l \right)^{2(j-1)}
\]

The above equations show that for common AWGN case, one can easily compute the noise variance in the wavelet domain, provided that the input noise variance \( \sigma^2 \) is known.

### 2.5.2 Noise variance estimation

In some applications of image denoising, the value of the input noise variance \( \sigma^2 \) is known, or can be measured based on information other than the corrupted data.
If this is not the case, one has to estimate it from the input data, eliminating the influence of the actual signal. Wavelet based methods commonly use the highest frequency subband of the decomposition for this purpose. In the DWT of an image, the $HH_1$ subband contains mainly noise. A robust estimate $\hat{\sigma}$ is obtained with a median measurement, which is highly insensitive to isolated outliers of potentially high amplitudes. In [49], it was proposed

$$\hat{\sigma} = \frac{\text{Median}(|w_i^{HH}|)}{0.6745}$$

The motivation is: if $u_n$ are $N$ independent Gaussian random variables of zero mean and variance $\sigma^2$, then $E(\text{Median}|u_n|_{0\leq n<N}) \approx 0.6745\sigma$. One often denotes $\text{Median}(|w|) = \text{MAD}(w)$, where MAD stands for Median Absolute Deviation. The estimate in equation (2.45) is commonly used in image denoising [54], [77] and we use it in this work as well. Other ways of estimating $\sigma$ in the wavelet domain is explained in [43].

### 2.5.3 Problems in Wavelet Domain

The wavelet decomposition and thresholding techniques have some underlying disadvantages. For instance, the estimated wavelet coefficients obtained after hard-thresholding are not continuous at the threshold point $\lambda$, which may lead to the oscillation of the reconstructed signal. In the soft-thresholding case, there are deviations between image coefficients and thresholded coefficients, which directly influence the accuracy of the reconstructed signal.

Retention of the edges is also a problem here because of smoothing. Different edge detection algorithms may be used to extract the contour feature of cell images.

Bilateral filter may help to achieve the objective of edge retention.

### 2.6 Bilateral Filtering

The bilateral filter [7] is a nonlinear filter that does spatial averaging without smoothing edges. It has shown to be an effective image denoising technique. An important issue with the application of the bilateral filter is the selection of the
2.6. Bilateral Filtering

filter parameters, which affect the results significantly. Another research interest of bilateral filter is acceleration of the computation speed.

Bilateral filter [7] is firstly presented by Tomasi and Manduchi in 1998. The concept of the bilateral filter was also presented in [78] as the SUSAN filter and in [79] as the neighborhood filter. It is worth mentioning that the Beltrami flow algorithm is considered as the theoretical origin of the bilateral filter [80]-[82], which produces a spectrum of image enhancing algorithms ranging from the 2L linear diffusion to the 1L non-linear flows. The bilateral filter takes a weighted sum of the pixels in a local neighborhood; the weights depend on both the spatial distance and the intensity distance. In this way, edges are preserved well while noise is averaged out. Mathematically, at a pixel location \( x \), the output of a bilateral filter is calculated as follows

\[
\hat{I}(x) = \frac{1}{C} \sum_{y \in N(x)} e^{-\frac{(x-y)^2}{\sigma^2}} e^{-\frac{(I(y) - I(x))^2}{2\sigma_f^2}} I(y)
\]  

(2.46)

where \( \sigma_d \) and \( \sigma_r \) are parameters controlling the fall-off of weights in spatial and intensity domains, respectively, \( N(x) \) is a spatial neighborhood of pixel \( I(x) \), and \( C \) is the normalization constant:

\[
C = \sum_{y \in N(x)} e^{-\frac{(x-y)^2}{\sigma^2}} e^{-\frac{(I(y) - I(x))^2}{2\sigma_f^2}}
\]  

(2.47)

Fig. 2.13 shows the illustration of 1D bilateral filter. The top right image is the input noisy signal. The top left image shows the intensity Gaussian while the middle image shows the special Gaussian. The bilateral response is shown at the bottom.

Another parameter during the running of the bilateral filter is the window size of how many pixels should be computed on time. The window size is related to the spatial Gaussian. Basically, based on the property of the Gaussian distribution, window size should be around 2 to 3 times the standard deviation of the Gaussian, since when it’s over 3 times sigma, the output of Gaussian almost equals to zero.

In some reported works, it is shown that the bilateral filter is identical to the first iteration of the Jacobi algorithm (diagonal normalized steepest descent) with a specific cost function. Elad et al. [83] related the bilateral filter with the anisotropic diffusion.
objective quality or distortion assessment approaches are used. These are mean
squared error (MSE) and peak signal to noise ratio (PSNR). Both are mathemat­
ically defined measures.

The second criterion considers human visual system (HVS) characteristics in
an attempt to incorporate perceptual quality measures. In practice, however, the
HVS is more tolerant to a certain amount of noise than to a reduced sharpness.
Moreover, the visual quality is highly subjective [84], and difficult to express ob-
jectively. In addition, the HVS is also highly intolerant to various artifacts, like
"blips" and "bumps" in the reconstructed image [85].

The importance of avoiding those artifacts is also meaningful. For instance,
in certain applications (like astronomy, or medicine) such artifacts may cause
wrong data interpretations. Unfortunately, none of these complicated objective
metrics in the literature have shown any clear advantage over simple mathematical
measures such as MSE and PSNR under strict test conditions and different image
distortion environments.

Good MSE or PSNR does not imply that the visual quality of the image is
good. To overcome this problem Image Quality Index (IQI) [86] is considered as
the third parameter for judging the quality of denoised images.

Thus, to evaluate the existing and proposed algorithms experimented in this
thesis works, three objective assessment tools, MSE, PSNR, and Image Quality
Index (IQI), will be applied to obtain fair and complete performance evaluation.
These three image quality assessment methods are briefly introduced as follows:

1. Mean Squared Error (MSE)

Let \( f = f_i | i = 1,2, \ldots, M \) and \( g = g_i | i = 1,2, \ldots, M \) be the original and the
denoised images respectively. The ultimate objective of image denoising is
to produce an estimation of the original image \( f \) and noise free image \( g \), which
approximate it best, under given evaluation criteria. As in any estimation
problem, an important objective goal is to minimize the error of the result
as compared to the uncorrupted data. In this respect, a common criterion
is minimizing the mean squared error (MSE), that is defined in equation
\((2.48)\).

\[
MSE = \frac{1}{M} \sum_{i=1}^{M} (g_i - f_i)^2
\]  

(2.48)
2.7. Performance evaluation in image denoising

where $M$ is the number of elements in the image. For example, if we want to find the MSE between the denoised and the original image, then we would take the difference between the two images pixel-by-pixel, square the results, and average the results.

2. Peak Signal to Noise Ratio (PSNR)

In image processing, another common performance measurement is the peak signal to noise ratio (PSNR), which is for grey scale images defined in dB in equation (2.49) as

$$PSNR = 10 \log_{10} \left( \frac{(2^n - 1)^2}{MSE} \right)$$  \hspace{1cm} (2.49)

where $n$ is the number of bits per symbol.

3. Image Quality Index (IQI)

Wang and Bovik [86] proposed an index which is designed by modelling any image distortion as a product of three different factors: loss of correlation, luminance distortion, and contrast distortion. Although the new index is mathematically defined and does not explicitly employ the human visual system model, experiments on various image distortion types show that it exhibits surprising consistency with subjective quality measurement. The image Quality Index (Q) is defined in equation (2.50).

$$Q = \frac{\sigma_f \sigma_g}{\sigma_f \sigma_g} \frac{2 \bar{f} \bar{g}}{\bar{f}^2 + \bar{g}^2} + \frac{2 \sigma_f \sigma_g}{\sigma_f^2 + \sigma_g^2}$$  \hspace{1cm} (2.50)

where

$$\bar{f} = \frac{1}{M} \sum_{i=1}^{M} f_i,$$

$$\bar{g} = \frac{1}{M} \sum_{i=1}^{M} g_i,$$

$$\sigma_f^2 = \frac{1}{M-1} \sum_{i=1}^{M} (f_i - \bar{f})^2,$$

$$\sigma_g^2 = \frac{1}{M-1} \sum_{i=1}^{M} (g_i - \bar{g})^2,$$

$$\sigma_{fg}^2 = \frac{1}{M-1} \sum_{i=1}^{M} (f_i - \bar{f})(g_i - \bar{g}).$$

The first component of equation (2.50) is the correlation coefficient between $f$ and $g$, which measures the degree of linear correlation between $f$ and $g$ and its dynamic range is [-1,1]. The second component, with a value range of [0,1], measures how close the mean luminance is between $f$ and $g$. $\sigma_f$
and $\sigma_g$ can be viewed as estimate of the contrast of $f$ and $g$, so the third component with a value range of $[0,1]$ measures how similar the contrasts of the images are.

Thus, $Q$ can be rewritten as

$$Q = \frac{4\sigma_f \bar{f} \bar{g}}{(\sigma_f^2 + \sigma_g^2)(\bar{f}^2 + \bar{g}^2)}$$

(2.51)

The dynamic range of $Q$ is $[-1,1]$. The best value is 1 which means that the tested image is exactly equal to the original image. The best value is achieved, if and only if, $g_i = f_i$ for all $i = 1, 2, M$. The lowest value of -1 occurs when $g_i = 2\bar{f} - f_i$ for all $i = 1, 2, M$.

2.8 Conclusion

Brief introduction to different kinds of noises together with different denoising models are modestly described with special reference to multiresolution analysis. Advantages and limitations of the models in denoising images are also addressed and discussed.