Nonparametric Neighbourhood Based Multiscale Model for Image Analysis and Understanding

by

Aanchal Jain

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Systems Design Engineering

Waterloo, Ontario, Canada, 2012

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Image processing applications such as image denoising, image segmentation, object detection, object recognition and texture synthesis often require a multi-scale analysis of images. This is useful because different features in the image become prominent at different scales. Traditional imaging models, which have been used for multi-scale analysis of images, have several limitations such as high sensitivity to noise and structural degradation observed at higher scales. Parametric models make certain assumptions about the image structure which may or may not be valid in several situations. Non-parametric methods, on the other hand, are very flexible and adapt to the underlying image structure more easily. It is highly desirable to have efficient non-parametric models for image analysis, which can be used to build robust image processing algorithms with little or no prior knowledge of the underlying image content. In this thesis, we propose a non-parametric pixel neighbourhood based framework for multi-scale image analysis and apply the model to build image denoising and saliency detection algorithms for the purpose of illustration. It has been shown that the algorithms based on this framework give competitive results without using any prior information about the image statistics.

Acknowledgements

I would like to thank my supervisors Prof. Paul Fieguth and Prof. Alexander Wong for their constant guidance and support in my research and in the preparation of this document. Their insights and inputs have been invaluable. The knowledge and skills I have developed by working under their supervision will help me in my future endeavours, and I am grateful to them for the same. I would also like to thank my readers, Prof. Hamid Tizhoosh and Prof. Andrea Scott for reviewing my thesis.

Dedication

This is dedicated to the Vision and Image Processing (VIP) lab at the University of Waterloo, a lab which I am truly glad to be a part of.

Table of Contents

List of Figures

Chapter 1

Introduction

In this thesis, we attempt to study and compare different multi-scale frameworks for image analysis and propose our own non-parametric neighbourhood based image analysis model. Based on this model, image saliency detection and image denoising algorithms are developed and their results are discussed.

1.1 Motivation

Parametric multi-scale image analysis methods such as scale space model (explained in [1]) have several shortcomings (discussed in [2]) such as degradation of the image at higher scales and loss of structural details in the multi-scale representations. The parametric methods suffer from these limitations because they use fixed parameters for deriving the multi-scale representations without taking into account the underlying image statistics. This may or may not be valid in different situations. For example, in the scale space model, the given image is convolved with a series of Gaussian kernels with varying standard deviations, with the standard deviations behaving as the fixed scale parameters. This representation gives highly blurred images at the coarser scales (large values of standard deviations) leading to loss of structural details at these scales. It is desirable to have image analysis methods which adapt themselves to the underlying image content without requiring any prior information or assumptions about the image. We attempt to address this problem in our thesis by presenting an alternate framework based on non-parametric neighbourhood based methods. These methods are flexible and adapt themselves to the underlying image content more easily. It will be shown through experimental results that

the proposed techniques are very efficient for multi-scale image modelling and have also been shown to be robust to various noise characteristics.

1.2 Thesis Outline

In chapter 2, we present a brief summary of the conventional approaches which have been used in the recent years for multi-scale image understanding and analysis. We also give a theoretical overview of the wavelet theory, both decimated and undecimated wavelets and how they serve as a powerful tool in multi-scale image modelling. In chapter 3, we describe the development of the non-parametric neighbourhood based framework. In chapter 4, we discuss the application of the approach developed to saliency detection in natural images, including experimental results and performance comparison with other methods. In chapter 5, we discuss the application of the approach to image denoising. In chapter 6, we conclude and give some final remarks about the method including the possible further applications of the method proposed and the advantages presented by the method.

Chapter 2

Background

In this section, we will be discussing some multi-scale imaging models which have been proposed in recent years [1, 2, 3, 4, 5] and have been used in image processing applications like image denoising, texture synthesis, image segmentation and image saliency detection. In section 2.1, we will discuss parametric models, their development, advantages and disadvantages. In section 2.2, we will talk about the more recent non-parametric models and their advantages over parametric models.

2.1 Parametric Models for Image Processing

Parametric methods [1, 6, 7] require assumptions to be made about the image data which are implicitly captured in the "parameters" used to derive these models. For example in the scale space model proposed in [1], the images are convolved with a series of Gaussian kernels with varying standard deviations, with the standard deviations behaving as the scale parameters. This section describes the development of parametric methods which are popular and have been used in several image processing applications [8, 9, 10]. In section 2.1.1, we discuss the scale space models and in section 2.1.2, we discuss multi-scale models based on the wavelet decomposition of images.

2.1.1 Scale Space Model (linear)

In [1, 6, 7], linear scale space models have been discussed which can be used for describing the structural features in an image at different scales. In this technique, the original signal

is convolved with a series of Gaussian kernels with a set of different standard deviations. The scale-parametric representation is given by

$$L_\sigma(x) = f(x) * g(x, \sigma) \tag{2.1}$$

where $f(x)$ is the original image, $g(x, \sigma)$ is the Gaussian kernel with standard deviation σ, and L_σ is the image representation corresponding to the scale parameter σ. The Gaussian convolution of a signal $f(x)$ is defined as

$$L_\sigma(x) = f(x) * g(x, \sigma) = \int_{-\infty}^{\infty} f(u) \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-u)^2}{2\sigma^2}} du \tag{2.2}$$

The (x, σ) plane is called the scale space plane [1], and each value of σ gives us a Gaussian smoothed version of the original image $f(x)$, the amount of smoothing increasing with σ. This set of Gaussian smoothed functions is called the scale space image [1].

In this case, the standard deviation of the Gaussian kernels behaves as the scale parameter, which proves to be an important parameter for the description of the given image as different features in the image are observed at different scales, giving us an idea about the underlying structure of the image. Large values of σ give a smoothed signal and small values give signals with more local structural variations. Gaussian kernels are used for convolution because of their useful properties. The Gaussian kernels satisfy a "well-behavedness" criterion (explained in [1]), i.e., they are symmetric and strictly decreasing about the mean, which implies that the weighting assigned to pixels decreases smoothly with distance from the central pixel.

The approach discussed above uses fixed parameters, independent of the image structure, hence local structural information in the signal may be ignored. The method also gives high smoothing at higher (coarser) scales. This gives multi-scale image models where we observe structural degradation of images, particularly at coarser scales. The method has also been shown to be sensitive to noise in [2]. An example of the application of this model has been shown in Figure 2.1. The first row of the figure shows the scale-space multi-scale representation for a noiseless image and the second row shows the same for an image with additive Gaussian noise ($\sigma = 0.3$). We can see in this example that there is more smoothing observed at higher scales, which results in the blurring of edges. At these scales, the spatial location of the edges does not sharply coincide with their corresponding locations in the original image. It is important to observe here that the linear scale space model is sensitive to noise (as we observe more degradation of the scale-space structure in the noisy image). These factors motivate us to study other non-parametric models for image analysis. In section 2.2.1, we discuss a non-linear scale space approach which takes the local image structure into account while deriving the multi-scale model for the image.

2.1.2 Wavelet Based Multi-scale Models

In recent years, wavelets have become an important part of transform based image processing [11, 12, 13] and have been used in applications like image compression [14, 15], image segmentation [16, 17] and image denoising [9, 10]. Wavelets are the foundation of multi-resolution theory [11, 12], which makes them important for the analysis of images at multiple scales. In chapter 4 and chapter 5, where we propose saliency detection and image denoising algorithms based on our proposed framework, we have used wavelet decompositions extensively. This chapter explains the mathematics behind the wavelet theory and serves as a foundation for the subsequent chapters.

We define an inner product between two real signals $f(x)$ and $g(x)$ as

$$<f, g> = \int_{-\infty}^{\infty} f(x)g(x)dx \tag{2.3}$$

Let V_j denote a vector space, which is a set of all possible signals at resolution 2^j. Then $[V_j]_{j\in Z}$ denotes a multi-resolution vector space [11]. The multi-resolution vector space satisfies the causality property, i.e., $\forall j \in Z$, $V_j \subset V_{j+1}$. We define the orthonormal basis for a vector space V_j, which characterizes the multi-resolution vector space sequence $[V_j]_{j\in Z}$. For the vector space sequence $[V_j]_{j\in Z}$, there exists a unique scaling function $\phi(x)$ such that for any $j \in Z$,

$$\phi^j(x) = \sqrt{2^j}\phi(2^j x) \tag{2.4}$$

then $[\phi^j(x - 2^{-j}n)]_{n\in Z}$ is an orthonormal basis of vector space V_j. The proofs for these derivations are given in [11]. From the scaling function $\phi(x)$, an orthonormal basis can be derived for V_j, for any $j \in Z$, by scaling and translating $\phi(x)$. An approximation of the signal $f(x)$ at resolution of 2^j is essentially a projection of the signal $f(x)$ on the vector space V_j, which can be defined in terms of the orthonormal basis vectors of the vector space V_j. Let $f^{V_j}(x)$ denote a projection of signal $f(x)$ on the vector space V_j

$$f^{V_j}(x) = \sum_{n=-\infty}^{\infty} < f, \phi_n^j > \phi_n^j(x) \tag{2.5}$$

Thus $f^{V_j}(x)$, i.e., the approximation at resolution 2^j, is characterized by the coefficients $[< f, \phi_n^j >]_{n\in Z}$. $< f, \phi_n^j >$ simplifies to

$$< f, \phi_n^j >= \big(f(x) * \phi^j(x)\big)(2^{-j}n) \tag{2.6}$$

This is equivalent to the convolution of $f(x)$ with $\phi^j(-x)$ followed by sampling with a factor of 2^{-j}. Let A_j denote the approximation coefficients for the resolution 2^j. A_j can be defined as

$$A_j = \left[\sqrt{2^j}\big(f(x) * \phi^j(-x)\big)(2^{-j}n)\right]_{n\in Z} \tag{2.7}$$

$\left[\phi_n^{j+1}\right]_{n\in Z}$ defines an orthonormal basis for the vector space V_{j+1} and $\left[\phi_n^j\right]_{n\in Z}$ defines an orthonormal basis for the vector space V_j. Now $V_j \subset V_{j+1}$. This implies that for any $p(x) \in V_j$, $p(x) \in V_{j+1}$ and hence $p(x)$ can be defined in terms of the orthonormal basis of the vector space V_{j+1}.

$$p(x) = \sum_{k=-\infty}^{\infty} < p, \phi_k^{j+1} > \phi_k^{j+1}(x) \tag{2.8}$$

This is a useful property as it implies that an approximation at a lower resolution, i.e., A_j can be derived from an approximation at a higher resolution, i.e., A_{j+1}. The approximation A_j at resolution 2^j is given by an orthogonal projection of the signal on the vector space V_j. The detail at resolution 2^j is given by an orthogonal projection of the signal on O_j where O_j is orthogonal to V_j [11] such that

$$O_j \oplus V_j = V_{j+1} \tag{2.9}$$

The projection of $f(x)$ on O_j is denoted by $f^{O_j}(x)$. This projection describes the differences between the approximations at resolutions 2^j and 2^{j+1}. Let $\psi(x)$ be a wavelet function. Then $\left[\psi^j(x - 2^{-j}n)\right]_{n\in Z}$ is an orthonormal basis of O_j.

$$f^{O_j}(x) = \sum_{n=-\infty}^{\infty} < f, \psi_n^j > \psi_n^j(x) \tag{2.10}$$

In a way similar to the one described for the approximation coefficients A_j above, the detail coefficients D_j at resolution 2^j can be defined as

$$D_j = \left[\sqrt{2}\big(f(x) * \psi^j(-x)\big)(2^{-j}n)\right]_{n\in Z} \tag{2.11}$$

The wavelet representation is given by

$$\left[A_J, (D_j)_{J\le j\le -1}\right] \tag{2.12}$$

where L is the number of levels being considered in the decomposition and $J = -L$. In the wavelet decomposition the approximation coefficients are defined only for the lowest resolution and for all other resolutions, detail coefficients are defined to describe the difference between the resolution levels. A_j and D_j can be obtained from A_{j+1} by filtering

and downsampling by a factor of 2, as shown in Figure 2.2. Filtering A_{j+1} with a band pass filter $\widetilde{G}$ and downsampling gives us the detail coefficients at resolution 2^j, i.e., D_j. Filtering A_{j+1} with a low pass filter $\widetilde{H}$ and downsampling gives the approximation coefficients at resolution 2^j, i.e., A_j. The discrete filters $\widetilde{G}$ and $\widetilde{H}$ can be defined in terms of the wavelet function $\psi(x)$ and the scaling function $\phi(x)$. H is a discrete filter with the impulse response

$$h(n) = \frac{1}{\sqrt{2}} < \phi_0^{-1}, \phi_n^0 > \tag{2.13}$$

and $\widetilde{H}$ has an impulse response $\widetilde{h}(n) = h(-n)$. G is a discrete filter with the impulse response

$$g(n) = \frac{1}{\sqrt{2}} < \psi_0^{-1}, \phi_n^0 > \tag{2.14}$$

and $\widetilde{G}$ has an impulse response $\widetilde{g}(n) = g(-n)$. H and G are quadrature mirror filters [11]. Just as A_{j+1} can be decomposed to obtain A_j and D_j, A_{j+1} can be resynthesized from A_j and D_j as shown in Figure 2.3.

The wavelet decomposition model discussed above can be extended to two dimensions using separable filters [11]. We now work on a two dimensional signal $f(x, y)$. $[V_j]_{j \in Z}$ is still a series of vector spaces, but in this context contains two dimensional signals instead of one dimensional signals. In the 2-D case, the scaling function is defined as

$$\Phi(x, y) = \phi(x)\phi(y) \tag{2.15}$$

where $\phi(x)$ is a 1-D scaling function. In the 2-D case, the approximation coefficients (A_j) are defined as

$$A_j = \left[< f(x, y), \phi_n^j(x)\phi_m^j(y) > \right]_{(n,m) \in Z^2} \tag{2.16}$$

We have a 1-D wavelet function $\psi(x)$. The orthonormal basis of O_j is given by the family of functions

$$\left[\phi_n^j(x)\psi_m^j(y), \psi_n^j(x)\phi_m^j(y), \psi_n^j(x)\psi_m^j(y) \right]_{(n,m) \in Z^2} \tag{2.17}$$

In the 2-D case, the detail coefficients are given by three detail signal images. The horizontal detail coefficients H_j at resolution 2^j are defined as

$$H_j = \left[< f(x, y), \psi_n^j(x)\phi_m^j(y) > \right]_{(n,m) \in Z^2} \tag{2.18}$$

The horizontal detail coefficients represent the variations along columns (horizontal edges). Similarly the vertical detail coefficients V_j represent the changes in variations along rows

(vertical edges)

$$V_j = \left[< f(x,y), \phi_n^j(x)\psi_m^j(y) > \right]_{(n,m)\in Z^2} \tag{2.19}$$

and the diagonal details D_j represent the changes in the diagonal direction.

$$D_j = \left[< f(x,y), \psi_n^j(x)\psi_m^j(y) > \right]_{(n,m)\in Z^2} \tag{2.20}$$

Thus a wavelet decomposition for a 2-D signal for L levels is given by

$$\left[A_J, (H_j)_{J\leq j\leq -1}, (V_j)_{J\leq j\leq -1}, (D_j)_{J\leq j\leq -1} \right] \tag{2.21}$$

where $J = -L$. The coefficients A_j, H_j, V_j and D_j can be derived from A_{j+1} using separable functions $\widetilde{H}$ and $\widetilde{G}$. The filters are successively applied in the two different dimensions to get the set of coefficients as shown in Figure 2.4. The reconstruction of A_{j+1} from the coefficients A_j, H_j, V_j and D_j is also achieved by the application of upsampling and successive filter operations as shown in the Figure 2.5.

Decimated Wavelet Based Models

The two-dimensional DWT (Discrete Wavelet Transform) can be implemented using digital filters and downsamplers as shown in Figure 2.4. The sub-bands obtained after applying the wavelet transforms can be used for the reconstruction of the original image by applying an inverse wavelet transform as shown in the diagram in Figure 2.5. It is important to observe that in a DWT, the filtering operations are followed by downsampling, hence it is a decimated multi-scale representation. Starting from an original signal $f(x,y)$ we obtain a set of coefficients as which defines the wavelet decomposition of the signal completely

$$\left(A_J, (H_j)_{J\leq j\leq -1}, (V_j)_{J\leq j\leq -1}, (D_j)_{J\leq j\leq -1} \right) \tag{2.22}$$

where L is the number of resolution levels for which the wavelet decomposition is defined and $J = -L$. When we apply the transform on the approximation $A_{j+1}(x,y)$ at resolution level 2^j we get a set of four signals, $A_j(x,y), H_j(x,y), V_j(x,y)$ and $D_j(x,y)$, and the number of pixels in each of these signals is $\frac{1}{4}^{th}$ the number of pixels in the signal $A_{j+1}(x,y)$. Thus, the number of samples in all the signals together at resolution level 2^j is the same as the number of samples in the approximation at the level 2^{j+1}. At end of the decomposition, the number of samples in the wavelet decomposition, as shown in equation 2.22 is equal to the total number of samples in the original signal $f(x,y)$. Thus, this is a non-redundant transform on a 2-D signal where the number of samples remain preserved.

Figure 2.6 shows the result of applying the 2-D DWT (at one level) for a natural image. The four coefficient sets obtained after the decomposition have been shown. We can see that the approximation coefficients largely resemble the original image whereas the detail coefficients highlight the directional changes in energy. The horizontal detail coefficients represent the horizontal edges in the image, and the vertical detail coefficients and the diagonal detail coefficients represent the vertical and diagonal edges respectively.

The two-dimensional DWT has been used in several image processing algorithms [9, 10, 14, 15, 16, 17]. The conventional DWT, in which every filtering operation is followed by a downsampling step suffers from some limitations. It exhibits aliasing and there might be information loss while reconstructing the original signal from the decomposition [18]. The transform is rotation-variant and lacks directional sensitivity [18]. These properties make it difficult to build robust image processing algorithms based on the DWT model. It is desirable to use a wavelet decomposition model in which the advantages of wavelets can be retained while resolving the mentioned limitations.

Undecimated Wavelet Based Models

The Undecimated Wavelet Transform or the Stationary Wavelet Transform (SWT) is a modified version of the Discrete Wavelet Transform which leads to an overdetermined or redundant representation of the original data. The filtering step in SWT is similar to that in the DWT, however we do not decimate or downsample the sub-bands after the filtering step and each of the resulting decompositions has the same size as the input images. The number of samples in all the decompositions together is more than the number of samples in the original image $f(x, y)$. Though this transform introduces redundancy and increases the computational intensity of the algorithms, it leads to much better structural preservation, fewer artefacts and rotation invariance [19, 20, 21, 22]. These properties are highly desirable in many applications where high accuracy is required.

Figure 2.7 shows the effect of application of 2-D DWT and SWT (on 2 decomposition levels) to the same image used in Figure 2.6. The figure shows the detail coefficients for both, DWT and SWT at the two decomposition levels. We can see that there is more structural preservation (more well defined edges) in the detail coefficients in the case of the SWT. The difference between the coefficients is more significant at level 2, which is in accordance with expectations as the DWT level 2 coefficients have been obtained after downsampling and the coefficients for SWT are at the same resolution as the original image. Thus for multi-scale non-parametric models, the stationary wavelets are an excellent tool as they provide multi-scale information about the image and the structural preservation is significantly better than the DWT. It will be shown in chapter 3 that SWT decomposition

gives more redundancy in terms of pixel neighbourhoods. In chapter 4 and 5, it will be illustrated how the SWT based algorithms (saliency detection and image denoising) give much better performance than the corresponding DWT based algorithms.

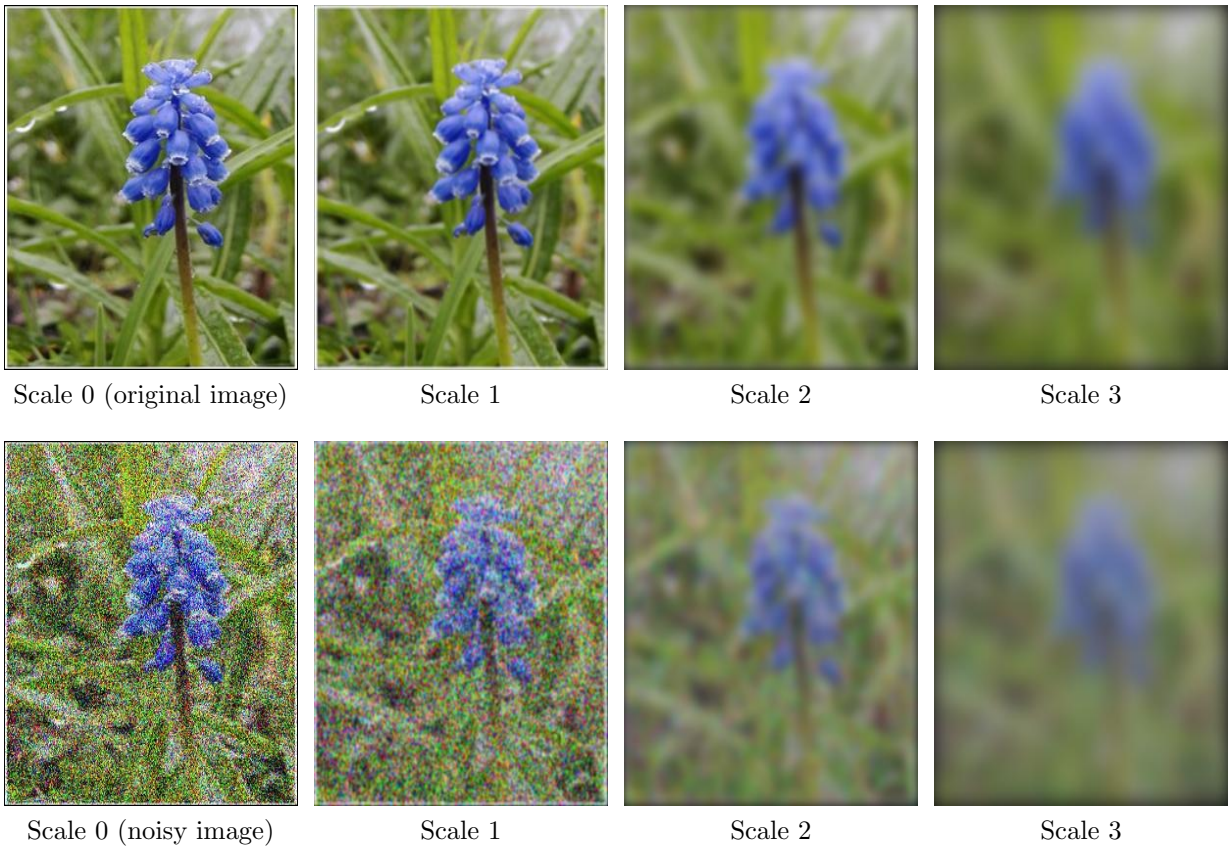


Figure 2.1: Multi-scale representation of an image without noise (top) and with noise (bottom) (additive Gaussian noise, $\sigma = 0.3$). It can be seen that at higher scales, the edges become blurred. In the case of the noisy image, the spatial distortion of the image is more prominent.

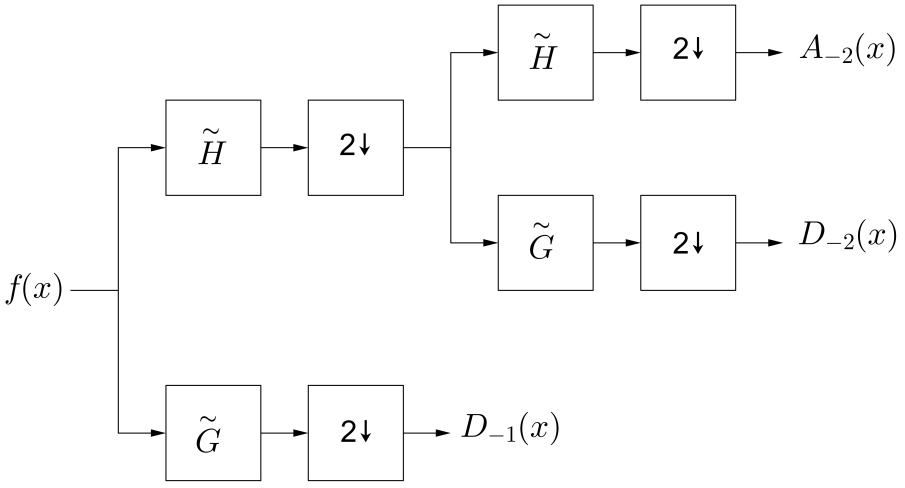


Figure 2.2: A two-level discrete wavelet decomposition is shown. The filtering operations are followed by downsampling to get the appropriate wavelet coefficients, i.e., approximation and detail coefficients.

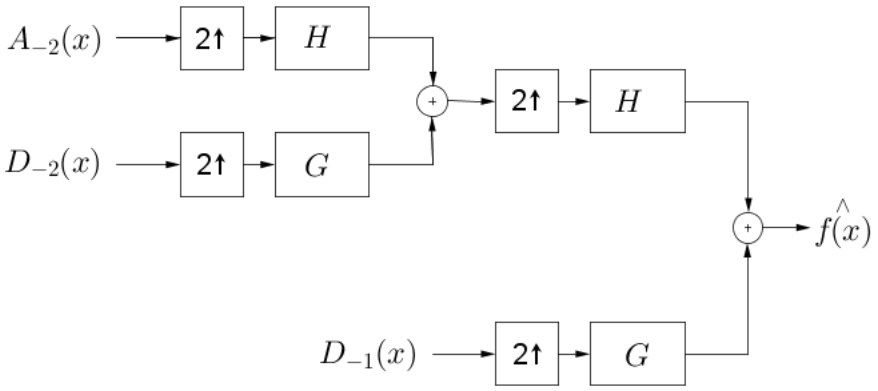


Figure 2.3: A two-level discrete wavelet reconstruction is shown. The wavelet coefficients (which are obtained using the process depicted in Figure 2.2) undergo upsampling (inserting zeros after every sample in the original signal) and filtering to reconstruct an estimate of the original signal.

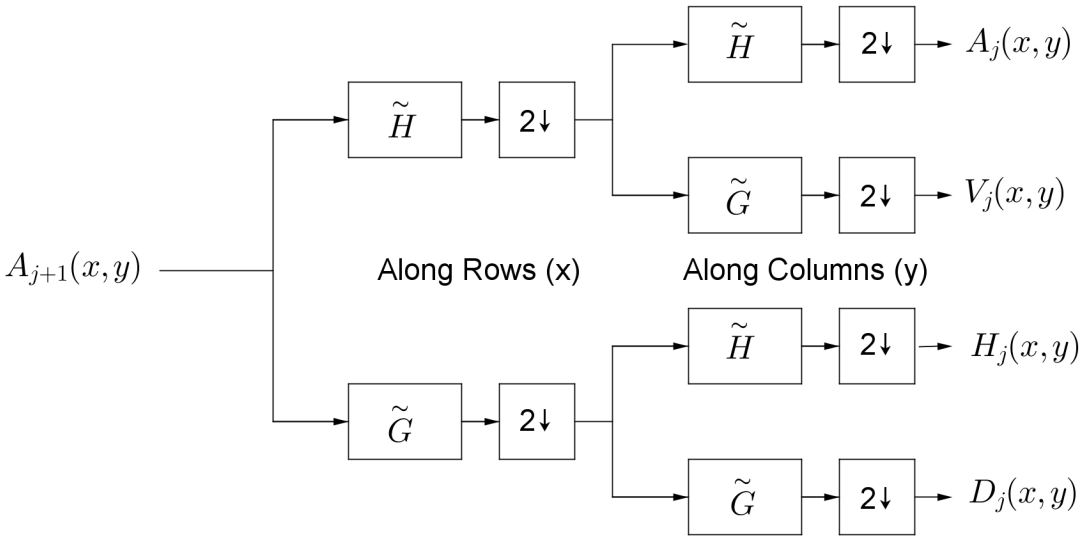


Figure 2.4: This figure shows the decomposition of an approximation coefficient set at a higher resolution level (A_{j+1}) to obtain the set of four coefficients (approximation coefficients and the three direction detail coefficients) at a lower resolution level.

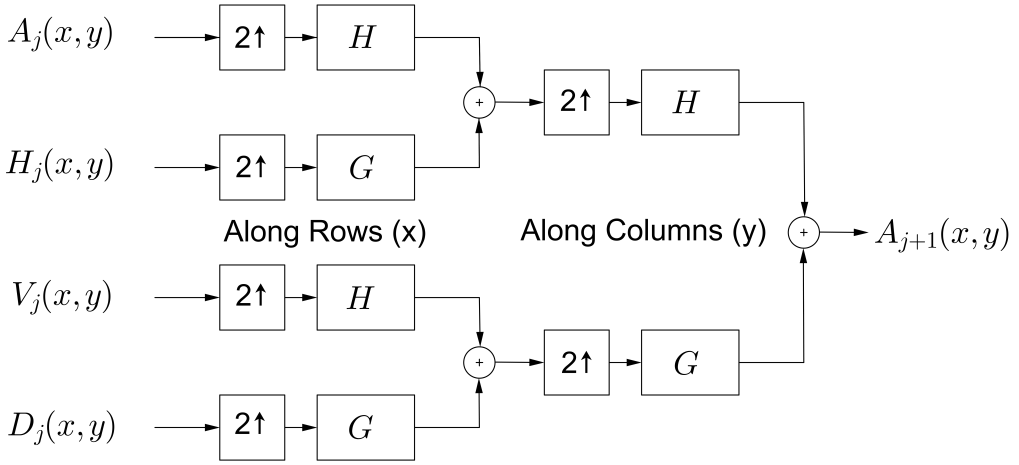


Figure 2.5: This figure shows the reconstruction of the higher resolution level approximation coefficient set from the lower resolution level coefficients. This process is the reverse of the process illustrated in Figure 2.4.

Original Image

Approximation Coefficients

Horizontal Detail Coefficients

Vertical Detail Coefficients

Diagonal Detail Coefficients

Figure 2.6: DWT (1 level) decomposition for a natural image. The detail coefficients (horizontal, vertical and detail) highlight the horizontal, vertical and diagonal edges respectively (the detail coefficients have been scaled (range 0 to 1) and inverted for visual representation).

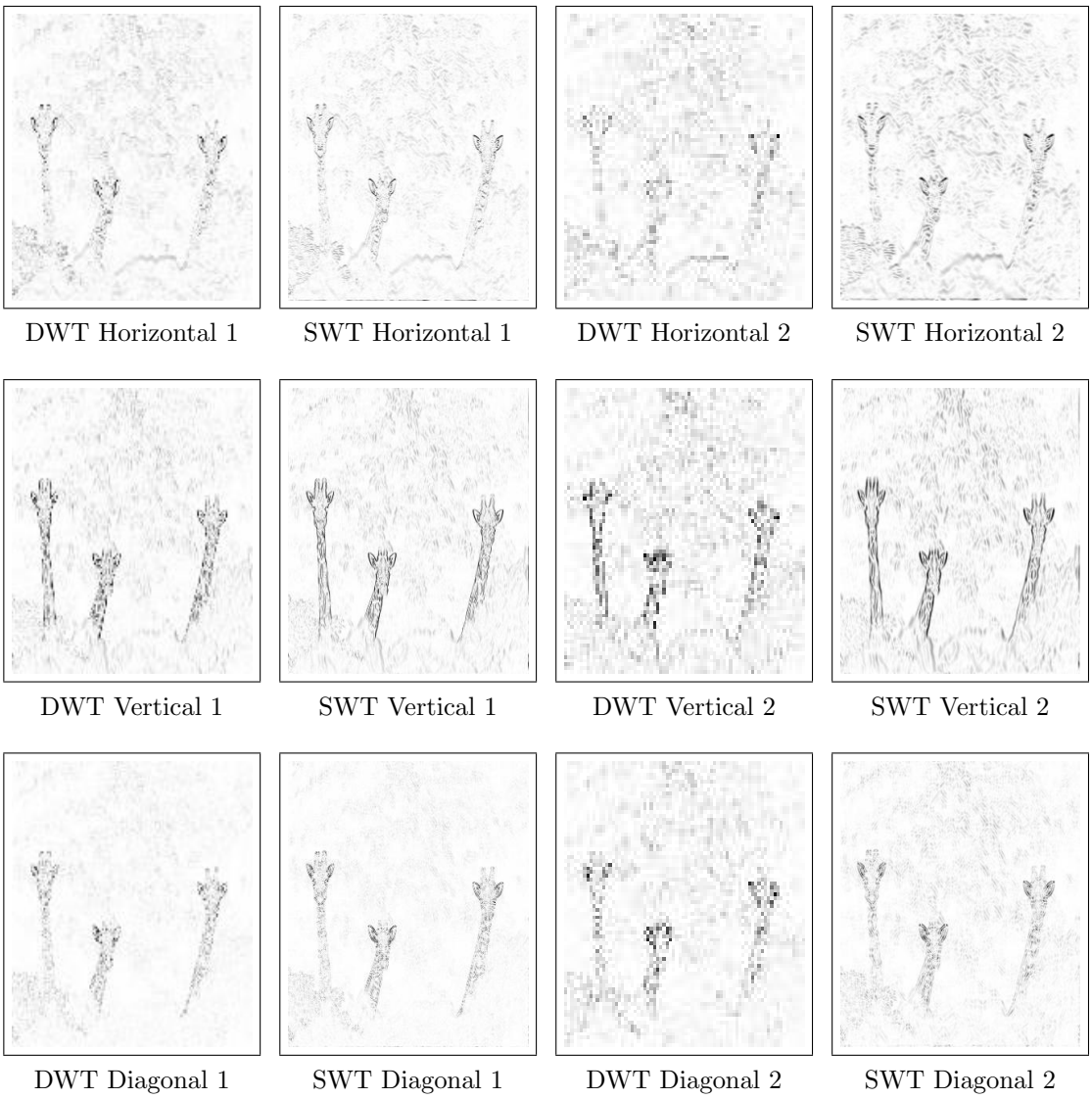


Figure 2.7: Comparison of level 2 detail coefficients in DWT and SWT. It can be clearly noticed that in the detail coefficients corresponding to the SWT, the structural features (edges in this case) are more clearly defined as compared to the coefficients corresponding to the DWT, in all directions (the detail coefficients have been scaled (range 0 to 1) and inverted for visual representation).

2.2 Non-parametric Models for Image Processing

2.2.1 Anisotropic (non-linear) Scale Space Model

A major drawback of the linear scale space model discussed in section 2.1.1 is that at higher scales significant blurring is observed, as can be seen in Figure 2.1. This results in loss of structural details at these scales. For the preservation of structural details at all scales in the multi-scale representation, it is desirable to have a filtering process where these details would not have to undergo blurring. Also in Figure 2.1, the spatial location of edges in the image at higher scales does not coincide with their corresponding location in the original image. Thus, in addition to blurring, spatial distortion is also observed at higher scales in the linear scale space representation [2, 3]. These details have to be traced from their coarse scale representations to the original image through the entire scale space image, making the process computationally intensive and inefficient [3]. In a multi-scale representation of a signal, it is expected that while moving from the finer scales to coarser scales, no new features would be observed [7], while it is acceptable for existing features to disappear. This condition, known as the causality condition has been stated as an important feature in multi-scale image models [3, 6, 7], and is satisfied in the linear scale space approach model in [1, 6, 7]. The nonlinear multi-scale model described in [3] satisfies this causality condition along with two other conditions which have been shown to contribute to the structural preservation in images, and which are discussed below.

We want region boundaries or edges to be preserved at all scales, i.e., when we move from finer scales to coarser scales, we want blurring to take place within the region boundaries, without blurring the boundaries themselves (piecewise smoothing) [3, 23, 24, 25]. Also, the position of the structural details at all scales should coincide with their positions in the original image (immediate localization). In [3, 24], nonlinear scale space models have been proposed which satisfy these two properties along with the causality property mentioned above.

The convolution kernels ($g(x, \sigma)$) as discussed in equation 2.2 used in the linear scale space approach are based on fixed parameters which do not take into account the underlying structure of the original image. This is responsible for the structural degradation of the images in that model [2, 3, 7]. We consider the diffusion equation as explained in [3]

$$\frac{\partial L_t}{\partial t} = c_t(x)\nabla^2 L_t + \nabla(c\nabla L_t) \tag{2.23}$$

where $c_t(x)$ is the diffusion coefficient, ∇ is the gradient operator and L_t is the representation of the image at scale t.

If at a particular pixel, $c_t(x) = 0$, we would not observe blurring at that pixel and if $c_t(x) = 1$, we would observe blurring. In an ideal situation, we want $c_t(x)$ to take a value of 0 at the pixels corresponding to the boundaries and a value of 1 for pixels corresponding to interior regions in the image. Hence, an appropriate function has to be chosen for $c_t(x)$ which would satisfy this property. In [3], it has been shown that the diffusion coefficient should be a binary valued function of the gradient of $L_t(x)$, i.e.,

$$c_t(x) = g(\|\nabla L_t(x)\|) \tag{2.24}$$

The gradient of intensity values of the pixels has been used for the estimation of boundary pixels in the image, under the assertion that the magnitude of the gradient would be large at boundaries in the image, and small within the regions in the image. Since this is an approach based on non-linear or anisotropic filters, the blurring which happens in each pixel region in the image depends on the local structure around the pixel region, i.e., the anisotropic filter "adapts" to the local structure hence preserving the structural details. In this approach the boundaries / edges remain preserved even at coarser scales, without undergoing any spatial shift. The multi-scale representation can be described as follows

$$L_t(x) = L_{t-1}(x) + d_t(x) \tag{2.25}$$

where $L_0(x) = f(x)$. The relationship between two successive scales is defined by $d_t(x)$ which is defined in terms of the gradient ∇ and the diffusion coefficient $c_t(x)$ as given below

$$d_t(x) = \sum_i c_t(x+i)\nabla L_{t-1}(x+i) \tag{2.26}$$

Although the non-linear scale space approach satisfies the three conditions mentioned above and the structural preservation is better than the traditional linear scale space approach, there is still scope for improvement, as significant structural degradation and sensitivity to noise can still be observed, as has been illustrated in [2].

2.2.2 Non-parametric Sample Based Model for Image Understanding

We have seen in section 2.1.1 that the linear scale space approach does not use the underlying structural content whereas the non-linear approach uses the local structural information for "adapting" the anisotropic filter for blurring of images. In both cases, we observe structural degradation and sensitivity to noise, as has been illustrated in [2]. In

[2], a non-parametric scale space model has been proposed which uses global information in the original image and hence gives a model better satisfying the immediate localization property. The model uses global information from the original image without significantly increasing the computational complexity by using a sampling based approach. The multi-scale model is described as

$$L_t(x) = L_{t-1}(x) + c_t(x) \tag{2.27}$$

where $L_t(x)$ is the representation at scale t, $L_0(x) = f(x)$ i.e. the original signal and $c_t(x)$ is the residual inter-scale structure at scale t. The scale representations $L_t, L_{t-1}.....L_0$ are assumed to follow the Markov chain property where L_t (at an arbitrary scale t) depends only on the scale immediately below it, i.e., L_{t-1} and is independent of all other scale representations. $L_t(x)$ at an arbitrary scale t is estimated from $L_{t-1}(x)$ and a conditional probability distribution $p\big(L_t(x)/L_{t-1}(x)\big)$ using the Bayesian least squares solution as shown in equation below

$$\hat{L}_t(x) = \int L_t(x) p\big(L_t(x)/L_{t-1}(x)\big) dL_t(x) \tag{2.28}$$

The conditional probability distribution $p\big(L_t(x)/L_{t-1}(x)\big)$ at an arbitrary scale t is obtained from a quasi-random density estimation approach, for which samples from $L_{t-1}(x)$ have to be drawn. A set of m samples is drawn from a Sobol quasi-random sequence [2] with respect to x. From these samples, a distribution $p\big(L_{t-1}(x)\big)$ is created. These distributions for different scales t are used for deriving a Gaussian mixture distribution, which is used for determining samples to be used to estimate $p\big(L_t(x)/L_{t-1}(x)\big)$. From the Gaussian mixture model, the samples which belong to the Gaussian distribution corresponding to $L_{t-1}(x)$ form the sample set S. Given this set S, the estimate of the conditional probability distribution can be computed as shown below

$$\hat{p}\big(L_t(x)/L_{t-1}(x)\big) = \frac{q\big(L_t(x)/L_{t-1}(x)\big)}{\int_0^1 q\big(L_t(x)/L_{t-1}(x)\big) dL_t(x)} \tag{2.29}$$

where $q\big(L_t(x)/L_{t-1}(x)\big)$ quantifies the sample relevancy. It can be defined as

$$q\big(L_t(x)/L_{t-1}(x)\big) = \frac{1}{\sqrt{2\pi}\sigma L_t} \sum_{k \in S} f_T(k) \exp\left(\frac{-1}{2} \left(\frac{L_t - L_{t-1}(x_k)}{\sigma_{L_{t-1}}} \right)^2 \right) \tag{2.30}$$

$f_T(k)$ is an assessment measure of sample relevancy, which includes three important factors; image intensities, image gradient magnitudes and spatial closeness. Taking all these factors into account while assuming the sample relevancy is important. Taking image intensities as

a factor takes care of the immediate localization and piecewise smoothing properties. Using image gradients contributes to the preservation of structural details. This non-parametric multi-scale model takes global image information and satisfies all three properties of a good multi-scale model, i.e., causality, immediate localization and piecewise smoothing.

Chapter 3

Non Parametric Neighbourhood Based Model

In this chapter we develop a non-parametric framework for image understanding which is based on the pixel neighbourhoods in the image. In section 3.1, we define the neighbourhood based theory in image processing with a mathematical model. In section 3.2, we discuss the concept of statistical non-redundancy in image pixel intensities (spatial domain) and in section 3.3, we extend this concept of statistical non-redundancy to the multi-scale transforms. In section 3.4, we discuss some neighbourhood based statistics observed in natural images. In section 3.5, a non-parametric neighbourhood based model is then proposed for image understanding which is used in the following chapters for applications in image saliency detection (chapter 4) and image denoising (chapter 5).

3.1 Neighbourhood Based Theory in Image Modelling

Let χ be a set of pixels which define an image $f = \{f(i)|i \in \chi\}$. At any arbitrary pixel $i \in \chi$, the intensity at the pixel can be defined as $f(i)$. For any pixel $i \in \chi$, we define N_i, the neighbourhood around the pixel i, as the set of pixels surrounding pixel i in the image such that $N_i \subset \chi$. N_i^k is a pixel in the neighbourhood N_i at an offset k from the central pixel i. The offset k is a two dimensional spatial offset which specifies the horizontal and the vertical distance (in terms of the number of pixels) from the central pixel i. We denote this offset as $k = (x, y)$, where x denotes the horizontal offset and y denotes the vertical offset from the central pixel. The intensity at this arbitrary site N_i^k is denoted as $f(N_i^k)$. This is illustrated in Figure 3.1.

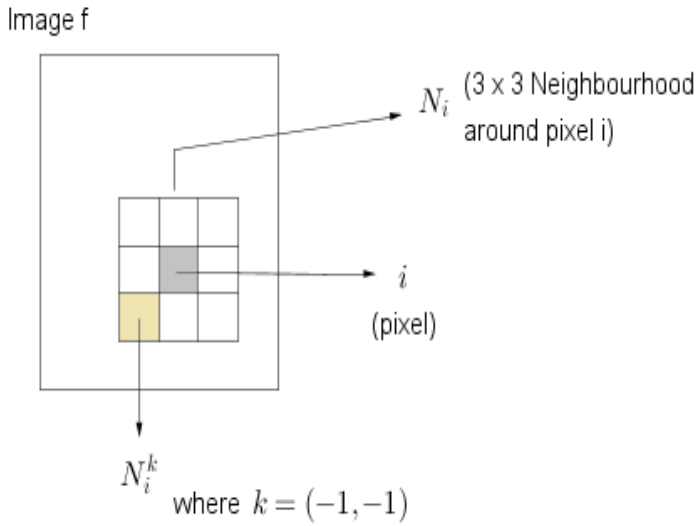


Figure 3.1: This figure illustrates the concept of a pixel neighbourhood and the offsets used to describe the sites in the pixel neighbourhoods.

3.2 Non-redundancy in Image Pixel Intensities (Spatial Domain)

We assert that in any image, a large number of pixel neighbourhoods are statistically redundant because they are similar to each other, and the pixel neighbourhoods which are statistically non-redundant define the image statistics. Thus, in this framework, we want to quantify the statistical non-redundancy of the pixel neighbourhoods to derive the image statistics. In chapter 4 and chapter 5, we use these derived image statistics to develop saliency detection and image denoising algorithms respectively. We define $P\big(f(N_i^k)|f(N_j^k)\big)$ as the probability of a site k in the pixel neighbourhood N_i being statistically redundant with respect to the corresponding site in the pixel neighbourhood N_j. The probability distribution for each such site is assumed to be independent of other sites, therefore, the probability $P\big(f(N_i)|f(N_j)\big)$ of a pixel neighbourhood N_i being statistically redundant with respect to the pixel neighbourhood N_j can be defined as

$$P\big(f(N_i)|f(N_j)\big) = \prod_k P\big(f(N_i^k)|f(N_j^k)\big) \tag{3.1}$$

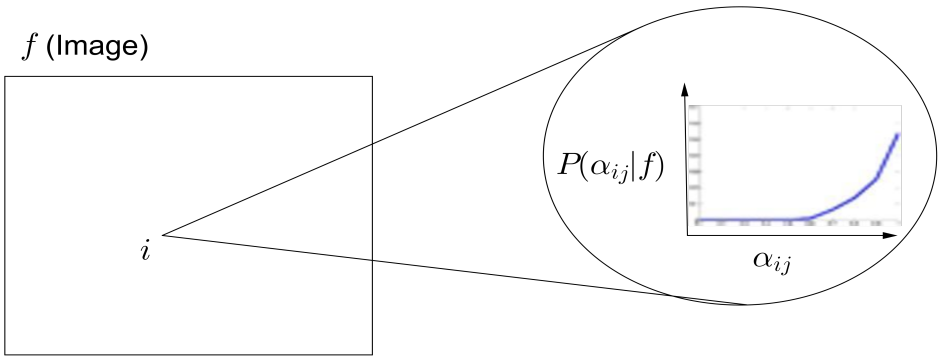


Figure 3.2: This figure shows the probability distribution ($P(\alpha_{ij}|f)$) of statistical non-redundancy (α_{ij}) of a pixel neighbourhood N_i with respect to all other pixel neighbourhoods N_j in the image. This gives an overall picture of the statistical non-redundancy of the site i with respect to the given image f.

We denote the statistical non-redundancy between two neighbourhoods N_i and N_j as α_{ij} which can be defined as

$$\alpha_{ij} = 1 - P\big(f(N_i)|f(N_j)\big) = 1 - \prod_k P\big(f(N_i^k)|f(N_j^k)\big) \tag{3.2}$$

We now wish to quantify the statistical non-redundancy of a pixel neighbourhood N_i with respect to a given image f. If for a particular pixel i we compute α_{ij} over all possible pixels j within the given image, we can compute the probability distribution of α_{ij} given the image f (denoted as $P(\alpha_{ij}|f)$), which gives an overall picture of the statistical non-redundancy of site i given the image f. This is represented in the Figure 3.2. We then further define the overall statistical non-redundancy S_i of a pixel i as the expected value of α_{ij} given an image f. That is

$$S_i = \underset{j}{E}(\alpha_{ij}|f) \tag{3.3}$$

Thus for any arbitrary pixel $i \in \chi$ in the image f, we can define a mapping S_i which gives the overall statistical non-redundancy of the pixel i with respect to the given image f. This non-redundancy map gives us statistical information about the image in a non-parametric

way, and can be used to derive important information about the structure of the image. Extending the concept of statistical non-redundancy to the wavelet domain can be useful in deriving multi-scale structural information of a given image.

3.3 Multi-scale Transform Non-redundancy

In the previous section we discussed the concept of statistical non-redundancy of a pixel with respect to a given image. If we extend this concept to the wavelet domain, and evaluate the statistical non-redundancy of wavelet coefficients at different sub-bands instead of pixels in the spatial domain, we can get different structural details present in the image, at multiple scales. The non-redundancy map discussed in section 3.2 can be obtained for each of the wavelet sub-bands, i.e., the approximation and detail coefficients at all resolution levels. Different structural characteristics of the image can be obtained from these non-redundancy maps. Non-redundancy map for the approximation coefficients corresponding to DWT, at resolution level j, is denoted by $nrDAj$. Similarly, $nrDHj, nrDVj$ and $nrDDj$ represent the non-redundancy maps for the horizontal, vertical and diagonal detail coefficients at resolution level j for DWT, and $nrSHj, nrSVj$ and $nrSDj$ represent the non-redundancy maps for the corresponding detail coefficients for SWT. In this context, the resolution levels denote the decomposition levels in the wavelet decomposition, i.e., coefficients at resolution level 1 are obtained on application of the wavelet transform to the original image, coefficients at resolution level 2 are obtained on application of the wavelet transform to approximation coefficients at level 1 and so on.

These maps can give us the structural features in the image at different scales. Figure 3.4 shows these non-redundancy maps for the wavelet coefficients (decimated as well as undecimated) for three successive resolution levels. The test image used is the same as that used in Figure 2.6. From the figure we can see that there is a significant difference between the non-redundancy maps for the decimated and the undecimated cases, particularly at resolution levels 2 and 3. The non-redundancy maps shown in this figure give us an idea of the distribution of pixel neighbourhoods in the wavelet transform domain. In these maps, the high intensity points correspond to the pixel neighbourhoods which are statistically non-redundant and vice versa. The non-redundancy maps for the approximation coefficients highlight the unique objects in the image, and the detail coefficients highlight the edges in the image. The maps for the lower resolution levels highlight the coarse features and those for higher resolution levels highlight the finer features in the image, e.g., in case of the vertical detail coefficients, $nrSV1$ highlights the edges of only the animal figures whereas $nrSV3$ also highlights the edges present in the background. It

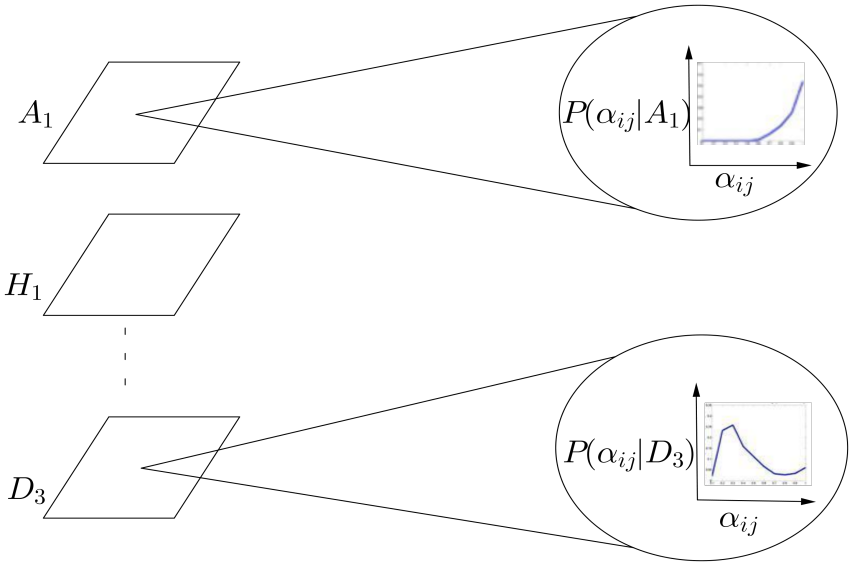


Figure 3.3: Statistical Non-redundancy Distribution for Wavelet Sub-bands. This figure shows a general multi-scale non-redundancy model and the wavelet coefficients could correspond to DWT or SWT.

can be seen from the figure that these features are very well defined in the maps for the undecimated wavelet coefficients, and are very blurred in case of the decimated wavelets. Clearly, the decimation step in the decimated wavelets affects the statistical distribution of pixel neighbourhoods of the wavelet coefficients. Thus, for a multi-scale model based on non-redundancy of coefficient neighbourhoods, undecimated wavelets are more useful.

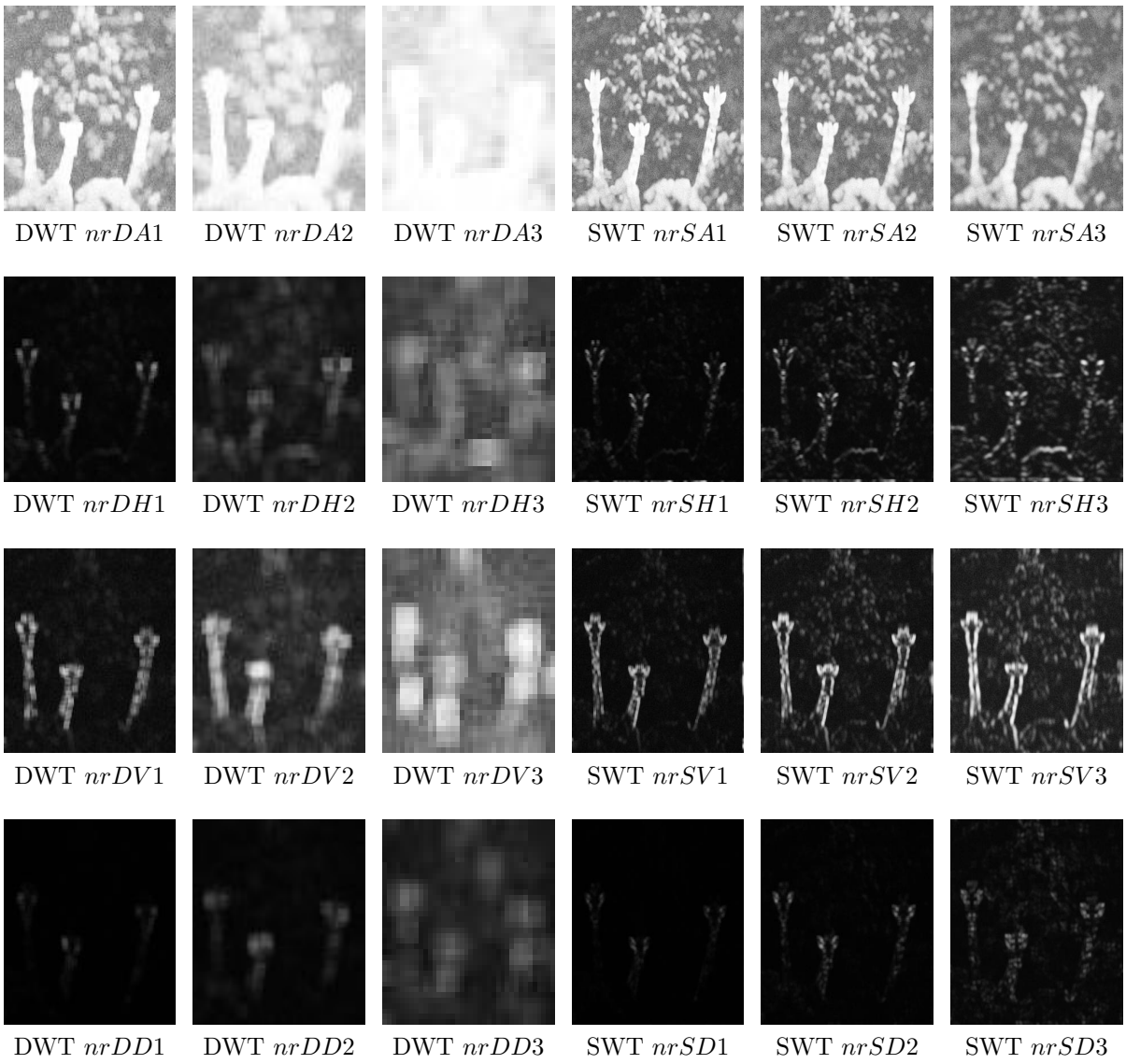


Figure 3.4: This figure compares the non-redundancy distribution for the decimated as well as undecimated wavelet coefficients at three different decomposition levels. We can see more structural details in the non-redundancy maps for the undecimated wavelets.

3.4 Redundancy Statistics in Multi-scale Transform Domain

We want to develop a non-parametric neighbourhood based model for image understanding at multiple scales. For this, we need to understand the statistical distribution of neighbourhoods in natural images. The neighbourhood similarity or dissimilarity statistics can be used for extracting useful information about the image statistics without any prior knowledge. To understand neighbourhood statistics, we need to define the concept of neighbourhood similarity. We denote a distance metric between two arbitrary neighbourhoods N_i and N_j as $\rho(N_i, N_j)$. The definition of the distance metric could depend on the application. For the case of natural images, we have selected the distance metric as

$$\rho(N_i, N_j) = 1 - e^{-\frac{\sum_k (f(N_i^k) - f(N_j^k)^2}{\sigma^2}} \tag{3.4}$$

where σ^2 is the variance of all combinations of $\sqrt{\sum_k (f(N_i^k) - f(N_j^k))^2}$. Two neighbourhoods N_i and N_j are said to be similar if

$$\rho(N_i, N_j) < \tau \tag{3.5}$$

where τ is a threshold value. The distance metric ρ gives a measure of how similar two image neighbourhoods are. Using this we can define a similarity index between two neighbourhoods N_i and N_j as

$$\eta_{ij} = 1 \quad if\ \rho(N_i, N_j) < \tau \tag{3.6}$$

$$= 0 \quad otherwise \tag{3.7}$$

$\eta_{ij} = 1$ implies that neighbourhood N_i is similar to neighbourhood N_j. τ is the cut-off threshold which is used for deciding how close the neighbourhoods need to be in order to be considered similar. Given the similarity metric for two arbitrary neighbourhoods N_i and N_j, for every pixel i in the image, we can determine how many neighbourhoods in the image are similar to the neighbourhood N_i corresponding to the pixel i. Let us define γ_i as the similarity ratio for a pixel neighbourhood N_i

$$\gamma_i = \frac{\sum_j \eta_{ij}}{T} \tag{3.8}$$

where T is the total number of pixels in the image under consideration. Thus, γ_i gives us the expected number of neighbourhoods in the image which match with the given neighbourhood N_i. The similarity ratio also gives a measure of how redundant a neighbourhood

is, where higher values of γ_i imply higher redundancy of neighbourhood N_i. We now introduce a redundancy factor β which decides the cut-off for the similarity ratio beyond which a pixel neighbourhood would be considered statistically redundant, i.e., if $\gamma_i > \beta$, pixel neighbourhood N_i is considered as a redundant neighbourhood. We define a redundancy index

$$\delta_i = 1 \quad if\ \gamma_i > \beta \tag{3.9}$$

$$= 0 \quad otherwise \tag{3.10}$$

where a value of 1 for the redundancy index δ_i implies that the pixel neighbourhood N_i is statistically redundant. For any image, the redundancy ratio Θ can be defined as the ratio of redundant pixel neighbourhoods to the total number of pixel neighbourhoods in the image, i.e.,

$$\Theta = \frac{\sum_i \delta_i}{T} \tag{3.11}$$

where T is the total number of pixel neighbourhoods in the image (which is the same as the total number of pixels in the image). The redundancy ratio Θ discussed above can be defined for a set of pixels χ. This set of pixels χ could represent the original image (spatial domain) or the wavelet coefficient sub-bands (at different decomposition levels and different orientations). The redundancy ratio Θ_χ is a function of τ (neighbourhood similarity threshold) and β (neighbourhood redundancy factor). To understand the neighbourhood redundancy statistics of an image, we want to see how $\Theta_\chi(\tau, \beta)$ varies

1. along τ, with fixed β, across different pixel sets χ
2. along β, with fixed τ, across different pixel sets χ

We want to see how $\Theta(\tau, \beta)$ varies for different sets of pixels, for comparison across different wavelet decomposition levels and also for comparison between decimated and undecimated wavelets. In this section, we discuss results of an experiment where we have considered several natural images. Each of these natural images was subjected to 3 level wavelet decompositions, both decimated as well as undecimated. For the rest of the section, the decimated wavelet coefficients will be denoted as $DA1$ (decimated, approximation coefficients, decomposition level 1), $DH1$ (decimated, horizontal coefficients, decomposition level 1) and so on. The undecimated wavelet coefficients will be denoted as $SA1$ (undecimated, approximation coefficients, level 1) and so on. For each of these sets of coefficients, we compute $\Theta(\tau, \beta)$ varying τ and β. We then take an average of these $\Theta(\tau, \beta)$ values across all the images considered, and use these average values to study the behaviour of

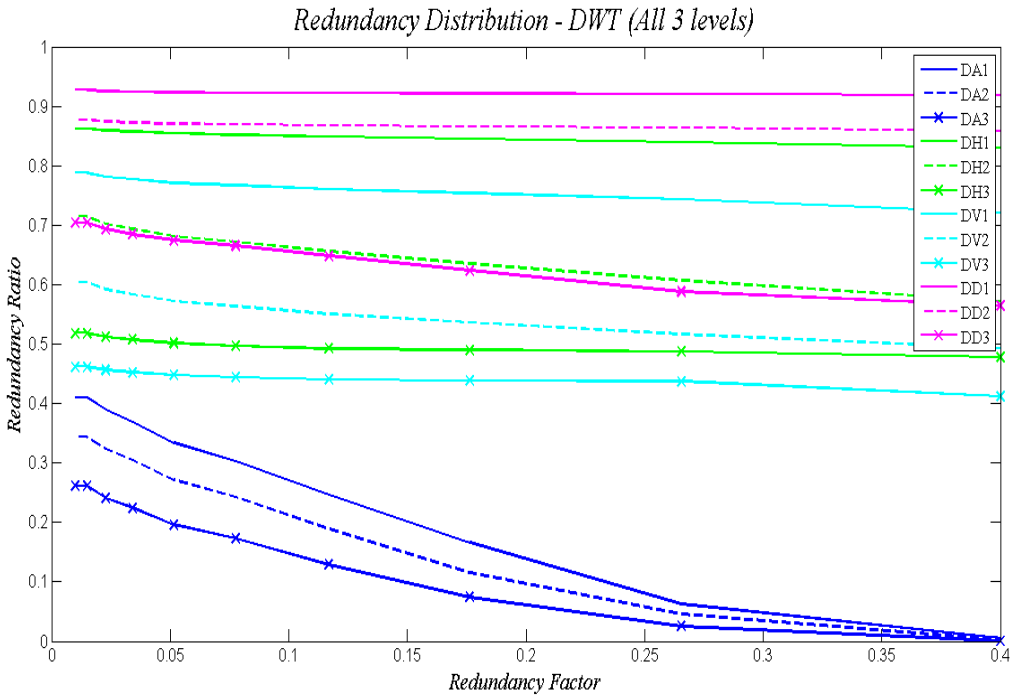


Figure 3.5: This figure shows the redundancy distribution for DWT. For a fixed neighbourhood similarity threshold, $\tau = 0.1294$, we plot the variation of the redundancy ratio $\Theta(\tau, \beta)$ with varying β. The redundancy ratio is a non-increasing function of β. The redundancy ratio decreases at higher decomposition levels.

neighbourhood redundancy in the wavelet domain. For the fixed τ, varying β case, we have fixed the value of τ to $\tau = 0.1294$, and for the fixed β, varying τ case, we have fixed the value of β tp $\beta = 0.0774$. In our experiments, the values of τ were varied exponentially (10 values were considered) from 0 to 1 and the β was varied exponentially (again, 10 values were considered) from 0 to 0.4. We selected the values of τ and β which were in the middle of their respective exponential ranges.

In Figure 3.5, we show the redundancy distribution for decimated wavelet coefficients, for all the three decomposition levels, for all orientations (A,H,V,D). In this case, the neighbourhood similarity threshold, i.e., τ is kept constant at $\tau = 0.1294$ and $\Theta(\tau, \beta)$ is varied along β (varying exponentially in the range $[0.01, 0.4]$). We can see that $\Theta(\tau, \beta)$ is a non-increasing function of β, which is in accordance with the theoretical expectation, i.e., with a stricter definition of redundancy, we cannot have more neighbourhoods satisfying the

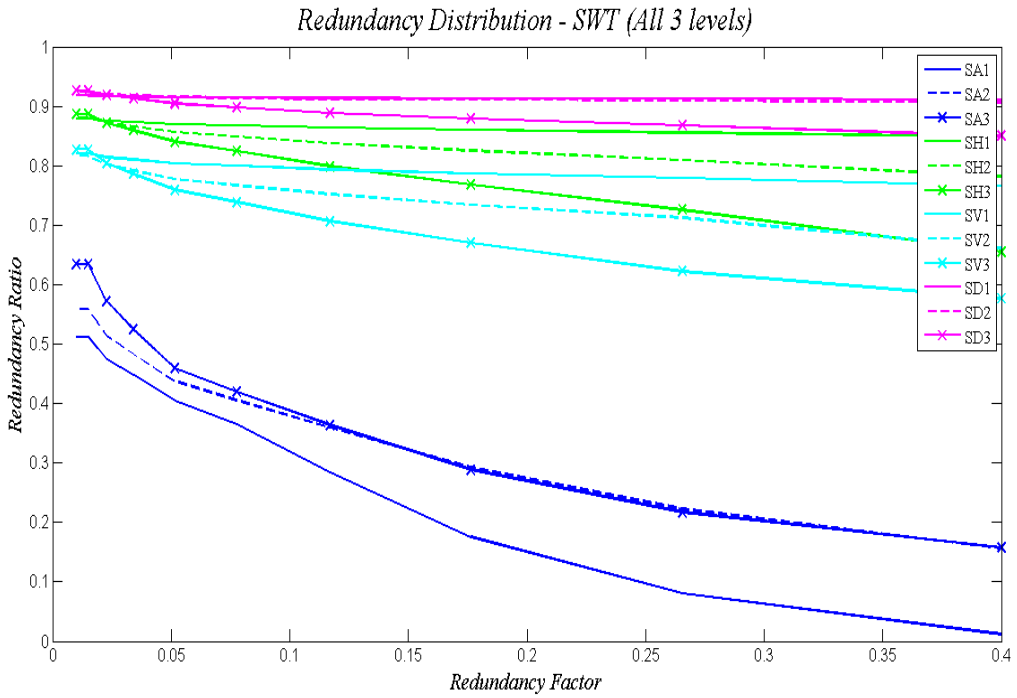


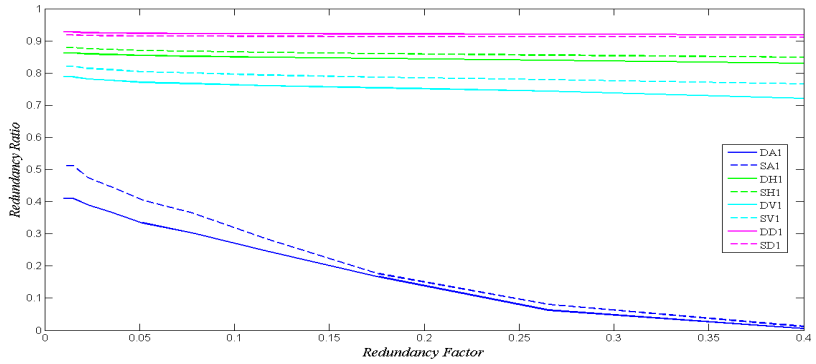
Figure 3.6: This figure shows the redundancy distribution for SWT. For a fixed neighbourhood similarity threshold, $\tau = 0.1294$, we plot the variation of the redundancy ratio $\Theta(\tau, \beta)$ with varying β. It can be seen that the neighbourhood redundancy does not decrease at higher decomposition levels, unlike the DWT case shown in Figure 3.5.

redundancy criterion. We can see that for every orientation (A,H,V,D), the redundancy ratio decreases as we progress along the decomposition levels.This makes sense because when we decimate or downsample the coefficients after applying the respective filters, we eliminate every alternate pixel in the pixel set, and this effects the neighbourhood similarity statistics leading to fewer matching neighbourhoods. Also, if we compare across the different orientations (A,H,V,D), we see more redundancy in detail coefficients as compared to the approximation coefficients. This is reasonable because detail coefficients essentially have a dark (and largely homogeneous) background with relatively few pixels in the pixel set representing the orientation specific details. This can be seen in Figure 3.4. Hence, we get a large number of matching neighbourhoods giving higher redundancy. Also, the detail coefficients show more stability in their redundancy characteristics with respect to β.

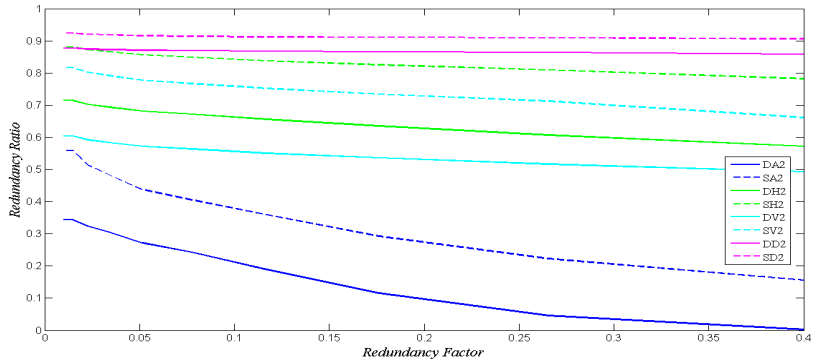
In Figure 3.6, we show the redundancy distribution for the undecimated wavelet coefficients, for all orientations and for all decomposition levels. Just like in the case of Figure 3.5, the threshold τ has been fixed to $\tau = 0.1294$ and β is varied. In this case, we observe that for the approximation coefficients, unlike in the case of decimated wavelets, the redundancy does not decrease with progressing decomposition levels. However, for the detail coefficient sub-bands, the redundancy decreases with the progressing decomposition levels. For the undecimated wavelet coefficients, we perform the same filtering operations as in the decimated case, but skip the downsampling step. The approximation coefficients are obtained by successive low pass filtering operations and the directional detail coefficients are obtained by using high pass filters in the corresponding orientation. This low pass filtering makes the pixel neighbourhoods in the approximation coefficients more similar, hence we observe more redundancy with progressing decomposition levels. Also, there is a significant difference in the redundancy distribution observed in approximation coefficients and detail coefficients.

In Figure 3.7, we compare the redundancy characteristics of the decimated and undecimated wavelet coefficients for all decomposition levels. Similar to the case of Figure 3.5 and Figure 3.6, the neighbourhood similarity threshold τ is fixed at $\tau = 0.1294$. In these figures it can be seen that at all decomposition levels, more redundancy is observed in the case of undecimated wavelet coefficients as compared to the decimated wavelet coefficients. The difference in these redundancy characteristics becomes more significant with progressing decomposition levels.

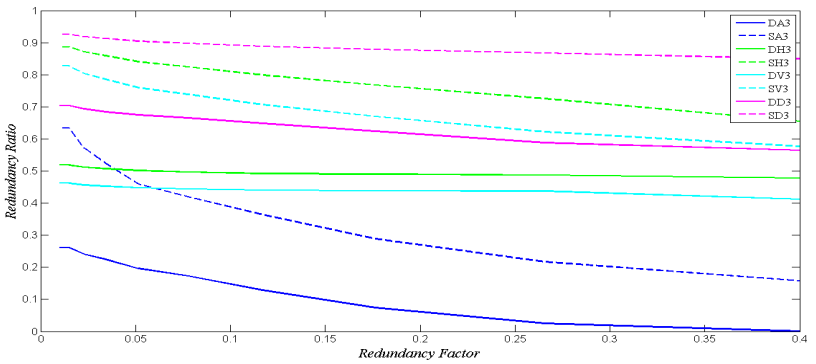
In Figure 3.8a, we fix the redundancy ratio β to a value $\beta = 0.0774$ and vary τ exponentially in the range $[0.01, 1]$ to get corresponding values of $\Theta(\tau, \beta)$ for the decimated wavelet coefficients. The redundancy of the pixel sets increases with increasing neighbourhood similarity threshold τ, which is reasonable because more pixel neighbourhoods would now satisfy the neighbourhood similarity criterion. Similar to the case of Figure 3.5, we observe that for the decimated wavelet coefficients the redundancy decreases with progressing decomposition levels. Also, the detail coefficients exhibit more redundancy as compared to the approximation coefficients. Figure 3.8b shows redundancy distribution (with varying τ and fixed β) for the undecimated wavelet coefficients. Again, similar to Figure 3.6, the redundancy of approximation coefficients increases with decomposition levels and that of the detail coefficients decreases with the levels. Figure 3.9 compares the DWT redundancy characteristics with the SWT redundancy characteristics (with fixed β and varying τ) for the three different decomposition levels. The results are similar to those in Figure 3.7, i.e., more redundancy is seen in case of the undecimated wavelet coefficients with the difference becoming more significant with the increasing decomposition levels. This is in accordance with expectations; the decimated wavelet coefficients undergo downsampling with each



Level 1 Decompositions (DWT-SWT Comparison)



Level 2 Decompositions (DWT-SWT Comparison)



Level 3 Decompositions (DWT-SWT Comparison)

Figure 3.7: The undecimated wavelet coefficients exhibit more redundancy than the decimated wavelet coefficients, with the difference increasing with the levels ($\tau = 0.1294$).

successive wavelet decomposition level, losing redundancy unlike the undecimated wavelet coefficients. Thus for image processing applications where higher pixel neighbourhood redundancy is useful, the undecimated wavelet transform serves as a more powerful tool. We will illustrate this in chapter 5, where we use a Markov-Chain Monte Carlo sampling based technique for image denoising.

3.5 Non-parametric Neighbourhood Based Framework

In section 3.1, we defined the neighbourhood theory in image modelling. In this section, we use this theory to propose a neighbourhood based framework which exploits the relation between pixel neighbourhoods in the image to derive statistical information about the image. For two arbitrary neighbourhoods N_i and N_j in an image $f = \{f(i)|i \in \chi\}$, we define the distance or the degree of dissimilarity between the neighbourhoods (or neighbourhoods) using as $\rho(N_i, N_j)$. The choice of ρ could vary on several factors including expected image statistics, the application for which the framework is being used, etc. This degree of dissimilarity between two pixels can be used for deriving statistical information from the image based purely on the image measurement given to us, without using any other information, as will be shown with a simple example below where the pixel neighbourhood theory is applied for texture synthesis.

To explain the pixel neighbourhood based non-parametric framework in detail, we use the example of texture synthesis for illustration [26, 27]. In this approach a neighbourhood similarity metric has been defined and has then been used to generate textures from a given sample texture. We start with a sample texture

$$f = \{f(i)|i \in \gamma\} \tag{3.12}$$

where γ denotes the set of pixels forming the original sample texture f from which we try and synthesize a complete texture image

$$g = \{g(j)|j \in \chi\} \tag{3.13}$$

where χ denotes the set of pixels representing the texture image g which we wish to synthesize. To synthesize an arbitrary pixel $i \in \chi$, in the texture image g, we consider a neighbourhood N_i (which is an NxN window around pixel i in this context). The process involves creation of a conditional probability distribution $P(i|N_i)$ [27] from which sampling is done [27] to obtain a value for the pixel i which we want to synthesize. The image is assumed to be a Markov Random Field [5, 27, 28] such that the value of an arbitrary pixel

$i \in \chi$ depends only on the neighbourhood N_i, and is independent of $\forall j \in \chi, j \notin N_i$. From the given sample texture f, we find an optimum neighbourhood, i.e., a neighbourhood $N_{i_{opt}}$ s.t.

$$N_{i_{opt}} = arg \min_{j \in \gamma} \rho(N_i, N_j) \tag{3.14}$$

Using this optimum neighbourhood from the sample texture f, we create a set of pixels

$$\zeta = \left\{ j | \rho(N_{i_{opt}}, N_j) < \varepsilon, \forall j \in \gamma \right\} \tag{3.15}$$

from which sampling is done (as explained in [26]), and the value of the sampled pixel is assigned to pixel i, which had to be synthesized. In this algorithm, the distance metric ρ has been defined as follows -

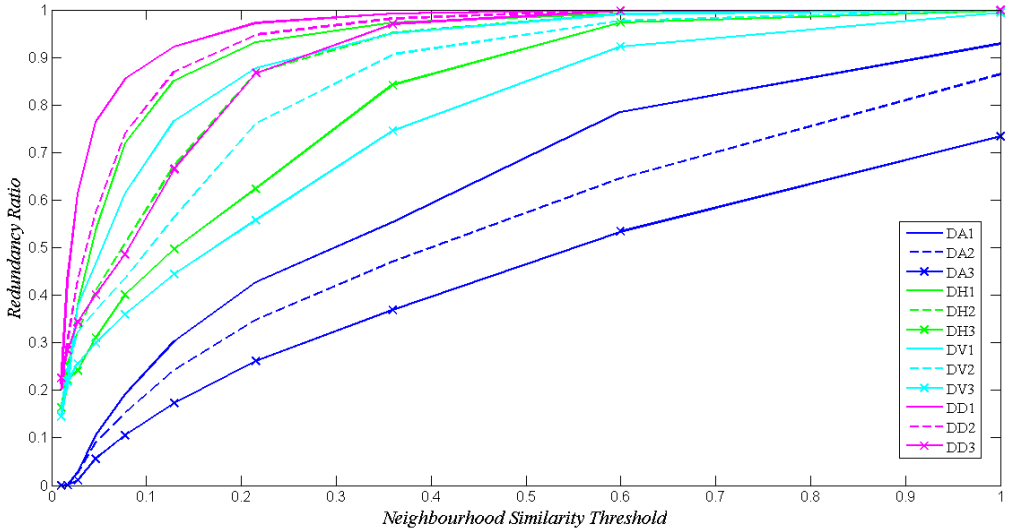
$$\rho(N_i, N_j) = d_{SSD}(N_i, N_j) * G \tag{3.16}$$

where d_{SSD} is the normalized sum of squared differences

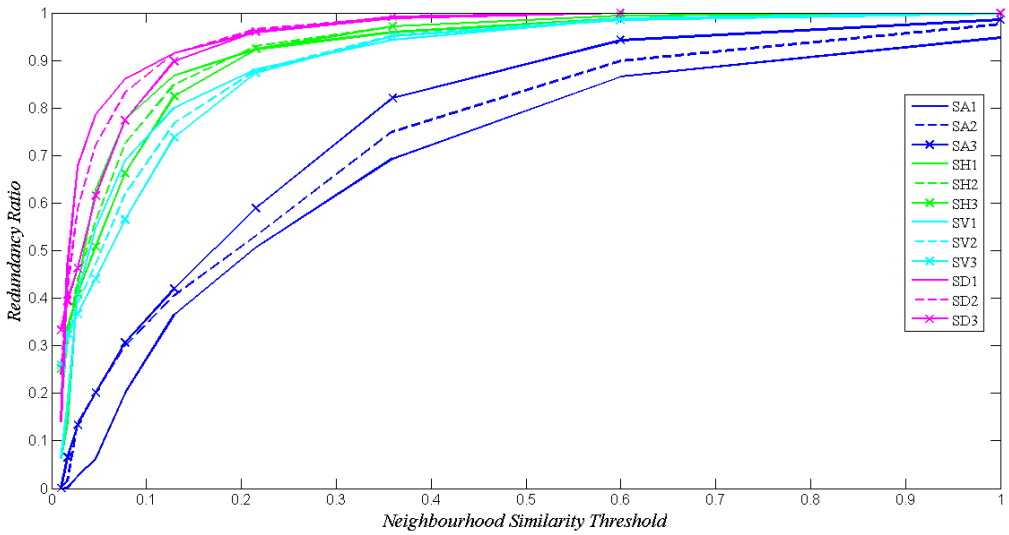
$$d_{SSD}(x, y) = \sum_k (x_k - y_k)^2 \tag{3.17}$$

and G represents a Gaussian kernel, which was defined in equation 2.2. The convolution with a Gaussian kernel is done to give more importance to neighbourhoods which are closer to each other [27].

The approach discussed above is an example of how neighbourhood-based methods can be used in non-parametric image processing. The important thing to note is that neighbourhood similarity / dissimilarity in images can be used for capturing the image statistics and this makes the approach very attractive for non-parametric methods in image processing.

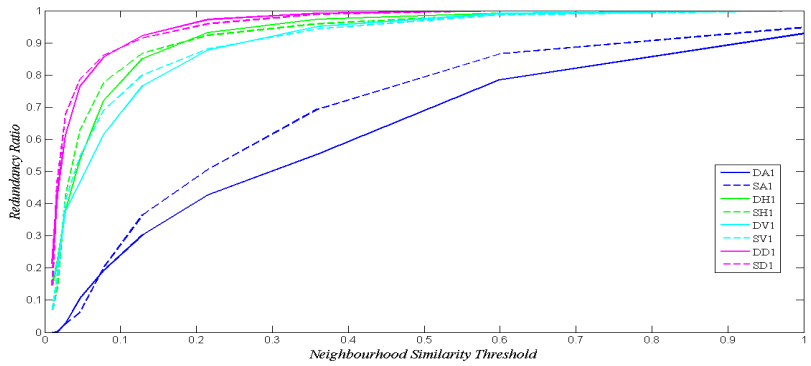


Redundancy Distribution for DWT (fixed redundancy factor)

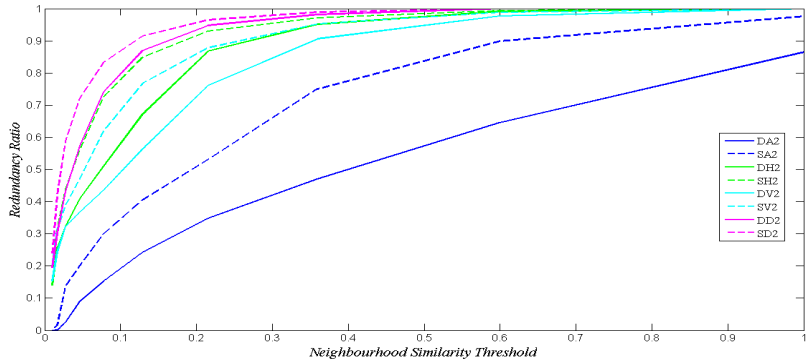


Redundancy Distribution for SWT(fixed redundancy factor)

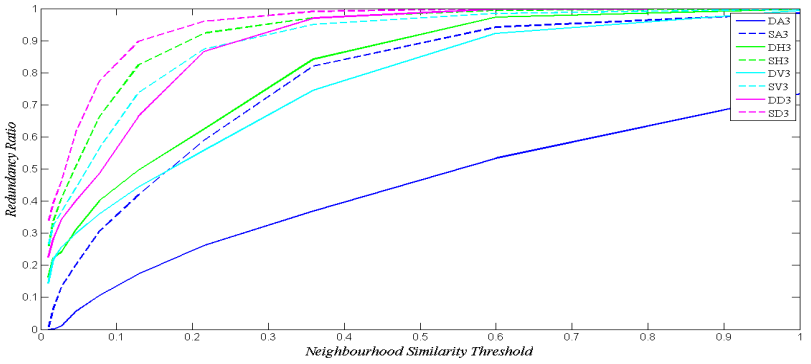
Figure 3.8: For a redundancy factor, $\beta = 0.0774$, we plot the variation of the redundancy ratio $\Theta(\tau, \beta)$ with varying τ. It can be seen that the neighbourhood redundancy decreases with higher decomposition levels for the DWT approximation coefficients, but for the SWT approximation coefficients this is not the case.



Level 1 Decompositions (DWT-SWT Comparison)



Level 2 Decompositions (DWT-SWT Comparison)



Level 3 Decompositions (DWT-SWT Comparison)

Figure 3.9: The undecimated wavelet coefficients exhibit more redundancy than the decimated wavelet coefficients, with the difference increasing with the levels ($\beta = 0.0774$).

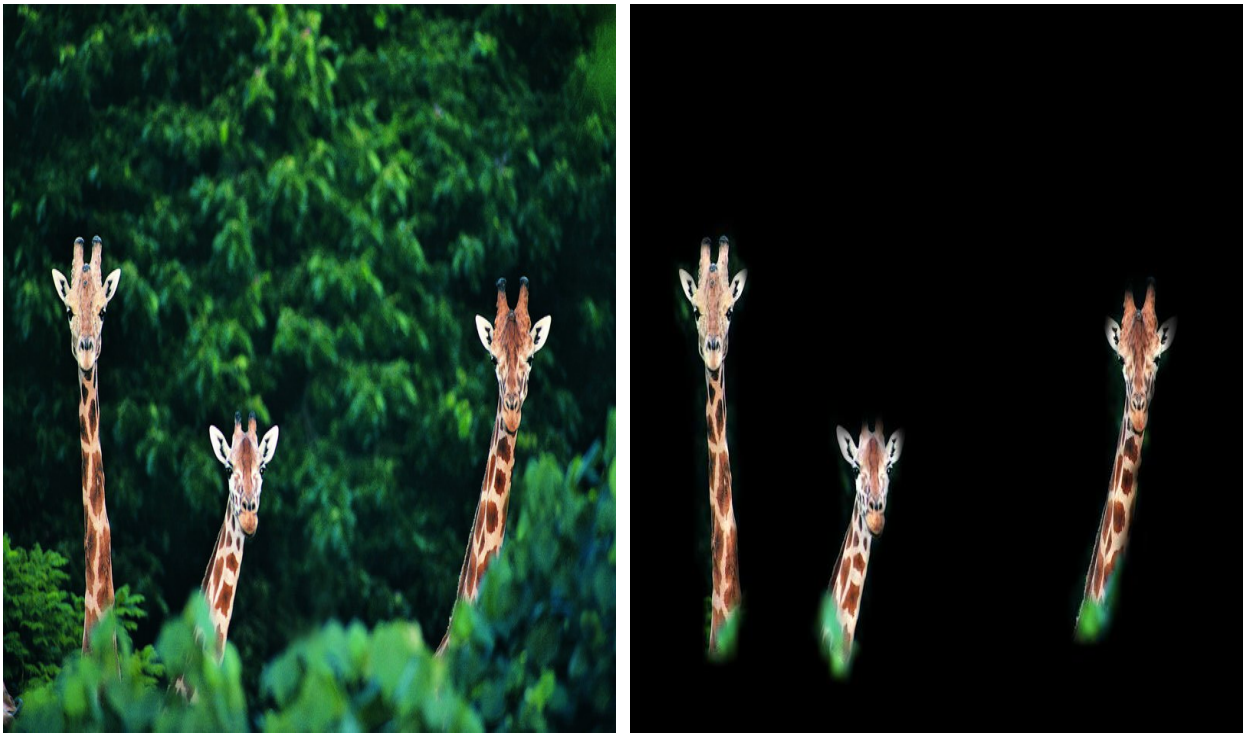
Chapter 4

Application to Saliency Detection in Natural Images

The automatic detection of saliency in images forms an important step in computer vision tasks such as image segmentation [29] and object recognition [30]. Traditional models of saliency detection are based on training, which makes them unsuitable for generalized tasks. As such, it is much desired to develop automatic methods for detecting saliency in images, without the need for prior knowledge of image content.

An observation that can be made with regards to saliency in images is that it is closely related to visual uniqueness in image attributes such as colours, edges, and boundaries. For example, in natural images such as that shown in Figure 4.1a, the salient objects of interest (e.g., giraffes) exhibit very different visual characteristics than the rest of the scene (the forest). From an information-driven perspective, one can also say that the object of interest is unique as it has low information redundancy within the given image compared to the rest of the scene, which is highly redundant. Motivated by this observation, our proposed work aims to detect saliency within an image without prior knowledge regarding the underlying image content by quantifying this non-redundant nature of salient objects using a statistical modelling approach. This quantification is accomplished in the proposed method by computing the statistical non-redundancy of each site in the image, given the other sites in the image based on site neighbourhoods. The resulting non-redundancy gives a saliency map that provides a rough classification of the image into regions which are salient (high statistical non-redundancy) and non-salient (low statistical non-redundancy), and can be used to identify the salient objects of interest.

This chapter is organized as follows. Previous methods used in saliency detection are



original image object image

Figure 4.1: In this example, the giraffes are the salient (unique) objects of interest that we wish to detect, as shown in the object image, relative to the background. In this figure the object image is based on ground truth.

discussed in section 4.1. The proposed method is described in section 4.2. Experimental results are presented in section 4.3 with conclusions drawn in section 4.4.

4.1 Previous Methods in Image Saliency

In recent years many general purpose saliency detection algorithms have been proposed. Itti and Koch proposed a saliency model based on the human visual search process [31, 32, 33, 34, 35]. Hou and Zhang proposed a spectral residual model [36], where the log-spectrum of an input image is analysed and the residual of the image is extracted in the spectral domain, which is then used to construct a saliency map in the spatial domain. This method simulates the behaviour of pre-attentive visual search and is an efficient method for early

stage visual processing. Wang et al. proposed a context-based model [37], where saliency is viewed as an anomaly relative to a context, which could be local or global, however this method needs a database of images for obtaining the extrinsic saliency map. Murray et al. [38] proposed an efficient model of colour appearance in human vision, which contains a principled selection of parameters as well as innate spatial pooling mechanism and can be generalized to obtain a saliency model, with scale integrated via a simple inverse wavelet transform over the set of weighted center-surround outputs and the reduction of ad-hoc parameters. Other methods include the frequency-tuned approach proposed by Achanta et al. [39], and the context-aware approach proposed by Goferman et al. [40]. In this section, we will discuss in detail the saliency detection methods which we have used for performance comparison with the proposed method.

4.1.1 Frequency Tuned Salient Region Detection

In [39], a frequency tuned saliency detection approach has been proposed, which will be referred to as the FT approach in subsequent parts of this chapter. The approach uses Difference of Gaussian filters for band pass filtering. This method retains the frequency content of the original image in the saliency maps. It gives full resolution saliency maps and the boundary information of the image is preserved. A DoG filter can be defined as

$$DoG(x, y) = G(x, y, \sigma_1) - G(x, y, \sigma_2) \tag{4.1}$$

where

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{\frac{-(x^2+y^2)}{\sigma^2}} \tag{4.2}$$

In equation 4.1, σ_1 and σ_2 are the standard deviations of the Gaussians. The band pass width for the DoG filter is determined by the σ_1/σ_2 ratio. It has been shown in [39] that the combined result of applying several band pass filters can be equivalently obtained by choosing a DoG with a large σ_1/σ_2 ratio. We define $\rho = \sigma_1/\sigma_2$, $\sigma_2 = \sigma$ and $\sigma_1 = \rho\sigma$. A summation over DoG with standard deviations in the ratio ρ results in

$$G_{combined} = \sum_{n=0}^{N-1} \left[G(x, y, \rho^{n+1}\sigma) - G(x, y, \rho^n\sigma) \right] \tag{4.3}$$

which results in

$$G_{combined} = G(x, y, \rho^N\sigma) - G(x, y, \sigma) \tag{4.4}$$

for an integer $N \geq 0$, which is equivalent to difference of two Gaussians, i.e., a band pass filter with σ_1/σ_2 ratio equal to ρ^N. This is essentially a parametric method where

parameter selection of σ_1 and σ_2 has to be done to retain the desired frequency range in the saliency maps. The saliency map S can be defined as

$$S(x, y) = |\mu - G_{combined}(x, y)| \tag{4.5}$$

where μ is the mean value of the image pixel intensities. There is no downsampling involved in the computation of saliency maps in this method. Hence we obtain full resolution saliency maps.

4.1.2 Global Contrast Based Salient Region Detection

Local contrast saliency extraction methods evaluate saliency of a region with respect to small neighbourhoods [33, 40, 41, 42, 43, 44]. These methods essentially focus on the edges in the image rather than uniformly highlighting the salient region. The global contrast method proposed in [45] computes saliency of an image region using its contrast with respect to the entire image. In the histogram-contrast based method (the method will be referred to as the HC method), colour statistics of the input image are used to define saliency values for image pixels. The colour of any arbitrary pixel i in the image is compared and contrasted to all other pixels in the image. The saliency map is defined as

$$S_i = \sum_{j \in \chi} D(i, j) \tag{4.6}$$

where S_i denotes the saliency value at pixel i, χ is the set of pixels comprising the image and $D(i, j)$ represents the colour distance metric between pixels i and j in the L*a*b* colour space [46]. From equation 4.6, it can be seen that pixels with the same colour would have same saliency value since spatial relationships between the pixels are not being taken into account at this stage. Grouping together pixels with the same colour value we get

$$S_i = S(c_l) = \sum_{j=1}^{n} D(c_l, c_j) \tag{4.7}$$

where n is the number of possible colours in the image. In [45], a quantization and smoothing process is employed to refine the saliency value for each pixel, where the weighted average of saliency values of similar colours is used. This saliency algorithm is closely related to an algorithm proposed in [46], where only luminance values of the pixels are considered, and the colour information is not taken into account. The method proposed in [46] is known as the luminance contrast method and will be referred to as the LC method for the remainder of this chapter.

In the region-contrast approach (which will be referred to as the RC approach), in addition to contrast, the spatial relationships of pixels are also taken into account. In this approach, the authors first segment the input image into regions, then compute the colour contrast at the region level, and define the saliency for each region as the weighted sum of region's contrasts to all other regions in the image. The weights are assigned in order of spatial distances with closer regions being assigned more weight and vice versa. The mathematical details for both the methods are given in [45].

4.1.3 Spectral Residual Based Saliency Detection

In [36], the authors have claimed that natural images are not completely random, they obey highly predictable distributions. If these statistically redundant parts of the image are removed, the "unique" part of an image can be highlighted. It has been observed that log spectra of different images share similar trends though they do contain statistical singularities [36]. These singularities in the spectrum may be responsible for unique or salient regions in the image. In the saliency detection method proposed in [36], the input image is downsampled to a 64 x 64 size and its log spectrum $L(f)$ is computed. The general shape of the log spectra $A(f)$ is computed by convolution of the log spectra $L(f)$ with an N x N averaging filter $h_n(f)$. $A(f)$ and $L(f)$ can be used to extract $R(f)$, i.e., the statistical singularities particular to the input image. $R(f)$ has been defined as the spectral residual of the image

$$R(f) = L(f) - A(f) \tag{4.8}$$

Using an Inverse Fourier Transform on $R(f)$ gives a saliency map in the spatial domain, which contains the unique part of the scene. In [36], the saliency maps have been smoothed with a Gaussian filter. The saliency detection algorithm discussed is known as the spectral residual algorithm, and will be referred to as the SR method in the rest of the chapter.

4.2 Image Saliency Based on Non-parametric Neighbourhood Based Methods

In natural images, salient objects of interest can be characterised by their statistical non-redundancy with respect to a given image. Our method aims to quantify the statistical non-redundancy (sections 4.2.1 and 4.2.2) to obtain a saliency map (section 4.2.3) which divides the image into salient and non salient regions, which can then be used to obtain object maps (section 4.2.4). The method has also been discussed in [47].

4.2.1 Statistical Non-redundancy Between Sites

Let N_i be a set of image pixels in the neighbourhood of site i. Suppose that the neighbourhood around an arbitrary site i in an image can be seen as a realization of the neighbourhood around some other site j:

$$N_i = N_j + \eta_{ij}, \tag{4.9}$$

where η_{ij} is a noise process between the two site neighbourhoods following some distribution $P(\eta_{ij})$. Suppose we model $P(\eta_{ij})$ as an independent and identically distributed Gaussian process with zero-mean and variance σ. Therefore, the probability of an arbitrary site k in the neighbourhood N_i being a realization of the corresponding site in the neighbourhood N_j (can also be defined as the probability of a site in N_i being statistically redundant with respect to N_j, as was defined in section 3.2) can be defined as:

$$P\big(f(N_i^k)|f(N_j^k)\big) = e^{\frac{-\left(f(N_i^k)-f(N_j^k)\right)^2}{\sigma^2}} \tag{4.10}$$

Therefore, one can quantify the statistical redundancy between the neighbourhoods as:

$$P\big(f(N_i)|f(N_j)\big) = \prod_k P\big(f(N_i^k)|f(N_j^k)\big). \tag{4.11}$$

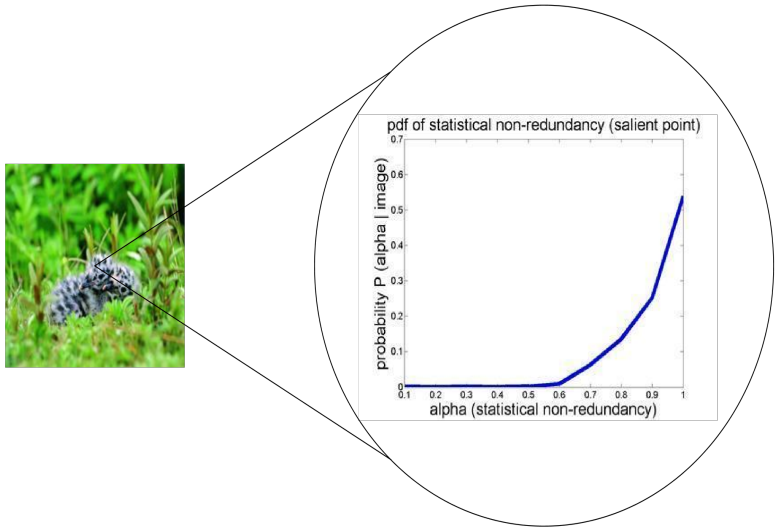
The statistical non-redundancy α_{ij} between two neighbourhoods N_i and N_j has been defined in equation 3.2 in chapter 3.

4.2.2 Statistical Non-redundancy Within the Image

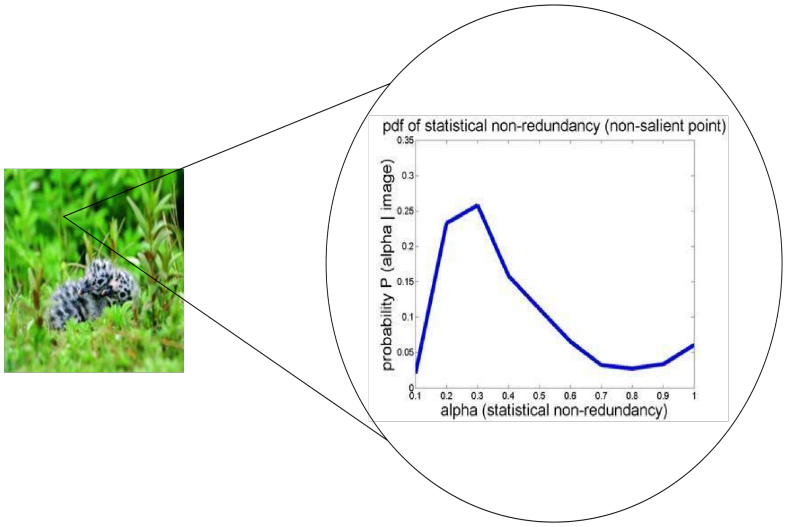
We had discussed in section 3.2 that $P(\alpha_{ij}|f)$ gives us an overall picture of the statistical non-redundancy of site i given the image f. Figure. 4.2 shows $P(\alpha_{ij}|f)$ for two different sites within the same image. We can see that the salient site has a higher degree of non-redundancy as compared to the non-salient site. This observation agrees with our expectation as the salient sites would have fewer matching neighbourhoods. Therefore, it can be observed that the overall statistical non-redundancy trend of a site gives us a good indicator of image saliency.

4.2.3 Saliency Map

We have shown that salient sites in the image have a higher overall statistical non-redundancy than non-salient sites. Salient regions in the image are unique and the neigh-



salient site



non-salient site

Figure 4.2: $P(\alpha_{ij}|f)$ for two sites in the same image. As expected, the distribution shows a significant saliency sensitivity.

bourhoods centred around the salient sites have fewer matching neighbourhoods as compared to the non-salient sites. We can use this fact to quantify the statistical non-redundancy of every site in the image and thus generate a saliency map, where the intensity at each site is proportional to the statistical non-redundancy of the corresponding site in the original image. In the proposed method, for each site i in image f, we compute $P(\alpha_{ij}|f)$ based on all sites j in f. The non-redundancy map defined in equation 3.3 can be used as a saliency map for a given image. The saliency map divides the image into high intensity and low intensity sites. The high intensity points in the saliency map correspond to the salient sites in the original image. This intuitively makes sense as a site corresponding to a unique region would have low information redundancy within the image and hence a higher value for statistical non-redundancy. For colour images, the saliency map is based on the aggregate expectation across all colour channels. While deriving the saliency maps for images of high resolution, using the proposed method, it was observed that computing the expected saliency value over all possible pixel pairs makes the algorithm computationally intensive. Instead of averaging α_{ij} over all possible j, if we sample a set of pixels randomly and average α_{ij} over the sampled set of pixels, we get saliency maps without much loss of accuracy with significantly better computational intensity.

$$S_i = \underset{j \in \psi}{E}(\alpha_{ij}|f) \tag{4.12}$$

where ψ represents the sampled set of pixels. The algorithm proposed above will be referred to as the NR (Non-redundancy) algorithm in the rest of this chapter.

4.2.4 Generating the Object Image

Given the computed saliency map, one can get a rough idea of where the salient objects of interest are by applying an appropriate threshold on a saliency map, which can then be used to generate an object image. Choosing a low threshold would unnecessarily include sites from the background (false positives) in the salient object region of the object image. On the other hand choosing a high threshold would lead to sites from the object region being neglected in the object image (false negatives). For illustrative purposes, we generate an object image by applying a threshold to the original image using the mean absolute deviation of the saliency map as the threshold (denoted by T):

$$T = E\big(|X - E(X)|\big) \tag{4.13}$$

For our purposes, X is the vector containing all sites in the saliency map. Figure 4.3 shows the object images derived from saliency maps for some natural images.

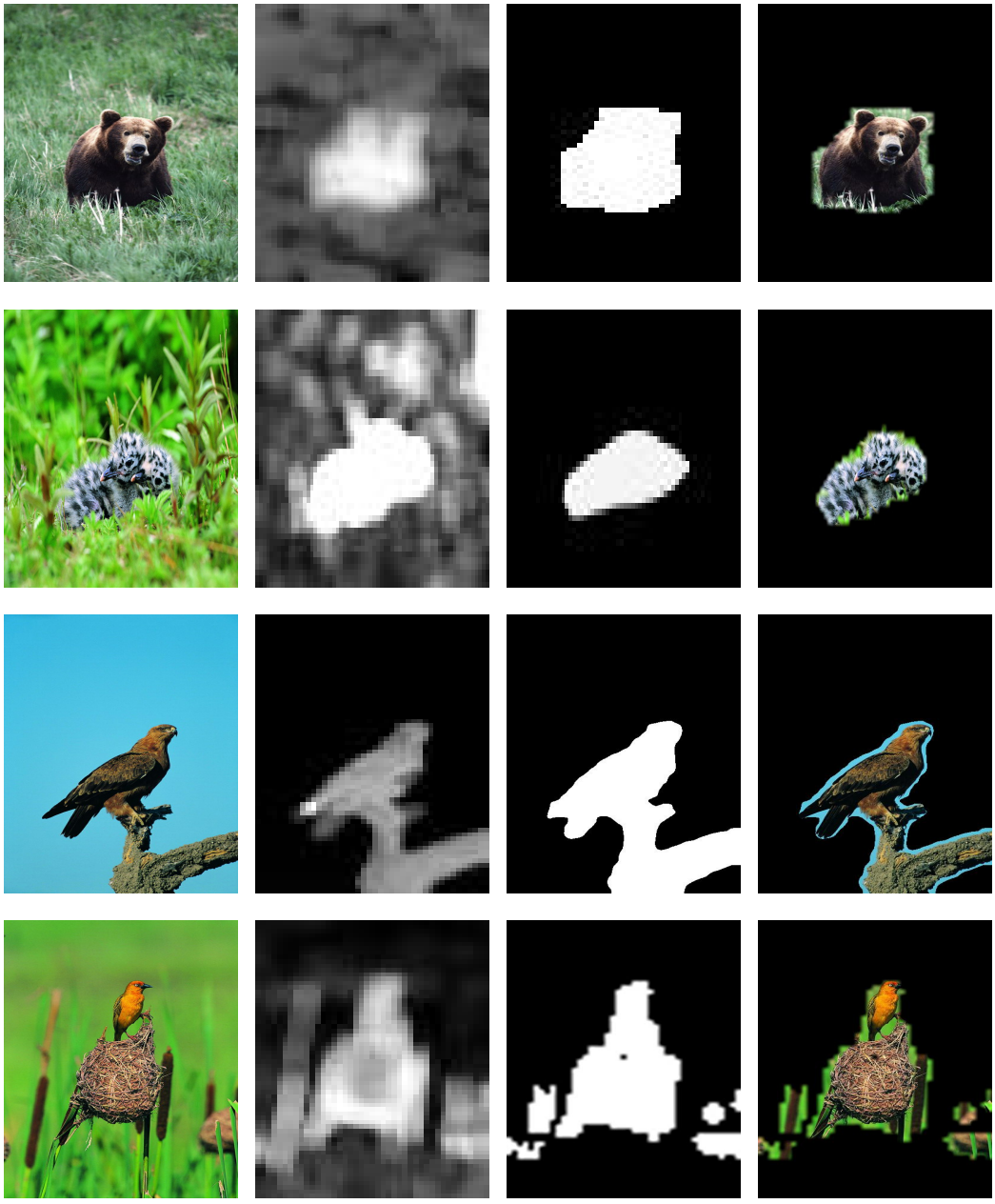


Figure 4.3: A threshold has been applied to the saliency maps to obtain the object images for the natural images. The algorithm marks out the salient objects in the images with a good precision.

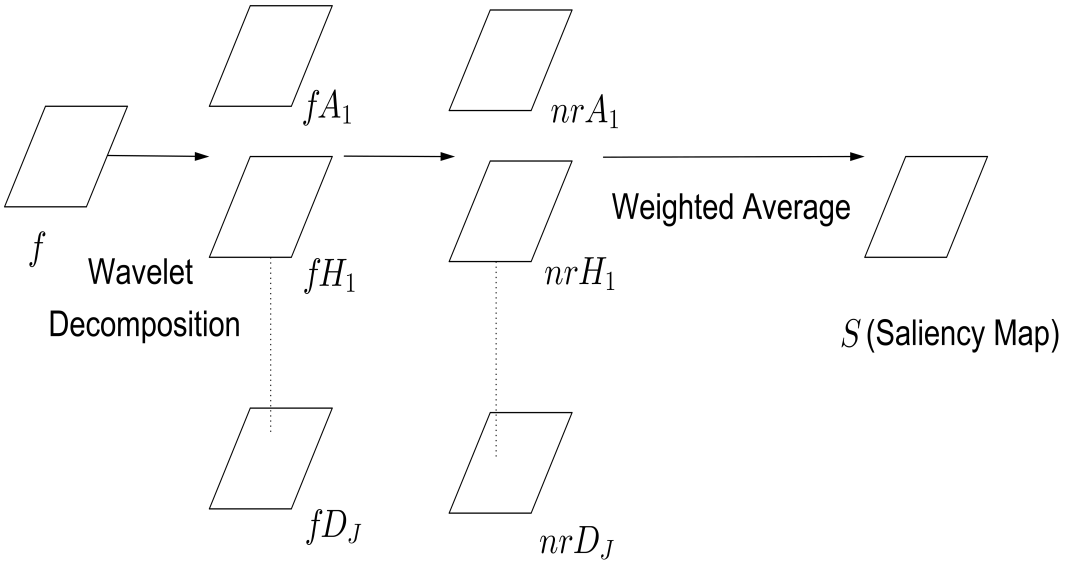


Figure 4.4: This figure illustrates the steps involved in the Multi-scale saliency detection algorithm.

4.2.5 Multi-scale Saliency Detection

In section 4.2.3, we proposed a saliency detection algorithm which exploits the non-redundancy statistics of the pixel neighbourhoods to obtain saliency maps. We now extend this concept to the multi-scale non-parametric model discussed in chapter 3. In the multi-scale saliency detection algorithm, we exploit the non-redundancy statistics of the wavelet sub-band coefficients to obtain non-redundancy maps for the different sub-bands. We then compute a weighted average of these non-redundancy maps to get our saliency map. This process is illustrated in Figure 4.4. In our experiments, we have considered the sub-bands A_1, A_2 and A_3 (the approximation coefficients at three decomposition levels) as the approximation coefficients give us the overall picture of the salient objects in the image, whereas the detail coefficients would highlight the edges as can be seen in Figure 3.4. From Figure 3.4, we can see that the non-redundancy maps at higher decomposition levels highlight salient regions at coarse scales. Computing a weighted average of the approximation sub-bands at all the levels and assigning more weights to sub-bands at higher decomposition

levels ensures that the entire salient object is highlighted in the saliency maps, and salient regions at all scales are considered.

The multi-scale approach helps capture saliency in the image at different scales, where the fine sub-bands give fine saliency maps and the coarse sub-bands give coarse saliency maps. Also, the multi-scale saliency approach based on weighted averages discussed above is very robust to noise. This can be attributed to the fact that noise components of an image are filtered out in the approximation coefficients at higher decomposition levels. This fact is validated in section 4.3 where we discuss the performance comparison of the saliency algorithms proposed, with each other and with other popular saliency detection algorithms.

4.3 Performance Comparison with Other Methods

In this section we use a quantitative measure for performance evaluation of the proposed algorithm against the current state-of-the-art in saliency detection. We also compare the different configurations of our proposed algorithm. We use the precision-recall curves [48] for evaluating the quality of saliency maps produced by these algorithms. In our tests, we have taken images from the Achanta database [39] and from the database used in the spectral residual paper [36]. For each experiment, we have considered 20 images from these databases and have computed and plotted the average precision-recall characteristics for these images. In the rest of the section, we refer to our non-redundancy based saliency detection method as the NR method, and the other methods follow the same abbreviations as discussed in section 4.1. In these experiments, we have used 300 random samples (uniformly distributed over the whole image) for generating the saliency maps. Also, for computing the saliency maps of coloured images using our NR method, we calculate the NR based saliency maps for each of the channels (Hue, Saturation and Value) and then take the average of these maps.

4.3.1 Comparison of Non-redundancy Algorithms

The statistical non-redundancy algorithm proposed in section 4.2 can have different configurations. The saliency map might be produced directly from the image in the spatial domain, as is done in section 4.2.3, or it can be produced from the weighted average of the non-redundancy maps of the wavelet sub-bands, as discussed in section 4.2.5. When we consider the multi-scale case, the wavelet sub-bands chosen for the generation of the

saliency maps affect the performance of the algorithm. In this experiment, we consider three configurations for the decimated wavelet case

1. D1: In this configuration we take a level 1 DWT decomposition of an image and use the DWT level 1 approximation coefficients ($DA1$) for getting the saliency map, i.e., the saliency map in this case is $nrDA1$, i.e., the non-redundancy map for $DA1$.

2. D2: In this configuration, we take a level 2 DWT decomposition of the image and use the weighted average of the non-redundancy maps of level 1 approximation coefficients ($nrDA1$) and level 2 approximation coefficients ($nrDA2$) for producing the saliency map, i.e., the saliency map can be defined as

$$D2 = (w1 * nrDA1 + w2 * nrDA2) / (w1 + w2) \tag{4.14}$$

   where $w1$ and $w2$ are the weights assigned to the non-redundancy maps obtained from $DA1$ and $DA2$ respectively. In our experiments, we have fixed $w1 = 1$, and $w2 = 100$, to give more importance to the non-redundancy maps corresponding to coarser scales.

3. D3: Similar to the above cases, the saliency map for this case is the weighted average of $DA1$,$DA2$ and $DA3$. It can be defined as

$$D3 = (w1 * nrDA1 + w2 * nrDA2 + w3 * nrDA3) / (w1 + w2 + w3) \tag{4.15}$$

   where $w1, w2$ and $w3$ are the weights assigned to the non-redundancy maps obtained from $DA1, DA2$ and $DA3$ respectively. In our experiments, we have fixed $w1 =$ 1,$w2 = 100$ and $w3 = 1,000,000$ to give more importance to the wavelet sub-bands corresponding to the coarser scales.

Similar to the case of decimated wavelets, we have three configurations of our algorithm for the undecimated wavelet case; S1, S2 and S3 which represent the saliency maps corresponding to the level 1, level 2 and level 3 approximation coefficients of the SWT decomposition of the image. We compute the saliency maps for these configurations for 20 different images for performance comparison. In Figure 4.5, we have shown the saliency maps for the NR based algorithms for seven selected images. In Figure 4.6, we can see the precision-recall characteristics for the different configurations of the NR algorithm. The precision-recall characteristics shown in this image are the average of 20 different images taken from the Achanta [39] database. It can be seen that the S3 configuration of the NR algorithm shows the best performance. In the next section, we will see the effect of additive noise on the performance of the NR algorithm in its various configurations.

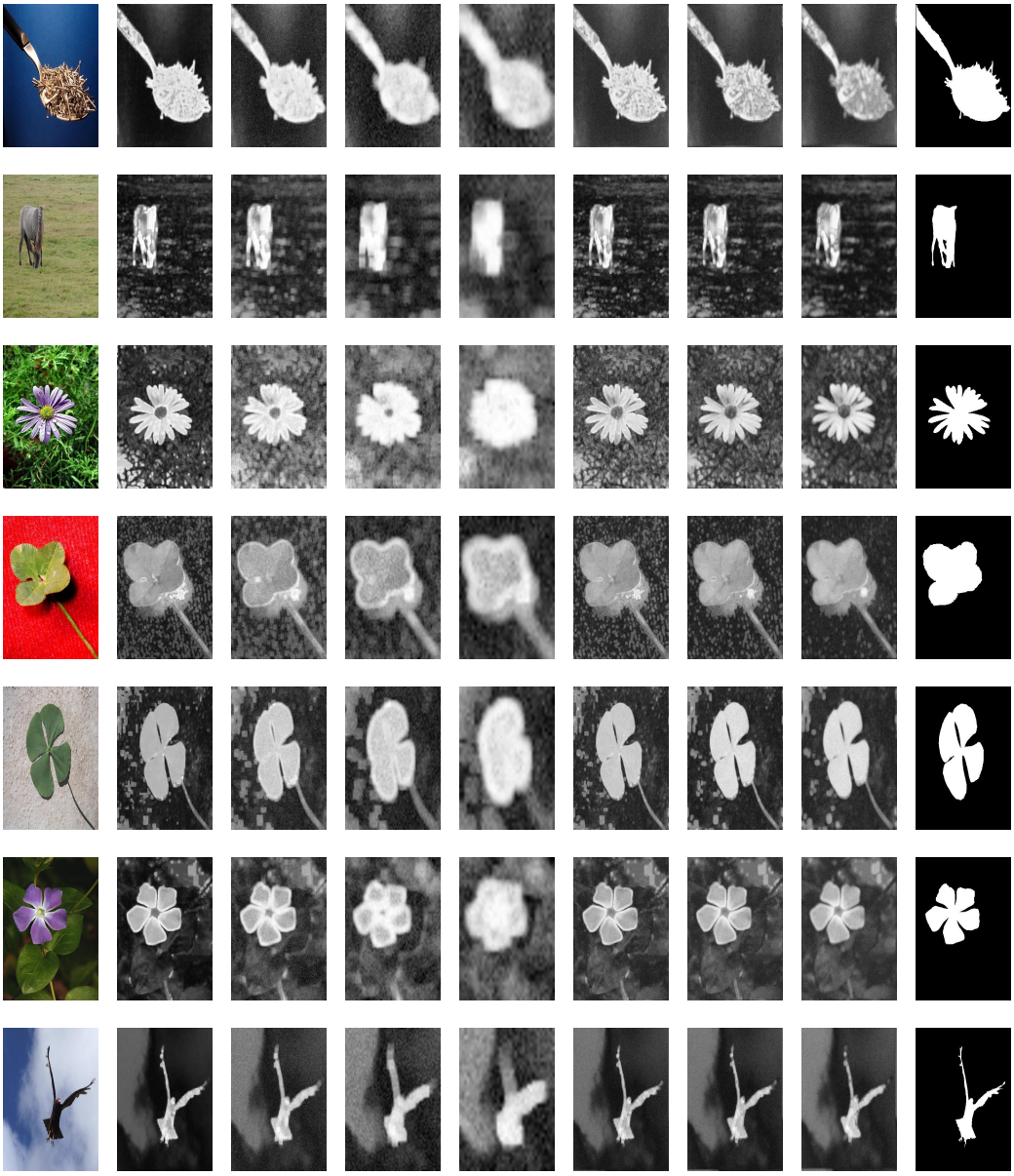


Figure 4.5: This figure shows the saliency maps for seven different images produced using the different configurations of the NR algorithm.

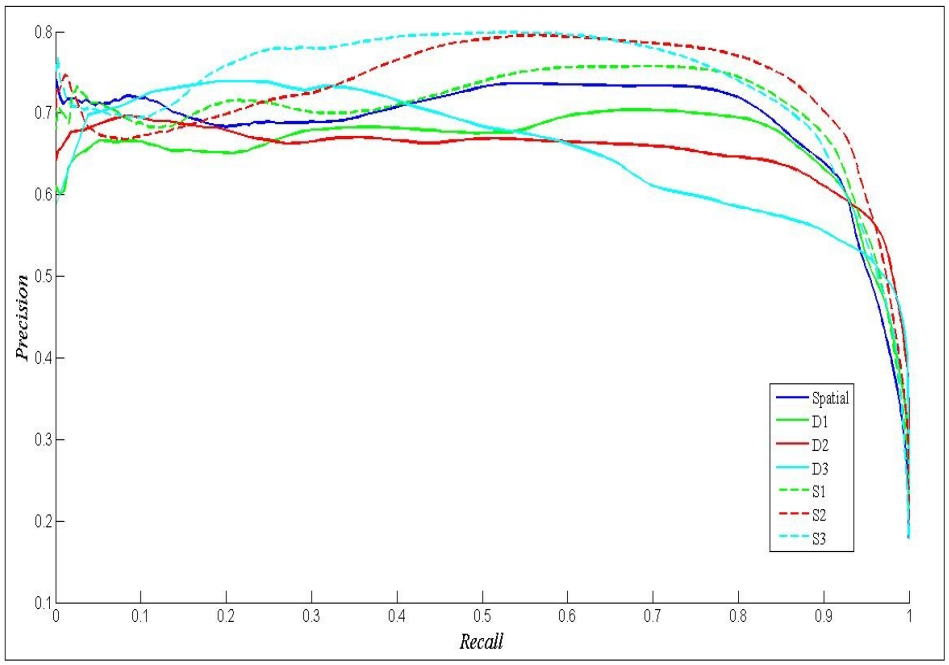


Figure 4.6: This figure shows the precision-recall characteristics for the NR saliency algorithms taken over 20 different images. We can see that the S3 configuration of the NR algorithm gives the best performance.

4.3.2 Effect of Noise to Non-redundancy Algorithms

In this section, we use the same image set of 20 images used in the previous section and corrupt the images with an additive white Gaussian noise of standard deviation $\sigma = 0.05$ (on a scale of 1). The NR saliency detection algorithm, in all its configurations, is then applied to these noisy images and the performance is evaluated. In Figure 4.7, we can see the saliency maps for seven images obtained using the different configurations of the NR algorithms. It can be seen that the saliency maps obtained from the undecimated wavelet based approach give sharper and more well defined saliency maps. In Figure 4.8, we see the corresponding precision-recall characteristics. It can be seen that the wavelet based configurations have a significantly better performance in presence of noise. The undecimated wavelet based configuration outperforms the decimated wavelet based configuration in the noisy as well as the noiseless cases. This proves that undecimated wavelet based saliency approach is robust to noise and also gives more detailed saliency

maps.

4.3.3 Performance Comparison with State-of-the-art

In the above sections we concluded that the S3 configuration of the proposed NR algorithm gives the best performance in terms of precision-recall characteristics. In this section, we will consider the saliency maps produced by this configuration and compare them to the saliency maps produced by the current state-of-the-art in saliency detection; Frequency Tuned (FT) [39], Histogram Contrast (HC) and Region Contrast (RC) [45], Luminance Contrast (LC) [46] and Spectral Residual (SR) [36].

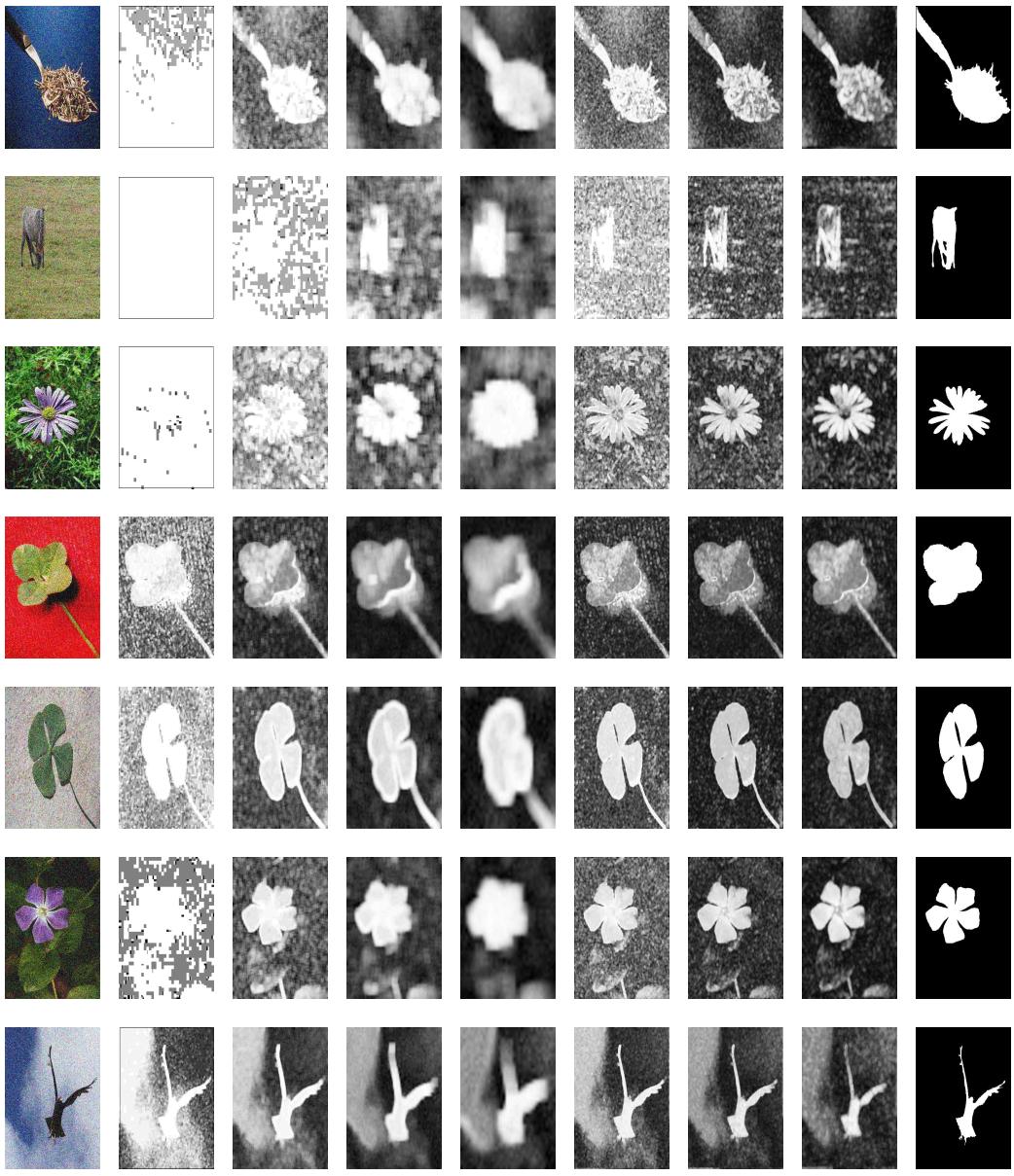


Figure 4.7: This figure shows the saliency maps for the seven noisy images (additive Gaussian noise, $\sigma = 0.05$) produced using the different configurations of the NR algorithm. The image database is same as the one used in Figure 4.5. We can see the undecimated wavelet based saliency maps (S1,S2 and S3) mark out the salient regions from the background much more clearly.

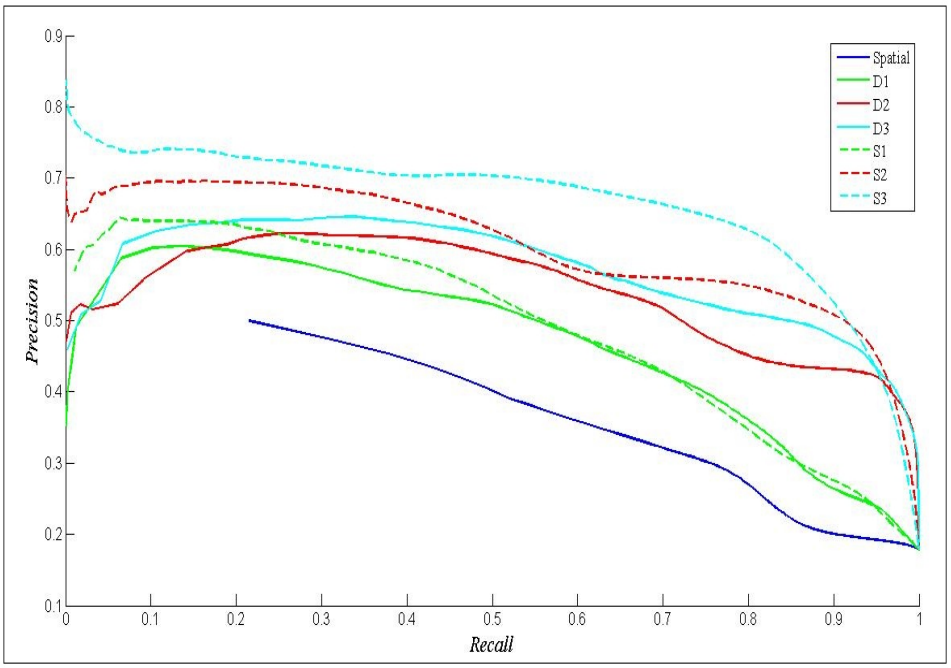


Figure 4.8: This figure shows the precision-recall characteristics for the NR saliency algorithms taken over 20 different noisy (additive Gaussian noise, $\sigma = 0.05$) images. The image set is the same as that considered in the case of Figure 4.6. We can see that the S3 configuration of the NR algorithm gives the best performance, and there is significant difference in the performance of the spatial and wavelet based configurations.

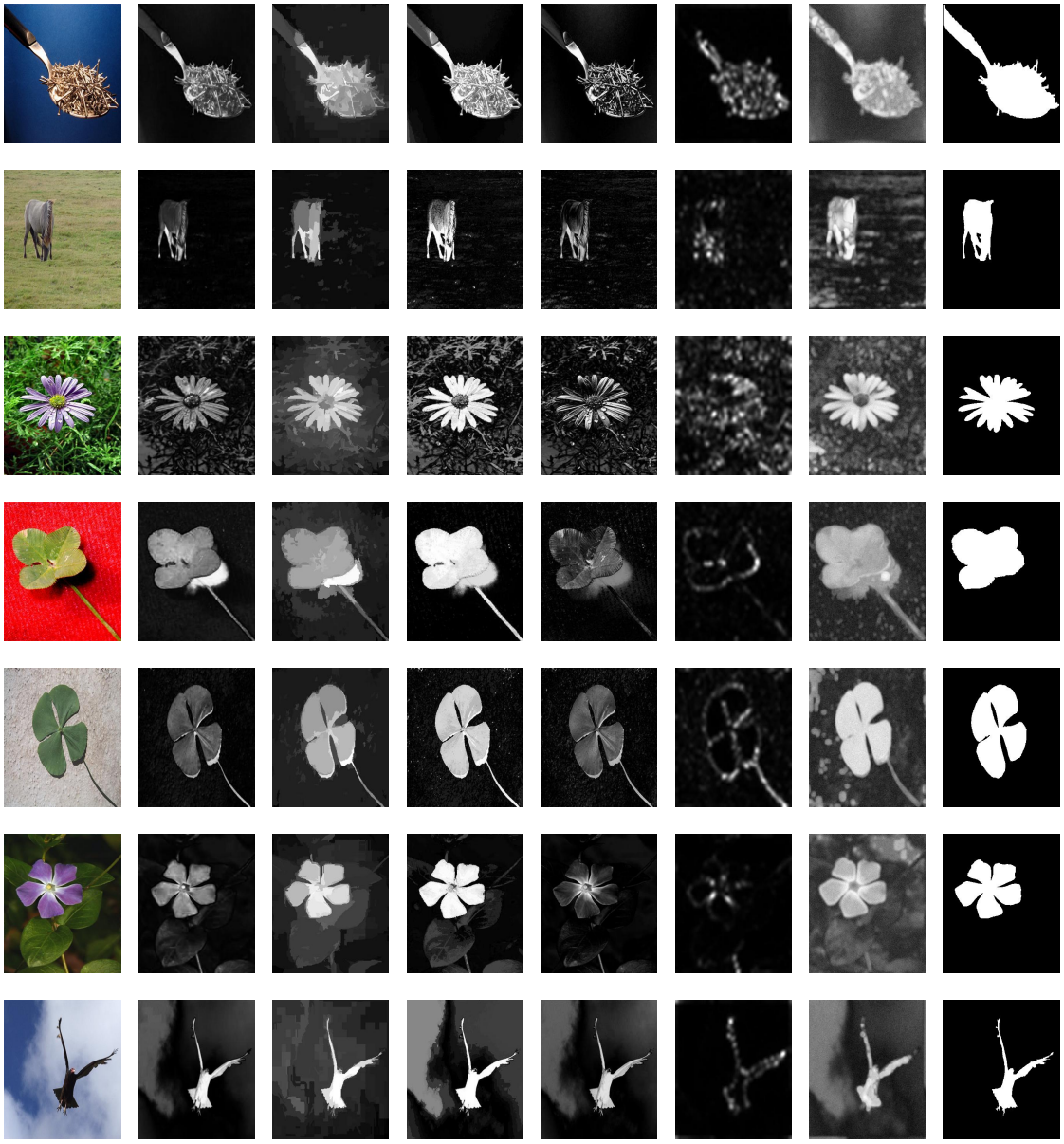


Figure 4.9: This figure shows the saliency maps for the state-of-the-art in saliency detection algorithms and the S3 configuration of our proposed NR algorithm (denoted as SWT3).

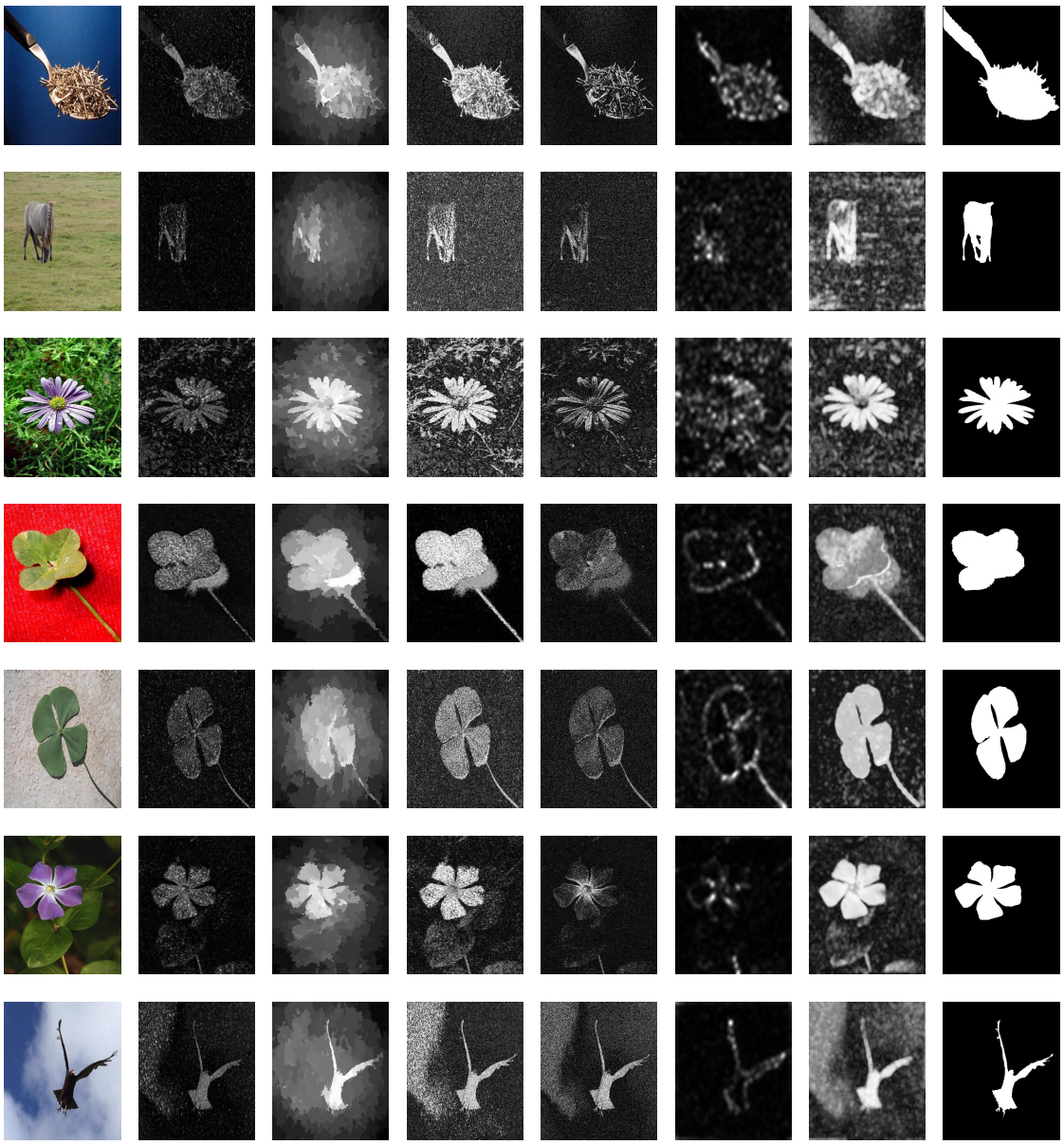


Figure 4.10: This figure shows the saliency maps for the state-of-the-art in saliency detection algorithms and the S3 configuration of our proposed NR algorithm (denoted as SWT3) for noisy images (additive Gaussian noise, $\sigma = 0.05$). For most algorithms, we can see a significant degradation in saliency maps when compared to Figure 4.5.

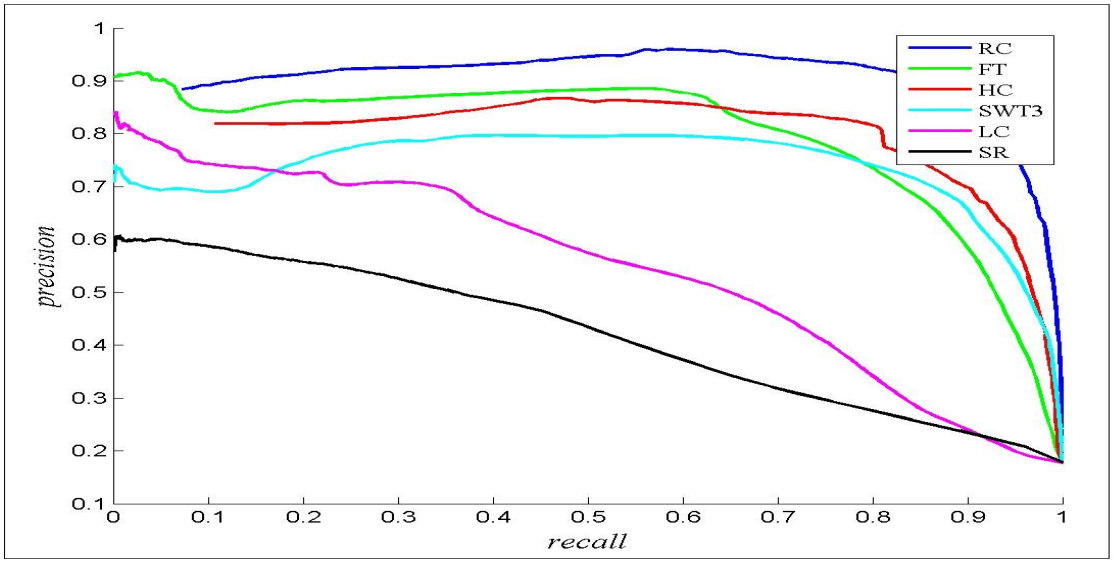


Figure 4.11: This figure shows the average precision-recall characteristics (taken over 20 images) of some popular saliency detection algorithms along with the proposed NR algorithm with S3 configuration (labelled as SWT3).

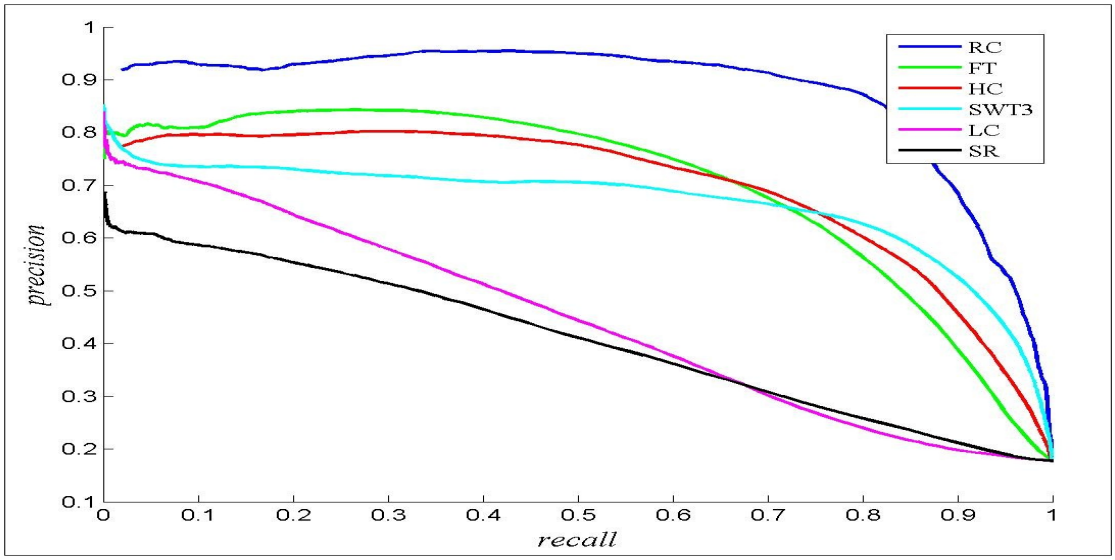


Figure 4.12: Similar to Figure 4.11, this figure shows precision-recall characteristics for 20 noisy images (additive Gaussian noise, $\sigma = 0.05$).

4.4 Conclusions and Further Work

We have used the non-parametric multi-scale framework proposed in chapter 3 to develop a statistical non-redundancy based saliency detection algorithm. We have compared the different configurations of the proposed algorithm and have observed that the S3 configuration of the algorithm gives the best results in terms of precision-recall characteristics. This can be attributed to the fact that in this configuration, we are using full resolution non-redundancy maps which captures the non-redundancy statistics of the image at multiple scales. Also, this configuration is the most robust to noise, which can be attributed to the fact that most of the noise is captured in the detail coefficients of the wavelet decompositions, and the level 2 and level 3 approximation coefficients do not have much noise.

The proposed algorithm outperforms the SR algorithm and the LC algorithm. However, in the present configuration it does not outperform the other methods, i.e., FT, HC and LC methods. The performance of the proposed algorithm can be improved in several ways

1. Instead of the random sampling approach which we have used in equation 4.12, we can use guided sampling approaches
2. Different weights can be assigned to the non-redundancy maps while computing the saliency maps.
3. In the proposed approach, we take a simple average of the saliency maps obtained from the different channels (Hue, Saturation and Value). Instead of taking a simple average, a weighted average approach could give better saliency maps.
4. Use of rotation invariant filters while computing the statisical non-redundancy can give better results.

The proposed approach for saliency detection has scope for improvement but it gives promising results for the non-parametric multi-scale framework developed in chapter 3, as it shows that the application of the framework can give better algorithms with more detail preservation and robustness to noise.

Chapter 5

Application to Image Denoising

Digital images could be corrupted by noise due to various reasons which are affected by the way images are captured and transmitted, e.g., images transmitted through a wireless medium might be corrupted by noise due to atmospheric conditions. In all applications involving images, it is desirable to extract images in their denoised form, i.e., to get an estimate of the original image. This requires an understanding of the image statistics and the noise statistics. The understanding can be used to model the degradation process and apply the inverse process in order to recover the original image. The non-parametric multi-scale framework developed in chapter 3 of the thesis has been applied to the problem of image denoising. In section 5.1, we discuss some denoising algorithms which employ the multi-scale wavelet transforms and have been proposed in recent years. In section 5.2, we discuss the proposed denoising algorithm. In section 5.3, we compare the proposed method with the state-of-the-art in image denoising, followed by results and conclusions in section 5.4.

5.1 Use of Wavelets in Image Denoising

Wavelet based image denoising methods [10, 19, 49, 50, 51] have recently gained a lot of popularity. Wavelets are powerful signal processing tools as they enable efficient removal of noise without losing much of the signal information. This can be done because in a wavelet decomposition, in the detail coefficients, the small coefficients are mostly due to noise and the large coefficients are due to important signal features. Modifying these small detail coefficients can suppress the noisy component of the image without significant effect on the underlying structure of the image.

Wavelet shrinkage methods are very simple to implement. We consider an image f and apply a J level wavelet transform to it. We thus obtain the approximation coefficients A_J and the detail coefficients $\{H_i, V_i, D_i\}_{i=\{1,2..J\}}$. To each detail coefficient, a threshold rule is applied. Soft thresholding [49] involves shrinkage of each coefficient and can be represented as

$$\eta_s(w,t) = w - t \qquad w \geq t \tag{5.1}$$
$$\eta_s(w,t) = 0 \qquad |w| < t \tag{5.2}$$
$$\eta_s(w,t) = w + t \qquad w \leq -t \tag{5.3}$$

where w represents an individual detail coefficient and η represents the value of the detail coefficient after the application of the thresholding rule. Hard thresholding [49] shrinks only those coefficients which are lower than the specified threshold and leaves other coefficients unaffected. It can be represented as follows

$$\eta_h(w,t) = w \qquad |w| \geq t \tag{5.4}$$
$$\eta_h(w,t) = 0 \qquad |w| < t \tag{5.5}$$

The inverse wavelet transform is then applied on these modified wavelet coefficients to obtain a denoised image. The threshold chosen could be the same for all the resolution levels (1, 2...J) or could be different for each level. The way the threshold is selected gives rise to a number of wavelet shrinkage methods such as SureShrink [49], VisuShrink [49], NeighShrink etc. [49, 52, 53, 54].

5.1.1 VisuShrink

In the VisuShrink thresholding method [49], a common threshold t is proposed for all resolution levels.

$$t = \sqrt{2logn} \tag{5.6}$$

where n is the total number of transform coefficients in the decomposition [49]. This approach assumes the noise in the image to be additive white Gaussian noise. If soft thresholding is done with this threshold, the noise in the image is removed, but at the cost of important features in the image [49]. Hard thresholding is better at feature preservation but does not give smooth fits [49]. This can be attributed to the fact that a limited number of coefficients are pushed down to zero without affecting the other coefficients.

5.1.2 SureShrink

SureShrink [49] is a level-dependent thresholding technique in which t_j is estimated from the detail coefficients at the corresponding resolution level.

5.1.3 NeighShrinkSURE

NeighShrinkSURE [55] is an improvement over the NeighShrink [56] method proposed for wavelet shrinkage. The NeighShrink method uses universal thresholds and identical neighbourhood window sizes for all wavelet sub-bands. In the NeighShrinkSURE method, an optimal procedure is used for the selection of neighbourhood window sizes and thresholds for each wavelet subband. The optimal threshold and neighbouring window size for every wavelet sub-band is determined using Stein's Unbiased Risk Estimate (SURE) [57]. This method gives the best PSNRs among all the current wavelet shrinkage methods [55].

5.2 Non-parametric Approach Based Denoising

In this chapter, we propose and discuss a wavelet based non-parametric multi-scale approach for image denoising. Given a noisy image, we want to obtain an estimate of the denoised image, without using any prior information about the image. In the proposed approach, we exploit MCMC (Markov-Chain Monte Carlo) sampling of wavelet coefficients at each sub-band to derive a non-parametric multi-scale description of the noisy image, which is then used for the estimation of the denoised image. In section 5.2.1 we the stochastic denoising approach based on the MCMC sampling. In section 5.2.2 we apply this concept to the wavelet domain based on the non-parametric framework developed in chapter 3.

5.2.1 MCMC Sampling Based Non-parametric Image Denoising

The image denoising problem can be seen as the estimation of the original image given the noisy image measurements and some image or noise statistics, based on which a mathematical model can be developed [58]. Information redundancy in small local neighbourhoods has proved to be useful in several image denoising algorithms, e.g, box filtering and Gaussian filtering [59], Lee filtering [60], total variation [61], Bayesian Estimation [62], anisotropic filtering [63], bilateral filtering [64] and trilateral filtering [65]. A number of

algorithms which exploit information redundancy over the whole image have also been proposed [66, 67, 68, 69].

Let χ denote a set of pixel (locations) representing an image g such that $g = \{g(i)|i \in \chi\}$, where $i \in \chi$ represents any arbitrary pixel in the set. Let $f = \{f(i)|i \in \chi\}$ denote the noise contaminated image, and $n = \{n(i)|i \in \chi\}$ denote the noise. In an additive image noise model, the relationship between g, n and f can be represented as

$$f(i) = g(i) + n(i) \quad \forall\, i \in \chi \tag{5.7}$$

In chapter 3, we had proposed a multi-scale non-parametric model in which pixel neighbourhood interactions of the image are used for capturing the image statistics. In section 3.2, we use the pixel neighbourhood interactions to derive the non-redundancy statistics of an image. For each pixel in the image, we derive a probability distribution of the statistical non-redundancy of that pixel neighbourhood with respect to other pixel neighbourhoods in the image, which is used for deriving the saliency map. The denoising algorithm uses the same framework, but instead of using the non-redundancy distributions of the pixel neighbourhoods in the image, we use the redundancy distributions of the pixel neighbourhoods. In equation 3.1 we define the probability $P\big(f(N_i)|f(N_j)\big)$ of a pixel neighbourhood N_i being a realization of another pixel neighbourhood N_j. This gives a measure of the redundancy β_{ij} of the pixel neighbourhood N_i with respect to the pixel neighbourhood N_j

$$\beta_{ij} = P\big(f(N_i)|f(N_j)\big). \tag{5.8}$$

Computing β_{ij} for all possible values of j gives an overall redundancy distribution of the pixel i with respect to the given image f. This redundancy distribution of each pixel is used for the generation of MCMC samples [58], which can be used for deriving a conditional distribution $p\left(g(i)|f(i)\right)$ of the spatial intensity for each pixel $g(i),\ i \in \chi$ in the denoised image g, given the noisy image f . The expected value of each of these conditional distributions gives an estimate $\hat{g}(i)$ of the pixel i in the noise-free image. The process is explained in detail in [58]. The pixel neighbourhood redundancy is captured in the objective function used in the MCMC sampling process in [58]. The objective function used in [58] is defined as

$$\phi(i'_k|i_1) = \prod_j exp\left[-\frac{\rho(i'_k, i_1)}{\frac{\sigma_n^4}{\sigma_l^2} + \rho(i'_k, i_1)}\right] \tag{5.9}$$

where σ_n is the noise variance and σ_l is the local variance around site i_1 respectively. We define ρ as

$$\rho(i'_k, i_{k-1}) = \left(N^j_{i'_k} - N^j_{i_1}\right)^2. \tag{5.10}$$

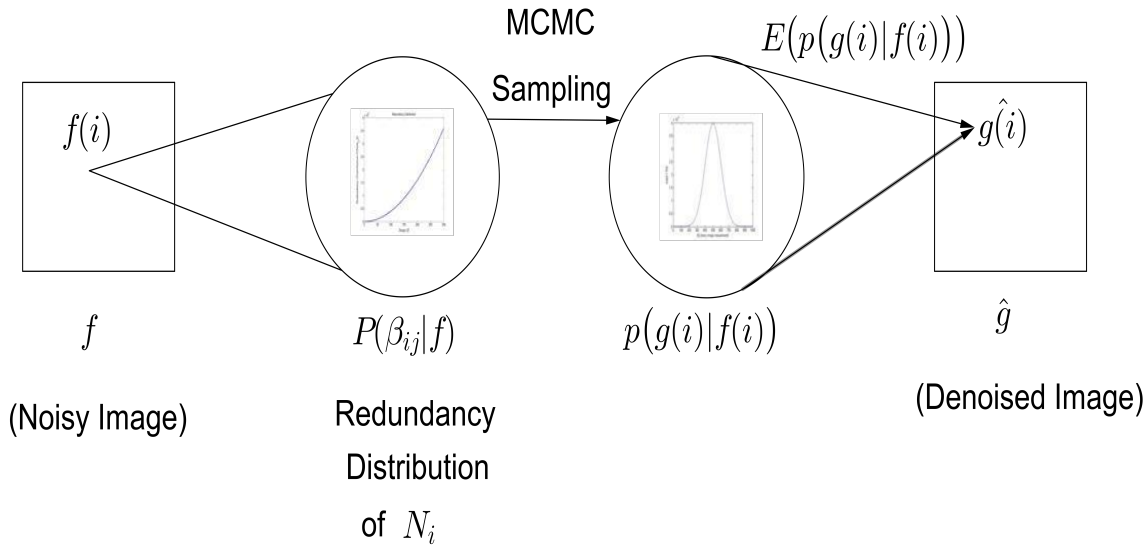


Figure 5.1: The figure shows the model used for the denoising of images using the MCMC sampling. For each site, MCMC sampling is done to obtain a distribution which gives the expected value for the corresponding site in the noise free image.

The objective function discussed in equation 5.9 ensures that sites which have close spatial interactions with a given site have a higher probability of being accepted as samples for MCMC sampling [58]. Thus, redundancy of pixel neighbourhoods implies more sites with close spatial interactions and higher density of samples. For each site $i \in \chi$ in the noisy image measurement f, we have an estimated distribution $\hat{p}(g(i)|f(i))$, and the expected value of this distribution gives us the estimate of the intensity $\hat{g}(i)$ of the corresponding site in the noise free image g. This process is represented in Figure 5.1.

5.2.2 Multi-scale Framework based Denoising

In section 5.2.1, we discussed stochastic denoising based on MCMC sampling. In this section, we discuss the extension of this concept to the non-parametric multi-scale framework developed in chapter 3, for the development of a new image denoising algorithm. In the

proposed denoising algorithm, which we refer to as DWTMCMC (MCMC sampling based denoising with decimated wavelets) and SWTMCMC(MCMC sampling based denoising with undecimated wavelets), we perform estimation of the wavelet coefficients of the noise free image given the wavelet coefficients of a noisy image using the MCMC approach explained in section 5.2.1. Thus, in this approach, we use the MCMC method to learn a non-parametric distribution for each pixel in each sub-band and use this to obtain an estimate of the denoised image. This process is illustrated in Figure 5.2 and is summarized below.

1. Apply a J-level wavelet transform (DWT or SWT) to the given noisy image f to obtain the wavelet sub-bands fA_J and $\{fH_i, fV_i, fD_i\}$.

2. Use the noisy image wavelet coefficients and the MCMC estimation approach discussed in section 5.2.1 to estimate the wavelet coefficients for the denoised image. This gives wavelet sub-bands $\hat{g}A_J$ and $\{\hat{g}H_i, \hat{g}V_i, \hat{g}D_i\}$.

3. Apply the inverse wavelet transform on these estimated wavelet coefficients to obtain the estimate $\hat{g}$ of the noise free image.

The main difference in these two methods is that instead of exploiting the local spatial intensity interactions of the pixels in the image, we now exploit the local interactions of the wavelet coefficients. The wavelet transform used for this algorithm could be the decimated (DWT) or the undecimated (SWT) wavelet transform. The undecimated wavelets exhibit more redundancy, hence it can be expected that the coefficients would show more local spatial interactions, and we can expect to observe more structural preservation. This will be validated in the next section where we discuss and compare the results of the denoising experiments based on the algorithm proposed.

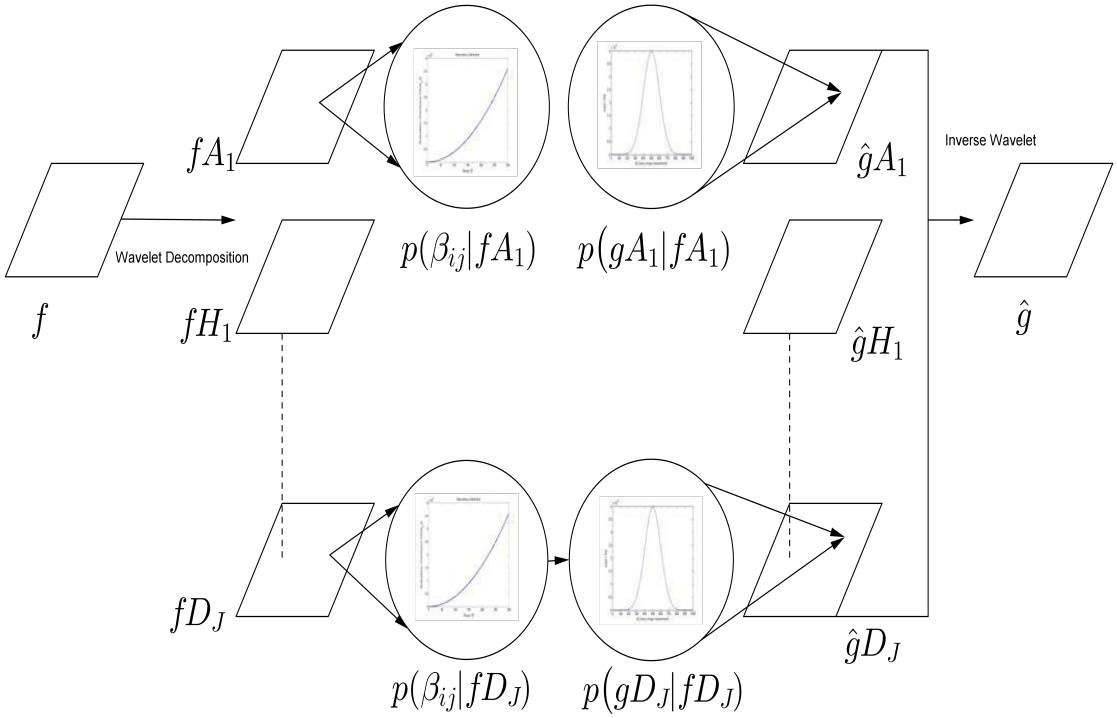


Figure 5.2: The stochastic denoising process illustrated in Figure 5.1 is applied to the multi-scale wavelet transforms in this figure. For each wavelet coefficient, MCMC sampling is done to obtain a distribution which gives the expected value for the corresponding wavelet coefficient for the noise free image. Applying the inverse wavelet transform to these estimated wavelet coefficients gives an estimate of the noise free image.

5.3 Performance comparison across different methods

We have compared the performance of our proposed denoising algorithms (both DWT based and SWT based) with the spatial MCMCD method [58], the NeighShrinkSURE method [55] (denoted as NSS), which is the current state-of-the-art in wavelet based image denoising and BM3D [69] which is the current state-of-the-art in image denoising. The quantitative measures we have used for performance comparison are PSNR [70] and MSSIM [71, 72]. For comparing the robustness of the denoising algorithms we use several different noise characteristics, i.e., the simple additive white Gaussian noise, mixed additive Gaussian noise, speckle noise and convoluted Gaussian noise. In these experiments, we have used a neighbourhood size of 7 x 7 and a search spread size of 21 for the MCMC based denoising algorithms.

5.3.1 Simple Additive Gaussian Noise

In this section, we have considered the case of simple additive Gaussian noise. The noise model can be represented as discussed in equation 5.7 where n represents the Gaussian noise. The additive Gaussian noise is assumed to be zero mean and is characterised by its standard deviation (σ). In our experiments, we have added noise to the image in varying degrees and have applied denoising algorithms to the noisy images. The σ has been varied from $\sigma = 20$ to $\sigma = 200$ (on a scale of 255) in steps of 20. For all these denoised images, we have computed the PSNR (Peak Signal to Noise Ratio) and MSSIM (Mean Structural Similarity Index) values, and have computed the average PSNR and MSSIM values for 10 natural images. The images denoised with the different denoising algorithms in consideration are shown in Figure 5.3. In Figure 5.3, NSS refers to the NeighShrinkSURE algorithm, MCMCD refers to the Stochastic Denoising algorithm based on MCMC sampling [58], DWTMCMC and SWTMCMC refer to the DWT and SWT based algorithms proposed in section 5.2.2 respectively. It can be seen that the SWTMCMC is successful in denoising the images effectively without any introduction of artefacts. Particularly in images with textured backgrounds, we can see more well defined structure in case of SWTMCMC as compared to NSS, MCMCD and DWTMCMC. This is a promising result as we can remove noise without losing important structural details of the image. Figure 5.4 shows the performance comparison of the denoising algorithms based on the PSNR and MSSIM measures. It can be seen that after the BM3D (which has the best performance for all noise levels), the SWTMCMC shows the best performance. Also, the performance of SWTMCMC is significantly better than NSS which is the current state-of-the-art in wavelet based denoising of images.

5.3.2 Mixed Additive Noise

The simple additive Gaussian noise model discussed above is a very simple model which may not be valid for various situations [73, 74, 75]. We have considered different types of noise distributions to comprehensively evaluate the denoising performance of the algorithms in consideration. In this section we discuss the mixed additive Gaussian noise in which noise is assumed to come from two different sources, giving two Gaussian noise components with their own noise levels. The mixed additive Gaussian noise model which we have used in our experiments can be represented as

$$f(i) = g(i) + n_1(i) + n_2(i) \quad \forall i \in \chi \tag{5.11}$$

where n_1 is a Gaussian noise component with noise level $\sigma_1 = 2\sigma/3$ and n_2 is the second noise component with noise level $\sigma_2 = \sigma/4$. For our experiments, we have varied σ from $\sigma = 20$ to $\sigma = 180$ in steps of 40. Figure 5.5 shows the denoised images for the case in which we set $\sigma = 100$. In Figure 5.5, we can see that the NSS algorithm is not able to remove the noise effectively. The MCMCD and DWTMCMCD remove the noise components to a large extent but result in blurring of the backgrounds, i.e., we lose structural details in the background. BM3D and SWTMCMC both give good denoising results without blurring the background or introducing artefacts. Figure 5.6 shows the denoising performance of the algorithms for the mixed additive noise. In terms of PSNR, the performance of the SWTMCMC algorithm closely follows that of BM3D and outperforms all other algorithms. An important observation is that for this type of noise, the DWTMCMC outperforms the NSS. This can be attributed to the fact that NSS assumes the noise to be simple additive Gaussian noise whereas the stochastic denoising algorithms, i.e., MCMCD, DWTMCMC and SWMCMC make no such assumptions. In terms of MSSIM, the performance trends are similar to those of PSNR based performance trends.

5.3.3 Rayleigh Noise

In many applications of image processing [76, 77, 78], we come across noisy images where the noise component follows a Rayleigh distribution [79]. The additive Rayleigh noise model can be represented as

$$f(i) = g(i) + n(i) \quad \forall i \in \chi \tag{5.12}$$

where n follows a Rayleigh distribution with mean $\mu = 0$ and standard deviation σ. In our experiments we have varied σ from $\sigma = 20$ to $\sigma = 200$ in steps of 20 (on a scale of 255).

Figure 5.7 shows the denoised images for $\sigma = 140$. It can be seen that NSS is not able to remove noise from the noisy images. There is significant blurring of the background in case of MCMCD and DWTMCMC. The BM3D and SWTMCMC give good denoising results without loss of structural details in the image. In terms of PSNR and MSSIM, the denoising performance follows similar trends as in the case of mixed additive Gaussian noise.

5.3.4 Speckle Noise

Speckle noise is multiplicative noise which can be represented by the following model

$$f = g + g * n \tag{5.13}$$

where g is the original image, f is the noise contaminated image and n is a uniformly distributed noise with mean $\mu = 0$ and standard deviation σ. Figure 5.9 shows denoised images for $\sigma = 100$. Again, we see that the SWTMCMC and BM3D give the best denoising performance and preservation of structural details. Figure 5.10 shows the denoising performance in terms of PSNR and MSSIM. It can be seen that for low noise levels, SWTMCMC outperforms BM3D, which is a promising result.

5.3.5 Convoluted Gaussian Noise

The convoluted Gaussian noise model can be represented as

$$f = g \star n \tag{5.14}$$

where g is the original image, f is the noise contaminated image and n is the noise component, i.e., the noise is convoluted with the original image. The noise component n follows a Gaussian distribution. The noise is characterized by the peak amplitude of the noise signal β which we call the noise level and the standard deviation σ which we call the blur level. Figure 5.11 shows the denoising performance of the algorithms when each of these quantities is varied keeping the other one fixed.In the first plot, the blur level is fixed at $\sigma = 100$ and the noise level β is varied from $\beta = 20$ to $\beta = 200$ in steps of 20. We can see that the SWTMCMC denoising performance very closely follows that of BM3D and outperforms all the other algorithms. In the second plot, the noise level is fixed at $\beta = 100$ and the blur level σ is varied from $\sigma = 20$ to $\sigma = 200$ in steps of 20. As the blur is increased, the denoising performance of SWTMCMC gets better than that of BM3D. This

can also be attributed to the fact that increasing the blur level while keeping noise level constant reduces the effect of noise, which can be validated by the increase in PSNR of the noisy image with the increase in blur level.

5.4 Conclusions

The denoising results which we have discussed above show us that the proposed algorithms give very promising results. The performance of the SWTMCMC algorithm closely follows that of the BM3D which is the current state-of-the-art in image denoising and significantly outperforms the NSS, which is the current state-of-the-art in wavelet based image denoising. There is a significant difference in the performance of the MCMCD and SWTMCMC algorithm, which proves that the multi-scale framework proposed in chapter 3 gives a very effective multi-scale description of the image based on which effective image processing algorithms can be developed. The SWTMCMC shows significantly more image detail preservation when compared to the DWTMCMC and MCMCD. This can be attributed to the redundancy characteristics of the wavelet decompositions of the images discussed in chapter 3, i.e., the SWT coefficients exhibit close local spatial interactions, giving better MCMC sampling results and better denoising. The proposed algorithm performs well for a variety of noise characteristics. An important advantage of the proposed denoising technique is that it does not require any information about the image statistics or the noise statistics. The algorithm is very flexible and adapts itself to the underlying image statistics. This makes it a very useful general purpose algorithm which can be used in a variety of image processing applications, from remote sensing to medical imaging.

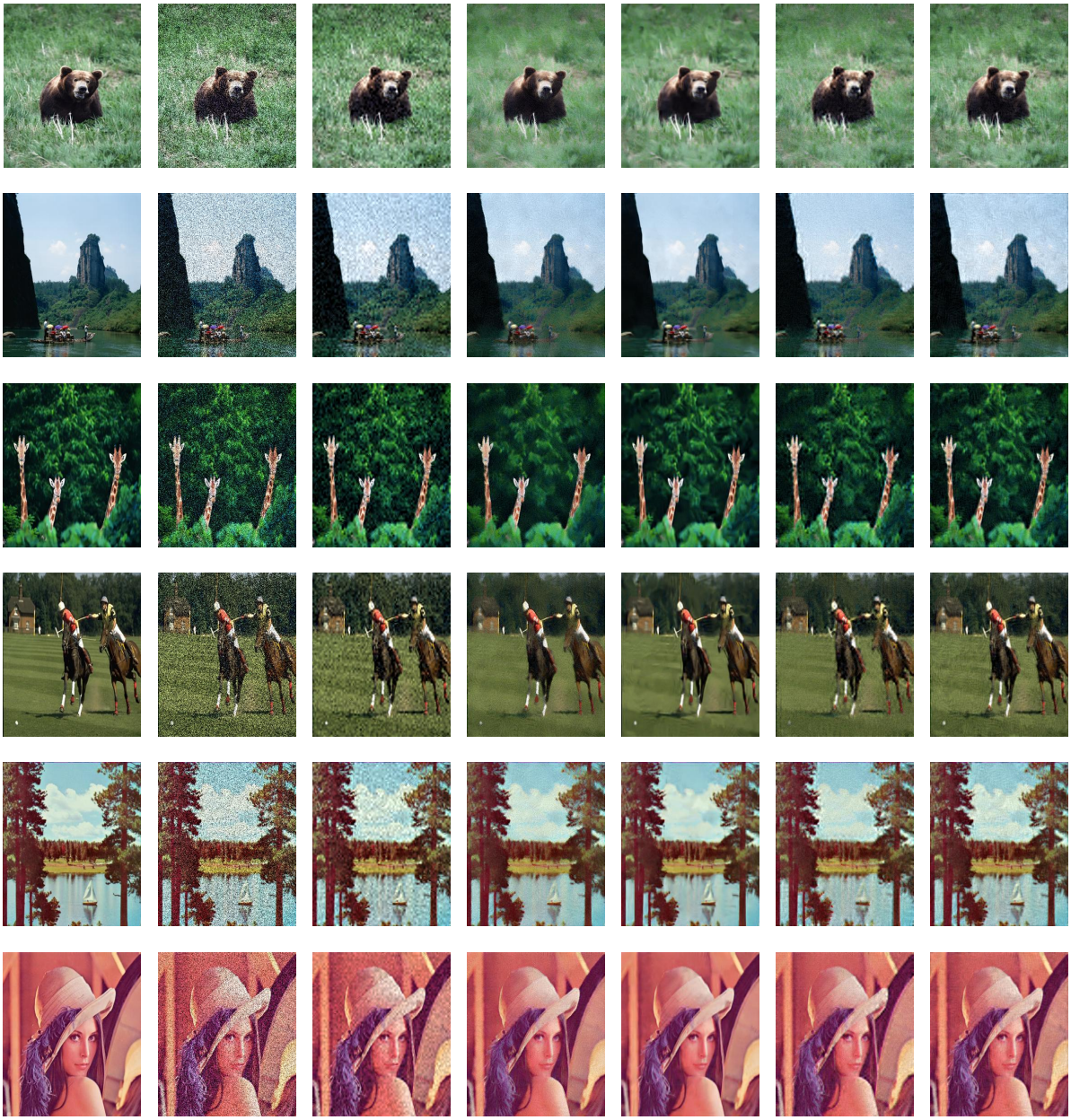
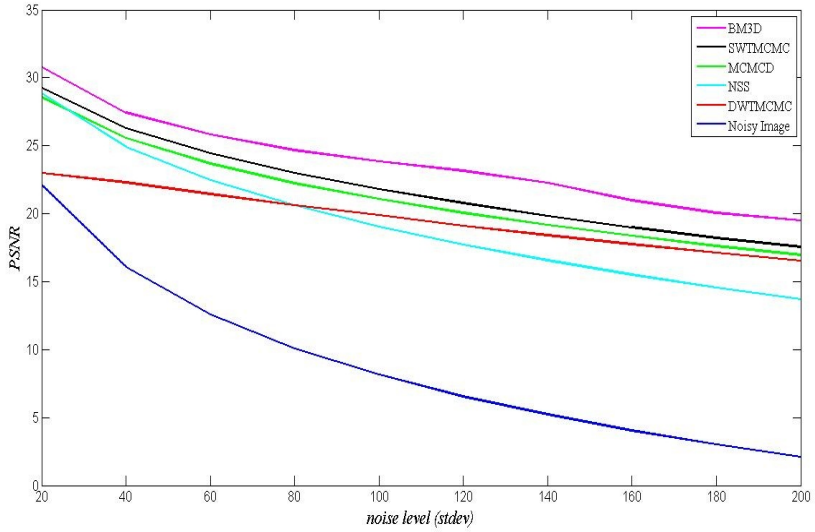
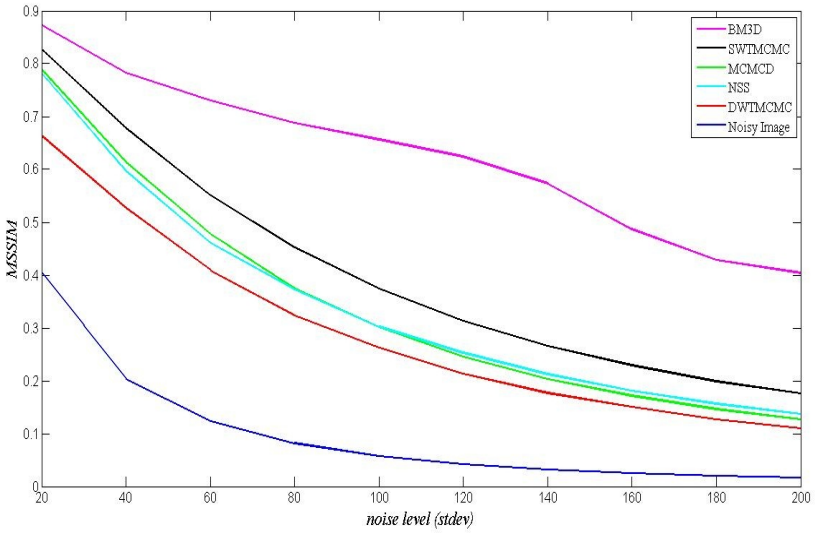


Figure 5.3: Simple additive Gaussian noise of $\sigma = 60$ is considered. It can be seen that the SWT based MCMC algorithm (proposed algorithm) is successful in denoising the images without the introduction of any artefacts in the images. For the textured images, the background structure is well preserved by the SWTMCMC algorithm.

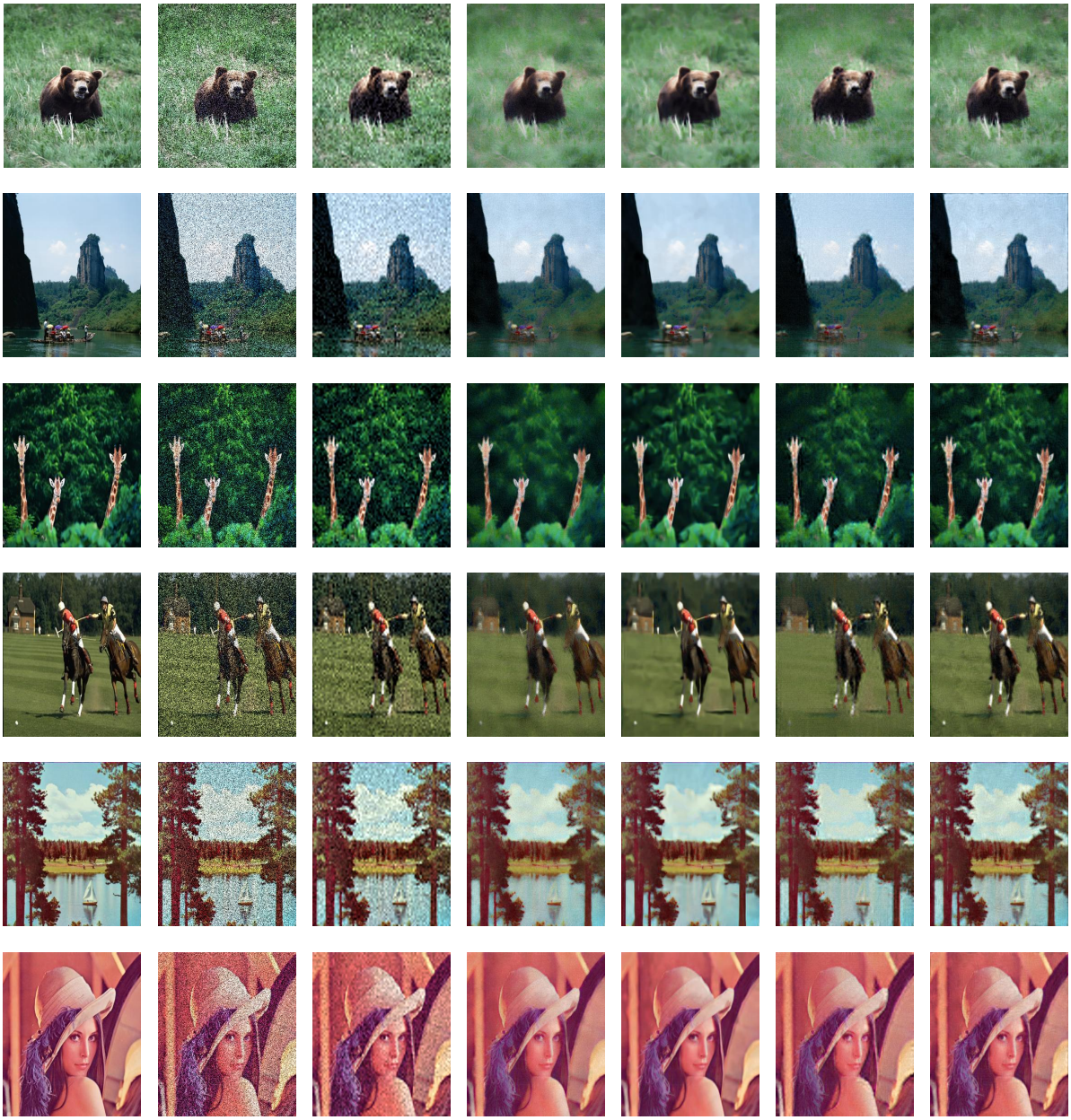


PSNR based Performance Comparison



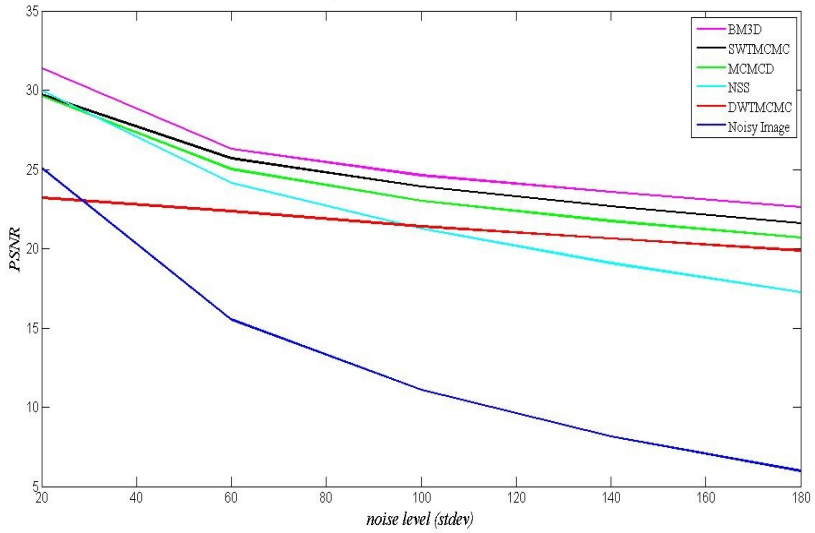
MSSIM based Performance Comparison

Figure 5.4: This figure shows performance of the denoising algorithms in case of simple additive Gaussian noise. In both the plots, we see that the SWTMCMC has the best performance after the BM3D and it outperforms the NeighShrinkSURE (NSS) algorithm (the current state of the art in wavelet based denoising) by a significant margin.

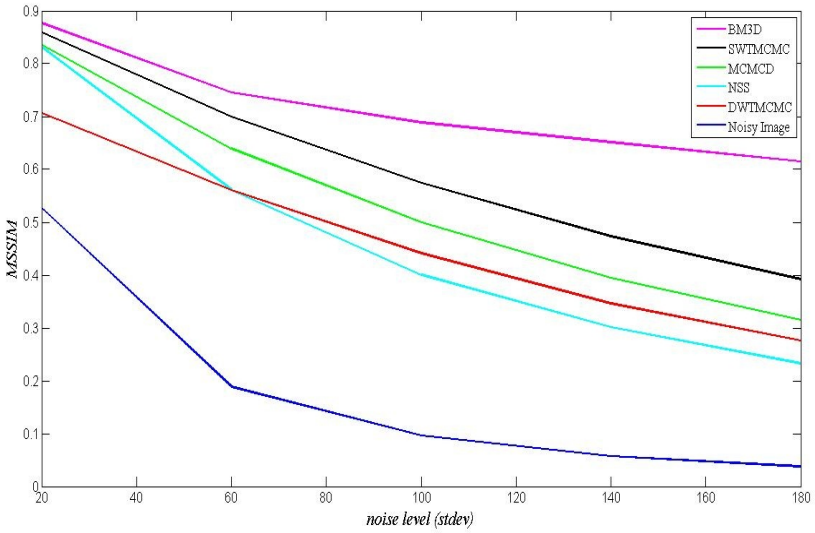


Original Image | Noisy Image | NSS | MCMCD | BM3D | DWTM-CMC | SWTMCMC

Figure 5.5: Mixed additive Gaussian noise of $\sigma = 100$ is considered. The SWTMCMC (proposed algorithm) and BM3D give denoised images without introduction of artefacts. The NSS algorithm is not able to remove the noise in the images.

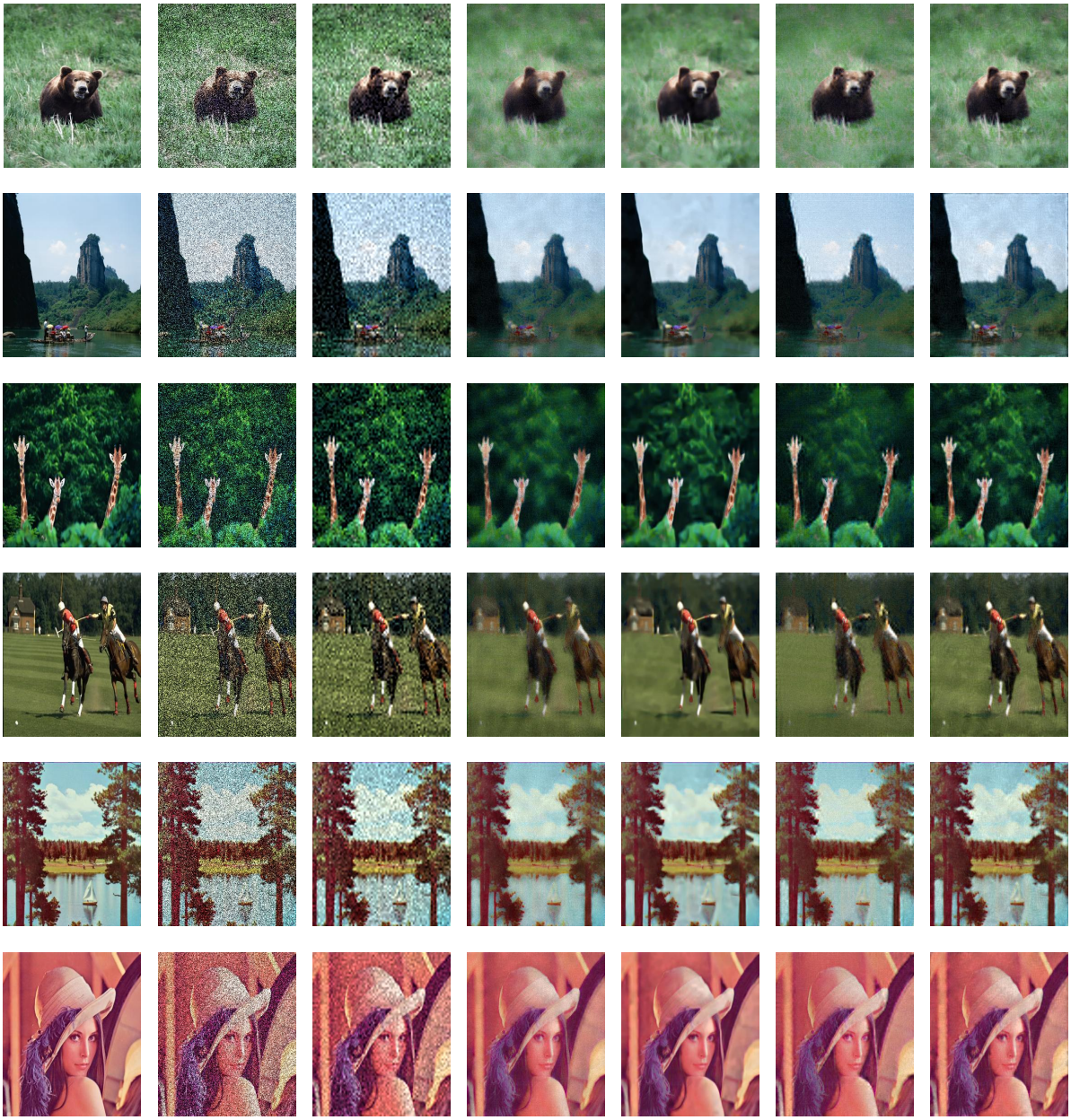


PSNR based Performance Comparison



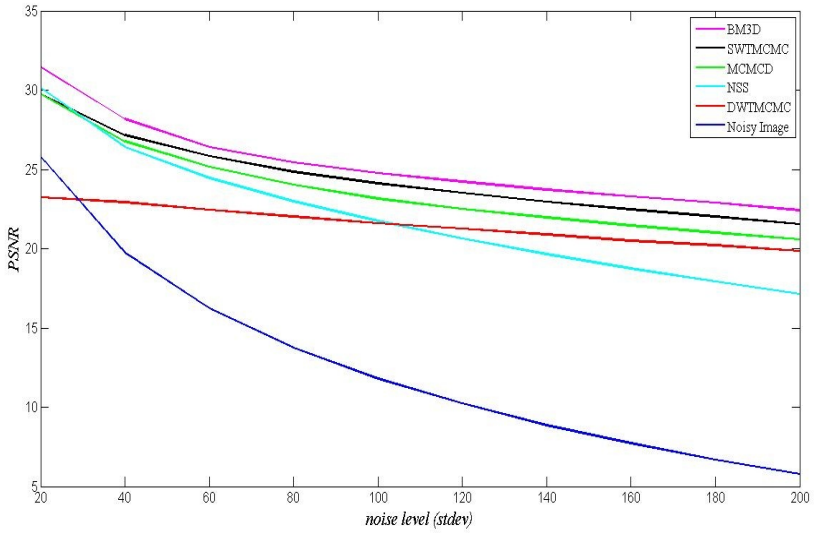
MSSIM based Performance Comparison

Figure 5.6: In this figure, mixed additive Gaussian noise is considered. The SWTMCMC outperforms the NSS by a more significant margin in this case as compared to the simple additive Gaussian noise case shown in Figure 5.4.

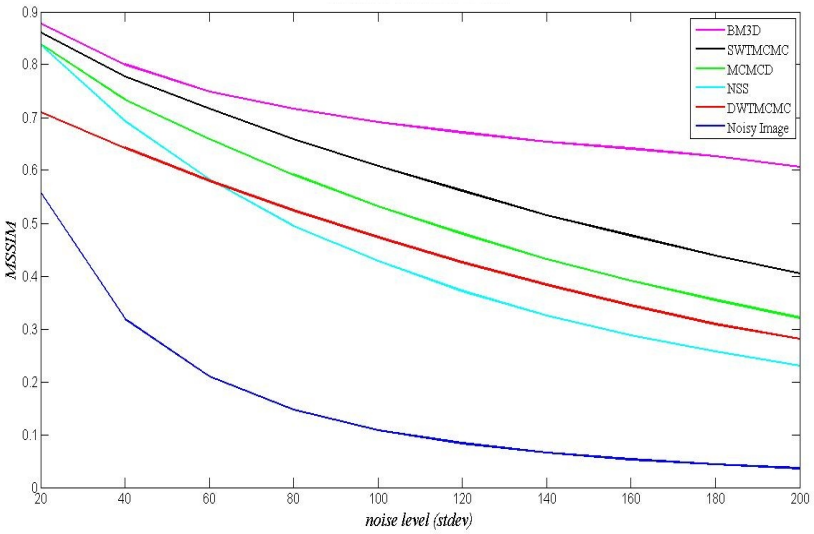


Original Image | Noisy Image | NSS | MCMCD | BM3D | DWTM-CMC | SWTMCMC

Figure 5.7: Additive Rayleigh noise of $\sigma = 140$ is considered. It can be seen that SWTM-CMC (proposed algorithm) is successful in denoising the images without the introduction of any artefacts in the images, but the NSS is not able to remove noise and MCMCD results in excessive blurring of the images.

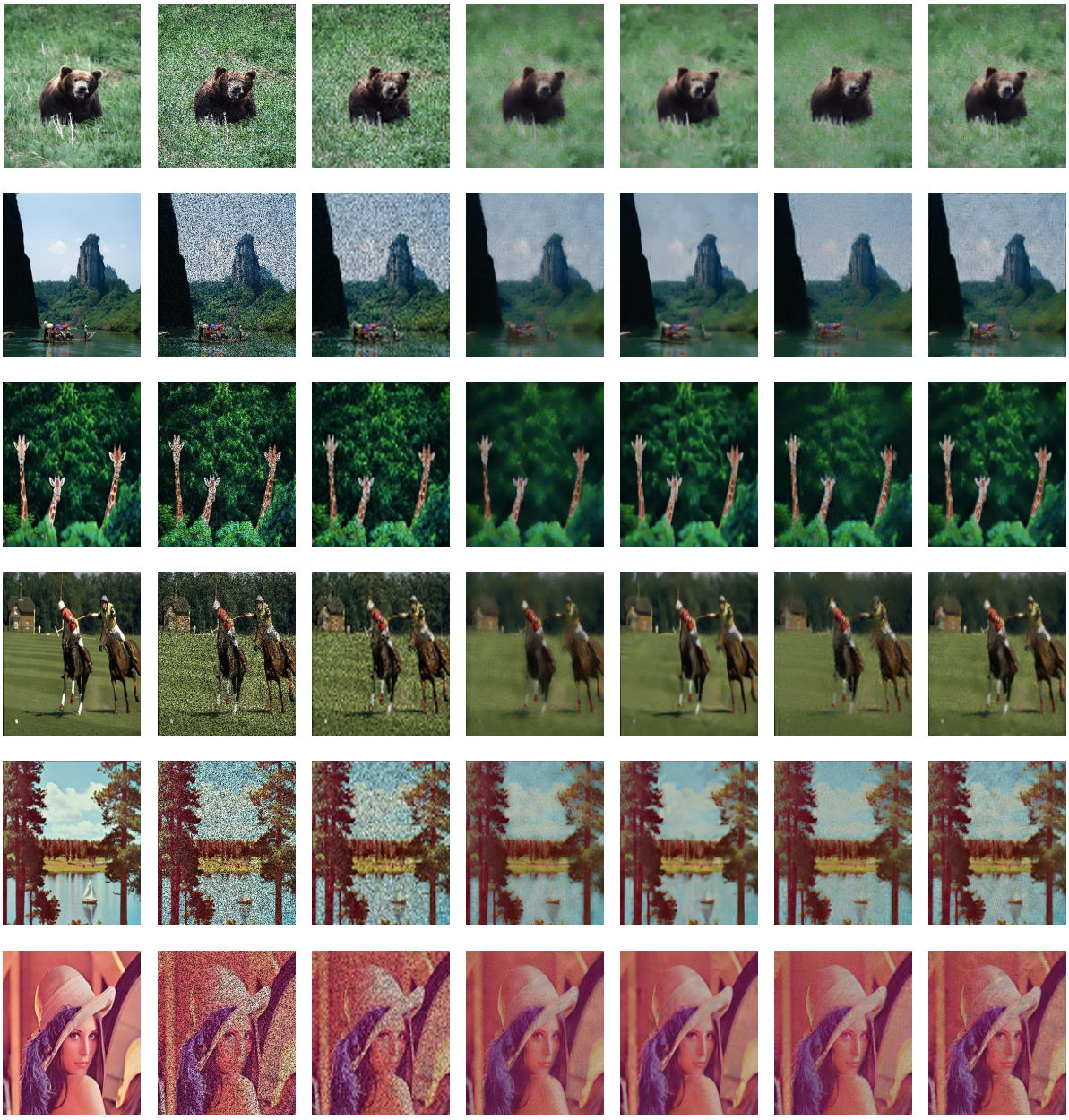


PSNR based Performance Comparison



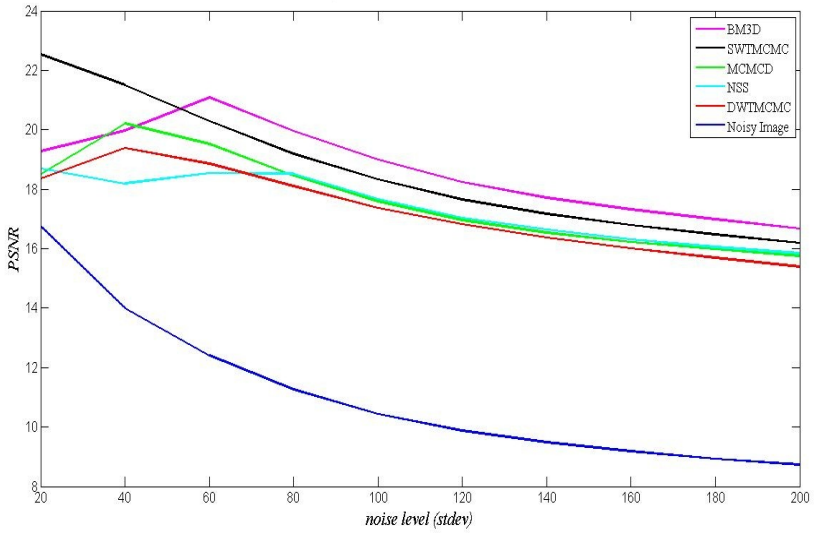
MSSIM based Performance Comparison

Figure 5.8: In this figure, additive Rayleigh noise is considered. The performance trends are the same as those in Figure 5.6.

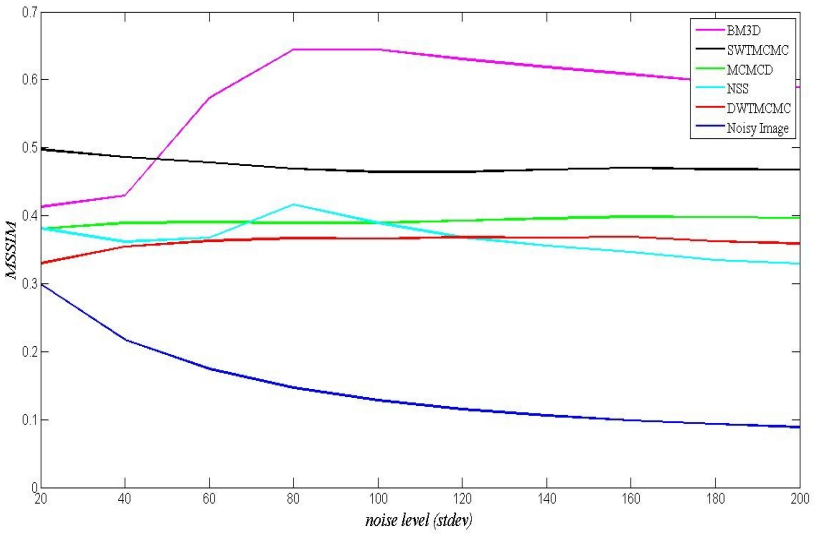


Original Image | Noisy Image | NSS | MCMCD | BM3D | DWTM-CMC | SWTMCMC

Figure 5.9: Speckle noise of $\sigma = 100$ is considered. In some cases (e.g., girraffe image), the structural preservation in SWTMCMC (proposed algorithm) is better than that in BM3D.

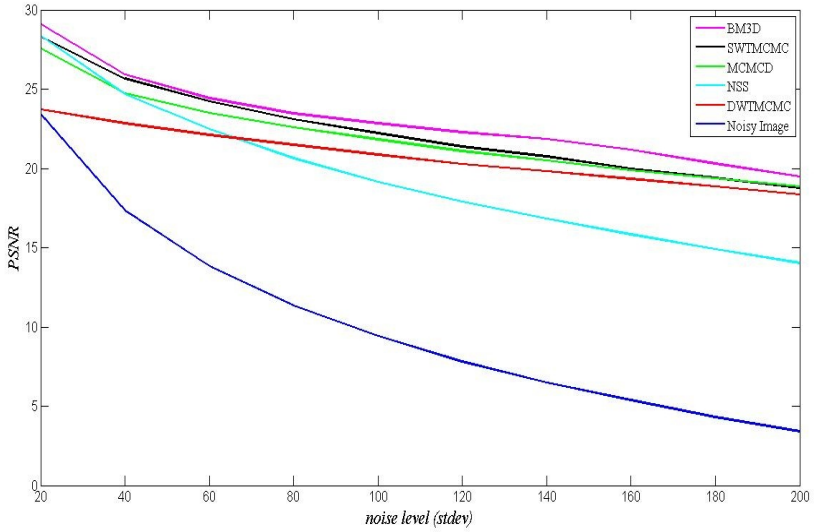


PSNR based Performance Comparison

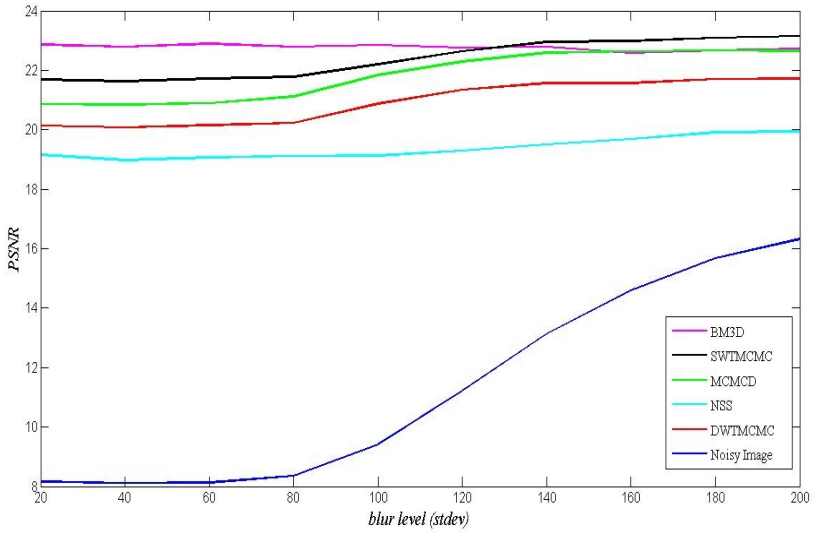


MSSIM based Performance Comparison

Figure 5.10: In this figure, speckle noise is considered. The SWTMCMC (proposed algorithm) outperforms the NSS, which assumes a simple additive Gaussian noise model.



PSNR vs Noise Levels (Convoluted Gaussian Noise)



PSNR vs Blur Levels (Convoluted Gaussian Noise)

Figure 5.11: The plot at the top shows the PSNR vs Noise level characteristics for images are corrupted with Convoluted Gaussian Noise with varying noise levels and fixed blur level $\sigma = 100$. The plot at the bottom shows the PSNR vs Blur characteristics for denoised images with varying blur levels and fixed noise level $\beta = 100$.

Chapter 6

Conclusions and Further Applications

In this thesis, a multi-scale non-parametric framework was proposed for image analysis and understanding. The framework was used for the development of two image processing algorithms; a saliency detection algorithm, i.e., the non-redundancy based saliency detection algorithm and an image denoising algorithm, i.e., the SWTMCMC algorithm. Both these algorithms give competitive results when compared with the state-of-the-art in their respective domains. Thus we can conclude that the framework is effective at multi-scale analysis of images and can be used for the development of several image processing applications like image segmentation, image compression, texture synthesis etc. The framework offers several advantages over the traditional multi-scale representations which were discussed in chapter 2.

1. Algorithms based on the framework do not require any prior knowledge about the image, and this makes it a good choice for many applications where automated image analysis is required and prior information about the image statistics is not known beforehand.

2. Algorithms based on the proposed framework are very flexible and adapt themselves to the underlying image statistics, as opposed to the parametric models discussed in section 2.1.1, where fixed parameters were used for the analysis of the image.

3. It is robust to a variety of noise characteristics as can be seen from the performance of the denoising algorithm discussed in section 5.3. The denoising algorithm developed based on this framework gives high PSNR values for a variety of noise characteristics. This makes the proposed framework suitable for use in a variety of image processing applications.

4. In the traditional multi-scale frameworks, significant structural degradation is observed at higher scales. The framework proposed in the thesis addresses this limitation as can be seen from the results of the denoising algorithm, where we achieve high PSNR values and retain the structural details in the denoised images.

5. The immediate localization property discussed in section 2.2.1 is well satisfied by the proposed multi-scale framework as can be seen from the results of the proposed algorithms.

The framework developed in the thesis is flexible and robust and can be used for solving many image processing challenges ranging from remote sensing to medical imaging. Future work related to the thesis would involve the application of the framework to address specific image processing problems.

The present framework has some limitations which need to be addressed to improve the framework. The algorithms consider the pixel neighbourhood interactions over the entire image which makes the algorithms computationally intensive. In chapter 4, section 4.2.3, we introduced the use of random sampling to improve the computational intensity of the saliency detection algorithm. However, the results of the algorithm still remain dependent on the sample selection to some extent, and using guided samples instead of random samples can potentially give better results. The saliency algorithm developed in chapter 4 was applied to a number of images from the Achanta database [39], and it was observed that they produce good saliency maps for images with homogeneous backgrounds. However, the saliency maps for images with cluttered backgrounds are not very good in terms of highlighting the object of interest. The way the different wavelet sub-bands are analysed and interpreted affects the performance of the algorithms based on this framework, hence more experimentation in this direction can give algorithms with better performance. In our experiments, for coloured images, we separated the image into the three channels, hue, saturation and value, and applied the framework independently to all these channels. For the generation of saliency maps for coloured images, we have taken the simple average of the saliency maps of individual channels. However, the spatial interactions between pixels in these different channels can be different, and if we adapt our analysis of sub-bands to the channel characteristics, we can get better results.

Future work in this direction should involve more experiments on the spatial interactions of the wavelet sub-bands, and development of algorithms which take the differences in the spatial interactions of the different sub-bands into account. The framework proposed has been applied to saliency detection and image denoising in this thesis. However, it is a generic multi-scale framework and can be used in a variety of applications in image processing, e.g., image segmentation, image compression, object detection, etc.

References

[1] A. Witkin, "Scale-space filtering," *Readings in computer vision: issues, problems, principles, and paradigms*, pp. 329–332, 1987.

[2] A. Wong, A. Mishra, D. A. Clausi, and P. Fieguth, *Nonparametric sample-based methods for image understanding*. World Scientific Publishing, 2011.

[3] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 7, pp. 629–639, 1990.

[4] J. De Bonet and P. Viola, "A non-parametric multi-scale statistical model for natural images," *Advances in neural information processing systems*, pp. 773–779, 1998.

[5] J. Romberg, H. Choi, and R. Baraniuk, "Bayesian tree-structured image modeling using wavelet-domain hidden markov models," *IEEE Transactions on Image Processing*, vol. 10, no. 7, pp. 1056–1068, 2001.

[6] L. Florack, B. ter Haar Romeny, J. Koenderink, and M. Viergever, "Linear scale-space," *Journal of Mathematical Imaging and Vision*, vol. 4, no. 4, pp. 325–351, 1994.

[7] J. Koenderink, "The structure of images," *Biological cybernetics*, vol. 50, no. 5, pp. 363–370, 1984.

[8] T. Lindeberg, "Edge detection and ridge detection with automatic scale selection," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 465–470, 1996.

[9] E. Balster, Y. Zheng, and R. Ewing, "Feature-based wavelet shrinkage algorithm for image denoising," *IEEE Transactions on Image Processing*, vol. 14, no. 12, pp. 2024–2039, 2005.

[10] F. Luisier, T. Blu, and M. Unser, "A new sure approach to image denoising: Interscale orthonormal wavelet thresholding," *IEEE Transactions on Image Processing*, vol. 16, no. 3, pp. 593–606, 2007.

[11] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, 1989.

[12] R. Gonzalez and E. Richard, "Woods, digital image processing," 2002.

[13] J. Walker, "Wavelet-based image processing," *Applicable Analysis*, vol. 85, no. 4, pp. 439–458, 2006.

[14] F. Meyer, A. Averbuch, and J. Stromberg, "Fast adaptive wavelet packet image compression," *IEEE Transactions on Image Processing*, vol. 9, no. 5, pp. 792–800, 2000.

[15] T. Ryan, L. Sanders, H. Fisher, and A. Iverson, "Image compression by texture modeling in the wavelet domain," *IEEE Transactions on Image Processing*, vol. 5, no. 1, pp. 26–36, 1996.

[16] X. Zhang and M. Desai, "Wavelet based automatic thresholding for image segmentation," in *IEEE International Conference on Image Processing*, vol. 1, pp. 224–227, 1997.

[17] C. Jung, "Unsupervised multiscale segmentation of color images," *Pattern Recognition Letters*, vol. 28, no. 4, pp. 523–533, 2007.

[18] I. Selesnick, R. Baraniuk, and N. Kingsbury, "The dual-tree complex wavelet transform," *IEEE Signal Processing Magazine,*, vol. 22, no. 6, pp. 123–151, 2005.

[19] X. Wang, R. Istepanian, and Y. Song, "Microarray image enhancement by denoising using stationary wavelet transform," *IEEE Transactions on NanoBioscience*, vol. 2, no. 4, pp. 184–189, 2003.

[20] G. Nason and B. Silverman, "The stationary wavelet transform and some statistical applications," *Lecture Notes in Statistics-New York-Springer Verlag-*, pp. 281–281, 1995.

[21] M. Unser, "Texture classification and segmentation using wavelet frames," *IEEE Transactions on Image Processing*, vol. 4, no. 11, pp. 1549–1560, 1995.

[22] M. Beaulieu, S. Foucher, and L. Gagnon, "Multi-spectral image resolution refinement using stationary wavelet transform," in *IEEE Geoscience and Remote Sensing Symposium*, vol. 6, pp. 4032–4034, 2003.

[23] J. Yu, Y. Wang, and Y. Shen, "Noise reduction and edge detection via kernel anisotropic diffusion," *Pattern Recognition Letters*, vol. 29, no. 10, pp. 1496–1503, 2008.

[24] G. Gilboa, N. Sochen, and Y. Zeevi, "Image enhancement and denoising by complex diffusion processes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 8, pp. 1020–1036, 2004.

[25] G. Gilboa, "Nonlinear scale space with spatially varying stopping time," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 12, pp. 2175–2187, 2008.

[26] A. Efros and T. Leung, "Texture synthesis by non-parametric sampling," in *The Proceedings of the Seventh IEEE International Conference on Computer Vision*, vol. 2, pp. 1033–1038, 1999.

[27] P. Fieguth, *Statistical image processing and multidimensional modeling*. Springer Verlag, 2010.

[28] M. Crouse, R. Nowak, and R. Baraniuk, "Wavelet-based statistical signal processing using hidden markov models," *IEEE Transactions on Signal Processing*, vol. 46, no. 4, pp. 886–902, 1998.

[29] J. Han, K. Ngan, M. Li, and H. Zhang, "Unsupervised extraction of visual attention objects in color images," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 1, pp. 141–145, 2006.

[30] U. Rutishauser, D. Walther, C. Koch, and P. Perona, "Is bottom-up attention useful for object recognition?," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. II–37, 2004.

[31] L. Itti and C. Koch, "A saliency-based search mechanism for overt and covert shifts of visual attention," *Vision research*, vol. 40, no. 10-12, pp. 1489–1506, 2000.

[32] L. Itti and C. Koch, "Computational modeling of visual attention," *Nature reviews neuroscience*, vol. 2, no. 3, pp. 194–203, 2001.

[33] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.

[34] L. Itti and C. Koch, "A comparison of feature combination strategies for saliency-based visual attention systems," in *SPIE human vision and electronic imaging IV*, vol. 3644, pp. 473–482, Citeseer, 1999.

[35] L. Itti and C. Koch, "Feature combination strategies for saliency-based visual attention systems," *Journal of Electronic Imaging*, vol. 10, p. 161, 2001.

[36] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, 2007.

[37] M. Wang, J. Konrad, P. Ishwar, K. Jing, and H. Rowley, "Image saliency: From intrinsic to extrinsic context," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 417–424, 2011.

[38] N. Murray, M. Vanrell, X. Otazu, and C. Parraga, "Saliency estimation using a non-parametric low-level vision model," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 433–440, 2011.

[39] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1597–1604, 2009.

[40] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2376–2383, 2010.

[41] C. Koch and S. Ullman, "Shifts in selective visual attention: towards the underlying neural circuitry.," *Hum Neurobiol*, vol. 4, no. 4, pp. 219–27, 1985.

[42] Y. Ma and H. Zhang, "Contrast-based image attention analysis by using fuzzy growing," in *Proceedings of the eleventh ACM international conference on Multimedia*, pp. 374–381, 2003.

[43] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," *Advances in neural information processing systems*, vol. 19, p. 545, 2007.

[44] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, and H. Shum, "Learning to detect a salient object," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 2, pp. 353–367, 2011.

[45] M. Cheng, G. Zhang, N. Mitra, X. Huang, and S. Hu, "Global contrast based salient region detection," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 409–416, 2011.

[46] Y. Zhai and M. Shah, "Visual attention detection in video sequences using spatiotemporal cues," in *Proceedings of the 14th annual ACM international conference on Multimedia*, pp. 815–824, 2006.

[47] A. Jain., A. Wong, and P. Fieguth, "Saliency detection via statistical non-redundancy," in *IEEE International Conference on Image Processing (accepted)*, 2012.

[48] M. Buckland and F. Gey, "The relationship between recall and precision," *Journal of the American society for information science*, vol. 45, no. 1, pp. 12–19, 1994.

[49] D. Donoho and I. Johnstone, "Threshold selection for wavelet shrinkage of noisy data," in *Engineering in Medicine and Biology Society. Engineering Advances: New Opportunities for Biomedical Engineers. Proceedings of the 16th Annual International Conference of the IEEE*, pp. A24–A25, 1994.

[50] F. Argenti and G. Torricelli, "Speckle suppression in ultrasonic images based on undecimated wavelets," *EURASIP Journal on Applied Signal Processing*, pp. 470–478, 2003.

[51] J. Portilla, V. Strela, M. Wainwright, and E. Simoncelli, "Image denoising using scale mixtures of gaussians in the wavelet domain," *IEEE Transactions on Image Processing*, vol. 12, no. 11, pp. 1338–1351, 2003.

[52] D. Donoho, "De-noising by soft-thresholding," *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 613–627, 1995.

[53] D. Donoho and I. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *Journal of the american statistical association*, pp. 1200–1224, 1995.

[54] D. Donoho and J. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.

[55] Z. Dengwen and C. Wengang, "Image denoising with an optimal threshold and neighbouring window," *Pattern Recognition Letters*, vol. 29, no. 11, pp. 1694–1697, 2008.

[56] G. Chen, T. Bui, and A. Krzyzak, "Image denoising using neighbouring wavelet coefficients," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, pp. ii–917, 2004.

[57] C. Stein, "Estimation of the mean of a multivariate normal distribution," *The annals of Statistics*, pp. 1135–1151, 1981.

[58] A. Wong, A. Mishra, W. Zhang, P. Fieguth, and D. Clausi, "Stochastic image denoising based on markov-chain monte carlo sampling," *Signal Processing*, 2011.

[59] M. Lindenbaum, M. Fischer, and A. Bruckstein, "On gabor's contribution to image enhancement," *Pattern Recognition*, vol. 27, no. 1, pp. 1–8, 1994.

[60] J. Lee, "Digital image enhancement and noise filtering by use of local statistics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, no. 2, pp. 165–168, 1980.

[61] L. Rudin and S. Osher, "Total variation based image restoration with free local constraints," in *IEEE International Conference on Image Processing*, vol. 1, pp. 31–35, 1994.

[62] S. Geman and D. Geman, "Stochastic relaxation, gibbs distributions, and the bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, no. 6, pp. 721–741, 1984.

[63] S. Greenberg and D. Kogan, "Improved structure-adaptive anisotropic filter," *Pattern Recognition Letters*, vol. 27, no. 1, pp. 59–65, 2006.

[64] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *IEEE Conference on Computer Vision*, pp. 839–846, 1998.

[65] W. Wong, A. Chung, and S. Yu, "Trilateral filtering for biomedical images," in *IEEE International Symposium on Biomedical Imaging: Nano to Macro*, pp. 820–823, 2004.

[66] M. Mahmoudi and G. Sapiro, "Fast image and video denoising via nonlocal means of similar neighborhoods," *IEEE Signal Processing Letters*, vol. 12, no. 12, pp. 839–842, 2005.

[67] A. Buades, B. Coll, and J. Morel, "Nonlocal image and movie denoising," *International Journal of Computer Vision*, vol. 76, no. 2, pp. 123–139, 2008.

[68] J. Orchard, M. Ebrahimi, and A. Wong, "Efficient nonlocal-means denoising using the svd," in *IEEE International Conference on Image Processing*, pp. 1732–1735, 2008.

[69] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080–2095, 2007.

[70] Q. Huynh-Thu and M. Ghanbari, "Scope of validity of psnr in image/video quality assessment," *Electronics letters*, vol. 44, no. 13, pp. 800–801, 2008.

[71] Z. Wang, E. Simoncelli, and A. Bovik, "Multiscale structural similarity for image quality assessment," in *Thirty-Seventh Asilomar Conference on Signals, Systems and Computers*, vol. 2, pp. 1398–1402, 2003.

[72] Z. Wang and A. Bovik, "Mean squared error: Love it or leave it? a new look at signal fidelity measures," *IEEE Signal Processing Magazine*, vol. 26, no. 1, pp. 98–117, 2009.

[73] R. Li and Y. Zhang, "A hybrid filter for the cancellation of mixed gaussian noise and impulse noise," in *Information, Communications and Signal Processing and the Fourth Pacific Rim Conference on Multimedia*, vol. 1, pp. 508–512, IEEE, 2003.

[74] E. Abreu, M. Lightstone, S. Mitra, and K. Arakawa, "A new efficient approach for the removal of impulse noise from highly corrupted images," *IEEE Transactions on Image Processing*, vol. 5, no. 6, pp. 1012–1025, 1996.

[75] L. Lemieux, U. Wieshmann, N. Moran, D. Fish, and S. Shorvon, "The detection and significance of subtle changes in mixed-signal brain lesions by serial mri scan matching and spatial normalization," *Medical Image Analysis*, vol. 2, no. 3, pp. 227–242, 1998.

[76] R. Wagner, M. Insana, and D. Brown, "Statistical properties of radio-frequency and envelope-detected signals with applications to medical ultrasound," *JOSA A*, vol. 4, no. 5, pp. 910–922, 1987.

[77] M. Lysaker, A. Lundervold, and X. Tai, "Noise removal using fourth-order partial differential equation with applications to medical magnetic resonance images in space and time," *IEEE Transactions on Image Processing*, vol. 12, no. 12, pp. 1579–1590, 2003.

[78] P. Shankar, J. Reid, H. Ortega, C. Piccoli, and B. Goldberg, "Use of non-rayleigh statistics for the identification of tumors in ultrasonic b-scans of the breast," *IEEE Transactions on Medical Imaging*, vol. 12, no. 4, pp. 687–692, 1993.

[79] D. Ruderman and W. Bialek, "Statistics of natural images: Scaling in the woods," *Physical Review Letters*, vol. 73, no. 6, pp. 814–817, 1994.

[80] E. Whitley, J. Ball, *et al.*, "Statistics review 6: Nonparametric methods," *Critical Care-London-*, vol. 6, no. 6, pp. 509–513, 2002.

[81] H. Tassignon, "Wavelets in image processing," *Image processing II: mathematical methods, algorithms and applications*, p. 228, 2000.

[82] N. Kingsbury, "Image processing with complex wavelets," *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 357, no. 1760, pp. 2543–2560, 1999.

[83] R. Coifman and D. Donoho, "Translation-invariant de-noising," *Lecture Notes in Statistics-New York-Springer Verlag-*, pp. 125–125, 1995.

[84] J. Portilla and E. Simoncelli, "A parametric texture model based on joint statistics of complex wavelet coefficients," *International Journal of Computer Vision*, vol. 40, no. 1, pp. 49–70, 2000.

[85] A. Bovik, N. Gopal, T. Emmoth, and A. Restrepo, "Localized measurement of emergent image frequencies by gabor wavelets," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 691–712, 1992.

[86] C. Chubb, J. Econopouly, and M. Landy, "Histogram contrast analysis and the visual segregation of iid textures," *JOSA A*, vol. 11, no. 9, pp. 2350–2374, 1994.

[87] S. Geman and D. McClure, "Statistical methods for tomographic image reconstruction," *Bull. Int. Stat. Inst*, vol. 4, pp. 5–21, 1987.

[88] M. Chen, "Importance-weighted marginal bayesian posterior density estimation," *Journal of the American Statistical Association*, pp. 818–824, 1994.