Improved Wavelet-based Image Denoising Algorithm Using Adaptive Center Weighted Median Filter

by

S. M. Mahbubur Rahman

A thesis submitted to the Department of Electrical and Electronic Engineering

of

Bangladesh University of Engineering and Technology

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN ELECTRICAL AND ELECTRONIC ENGINEERING

DEPARTMENT OF ELECTRICAL AND ELECTRONIC ENGINEERING

BANGLADESH UNIVERSITY OF ENGINEERING AND TECHNOLOGY

2002

The thesis entitled **"Improved Wavelet-based Image Denoising Algorithm Using Adaptive Center Weighted Median Filter"** submitted by S. M. Mahbubur Rahman Roll No.: 040006226P, Session: April 2000 has been accepted as satisfactory in partial fulfillment of the requirements for the degree of MASTER OF SCIENCE IN ELECTRICAL AND ELECTRONIC ENGINEERING.

BOARD OF EXAMINERS

1. 14/5/2002 — **Chairman**

(Dr. Md. Kamrul Hasan)
Associate Professor
Department of Electrical and
Electronic Engineering, BUET
Dhaka-1000, Bangladesh

2. — **Member** (Ex-officio)

(Dr. M. M. Shahidul Hassan)
Professor and Head
Department of Electrical and
Electronic Engineering, BUET
Dhaka-1000, Bangladesh

3. 14/05/02 — **Member**

(Dr. Md. Saifur Rahman)
Professor
Department of Electrical and
Electronic Engineering, BUET
Dhaka-1000, Bangladesh

4. — **Member** (External)

(Dr. Chowdhury Mofizur Rahman)
Professor
Department of Computer Science
and Engineering, BUET
Dhaka-1000, Bangladesh

Declaration

It is hereby declared that this thesis or any part of it has not been submitted elsewhere for the award of any degree or diploma.

S. M. Rahman

(S. M. Mahbubur Rahman)

Dedication

To my beloved parents

Acknowledgements

I would like to thank my advisor, Dr. Md. Kamrul Hasan, Associate Professor, Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology, Bangladesh, for his friendly supervision, constructive suggestions and constant support during this research. I am thankful to him for introducing me to the area of wavelet and Image Processing. I am grateful to my colleague Mr. Imamul Hassan Bhuiyan, to Mr. Newaz M. S. Rahim, now doing Ph. D. in Japan for supplying needed journal papers and my friends, Muhammad Shamsul Arefeen Zilany, Shaikh Anowarul Fattah for consulting the theoretical analysis. Special note of thanks goes to Dr. M. Rezwan Khan, Dr. Mohammad Ali Choudhury, Dr. Md. Aynal Haque for providing me their helping hand whenever required. Finally, I express my deep gratitude to my parents for their continuous support, love and encouragement.

Contents

List of Tables

List of Figures

Abstract

In this thesis, an improved wavelet-domain image denoising method using a new adaptive center weighted median filter (ACWMF) with the minimum mean square error (MMSE) estimator is proposed. The performance of an image denoising algorithm using the MMSE estimator is highly dependent on the accuracy of signal coefficients' variance field estimation from the noisy wavelet coefficients. The proposed ACWMF provides more accurate estimate of the variance field as compared to existing methods, by re-calculating a coefficient variance from a preliminary estimated one using an iterative smoothing process, for two reasons. First, in the statistical modeling of the wavelet coefficients, the reported methods so far neglects the effect of the high occurrence of very low magnitude coefficients in the distribution. Second, most of the cases the variance fields are estimated exploiting the global characteristics of the image wavelet coefficients. Therefore, the estimated variance field using the existing methods is not accurate enough because of the mismatch between the local statistics and the assumed one.

In this work, it is demonstrated that a straight median smoothing to the noise dominating low energy coefficients, and a new modified weighted median smoothing derived from the local neighborhood to the signal dominating high energy coefficients, improve the accuracy of the preliminary estimated variance field obtained from two popular recent methods. These two types of smoothing filter are integrated into one, termed as the ACWMF. As the smoothing process on the variance field by the ACWMF is different in noise and signal dominated window, an adaptation in accordance with the characteristics of the image coefficients under smoothing is incorporated. Simulation results on a number of standard images show better performance both in visual quality and in terms of peak signal to noise ratio (PSNR) as compared to other recent image denoising methods.

Chapter 1

Introduction

1.1 Image Denoising: Background

Visual perception is the major source (about 85 percent) of total information acquired by human [1]. In many cases images available to the users are in corrupted form. Then image denoising algorithms are an important field of image processing which has been widely studied and applied to many fields. In the practical situation an image can be corrupted by noise during its acquisition, digitization, transmission etc. Most in the cases imaging environment is not ideal. For example the vibration of the imaging instrument, e.g. spinning and tumbling of the space craft, modulation of the refractive index of the medium etc. cause noisy version of the true image [2]. Degradation may also occur due to quantization noise resulting from the digitization of the recorded images on photographic film [3]. Image may also be corrupted by the external spurious signal during its transmission. The aim of image enhancement by denoising is to remove the noise while retaining the important signal features as much as possible. Denoising also improve the perceptual quality of the corrupted image. The important applications of image denoising are in medical imagery, e.g. X-ray, magnetic resonance imagery (MRI), mammography, computer tomography, digital radiology, astronomical imagery in the space probe, remote-sensing imagery, e.g. synthetic aperture radar (SAR), weather survey, geological survey etc.

1.1.1 Image properties

The term image refers to a two dimensional intensity (it may be a light-intensity, temperature-intensity, or any other signal-intensity) function denoted by $f(x, y)$

where the value of f at a spatial coordinate (x, y) gives the intensity of the image at that point. Commonly the intensity is assumed to be light-intensity and simply it is termed as brightness. As light is a form of energy, the value of $f(x, y)$ is finite. That means, the brightness value is limited within a range. For example, for monochromatic images, the lowest value corresponds to black and the highest value corresponds to white. The interval between these two values is called gray scale and the brightness values bounded by these two limits are called gray levels. For computer processing, the continuous image function $f(x, y)$ is digitized and it is denoted as $f(i, j)$. It is sampled into a matrix of X rows and Y columns where the row and column indices (i, j) identify a point in the digitized image and the corresponding array value (which is an integer value assigned by image quantization for that image sample) gives the brightness value at that point. The elements of such matrix or array are called image pixels or shortly pels.

1.1.2 Image degradation model and definition of denoising

In image processing research group the term *denoising* means the retrieval of the true image that is corrupted by something else as we call it noise. Noise that corrupts the image may be signal dependent, or signal independent. If it is uncorrelated to the signal, the degradation model can simply be represented by a signal plus noise model. In most of the cases, the noise is assumed to be white Gaussian noise. Then the linear degradation model in discrete form can be written as

$$g(i, j) = f(i, j) + n(i, j) \tag{1.1}$$

where $f(i, j)$ is the true image, $n(i, j)$ is the additive white Gaussian noise (AWGN) and $g(i, j)$ is the degraded version of the true image. For signal dependent noise like Poisson, film-grain, speckle noise the degradation procedure cannot be modeled as additive in nature. But with some modification as in [3], we can re-model them as in Eqn. (1.1). Here we can show the re-modeling with some common signal dependent noise. For a low light imaging system the photon noise can be modeled as Poisson noise. If the observed image is $\tilde{g}(i, j)$, the degradation model can be written as in Eqn. (1.1) with the modification $g(i, j) = \tilde{g}(i, j)/\beta$ and noise being white Gaussian with a variance ${\sigma_n}^2(i, j) = E\{\tilde{g}(i, j)/\beta^2\}$. Here β is a factor of proportionality and $E\{\cdot\}$ denotes the statistical expectation operator. When images recorded on a photographic film are digitized a noise called film-grain noise corrupts the image. According to Froehlich *et. al.* in [3], [10] the corruption by film-grain noise can be modeled as

$$g(i,j) = f(i,j) + K[f(i,j)]^{\alpha} u_1(i,j) + u_2(i,j) \tag{1.2}$$

with $u_1(i,j)$ and $u_2(i,j)$ are the samples from two Gaussian distributed, totally uncorrelated, zero-mean random processes with variances ${\sigma_1}^2$ and ${\sigma_2}^2$, respectively. In Eqn. (1.2), K is a proportionality factor and α is an exponent usually having a value of 0.5. It can be easily shown as in [3] that the degradation by film grain noise can be modeled as presented in Eqn. (1.1) modifying the noise variance as ${\sigma_n}^2(i,j) = K^2 E\{g(i,j)\}{\sigma_1}^2 + {\sigma_2}^2$. Another type of noise encountered in the practical imaging system is the speckle noise. It is generated when system uses coherent radiation pattern like SAR, ultrasound, laser, sonar etc. The most widely used model for speckle noise is multiplicative [3], [11] and is given by

$$g(i,j) = f(i,j)u(i,j) \tag{1.3}$$

where $u(i,j)$ is a stationary noise uncorrelated with the image having a mean $\bar{u}$ and a variance ${\sigma_u}^2$. Depending on the application, distribution of the noise can be assumed Gaussian (for SAR images), exponential, Rayleigh etc. The multiplicative model in Eqn. (1.3) can be converted to additive model as in Eqn. (1.1), where $n(i,j)$ is zero-mean additive noise with variance given by

$${\sigma_n}^2(i,j) = \frac{{\sigma_u}^2}{1+{\sigma_u}^2}[v_g(i,j) + {m_g}^2(i,j)] \tag{1.4}$$

where $v_g(i,j)$ and ${m_g}^2(i,j)$ being the variance and the mean, respectively of the noisy image at the point (i,j). Therefore, for most of the practical noise encountered by the image processing society can be either modeled or re-modeled as additive noise. And for simplicity, we consider the synthetic noise be additive white Gaussian noise.

The terminology *image denoising* can be mathematically interpreted as the maximum possible retrieval of $f(i,j)$ from the observed image $g(i,j)$ without or using only partial information about the noise. The denoised version of the degraded image is very often represented by $\hat{f}(i,j)$ where '$\hat{\ }$' indicates an estimate. In many image processing applications the objective of the denoising technique is to minimize the mean square error (MSE) between the true and estimate image i.e. to minimize $\varepsilon = E\{||f - \hat{f}||^2\}$.

1.1.3 Literature review of the image denoising techniques

Images corrupted by noise can be denoised either in the pixel domain or in the transform domain. Common image denoising filters in the pixel domain are mean filter, median filter, Lee filter, Refined Lee filter, noise updating repeated wiener filter (NURW), adaptive neighborhood mean and median filter, weighted median filter etc. The mean filter smoothes the pixel data and it is a primitive one. The median filter performs better than mean filter, since it has a better chance that pixels those are not affected by noise will remain unaltered [4]. Lee filter estimate the original image using linear minimum mean square error (LMMSE) criterion assuming additive-noise model [5]. The Lee filter provides good noise attenuation over uniform areas, but it yields poor performance near the edges. The refined Lee filter improves its performance by self adjusting window on the basis of some criteria of the data set inside the window (like variance, entropy etc.) [6]. The NURW filter is nothing but the iterative application of the Lee filter, where in each iteration noise variance is updated by using locally LMMSE criterion [7]. To reduce blurring, filter having shape and size adaptive window is proposed in [3] and the method is termed as adaptive neighborhood filtering (ANF). In the ANF method, the window is grown around the seed pixel considering the local statistics. This shape adapted denoising method suffers from severe computational effort. Most of the linear filters in the literature encounter the blurring effect at the edges. Recently a novel nonlinear filter termed as the weighted median filter [8], [9] is used for image denoising. Since the difference of the data set inside the window represents the edge information, a linear combination of weighted median of the difference can improve the denoising performance [8]. In this case a major problem is to determine the weighting function. The performance of iterative median smoothing using Gaussian weighting function is described in some recent image processing articles [9]. This iterative smoothing process also suffers from huge computational burden.

To the signal processing community, a number of transform domains e.g. Fourier transform domain, discrete cosine transform (DCT) domain, generalized lapped orthogonal transform (GenLOT), wavelet domain, multiple-basis representation (MBR) and so on are preferable to process digital data depending on the application type [28]. The image denoising is not out of this territory. The choice of transform depends on a number of factors e.g. computational complexity, coding gain. Computational complexity is measured by the number of additions and multiplications necessary to implement the transform. Coding gain is a measurement that represents how well the transformation compacts the signal energy into a small number of coefficients. The comparison of some common transforms are available in [12], [16]. According to their study DCT shows a less computational complexity over wavelet transform. On the other hand wavelet is superior than DCT in respect of coding gain. Moreover wavelet transform gives additional freedom to choose different kinds of wavelet filter, in contrast to the only one in DCT. Due to high energy compaction efficiency of the wavelet transform, it shows significant success in the field of image denoising in the recent years [17]-[24]. Wavelet shrinkage method proposed by Donoho [17] is the pioneering work for image denoising and image compression using the wavelet transform. The described method in [17] provides a mini-max optimal solution. But many image denoising techniques prefer MSE as denoising criterion. The method reported in [19] is a spatially adaptive wavelet threshold (SAWT) method that uses MSE as a denoising performance index. The threshold parameter defined here is calculated by assuming that wavelet coefficients in each scale have generalized Gaussian (GG) distribution, which is seldom met in reality. The maximum likelihood (ML) method proposed in [21] shows that the MMSE estimator provides better result than the wavelet threshold method reported in [19]. To calculate the variance field for the MMSE estimator the maximum a posteriori (MAP) method [22] uses a Bayes-rule, modeling the wavelet coefficients in each subband as an independent Gaussian random variable with zero-mean and slowly varying variance field. An exponential prior for image coefficient variance exploited here, is motivated from the approximate histogram of true image coefficients. The image denoising algorithm proposed in [24] exploits both the inner and inter scale dependencies of image wavelet coefficients. In this method variance field is estimated as a two step procedure. First, an initial estimate of a variance field is calculated using the ML method proposed in [21]. Second, a finer scale coefficient variance is set to zero provided that its parent coefficient variance is less than a predefined threshold. This method shows a denoising performance close to that reported in [22].

1.1.4 Motivation and possible outcome

The prior assumed in [22] is derived from the approximate histogram of the image wavelet coefficients. The true histogram reveals that the method simply overlooks the effect of significantly high occurrence of very low magnitude wavelet coefficients in the distribution. Therefore, relatively low energy coefficient variances obtained from this method are overestimated in magnitude. We expect that a straight median smoothing on the low energy coefficient variances, which correspond to the noise dominated smooth regions of the image, will reduce the noise that remain in the preliminary estimated ones. Moreover, the histogram only represents a global picture of the coefficient statistics. So the prior devised from the histogram analysis as in [22] can not represent the local variation of the signal and noise perfectly. If the high energy coefficient variances those correspond to the signal dominated detail regions of the image, remain as they are in the the coefficient variance smoothing process, they will maintain the accuracy of the preliminary estimated ones. Our motive is to improve the estimation performance in these coefficients also. It is intuitive that an additional smoothing by a new modified weighted median filter on the preliminary estimated high energy coefficient variances, will provide more accuracy to their estimation. For this purpose, a smoothing term derived from the weighted median of the local neighborhood will be added to a high energy coefficient variance with a view that the gradient of the variance field be smoothed. The two possible outcomes are the overall improved denoising performance and faster convergence of the smoothing process. In our work, we want to unify these two types of smoothing in a single filter termed as adaptive center weighted median filter (ACWMF). Since the smoothing process of the variance field by the proposed ACWMF is different, depending on – signal dominant window or noise dominant window, it should be adapted according to the image wavelet coefficient characteristics. According to our methodology, the center value of the mask plays the key role for the adaptation of the proposed smoothing filter, hence, it is termed as the adaptive center weighted median filter (ACWMF). For inter band adaptation of this smoothing filter, we suggest an exponentially decaying inter scale model of image coefficients. And for inner band adaptation we prescribe a scale independent threshold parameter. The preliminary estimated variance field can be obtained from an existing method [21], [22].

The estimated variance field obtained from the ACWMF will then be used in the minimum mean square error (MMSE) estimator for denoising the noisy wavelet coefficients.

1.2 Organization of the Thesis

Chapter 1 describes the definition of denoising and the importance of the problem. This chapter also includes the brief history regarding different denoising strategies both in pixel and transform domains. The motivation of this work is also discussed in this chapter.

In Chapter 2, a brief review of wavelet theory and some of the recent wavelet-based image denoising algorithms are given. The contents of this chapter will provide some basic definitions and preliminary knowledge of wavelet transform that will facilitate the understanding of the main algorithm. This chapter also includes some recent wavelet domain image denoising methodologies.

Chapter 3 describes image denoising method using the proposed adaptive center weighted median filter (ACWMF). We start with presenting a statistical model (for both inter and inner scales) of the image wavelet coefficients. Finally, procedure of the calculation of variance field using the proposed ACWMF for the minimum mean square error (MMSE) estimator is described.

Chapter 4 represents the experimental results. We have tested the performance of the proposed algorithm using some of the standard images. Results are compared both in terms of mean square error (MSE), and visual quality. In addition convergence in the coefficient domain is also demonstrated.

Finally, we conclude in Chapter 5 with overall discussion of the presented work and some remarks for possible avenues of future investigation.

Chapter 2

Wavelets and Image Denoising

2.1 Introduction

The goal of image transformation is to represent the image in compact and complete form in any other domain from its pixel domain. The transform should decorrelate the spatially distributed energy to fewer data samples such that no information is lost. Due to the high energy compaction along with a special type of decorrelation, the wavelet transform surpasses the other transforms especially in the field of image enhancement by denoising or restoration, image compression etc. Wavelet transformed based or in general any kind transformed based image denoising scheme is basically a three-step procedure. Fig. 2.1 shows a block diagram of the transformed base image denoising scheme. First step is the transformation of spatial information into coefficient domain. The second step is to manipulate the transform coefficients. In general, some of the data samples with insignificant energy-levels are discarded in this step. The remaining coefficients also undergo some modification. In the last step, inverse transformation reconstructs the image from the restored coefficients. During the denoising scheme, not only a portion of noise is removed but also image features are lost. The objective of image denoising is to maximize the noise removal with least possible loss of the image features. To achieve this goal, wavelet shows significant success in the recent years [17]-[24]. Before introducing some already established wavelet based denoising algorithms, a brief introduction to wavelet theory will be presented here. Starting from the multiresolution analysis, first we will represent how 1-D wavelet coefficients can be obtained from a 1-D signal. The implementation of the filter bank theory to calculate the wavelet coefficients will be discussed later.

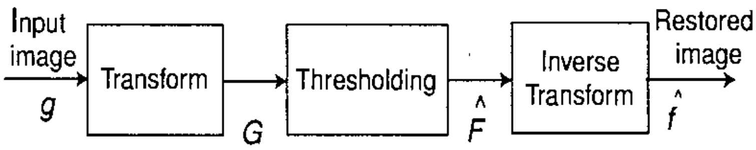


Figure 2.1: Block diagram of a transform based image denoising algorithm.

Finally, a procedure of calculating the 2-D wavelet coefficients using 1-D filter bank concept will be presented.

2.2 Wavelet Theory: A Brief Introduction

2.2.1 Wavelets and multiresolution analysis

The theory of wavelets starts with the concept of a multiresolution analysis. Multiresolution analysis enables us to represent a square integrable function $f(t)$, i.e. the function that exists in $L^2(-\infty, \infty)$ with $\int_{-\infty}^{\infty} |f(t)|^2 \mathrm{d}t < \infty$, as a linear sum of the weighted orthogonal bases (at different resolutions) having the vector spaces within L^2. In wavelet analysis, we go one step further and stipulate that every basis function be a dilation and translation of a single scaling function of unit norm, denoted as $\phi(t)$. Here it is required that integer translations $\{\phi(t - k)\}|_{k \in Z}$ be linearly independent and produce orthonormal basis for the subspace V_0. Likewise, for a fixed integer scale s, the translations $\{2^{-s/2}\phi(2^{-s}t - k)\}|_{k \in Z}$ form an orthonormal basis for the subspace V_s, such that the subspaces satisfy

$$\cdots \subset V_2 \subset V_1 \subset V_0 \subset V_{-1} \subset V_{-2} \subset \cdots \tag{2.1}$$

and

$$\overline{\bigcup_{s=-\infty}^{\infty} V_s} = L^2(\Re), \qquad \bigcap_{s=-\infty}^{\infty} V_s = \{0\} \tag{2.2}$$

Then each function $f(t) \in L^2(\Re)$ can be written as a linear combination of these basis functions with weights $\alpha_{s,k}$ given by

$$f(t) = \sum_{s=-\infty}^{\infty} 2^{-s/2} \sum_{k=-\infty}^{\infty} \alpha_{s,k}\phi(2^{-s}t - k) \tag{2.3}$$

where

$$\alpha_{s,k} = \int_{-\infty}^{\infty} f(t) 2^{-s/2}\phi(2^{-s}t - k)\mathrm{d}t. \tag{2.4}$$

The weights $\alpha_{s,k}$ are called scaling coefficients. From Eqn. (2.1) it is clear that $V_s \in V_{s-1}$, i.e., subspaces V_s are not mutually disjoint, therefore Eqn. (2.3) doesn't represent an orthogonal expansion of $f(t)$ into the subspaces V_s. Moreover, the functions $\{2^{-s/2}\phi(2^{-s}t-k)\}|_{s,k\in Z}$ do not represent a basis for $L^2(\Re)$ because they are not all linearly independent across scales.

The nesting property of the subspaces implies that the scaling function $\phi \in V_0$ also belongs to V_{-1} and thus satisfies the two-scale dilation equation

$$\phi(t) = \sqrt{2} \sum_{k=-\infty}^{\infty} h_k \phi(2t-k) \tag{2.5}$$

where the sequence $\{h_k\}$ represents the set of coefficients corresponding to the projection of the bases $\phi_{s-1,k}(t)$ of space V_{s-1} on to the bases $\phi_{s,k}(t)$ of space V_s, i.e.,

$$h_k \;=\; <\phi(t), \phi(2t-k)>, \quad k \in Z \tag{2.6}$$

Daubechies [14] has shown that $\{h_k\}$ may be also regarded as the lowpass filter coefficients in a filter bank.

The multiresolution nature of this analysis lends itself to a graphic interpretation. As the scale s decreases, the support of each basis function $2^{-s/2}\phi(2^{-s}t-k)$ decreases as well. Hence smaller scales may be regarded as representing a finer, or more detailed, resolution. Conversely, larger values of s indicate basis functions with wider support, suggesting coarser scale and less detail. This makes a sense in light of Eqn. (2.1), since $V_s \subset V_{s-1}$ implies that V_{s-1} represents a space of functions at a finer resolution than V_s.

Because we want an orthogonal decomposition of $f(t)$, we now define the "difference" space W_s as the complement of V_s in V_{s-1}, namely

$$V_{s-1} = V_s \bigoplus W_s, \qquad\qquad V_s \bigcap W_s = \emptyset \tag{2.7}$$

It can be shown that the W subspaces provide a decomposition of $L^2(\Re)$ into mutually orthogonal subspaces [14], i.e.

$$W_s \perp W_{s'} \qquad \text{if } s \neq s', \qquad \text{and} \qquad \bigoplus_{s=-\infty}^{\infty} W_s = L^2(\Re) \tag{2.8}$$

The basis for the subspace W_0 consists of translations of a new function, $\psi(t)$. Since the W spaces inherit the scaling properties of the V spaces, $\{2^{-s/2}\psi(2^{-s}t-k)\}|_{k\in Z}$ will be a basis for W_s. As with the V spaces, each function $f(t) \in L^2(\Re)$ can be written as a linear combination of these basis functions with weights $\beta_{s,k}$

given by

$$f(t) = \sum_{s=-\infty}^{\infty} 2^{-s/2} \sum_{k=-\infty}^{\infty} \beta_{s,k}\psi(2^{-s}t - k) \tag{2.9}$$

where

$$\beta_{s,k} = \int_{-\infty}^{\infty} f(t)2^{-s/2}\psi(2^{-s}t - k)\mathrm{d}t. \tag{2.10}$$

Here the function $\psi(t)$ is called the wavelet function, and the weights $\beta_{s,k}$ are called the wavelet coefficients. The difference between Eqn. (2.3) and Eqn. (2.9) is that the latter represents an orthogonal expansion of $f(t)$, also the functions $\{2^{-s/2}\psi(2^{-s}t - k)\}|_{s,k\in Z}$ do represent a basis for $L^2(\Re)$. Since $\psi \in W_0$ and $W_0 \subset V_{-1}$, we have

$$\psi(t) = \sqrt{2} \sum_{k=-\infty}^{\infty} g_k\phi(2t - k) \tag{2.11}$$

where $\{g_k\}$ is the set of coefficients corresponding to the projection of bases $\phi_{s-1,k}(t)$ of space V_{s-1} onto the bases $\psi(t)$ of space W_s and these can be represented as

$$g_k = < \psi(t), \phi(2t - k) >, \quad k \in Z \tag{2.12}$$

It is evident from Eqn. (2.8) that each subspace V_P is the direct sum of V_Q and some W subspaces, we can write for $Q > P$.

$$V_P = V_Q \bigoplus W_Q \bigoplus W_{Q-1} \bigoplus W_{Q-2} \bigoplus \cdots \bigoplus W_{P+1} \tag{2.13}$$

The another way of saying that a function belonging to the subspace V_P may be written as the linear combination of basis functions of mutually orthogonal subspaces $V_Q, W_Q, W_{Q-1}, \cdots, W_{P+1}$. Thus given the function $f(t) \in V_P$ we can write

$$f(t) = 2^{-Q/2} \sum_{k=-\infty}^{\infty} \alpha_{Q,k}\phi(2^{-Q}t - k) + \sum_{s=P+1}^{Q} 2^{-s/2} \sum_{k=-\infty}^{\infty} \beta_{s,k}\psi(2^{-s}t - k) \tag{2.14}$$

This equation is referred to as the *discrete wavelet transform* (DWT) of the corresponding input signal $f(t)$. The coefficients $\alpha_{Q,k}$ and $\beta_{s,k}$ can be obtained as

$$\begin{aligned} \alpha_{Q,k} &= < f(t), \phi_{Q,k}(t) > \\ \beta_{s,k} &= < f(t), \psi_{s,k}(t) > \end{aligned} \tag{2.15}$$

Here $\alpha_{Q,k}$ are generated by the $\phi_{Q,k}(t)$; thus they correspond to a coarse resolution in the multiresolution analysis (MRA). The coefficients $\alpha_{Q,k}$ are called *approximate coefficients* at scale Q. And $\beta_{s,k}$ are generated by the $\psi_{s,k}(t)$; therefore they represent the *details* at scale s. The coefficients $\beta_{s,k}$ are called *wavelet coefficients* at scale s.

2.2.2 Implementation of the DWT

It has been shown that the wavelet analysis relations given by the two-scale Eqn. (2.5) and Eqn. (2.11) can be expressed in terms of a two channel perfect reconstruction (PR) filter bank, where the lowpass and highpass filters have impulse responses $\{h_k\}$ and $\{g_k\}$, respectively [14]. Perfect reconstruction means that the output of the filter bank is identical to the input, except for a possible delay and overall scaling factor. A particular signal of interest, $x(t) \in V_0$, can be written as a linear combination of the basis functions $\{\phi(t-k)\}$ with weights $v_{0,k}$:

$$x(t) = \sum_{k=-\infty}^{\infty} v_{0,k}\phi(t-k) \tag{2.16}$$

If we pass the coefficients $v_{0,k}$ into a two-channel filter bank, after the first level of filtering and downsampling, we obtain the lowpass coefficients

$$v_{1,k} = \sum_{m=-\infty}^{\infty} h_{m-2k}v_{0,k} \tag{2.17}$$

and highpass coefficients

$$w_{1,k} = \sum_{m=-\infty}^{\infty} g_{m-2k}v_{0,k} \tag{2.18}$$

Similar iterations of the filter bank on the low pass channel result in the coefficients

$$\begin{aligned} v_{s,k} &= \sum_{m=-\infty}^{\infty} h_{m-2k}v_{s-1,k}, \\ w_{s,k} &= \sum_{m=-\infty}^{\infty} g_{m-2k}v_{s-1,k}, \end{aligned} \tag{2.19}$$

after s stages of the analysis filters, where $v_{s-1,k}$ is the output of the $(s-1)$-iteration of the filter bank. With the filters $\{h_k\}$ and $\{g_k\}$ as given in Eqn. (2.5) and Eqn. (2.11) respectively, it turns out that the numbers $\{v_{s,k}\}$ are in fact just the scaling coefficients $\{\alpha_{s,k}\}$, and the numbers $\{w_{s,k}\}$ are actually the wavelet coefficients $\{\beta_{s,k}\}$. Thus the wavelet decomposition of the signal $x(t)$ in the bases for the subspaces V_s and W_s, for $s > 0$, can be found by repeated filtering of

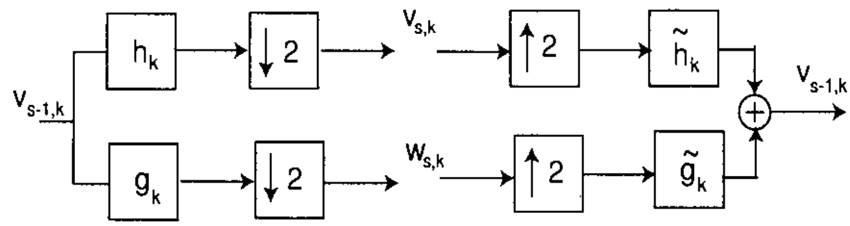


Figure 2.2: Analysis and synthesis stages of a 2-channel single-level orthogonal PR filter bank.

the scaling coefficients $\{v_{0,k}\}$ corresponding to the subspace V_0. The synthesis portion of the filter bank reconstructs the sequences $\{v_{s-1,k}\}$ from the sequences $\{v_{s,k}\}$ and $\{w_{s,k}\}$ via

$$v_{s-1,k} = \sum_{m=-\infty}^{\infty} \tilde{h}_{k-2m} v_{s,k} + \sum_{m=-\infty}^{\infty} \tilde{g}_{k-2m} w_{s,k} \tag{2.20}$$

as shown in Fig. 2.2. Note that the filters in the synthesis stage, with impulse responses $\{\tilde{h}_k\}$ and $\{\tilde{g}_k\}$, are not necessarily the same as those in the analysis stage, which have impulse responses $\{h_k\}$ and $\{g_k\}$. The filters $\{\tilde{h}_k\}$ and $\{\tilde{g}_k\}$ are called the dual filters to $\{h_k\}$ and $\{g_k\}$ [12]. The dual filters have corresponding dual scaling and wavelet functions, $\tilde{\phi}(t)$ and $\tilde{\psi}(t)$, which generate the dual spaces $\tilde{V}_s$ and $\tilde{W}_s$.

Orthogonality and symmetry of the basis functions are the two desirable characteristics of wavelet transform. Orthogonality decorrelates the transform coefficients, thereby minimizing the redundancy. Symmetry provides linear phase and permits simple symmetric boundary extension techniques that minimizes border artifacts. Since an FIR filter bank cannot have both orthogonal and symmetric filters with length greater than two, so for scaler wavelets, the filter bank cannot simultaneously have orthogonality and linear phase property (the only exception is the Haar wavelet with filter length two). For a truly orthogonal PR filter bank, $\{\tilde{h}_k\}$ and $\{\tilde{g}_k\}$ are just the time reversals of $\{h_k\}$ and $\{g_k\}$ respectively. Filter with a length greater than two, with symmetric property is called biorthogonal filter. Biorthogonal wavelets are preferable for image compression, since symmetric boundary extension of the transform coefficients improves compression performance. On the other hand, denoising requires higher decorrelation efficiency and therefore orthogonal wavelets are more preferable for this purpose. In our experiment, we examine the performances using 8-tap orthogonal Daubechies wavelet

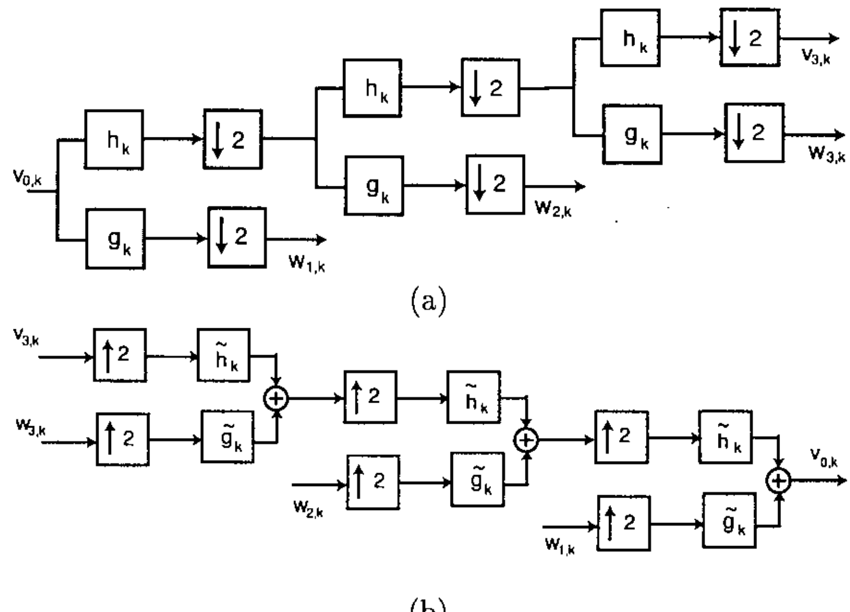


Figure 2.3: A three-level orthogonal PR filter bank; (a) Analysis stages , (b) Synthesis stages.

D_8. An example of three level analysis filter and its corresponding synthesis filter is shown in Fig. 2.3.

2.2.3 Vanishing moments of wavelets

As mentioned earlier, one of the desirable wavelet characteristics for image denoising is good energy compaction efficiency. The number of *vanishing moments* possessed by a wavelet provides a means for evaluating the energy compaction efficiency. The greater the number of vanishing moments, the better the energy compaction efficiency will be. A wavelet is said to have K vanishing moments if

$$\int_{-\infty}^{\infty} t^m \psi(t) \mathrm{d}t = 0 \tag{2.21}$$

where $0 \leq m \leq K-1$. This implies that polynomials up to order $K-1$ can be exactly represented by the approximate coefficients. The D_8 orthogonal filter bank has 8 vanishing moments, and we use this filter expecting high energy compaction.

2.2.4 DWT of 2-D signals

Since an image is a two dimensional function, it requires the 2-D wavelet transform. Separable wavelet transforms are common in two dimensional case. Sepa-

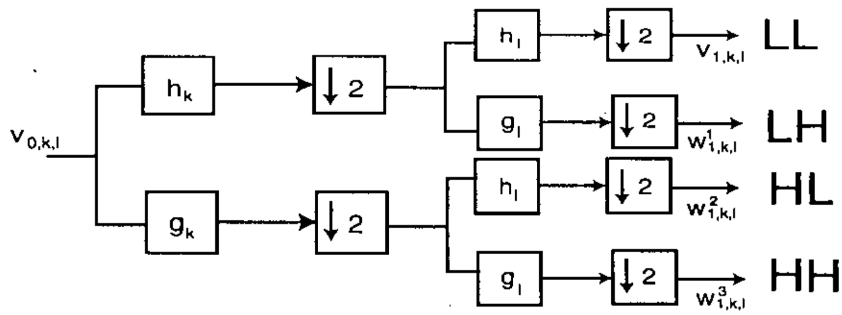


Figure 2.4: Analysis of a 2-channel single-level orthogonal PR filter bank for 2-D scaler wavelet transform.

rability of the wavelet transform enables one to separate the scaling and wavelet functions so that the 2-D transform is equivalent to two separate 1-D transforms. The 2-D DWT is implemented as a 1-D row transform followed by a 1-D column transform. The transform coefficients are obtained by projecting the 2-D input signal $f(r,t)$ on to a set of 2-D basis functions, expressed as the product of two 1-D basis functions as

$$\begin{aligned} \phi(r,t) &= \phi(r)\cdot\phi(t), & \psi_1(r,t) = \psi(r)\cdot\phi(t), \\ \psi_2(r,t) &= \phi(r)\cdot\psi(t), & \psi_3(r,t) = \psi(r)\cdot\psi(t) \end{aligned} \tag{2.22}$$

If the scaling function $\phi(r,t)$ and the wavelets are $\psi_1(r,t)$, $\psi_2(r,t)$ and $\psi_3(r,t)$, then the approximate coefficients, $\alpha(s,k,l)$ and the corresponding three types of detail coefficients, $\beta^1(s,k,l)$, $\beta^2(s,k,l)$ and $\beta^3(s,k,l)$ are

$$\begin{aligned} \alpha_{s,k,l} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(r,t)2^{-s/2}\phi(2^{-s}t-k)\phi(2^{-s}t-l)\mathrm{d}r\mathrm{d}t \\ \beta^1_{s,k,l} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(r,t)2^{-s/2}\phi(2^{-s}t-k)\psi(2^{-s}t-l)\mathrm{d}r\mathrm{d}t \\ \beta^2_{s,k,l} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(r,t)2^{-s/2}\psi(2^{-s}t-k)\phi(2^{-s}t-l)\mathrm{d}r\mathrm{d}t \\ \beta^3_{s,k,l} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(r,t)2^{-s/2}\psi(2^{-s}t-k)\psi(2^{-s}t-l)\mathrm{d}r\mathrm{d}t \end{aligned} \tag{2.23}$$

The coefficients $\alpha(s,k,l)$ are the coarse coefficients that constitutes the LL_s subband. The coefficients $\beta^1(s,k,l)$ are the vertical details corresponding to the LH_s subband, $\beta^2(s,k,l)$ are the horizontal details corresponding to the HL_s subband and $\beta^3(s,k,l)$ are the diagonal details corresponding to the HH_s subband. The procedure of orthogonal decomposition or reconstruction for 2-D sequence using a filter bank algorithm, is also a simple extension of the 1-D sequence. For simplicity, a single stage orthogonal decomposition of a 2-D signal is shown in Fig.

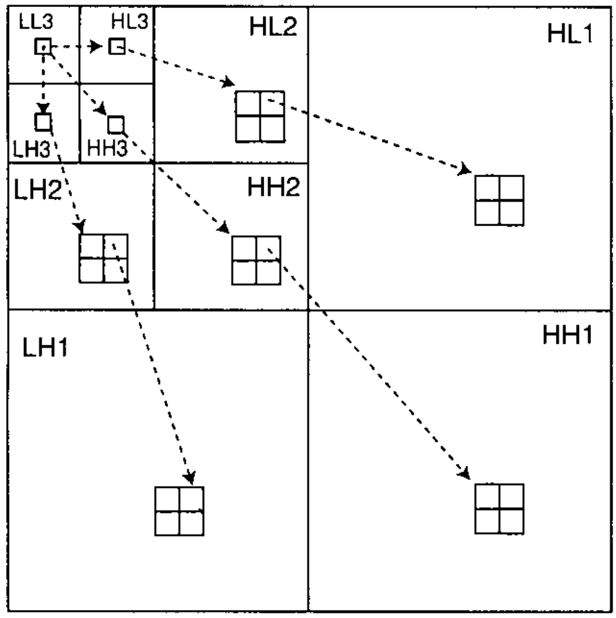


Figure 2.5: Spatial correlation of 2-D wavelet coefficients for three level decomposition.

2.4.

In the field of image denoising and image compression, the DWT stands as a unique transform for one of its very important features. One of the advantages of the wavelet transform over the other usual transforms is, there exist correlation between the discrete wavelet coefficients in each of the three spatial directions (i.e. HL_s, LH_s, and HH_s) [13], [15], [22]-[27]. The spatial correlation and the parent-child relationship among the wavelet coefficients at different resolutions are shown in Fig. 2.5 . In the next section, we present a brief review of some recent image denoising techniques in the wavelet domain.

2.3 Review of Wavelet-based Image Denoising

Research results for denoising images corrupted by additive white Gaussian noise (AWGN) in the wavelet domain have been proposed recently in [17]-[24]. Coefficient thresholding is one of the approaches [17]-[20] where noise dominant coefficients are forced to zero, by comparing the coefficients to an estimated threshold which is a function of the noise level. The wavelet shrinkage method in [17] provides a universal threshold parameter for image denoising and also gives an asymptotically minimax-optimal solution [23]. However, in many image processing applications minimum mean squared error (MMSE) metric is more preferable [19]-[23]. With an objective to minimize the MSE, spatially adaptive wavelet threshold (SAWT) method proposed in [19] defines an adaptive threshold parameter, which incorporates the changing characteristics of image. To estimate this adaptive threshold, the method assumes that the image coefficients within each subband are independently identically distributed (i.i.d.) random variables having a generalized Gaussian (GG) distribution. The maximum likelihood (ML) method [21], finite impulse response (FIR) Wiener filter method [23], and their variants claim that the MMSE estimator performs better than the wavelet threshold methods. To estimate the coefficient variances for the MMSE estimator, the methods proposed in [22]-[24] assume that coefficients are conditionally independent Gaussian random variables with zero-mean and slowly varying variance. The method proposed in [22] calculate the coefficient variances assuming an exponential prior distribution for the noise-free image coefficient variances in each subband. The assumption of the exponential prior is motivated by the resemblance of the approximate histogram of signal coefficients normalized by their standard deviation to the unit variance, zero-mean Gaussian probability density function (p.d.f). Incorporating this prior to the Bayes-rule, the maximum a posteriori (MAP) estimator proposed in [22] provides an appreciable result. The method proposed in [24] suggests an additional thresholding for the finer scale preliminary ML estimated variances exploiting the fact that the inter scale coefficients' power decay exponentially. In [23], an exponential decay model of local autocorrelation of image coefficients is exploited for estimating the FIR Wiener filter parameters. Both the methods proposed in [23] and [24] show results close to that reported in [22]. The performances of the algorithms reported in [22] and [23] are highly dependent on the degree of match between the actual distribution of the underlying image wavelet coefficients and the assumed prior distribution function. And the performance of the method described in [24] is highly dependent on the threshold parameter. Now a brief introduction of the image denoising techniques will be presented in terms of their chronological performances. For the simplicity, here we avoid the detail illustration of the methodology showing intermediate or similar results.

2.3.1 Denoising by Donoho's wavelet shrinkage method

In a sense Donoho is the pioneer to introduce wavelet shrinkage [17] as a successful tool for image denoising. Wavelet shrinkage refers to the reconstruction of the coefficients obtained by the wavelet transformation, followed by shrinking the empirical wavelet coefficients towards zero, followed by the inverse transformation. For denoising Donoho [17] has proposed a threshold value such that wavelet coefficients with magnitude less than the threshold are forced to zero. The reported threshold is often termed as Donoho's universal threshold.

Let $F_s(k)$ represent the orthonormal wavelet coefficients of the true image for a given scale s with 1−D representation. The corresponding wavelet coefficients of the noisy image can be written as $G_s(k) = F_s(k) + N_s(k)$, where $N_s(k)$ are the wavelet coefficients of the additive noise. If J_s is the total number of coefficients in a given scale s, and the noise variance is ${\sigma_n}^2$, then according to Donoho the threshold, T should be $\sigma_n\sqrt{2\log J_s}$. And using this threshold level the denoised image wavelet coefficients can be estimated either by hard-thresholding or by soft-thresholding. The popular hard-threshold function is given by

$$\hat{F}_s(k) = F_s(k).\mathbf{1}\{F_s(k) > T\} \tag{2.24}$$

which keeps the input if it is larger than the threshold T; otherwise, it is set to zero. The soft-threshold function also known as the shrinkage function given by

$$\hat{F}_s(k) = \mathrm{sgn}\,(F_s(k))\,(|F_s(k)| - T) \tag{2.25}$$

takes the argument and shrinks it toward zero by the threshold T. For denoising, soft-thresholding is chosen over the hard-thresholding for several reasons. First, soft-threshold has been shown to achieve near-optimal minimax solution over the large range of Besov spaces. Second, for the assumption of Gaussian prior which is common in many of the algorithms concerned, the optimal soft-thresholding yields smaller risk than optimal hard-thresholding. Lastly, in practice, the soft-thresholding provides more visually pleasant images over hard-thresholding because the latter is more discontinuous and yields abrupt artifacts when noise energy is significant.

Donoho [17] has also introduced a noise variance estimator from the highest band of the wavelet coefficients. Its practical importance is recognized with high appreciation to the image processing community, since in the practical case noise variance is totally unknown and it provides a way to estimate it. If G_1 represents the noisy wavelet coefficients in the highest subband HH_1, then the estimated noise standard deviation may be obtained as

$$\hat{\sigma}_n = \frac{\text{med}(|G_1|)}{0.6745} \tag{2.26}$$

2.3.2 Denoising by spatially adaptive wavelet threshold

Donoho's universal threshold [17] and Johnstone's SureShrink [18] methods define the threshold parameter which are uniform over the subband coefficients. It is obvious that the threshold value should be adaptive to the spatially changing statistics of the image. Spatially adaptive wavelet threshold (SAWT) proposed by S.G. Chang *et. al.* [19] incorporates this adaptation. The method assumes that image wavelet coefficients in each subband can be modeled as zero-mean random variables with generalized Gaussian (GG) distribution. The authors term the method as BayesShrink method, since the proposed algorithm minimizes MSE in a Bayes framework. According to the method the optimal threshold T^* is defined to be the argument which minimizes the expected error

$$T^* = \arg\min_T E_{G_s|F_s,F_s}(\hat{F}_s - F_s)^2 \tag{2.27}$$

where $G_s|F_s \sim \varphi(g - f, {\sigma_n}^2)$ and $F_s \sim GG_\tau, \sigma_F(f)$. Here, the parameter σ_F is the standard deviation of true image coefficients in a subband that determines the spread of the density function, and the parameter τ is called the shape parameter. According to them, for a shape parameter having a range of $\tau \in [0.5, 4]$, the optimal threshold T^* can be well approximated by

$$T_B(k) = \frac{{\sigma_n}^2}{\hat{\sigma_s}(k)} \tag{2.28}$$

allowing at most 5% difference from the minimum expected square error, with $\hat{\sigma_s}(k)$ be the estimated coefficient standard deviation. The estimated value of $\hat{\sigma_s}(k)$ plays a key role in image denoising as the threshold T_B depends on it. Assuming that common images have shape parameter within the range mentioned above, the proposed method incorporates the spatial adaptation to the threshold parameter only by estimating the coefficients' local standard deviations $\hat{\sigma_s}(k)$. To estimate the coefficient standard deviation $\hat{\sigma_s}(k)$ or simply coefficient variance $\hat{\sigma_s}^2(k)$, clustering of the coefficients with similar context is adopted in this algorithm. To calculate the activity level or simply the context of a current coefficient $G_s(k)$, in [19] a weighted average of the absolute value of the neighbors defined as

$$\mathbf{v}(k) = \hat{\mathbf{w}}^t \mathbf{u_k} \tag{2.29}$$

has been proposed. Here, $\mathbf{u_k}$ is a $M \times 1$ vector with elements being the absolute value of the coefficient itself and its square shaped local neighborhood. The weight $\hat{\mathbf{w}}$ is obtained by using the least-square method as

$$\begin{aligned} \hat{\mathbf{w}} &= \arg\min_{\mathbf{w}} \sum_k (|G_s(k)| - \mathbf{w^t u_k})^2 \\ &= (\mathbf{U^t U})^{-1} \mathbf{U^t} |\mathbf{G_s}| \end{aligned} \tag{2.30}$$

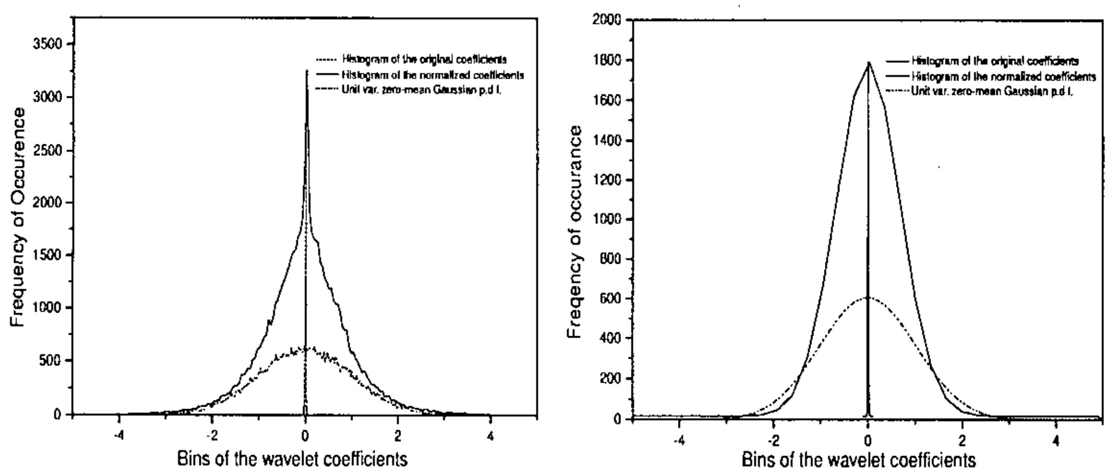
where $\mathbf{U}$ is a $J_s \times M$ matrix with each row being $\mathbf{u_k}$ for all k, and $\mathbf{G_s}$ is the $J_s \times 1$ vector containing all coefficients $G_s(k)$. Let $\mathbf{V_s}$ is a $J_s \times 1$ context vector with $\mathbf{v}(k)$ being its elements, for the corresponding coefficient vector $\mathbf{G_s}$. For any point k considering D closest points (in value) above $\mathbf{v}(k)$ and D closest points below, let A_k is the set of points having the similar context in $\mathbf{V_s}$. If B_k denotes the set of points in $\{G_s(k)\}$ that correspond to the set A_k, then the estimated variance field $\hat{\sigma_s}^2(k)$ with similar context can be obtained as

$$\hat{\sigma_s}^2(k) = \max\left(\frac{1}{2D+1} \sum_{j \in B_k} G_s(j)^2 - \sigma_n{}^2, 0\right) \tag{2.31}$$

Using the coefficient standard deviation $\hat{\sigma_s}(k)$ from Eqn. (2.31), the threshold $T_B(k)$ can be calculated from Eqn. (2.28). Finally, the common soft-thresholding technique as in Eqn. (2.25) can be used to estimate the denoised wavelet coefficients.

2.3.3 Denoising by statistical modeling of wavelet coefficients

Many researchers have incorporated statistical models for image wavelet coefficients in their own objectives. Mihcak *et. al.* [22] has modeled image wavelet coefficients as a doubly stochastic process. In [22], each subband is modeled as conditionally independent Gaussian random variables with slow varying variance rather than GG model [19], [20], and a stochastic prior is assumed to take the relation of the local variance. The assumed prior for the signal coefficient



(a) (b)

Figure 2.6: Histogram of the high-band wavelet coefficients of *Lena*, motivates the inner scale modeling; (a) Original, (b) Gaussian fitted.

variances within the scale is motivated by the histogram analysis of the image coefficients. Fig. 2.6(a) shows the true histogram of the original high-band HH_1 image coefficients of the original *Lena* image from the first scale , obtained by employing Daubechies-8 wavelets. Fig. 2.6(b) shows the Gaussian fitted approximated version of the true histogram. Both figures also show the histogram of the coefficients normalized by the true standard deviation. According to this algorithm, the normalized histogram can be approximated by a zero-mean, unit variance, Gaussian probability density function (p.d.f.), and hence an exponential prior can be assumed for each subband image coefficients.

In [22] the denoising algorithm operates in three steps as shown in Fig. 2.7. First, an initial estimate of signal coefficient variance $\tilde{\sigma_s}^2(k)$ for each point is calculated using an approximate maximum likelihood (ML) method as in [21]. From this estimated variance field, a decaying parameter λ_s for image prior is calculated. Second, using the parameter λ_s for each scale, more accurate variance filed $\hat{\sigma_s}^2(k)$ for each coefficient is re-estimated using an approximate maximum a posteriori (MAP) method. Third, the re-estimated coefficient variances are applied to a linear MMSE estimator to denoise the coefficients according to the relation

$$\hat{F}_s(k) = \frac{\hat{\sigma_s}^2(k)}{\hat{\sigma}_{g_s}^2(k)} G_s(k) = \frac{\hat{\sigma_s}^2(k)}{\hat{\sigma_s}^2(k) + \sigma_n^2} G_s(k) \tag{2.32}$$

where $\hat{\sigma_s}^2(k)$ represents the re-estimated variance of the k-th image wavelet coefficient and σ_n^2 is the known noise variance.

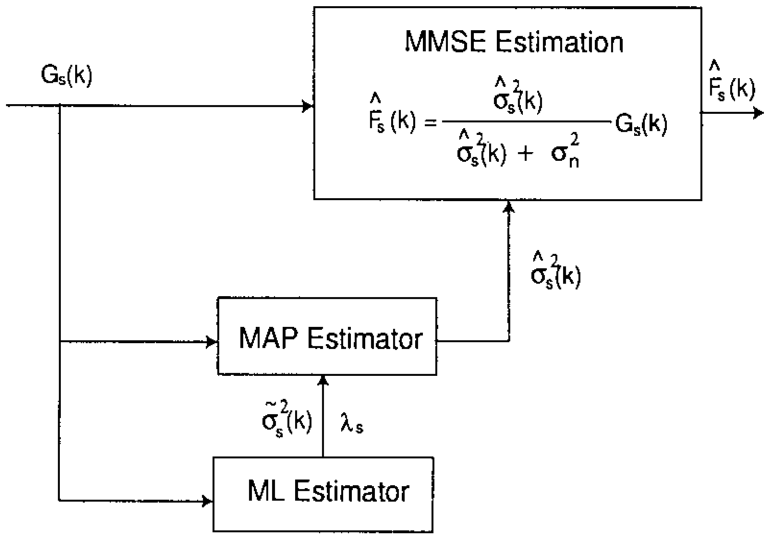


Figure 2.7: Block diagram of denoising algorithm using MAP estimator.

Assuming that the correlation between the variance of the neighboring coefficients is high in the given data field $G_s(k)$, they set $\tilde{\sigma_s}^2(k) \approx \tilde{\sigma_s}^2(j)$ for all $j \in \mathcal{N}(k)$. Here $\mathcal{N}(k)$ is a square shaped local neighborhood centered at $G_s(k)$. Variance of the estimated coefficients, $\tilde{\sigma_s}^2(k)$, can be estimated using the ML [21] estimator

$$\begin{aligned}\tilde{\sigma_s}^2(k) &= \arg\max_{\sigma_s{}^2 \geq 0} \prod_{j \in \mathcal{N}(k)} P\left(G_s(j) \mid \sigma_s{}^2\right) \\ &= \max\left(0, \frac{1}{M} \sum_{j \in \mathcal{N}(k)} G_s^2(j) - \sigma_n{}^2\right) \end{aligned} \tag{2.33}$$

where $P(\cdot|\tilde{\sigma_s}^2)$ represents the Gaussian distribution with zero-mean and variance $\tilde{\sigma_s}^2 + \sigma_n{}^2$ and M is the number of coefficients in $\mathcal{N}(k)$. Since approximate histogram of the true image coefficients as shown in Fig. 2.6(b) resembles to the zero-mean, unit variance Gaussian probability density function (p.d.f), a prior marginal near Gaussian distribution can be assumed for true image coefficients. Now, defining λ_s as an average of estimated signal coefficients' variances in each subband and assuming a prior marginal distribution $f_{\sigma_s}(\sigma_s{}^2) = (1/\lambda_s)\exp(-\sigma_s{}^2/\lambda_s)$ for each sub-band, $\hat{\sigma_s}^2(k)$ can be re-estimated using the MAP estimator as

$$\begin{aligned}\hat{\sigma_s}^2(k) &= \arg\max_{\sigma_s{}^2 \geq 0} \left[\prod_{j \in \mathcal{N}(k)} P\left(G_s(j) \mid \sigma_s{}^2\right)\right] f_{\sigma_s}(\sigma_s{}^2) \\ &= \max\left(0, \frac{M\lambda_s}{4}\left[-1 + \sqrt{1 + \frac{8}{\lambda_s M^2} \sum_{j \in \mathcal{N}(k)} G_s^2(j)}\right] - \sigma_n{}^2\right) \end{aligned} \tag{2.34}$$

For each subband the parameter λ_s can be calculated from the standard deviation of the estimated coefficients $\tilde{\sigma_s}^2$ using the ML estimator. The final estimate of the coefficient variances $\hat{\sigma_s}^2(k)$ are then used in Eqn. (2.32) to denoise the corrupted coefficients. The inverse transform of $\hat{F}_s(k)$ gives the denoised version of the given noisy image.

2.4 Conclusion

The importance of the wavelet transform in the field image denoising has been briefly discussed first. Next, a short introduction to the wavelet theory is presented. The 2-D extension of the 1-D wavelet transform is also discussed. Finally, some of the most recent wavelet domain denoising algorithms are presented. The algorithms are presented chronologically according to the degree of denoising performance. The methods showing similar results are only cited omitting the details.

Chapter 3

Wavelet-domain Adaptive Center Weighted Median Filter for Image Denoising

3.1 Introduction

Denoising images corrupted by additive white Gaussian noise is a classical problem to the signal processing community. The corruption of an image by noise is common during its acquisition or transmission. The aim of denoising is to remove the noise while keeping the signal features as much as possible. Traditional algorithms perform image denoising in the pixel domain [7]-[8]. However, in recent years, wavelet transform based image denoising algorithms have shown a remarkable success [17]-[24]. For the underlying problem, wavelet shrinkage method in [17] can provide an asymptotically minimax-optimal solution [23]. But for many image processing applications, e.g. medical imagery, astronomical imagery, remote sensing imagery, minimum mean-square error (MMSE) metric is more preferable [19]-[24]. Assuming transform coefficients within subbands are independent identically distributed (i.i.d.) random variables with generalized Gaussian (GG) distribution, an optimal MMSE estimator (Wiener filter) performs better than wavelet coefficient threshold methods [23]. However, such assumption may not be always realistic, considering the statistics of wavelet coefficients of natural images.

The wavelet coefficients within subbands are, infact, locally stationary and have dependency both in scale and across scales [23]-[25]. Considering such dependency with the neighbors, wavelet coefficients within each subband should be modeled as independent Gaussian random variables with zero-mean and slowly varying variance [21]-[25] rather than GG distribution [19], [20]. Estimation of the variance field using the Bayes-rule with a single exponential prior, inferred from the approximate histogram of true image coefficients in each scale, as in [22] is a crude one especially in the finest scale for two reasons. First, the assumed prior neglects the high frequency occurrence of very low magnitude coefficients in the histogram. The distinction of the approximate and actnal histogram is obvious in Fig. 2.6. Second, the local statistics of the signal or noise is not the same as that of the whole subband. Thus, the estimation of the variance field in [22] motivated by the histogram analysis of the wavelet coefficients is not accurate enough. The image denoising algorithm proposed in [24] adopts another approach considering both inner and inter scale dependency of the wavelet coefficients. In this method, if the parent value of an initial estimate of a coefficient variance is less than a predefined threshold, then it is set to zero. The denoising performance achieved by this method is close to that of reported in [22].

This work proposes a new smoothing process in the wavelet domain using an adaptive center weighted median filter (ACWMF) for better estimation of the variance field incorporating the local characteristics of the wavelet coefficients. For smoothing the variance field within the scale, the ACWMF does not assume any kind of particular distribution as in [19]-[24]. An exponentially decaying inter scale model for image wavelet coefficients is also suggested for adaptation of the ACWMF across scales. The key idea of the smoothing process is to suppress more noise from a preliminary estimate of the variance field without loosing image features that may degrade the quality of the denoised image. The preliminary estimation is performed assuming that the wavelet coefficients are conditionally i.i.d. as in [21], [22], [25]. For re-estimation of the variance field, first, we classify the overall scales as the noise dominant scales and the signal dominant scales. The ACWMF is operated only on the noise dominant scales. Due to the presence of a trace of signal component in these scales, the proposed ACWMF is designed as a special type of median filter that intelligently identifies the signal dominant regions from the noise dominated ones, and adapts itself in a predefined way during the smoothing process. Since high energy wavelet coefficients correspond to the signal features of sharp variation like edges and textures, and low energy corresponds to the smooth regions, during smoothing process we compare the center value within the mask to a prescribed threshold level to identify the signal or noise dominant regions in a scale. Because the adaptation of the proposed median filter depends on the magnitude of the center value within the mask, it is termed as the adaptive center weighted median filter (ACWMF). Finally, the variance field estimated from the ACWMF is used in a MMSE estimator for denoising wavelet coefficients.

3.2 Proposed ACWMF for Image Denoising

3.2.1 Overview of statistical models

Many researchers have exploited the statistics of wavelet coefficients for wavelet coding and denoising of natural images. Considering the spatial clusters, coefficients within each band can be globally modeled as independent Gaussian random variables with zero-mean and slowly varying variance field [21]-[25]. The slow variation of the variance field can roughly be incorporated by a stochastic prior as in [22]. To obtain better accuracy, by an additional smoothing, we consider the absolute value of the wavelet coefficients as a measure of the signal feature instead of assuming any global distribution throughout the scale concerned. Also the magnitude of the wavelet coefficients of a typical image are strongly correlated across scales. Embedded zero tree wavelet (EZW) coder by Shapiro [26], hidden Markov tree (HMT) model by Crouse [27] *et. al.* exploited the dependency between a parent coefficient and its children. Here we instead take a relation, that represents the average signal level distribution across scales. The standard deviation of the coefficients in each band represents the average signal level in that scale. Since most of the images have smooth regions with some edges and textures, it is expected that signal power decreases exponentially across the scales. That means, if $\bar{\lambda}_s$ is the average of three standard deviations of the wavelet coefficients of three high subbands HH_s, HL_s, LH_s, for a given scale s, then $\bar{\lambda}_s$ should have an exponential distribution across scales. The plot of $\bar{\lambda}_s$ across scales for different images is shown in Fig. 3.1. A normalized function $\Gamma(s) = C\beta^s, \beta > 1$ or $\Gamma(s) = Cs^p$ may be used to match such distribution, with C being a suitable constant, β denotes a base parameter, and p is an integer

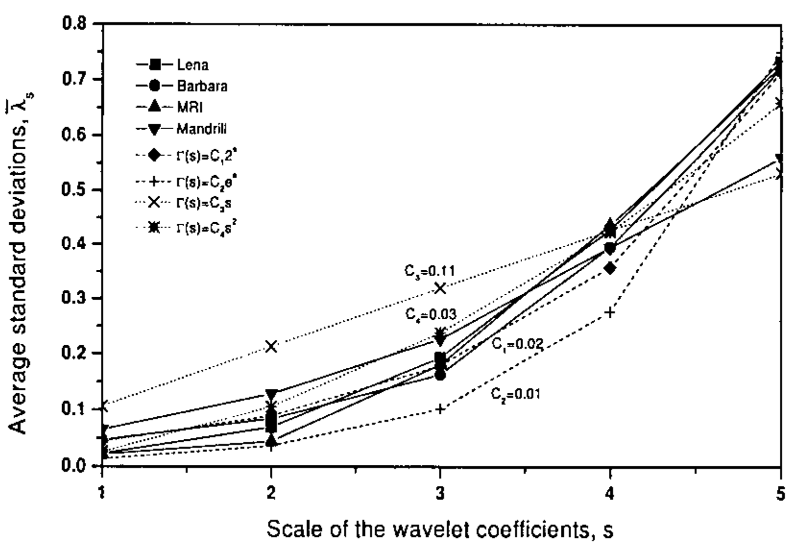


Figure 3.1: Average signal level distribution across scales for different images (solid line: distribution of $\bar{\lambda}_s$; dashed line: $\Gamma(s) = C\beta^s$ distribution; dotted line: $\Gamma(s) = Cs^p$ distribution).

exponent. For a number of images it is found that $\Gamma(s) = C\beta^s$ with $\beta = 2$ shows the best match in the least-square sense.

3.2.2 Denoising algorithm

In this work, we assume that image pixels are corrupted by additive white Gaussian noise (AWGN) with known variance σ_n^2. If σ_n^2 is unknown, it may be estimated by applying the median absolute deviation (MAD) method [17] in the highest subband of the wavelet coefficients according to Eqn. (2.26). Let $F_s(k)$ represent the wavelet coefficients of the true image for a given scale s. The corresponding wavelet coefficients of the noisy image are given by $G_s(k) = F_s(k) + N_s(k)$, where $N_s(k)$ are the wavelet coefficients of the additive noise.

The proposed denoising algorithm operates in three steps. First, we perform a crude estimation of the variance $\tilde{\sigma}_{g_s}^2(k)$ for the noisy coefficients, using the maximum likelihood (ML) estimator proposed in [21] or the maximum a posteriori (MAP) estimator proposed in [22]. In the second step, more accurate variance field $\hat{\sigma}_{g_s}^2(k)$ for the noisy coefficients, is re-calculated using the proposed ACWMF. Finally, the denoised wavelet coefficients, $\hat{F}_s(k)$ are estimated by the

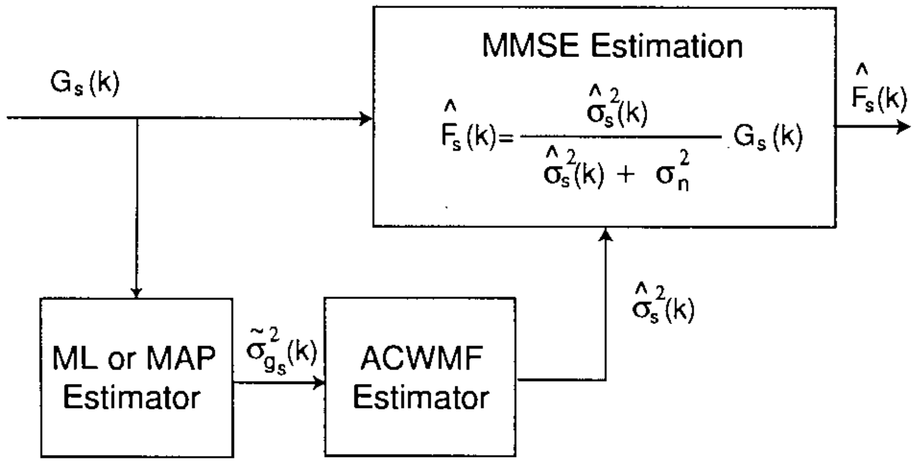


Figure 3.2: Block diagram of the denoising algorithm.

MMSE estimator given by

$$\hat{F}_s(k) = \frac{\hat{\sigma_s}^2(k)}{\hat{\sigma}_{g_s}^2(k)} G_s(k) = \frac{\hat{\sigma_s}^2(k)}{\hat{\sigma_s}^2(k) + \sigma_n^2} G_s(k) \tag{3.1}$$

Here $\hat{\sigma_s}^2(k)$ are the estimated variance of the noise-free coefficients. The block diagram of the proposed denoising algorithm is shown in Fig. 3.2.

3.2.3 Estimation of the underlying variance field

The performance of the MMSE type estimator depends highly on the quality of the estimation of $\hat{\sigma}_{g_s}^2(k)$ or simply $\hat{\sigma}_s^2(k)$. As mentioned in Sec. 3.2.2, preliminary estimation of the variance field is a prerequisite to the proposed smoothing process. For this estimation, assuming that the correlation between the variance of the neighboring coefficients is high in the given data field $G_s(k)$, one can set $\tilde{\sigma}_{g_s}^2(k) \approx \tilde{\sigma}_{g_s}^2(j)$ for all $j \in \mathcal{N}(k)$. Here $\mathcal{N}(k)$ is a square shaped local neighborhood centered at $G_s(k)$. Variance of the noisy coefficients, $\tilde{\sigma}_{g_s}^2(k)$, can be estimated using the ML estimator [21] as

$$\begin{aligned} \tilde{\sigma}_{g_s}^2(k) &= \arg\max_{\sigma_{g_s}{}^2 \geq 0} \prod_{j \in \mathcal{N}(k)} P\left(G_s(j) \mid \sigma_{g_s}{}^2\right) \\ &= \frac{1}{M} \sum_{j \in \mathcal{N}(k)} G_s^2(j) \end{aligned} \tag{3.2}$$

Here, it is assumed that noisy coefficients $G_s(j)$ have a Gaussian distribution with zero mean and variance $\tilde{\sigma}_{g_s}^2 = \tilde{\sigma_s}^2 + \sigma_n{}^2$, and M is the number of coefficients in $\mathcal{N}(k)$.

Proof of ML estimation 1 *The optimization problem can be written as*

$$\tilde{\sigma}_{g_s}^2(k) = \arg\max_{{\sigma_{g_s}}^2 \geq 0} F(\sigma_{g_s}) \tag{3.3}$$

where

$$F(\sigma_{g_s}) = \sum_{j \in \mathcal{N}(k)} \log\left(P\left(G_s(j) \mid {\sigma_{g_s}}^2\right)\right) \tag{3.4}$$

Note that

$$\begin{aligned} P\left(G_s(j) \mid {\sigma_{g_s}}^2\right) &= \frac{1}{\sqrt{2\pi{\sigma_{g_s}}^2}} \exp\left[-\frac{{G_s}^2(j)}{2{\sigma_{g_s}}^2}\right] \\ \log\left(P\left(G_s(j) \mid {\sigma_{g_s}}^2\right)\right) &= Const - \frac{1}{2}\log({\sigma_{g_s}}^2) - \frac{{G_s}^2(j)}{2{\sigma_{g_s}}^2} \end{aligned} \tag{3.5}$$

Hence

$$F(\sigma_{g_s}) = Const - \frac{M}{2}\log({\sigma_{g_s}}^2) - \frac{1}{2{\sigma_{g_s}}^2} S \tag{3.6}$$

where $S = \sum_{j \in \mathcal{N}(k)} {G_s}^2(j)$. *It can be easily verified that* F *is concave in* ${\sigma_{g_s}}^2$. *Hence from Kuhn-Tucker theorem, the maximum of* ${\sigma_{g_s}}^2$ *is unique. Now we can proceed by taking the first order derivatives and set zero to find* $\tilde{\sigma}_{g_s}^2(k)$.

$$\begin{aligned} \frac{\partial}{\partial \sigma_{g_s}} F(\sigma_{g_s}) &= 0 \\ \frac{S}{2{\sigma_{g_s}}^4} - \frac{M}{2{\sigma_{g_s}}^2} &= 0 \\ \tilde{\sigma}_{g_s}^2(k) &= \frac{1}{M} \sum_{j \in \mathcal{N}(k)} {G_s}^2(j) \end{aligned} \tag{3.7}$$

If the average power level of the signal coefficients in each subband, with a prior marginal distribution $f_{\sigma_s}({\sigma_s}^2) = (1/\lambda_s)\exp(-{\sigma_s}^2/\lambda_s)$, is λ_s, then $\tilde{\sigma}_{g_s}^2(k)$ can be estimated using the MAP estimator [22] as

$$\begin{aligned} \tilde{\sigma}_{g_s}^2(k) &= \arg\max_{{\sigma_{g_s}}^2 \geq 0} \left[\prod_{j \in \mathcal{N}(k)} P\left(G_s(j) \mid {\sigma_{g_s}}^2\right)\right] f_{\sigma_s}({\sigma_s}^2) \\ &= \frac{M\lambda_s}{4}\left[-1 + \sqrt{1 + \frac{8}{\lambda_s M^2} \sum_{j \in \mathcal{N}(k)} G_s^2(j)}\right] \end{aligned} \tag{3.8}$$

Proof of MAP estimation 1 *The optimization problem can be written as*

$$\tilde{\sigma}_{g_s}^2(k) = \arg\max_{{\sigma_{g_s}}^2 \geq 0} F(\sigma_{g_s}) \tag{3.9}$$

where

$$F(\sigma_{g_s}) = \sum_{j \in \mathcal{N}(k)} \log\left(P(G_s(j) \mid \sigma_s{}^2)\right) + \log f_{\sigma_s}(\sigma_s{}^2) \tag{3.10}$$

Note that

$$\begin{aligned}\log\left(P\left(G_s(j) \mid \sigma_{g_s}{}^2\right)\right) &= Const - \frac{1}{2}\log(\sigma_{g_s}{}^2) - \frac{G_s{}^2(j)}{2\sigma_{g_s}{}^2} \\ \log f_{\sigma_s}(\sigma_s{}^2) &= Const - \sigma_{g_s}{}^2/\lambda_s \end{aligned} \tag{3.11}$$

Hence

$$F(\sigma_{g_s}) = Const - \frac{\sigma_{g_s}{}^2}{\lambda_s} + \frac{M}{2}\log(\sigma_{g_s}{}^2) - \frac{1}{2\sigma_{g_s}{}^2}S \tag{3.12}$$

where $S = \sum_{j \in \mathcal{N}(k)} G_s{}^2(j)$. *Here also* F *is concave in* $\sigma_{g_s}{}^2$. *Hence by taking the first order derivatives and set zero to find* $\tilde{\sigma_{g_s}^2}(k)$.

$$\begin{aligned}\frac{\partial}{\partial \sigma_{g_s}{}^2}F(\sigma_{g_s}) &= 0 \\ \frac{S}{2\sigma_{g_s}{}^4} - \frac{M}{2\sigma_{g_s}{}^2} - \frac{1}{\lambda_s} &= 0 \\ \tilde{\sigma_{g_s}^2}(k) &= \frac{M\lambda_s}{4}\left[-1 + \sqrt{1 + \frac{8}{\lambda_s M^2}\sum_{j \in \mathcal{N}(k)} G_s{}^2(j)}\right]\end{aligned} \tag{3.13}$$

For renovation of the preliminary estimated variance field we introduce a new smoothing process that is adapted to both inter and inner scale, exploiting the statistics of wavelet coefficients. It is evident from Fig. 3.1 that approximate wavelet coefficients of practical images mostly represent the signal coefficients. We can neglect the smoothing of the noisy coefficients in the scales where signal coefficients dominate by defining a threshold in terms of λ_s. A subband having λ_s higher than the threshold value is assumed to have mostly signal coefficients with insignificant noise coefficients and vice versa. In the proposed method, we apply the modified median filter termed as the ACWMF to the noise dominated high bands. If the preliminary estimate of the local variances using an existing estimator, e.g. ML or MAP, is $\tilde{\sigma_{g_s}^2}$, then the re-estimated variance $\hat{\sigma_{g_s}^2}$ after passing through the proposed filter is given by

$$\hat{\sigma_{g_s}^2}(k) = \begin{cases} \tilde{\sigma_{g_s}^2}(k), & \text{if } \lambda_s \geq \lambda_{th} \\ \text{med}^i{}_{cw}\left(\tilde{\sigma_{g_s}^2}(j)\right), & \text{if } \lambda_s < \lambda_{th} \end{cases} \tag{3.14}$$

for all $j \in \mathcal{N}(k)$, where $\text{med}^i{}_{cw}$ represents the center weighted median value of the given set of data for the i-th iteration. The parameter λ_s can be calculated

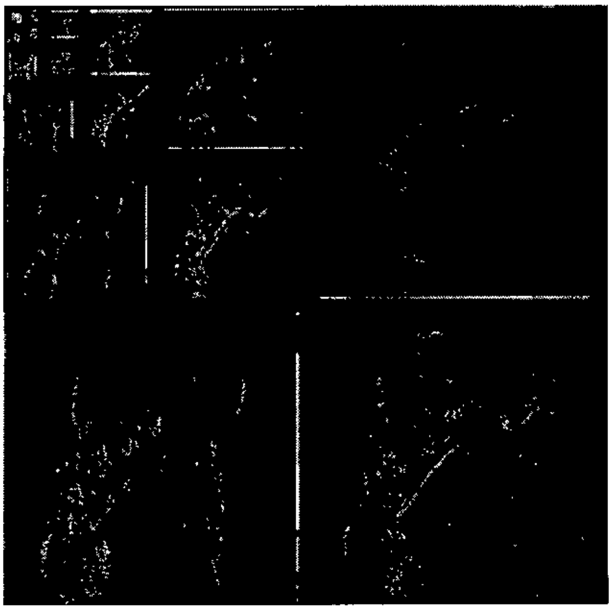


Figure 3.3: Five level wavelet decomposition of the image *Lena*. White pixel means the large magnitude coefficients, and black means the small magnitude.

from the standard deviation of the preliminary estimate of the signal coefficients' variances, $\tilde{\sigma_s}^2$ in each subband [22]. We define the threshold parameter λ_{th} as the weighted average of λ_s. Assuming that the inter scale average signal level decreases as a near exponential distribution following $\lambda_s = C2^s$, with C being a suitable constant, the λ_{th} can be calculated as

$$\lambda_{th} = \frac{\int \lambda_s 2^s \mathrm{d}s}{\int 2^s \mathrm{d}s} \tag{3.15}$$

In the high bands along with noise, signal component exists. Therefore, our objective is to remove noise from these bands with less possible distortion of the signal components. To achieve this goal, the proposed ACWMF first identify the signal or noise dominant regions within the subband and performs smoothing in a prescribed way. Within a subband to distinguish signal dominant regions from the noise dominant ones, we choose power level of a coefficient as a determining factor.

Magnitude of a wavelet coefficient and therefore its energy is a distinctive indicator of the image features like edge, texture etc. This can be verified from Fig. 3.3, which shows the wavelet decomposition of the original *Lena* image. For a noisy image, it is expected that the magnitude of the wavelet coefficients will be increased on average due to the addition of noise. This reveals that the low energy coefficients corresponding to the smooth regions in an image are dominated by the noise. We expect that, in this region smoothing of the preliminary estimated variance field by a straight median filter can produce better estimate of its noise-free value. On the contrary, high energy coefficients which corresponds to the regions of sharp transition in an image, are dominated by the signal. Using a straight median filter in this region would have adverse affect on the signal coefficients due to over smoothing. If we preserve the center value as it is in this region it ensures the correctness of the preliminary estimation. But our aim is to improve the performance in this region if possible. Let us take a hypothetical situation when center value of the variance field is less than the other value within mask. If we add a second term to the center value that is a weighted median of the difference of center to other value within the mask concerned, it is expected that center value will be boosted up and thereby enhance the edge information. In the other cases, the second term will simply smooth the gradient of the preliminary estimates and, hence, increase the convergence of smoothing process. To distinguish whether the variance is dominated by noise or by signal, we prescribe a scale independent threshold level which is a function of noise variance. Finally, our proposed adaptive median filter $\mathrm{med}^{i}{}_{cw}$ can be defined as

$$\mathrm{med}^{i}{}_{cw}\left(\tilde{\sigma_{g_s}}^2(j)\right) = \begin{cases} \mathrm{med}\left(\tilde{\sigma_{g_s}}^2(j)\right), & \text{if } \tilde{\sigma_{g_s}}^2(k) \leq \gamma {\sigma_n}^2 \\ \tilde{\sigma_{g_s}}^2(k) + \mathrm{med}_w\left(\tilde{\sigma_{g_s}}^2(j) - \tilde{\sigma_{g_s}}^2(k)\right), & \text{if } \tilde{\sigma_{g_s}}^2(k) > \gamma {\sigma_n}^2 \end{cases} \tag{3.16}$$

The parameter γ is a tuning factor that controls weight of the center value and the median value according to the mask size. The larger the mask size, the greater the chance that median value can affect the center value and hence the smaller the value of γ should be. Since a straight median operator on the second term of the second condition in this equation represents nothing but the first condition, it necessitates that the second term be weighted. We choose the weight space B such that

$$|B|_w = \int_{\mathcal{N}(k)} w(j)\mathrm{d}j = 1 \tag{3.17}$$

If the weighting function $w(j)$ is the same as in [22], i.e. $w(j) = \exp\left(-\tilde{\sigma}_s^2(j)/\lambda_s\right)$, then B can be defined as

$$B(j) = \frac{\exp\left(-\tilde{\sigma}_s^2(j)/\lambda_s\right)}{\int_{\mathcal{N}(k)} \exp\left(-\tilde{\sigma}_s^2(j)/\lambda_s\right)\mathrm{d}j} \tag{3.18}$$

for all $j \in \mathcal{N}(k)$. Finally, the weighted median operator med_w can be obtained as

$$\text{med}_w\left(\tilde{\sigma_{g_s}}^2(j)\right) = \text{med}\left(\tilde{\sigma_{g_s}}^2(j)B(j)\right) \tag{3.19}$$

for all $j \in \mathcal{N}(k)$. Though the ACWMF results in a smoother version of the data set, it is a contrast invariant operator [9] and also it shows an asymptotic convergence. As over-smoothing can affect the signal coefficients, the iterative smoothing process needs a terminating criterion that is a measure of smoothness of a data set. Here, we use "log-entropy" as an indicator of smoothness given by

$$H^i(\tilde{\sigma_{g_s}}^2) = -\sum_{k=1}^{J_s} \ln\left[\tilde{\sigma_{g_s}}^2(k)\right] \tag{3.20}$$

with J_s being the number of coefficients in a given subband s. If $|H^i - H^{i+1}|/|H^i|$ is less than a very small value ϵ, then the iteration is terminated. And the signal coefficients' variances can be estimated from $\hat{\sigma_{g_s}}^2$ as

$$\hat{\sigma}_s^2(k) = \max\left[0, \hat{\sigma_{g_s}}^2(k) - {\sigma_n}^2\right] \tag{3.21}$$

The estimated noise-free coefficients, $\hat{F_s}^2(k)$ are calculated by the estimated signal coefficient variance, $\hat{\sigma_s}^2(k)$ and the MMSE estimator as in Eqn. (3.1). Inverse wavelet transform of the denoised coefficients $\hat{F_s}^2(k)$ gives the denoised version of the image $\hat{f}(i,j)$.

3.3 Conclusion

In this chapter, we have presented a brief review of the present state of the art of different approaches for the problem and motivation behind the proposed algorithm. Then statistical modeling of image wavelet coefficients are discussed. A new exponential statistical model for the inter scale image wavelet coefficients is proposed here. Next we propose a wavelet domain adaptive center weighted median filter (ACWMF) for image denoising. The proposed filter operates only on the preliminary estimate of the variance field. The mathematical expressions for the preliminary estimation of the variance field are included for the simplicity. For re-estimation of the variance field the ACWMF adapts itself exploiting both the inner and inter scale statistics of image wavelet coefficients. The proposed exponentially decaying inter scale model is used to classify the scales into the signal and noise dominant ones. The ACWMF operates only on the noise dominant scales. For smoothing process within these scales, the proposed filter exploits inner scale statistics, where the coefficient energy is chosen as an index for classification of the signal and noise dominant regions. The explanations for the expected improvement of the denoising performance over some of the recent methods are also stated here.

Chapter 4

Simulation Results

4.1 Introduction

A relatively small selection of certain test images appears repeatedly in the published literature of image denoising. Using a few standard images has some benefits, such as allowing direct comparison of the performances of different denoising techniques. However, one drawback is that the so called "standard" images are not perfectly standard. For example, when a color image is converted to the grayscale image the algorithm may differ. This means that the authors citing results in different papers for the same image may be using slightly different images. For the best, the test images used in this work were collected from the different sources on the Internet. In addition to the most commonly used natural images, some synthetic images are tested to obtain a better overall performance of the proposed image denoising methods. The following list of sources indicates the origin of the test images used.

1. MATLAB image processing toolbox.

2. Waterloo BragZone
(http://links.uwaterloo.ca/bragzone.base.html)

3. Information Coding laboratory, UCSD ECE Department
(http://www.code.ucsd.edu/ sherwood/image examples/chan coded/chan coded.html)

4. Signal and image processing Institute, USC
(http://sipi.usc.edu/services/database/Database.html)

5. UICODER
(http://saigon.ece.wisc.edu/ waveweb/QMF/software.html)

We have tested the proposed denoising methods on three broad classes of images.

The test images are classified according to the mixture of sharp transitions, flat regions, shading, textures etc. contain in them. A good mixture of all these criteria is found in the very popular standard image *Lena* and most of the cited papers report on this image [29]. We choose it as the representative of the first class, having a good mixture of details and smooth regions. Detail simulation results on this test image is presented in this work. For the second class of images with very high details, our choice is the image *Mandrill.* The last class of images, having relatively a more smooth regions we choose the image *MRI.* We can further classify the image denoising methods on the basis of their applications. Denoising of *MRI* image will fall into the category of medical image denoising. To simulate the denoising performances on a typical astronomical image we prefer *Earthspace.* As an application on the low altitude remote sensing image, the image *Pentagon* is chosen.

The index used for comparing the performances of different denoising methods is the peak signal to noise ratio (PSNR). Since all the tests are performed on 8-bit grayscale images, the peak signal value is 255. Hence the PSNR values in dB for a $X \times Y$ test image $f(i,j)$ with its reconstructed image $\hat{f}(i,j)$ are calculated via

$$\mathrm{PSNR} = 10 \log_{10} \left(\frac{255^2}{\mathrm{MSE}} \right) \tag{4.1}$$

where the mean square error (MSE) is defined as

$$\mathrm{MSE} = \frac{1}{XY} \sum_{i=1}^{X} \sum_{j=1}^{Y} \left| f(i,j) - \hat{f}(i,j) \right|^2 \tag{4.2}$$

4.2 Result and Discussion

All the test images reported in this work are 512×512 grayscale images. The orthonormal wavelet Daubechies' of length-8 was used and decompositions were 5 levels deep. To estimate the signal coefficient variance, $\hat{\sigma}_s^2(k)$ centered square-shaped windows of size 3×3, 5×5 and 7×7 were employed. The parameter γ should be set empirically. In our experiment, we have observed that the suitable value of γ lies in the range 1 to 4. The value of this parameter is dependent on the mask size. In our presented result we have used $\gamma = 3$ for the mask size 3×3, 2 for the mask size 5×5, 1.5 for the mask size 7×7. We have compared six different denoising methods and the PSNR results are shown in the respective table of the test image. The first method is MATLAB's image denoising algorithm *wiener2*. The second method uses spatially adaptive wavelet thresholding (SAWT) method proposed in [19]. The third and fifth method show the results using locally adaptive window-based denoising using ML (LAWML) and locally adaptive window-based denoising using MAP (LAWMAP) methods reported in [22]. The results are compared using the same orthonormal wavelet transform. Results using the proposed ACWMF are presented as the fourth and sixth methods, and they are categorized according to the preliminary estimation process of the variance field. The method using an approximate ML method to compute the variance field will be termed as *adaptive center weighted median-based denoising using ML* (ACWMML). And the other method, using a approximate MAP estimator to compute the variance field will be termed as *adaptive center weighted median-based denoising using MAP* (ACWMMAP). Both the proposed methods are simulated for different terminating conditions by varying the value of ϵ as stated in the previous chapter. Simulations reveal that the arbitration of its value does not show any considerable degradation of the performance. The value of ϵ was set to $\sim 1.0\exp(-6)$ for the ACWMML method and $\sim 1.0\exp(-3)$ for the ACWMMAP method. The results for the different test images are presented in the next section.

4.2.1 Lena

We begin analysis with the widely used *Lena* image [29]. The image neither contains very large amount of high frequencies and oscillating patterns nor very large amonnt of flat patterns. Table 4.1 shows the denoising performance in terms of PSNR for different denoising methods as mentioned in the previous section. From this table, it is evident that the wavelet based denoising methods outperforms the usual pixel domain *winer2* method. Among the wavelet based denoising algorithms, the proposed ACWMMAP method shows the highest value of PSNR. It is also noticed that the result obtained by the ACWMML method is very close to the ACWMMAP method. This is due to the fact that we neglect smoothing of some of the approximate scales and for these scales initial estimate determines the amount of improvement. Since the ACWMMAP method takes initial estimate from the MAP estimator that is superior to the ML estimator, it performs bet-

Table 4.1: Results in PSNR (dB) for several denoising methods for the image *Lena*.

Denoising algorithms	Noise standard deviation σ_n			
	10	15	20	25
wiener2	33.59	31.13	29.01	27.20
SAWT	34.10	32.11	30.59	29.51
3×3 LAWML	33.77	31.37	29.59	28.11
3×3 ACWMML	34.42	32.34	30.81	29.59
3×3 LAWMAP	34.21	32.23	30.79	29.70
3×3 ACWMMAP	34.50	32.53	31.02	29.92
5×5 LAWML	34.17	32.00	30.40	29.09
5×5 ACWMML	34.45	32.45	31.01	29.87
5×5 LAWMAP	34.33	32.35	30.91	29.82
5×5 ACWMMAP	34.45	32.48	31.03	29.91
7×7 LAWML	34.18	32.07	30.53	29.31
7×7 ACWMML	34.28	32.29	30.82	29.68
7×7 LAWMAP	34.26	32.25	30.80	29.70
7×7 ACWMMAP	34.30	32.31	30.86	29.74

ter than the ACWMML method. It is also obvious that the performances of the LAWML and LAWMAP methods are highly dependent on the window size of the variance field estimation process, whereas both the ACWMML and ACWMMAP methods show close results independent of the mask size. Fig. 4.1 shows the original image *Lena*, a typical noisy version with $\sigma_n = 20$, denoised version using the *winer2* method, and spatially adaptive threshold method (SAWT). It is evident that the denoising by the SAWT method has more pleasing output than the normal pixel domain wiener filtering. Fig. 4.2 shows the comparison of the LAWML method reported in [21] to the proposed ACWMML method. Table 4.1 shows the results for the 512×512 image *Lena*. But for the space limitation Fig. 4.2 shows the comparison of a close-up view of *Lena* having a size of 256×256. The improvement in true size of the denoised image is difficult to understand without a careful observation. To show the details we select a section of the denoised image that contains both the smooth regions and sharp variations and magnify it to 4 times. The overall improvement in PSNR is ~ 0.65 dB, and it is clear in the zoomed-in section near the two eyes and the forehead of *Lena*. Fig. 4.3 shows the comparison of the LAWMAP method reported in [21] to the proposed ACWMMAP method.

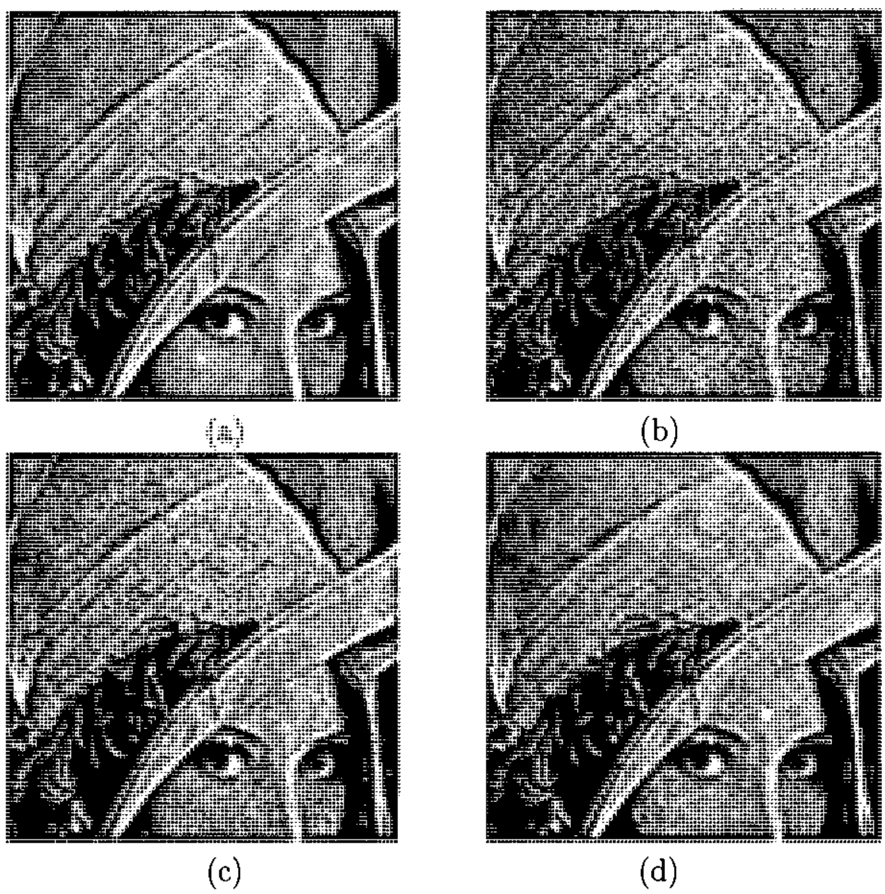


Figure 4.1: Comparing performances of some previous methods on *Lena* with $\sigma_n = 20$; (a) Original, (b) Noisy observation, (c) Denoising by *wiener2*, (d) Denoising by SAWT.

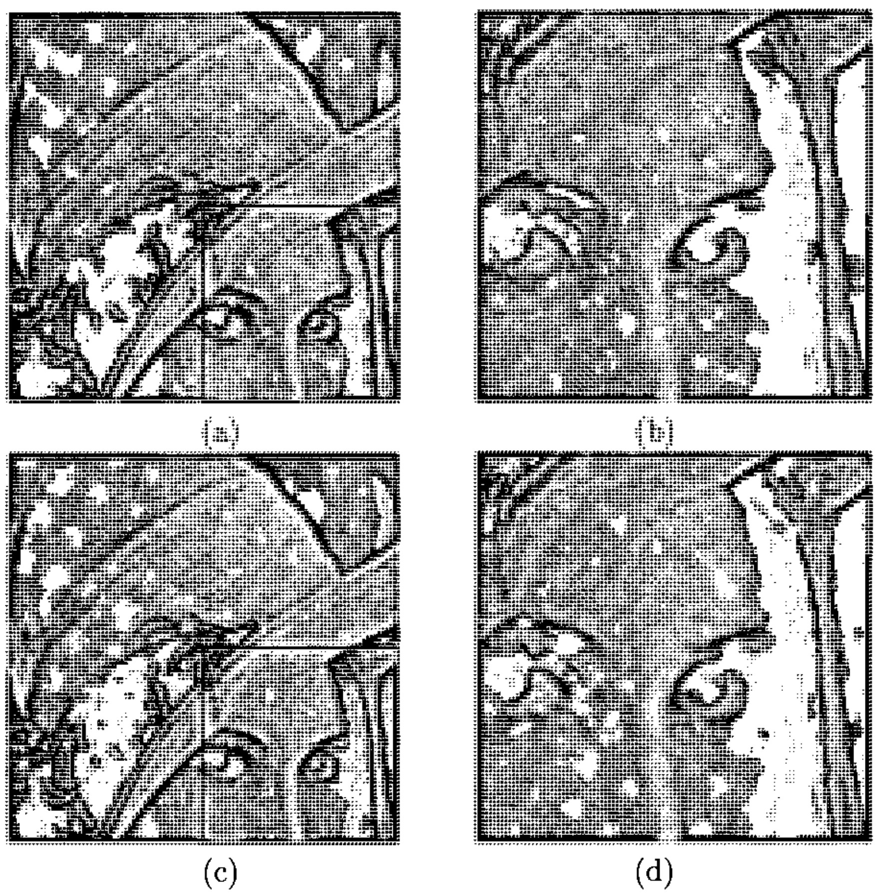
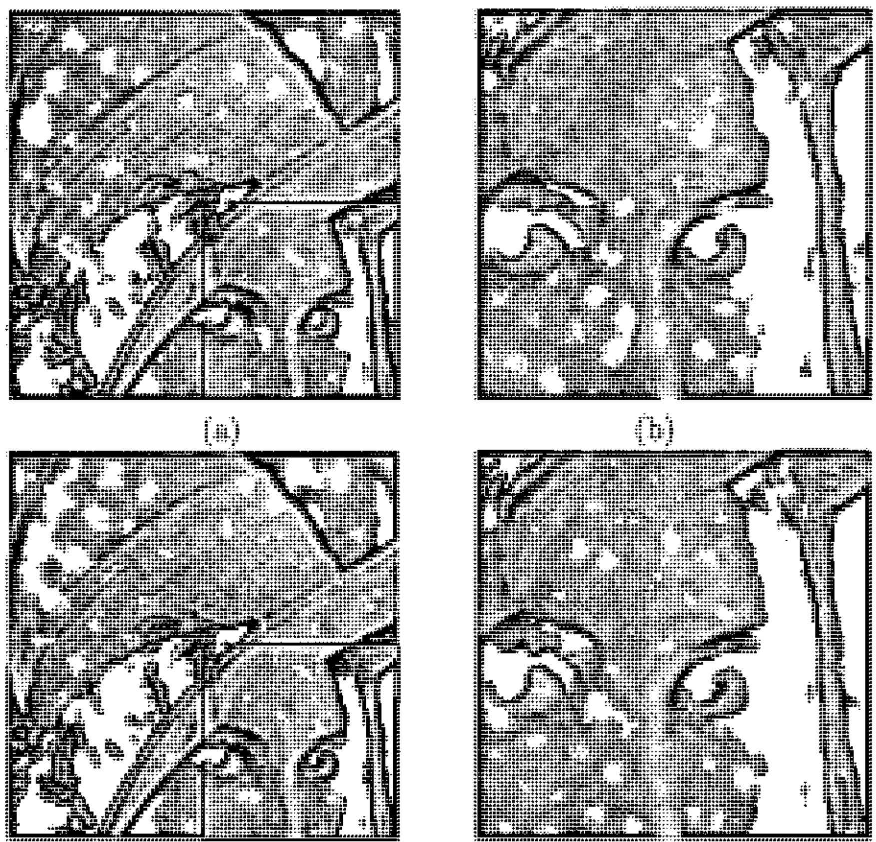


Figure 4.2: Comparing the performance of LAWML method with ACWMML method for *Lena* with $\sigma_n = 20$ and mask size 5×5; (a) Denoising by LAWML, (b) Zoomed-in section (inside the square) of (a), (c) Denoising by ACWMML, (d) Zoomed-in section of (c).



(a) (b) (c) (d)

Figure 4.3: Comparing the performance of LAWMAP method with ACWMMAP method for *Lena* with $\sigma_n = 20$ and mask size 3×3; (a) Denoising by LAWMAP, (b) Zoomed-in section (inside the square) of (a), (c) Denoising by ACWMMAP, (d) Zoomed-in section of (c).

Table 4.2: Results in PSNR (dB) for comparing the denoising performances of LAWML, LAWMAP, and ACWMML with mask size 5×5 for different scales of the image *Lena* corrupted by noise of standard deviation $\sigma_n = 15$.

Scales	Denoising performance in PSNR, dB		
	LAWML	LAWMAP	ACWMML
HL_1	34.37	34.90	35.33
LH_1	32.78	33.03	33.29
HH_1	36.62	38.58	38.67
HL_2	30.59	30.61	30.69
LH_2	29.34	29.35	29.38
HH_2	31.75	31.91	32.01

A close-up view and its magnified version are also shown for analyzing the details. The author cannot but mention that the overall improvement of PSNR of the order of ~ 0.25 dB is significant (as in this case) in the literature and this improvement is observable more clearly with a high quality printing.

Since the results in the coefficient domain may give more insight of the proposed technique, we simulate it also in the coefficient domain. For this comparison we choose three possible ways of smoothing by a median filter. The first one is the straight median filter. Second, is the proposed ACWMML as in Eqn. (3.16) except that the second term in the second condition is absent. Third, is the proposed ACWMML method. All the performances are simulated for denoising the coefficients of *Lena* image degraded by a noise level of $\sigma_n = 15$ using a mask size of 5×5. The performance comparison in terms of PSNR among these three possible ways of smoothing is shown in Fig. 4.4 for the coefficients in HL_1, in Fig. 4.5 for the coefficients in LH_1, in Fig. 4.6 for the coefficients in HH_1, in Fig. 4.7 for the coefficients in HL_2, in Fig. 4.8 for the coefficients in LH_2, and in Fig. 4.9 for the coefficients in HH_2. Here all the performance curves in Figs. 4.4, 4.5, 4.7, 4.8, 4.9, except in Fig. 4.6 reveal that the first one, a straight median filter degrades its performance in successive iterations after the first few iterations. The second method, which is a modified median filter as in Eqn. (3.16) except that the second term in the second condition is absent, shows consistent improvement and better result than the straight median filter.

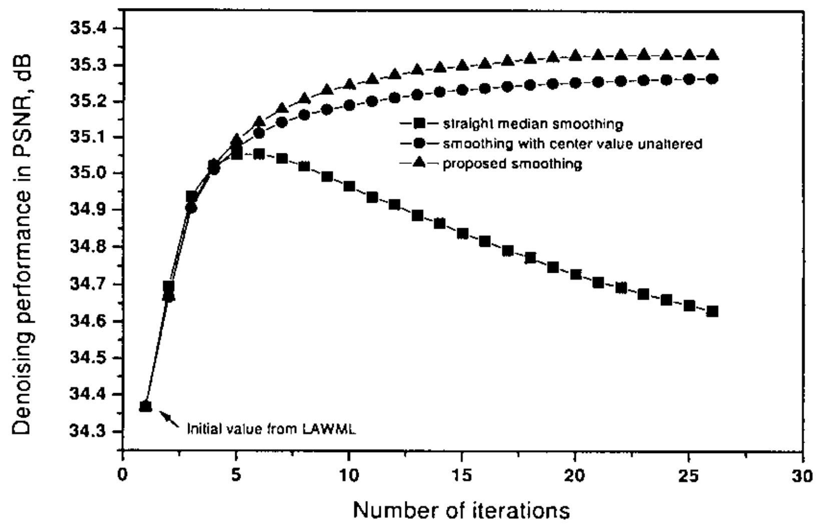


Figure 4.4: Comparing the performances of the straight median filter, ACWMML as in Eqn. (3.16) except that the second term in its second condition is absent, and the proposed method ACWMML for HL_1 coefficients of the image *Lena* with $\sigma_n = 15$ and mask size (5×5).

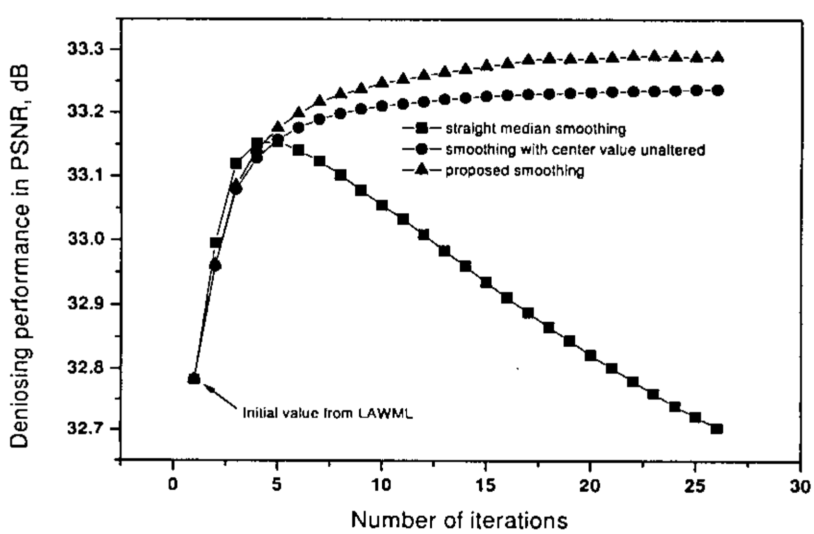


Figure 4.5: Comparing the performances of the straight median filter, ACWMML as in Eqn. (3.16) except that the second term in its second condition is absent, and the proposed method ACWMML for LH_1 coefficients of the image *Lena* with $\sigma_n = 15$ and mask size (5×5).

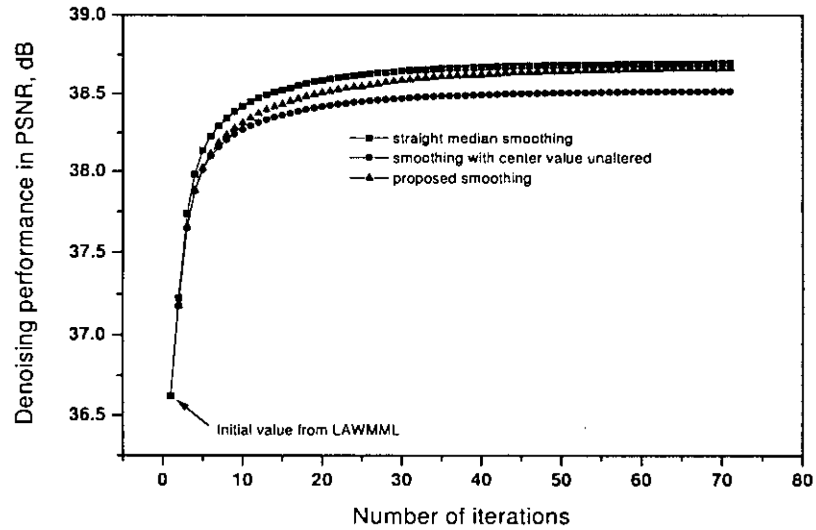


Figure 4.6: Comparing the performances of the straight median filter, ACWMML as in Eqn. (3.16) except that the second term in its second condition is absent, and the proposed method ACWMML for HH_1 coefficients of the image *Lena* with $\sigma_n = 15$ and mask size (5×5).

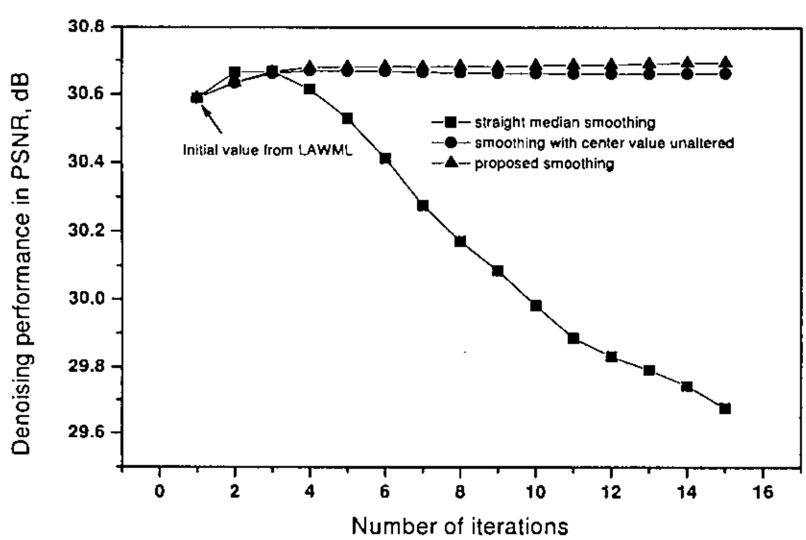


Figure 4.7: Comparing the performances of the straight median filter, ACWMML as in Eqn. (3.16) except that the second term in its second condition is absent, and the proposed method ACWMML for HL_2 coefficients of the image *Lena* with $\sigma_n = 15$ and mask size (5×5).

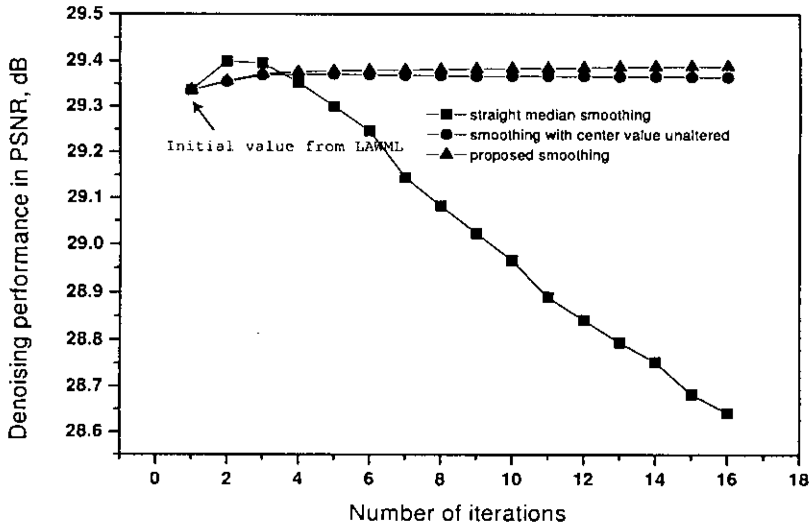


Figure 4.8: Comparing the performances of the straight median filter, ACWMML as in Eqn. (3.16) except that the second term in its second condition is absent, and the proposed method ACWMML for LH_2 coefficients of the image *Lena* with $\sigma_n = 15$ and mask size (5×5).

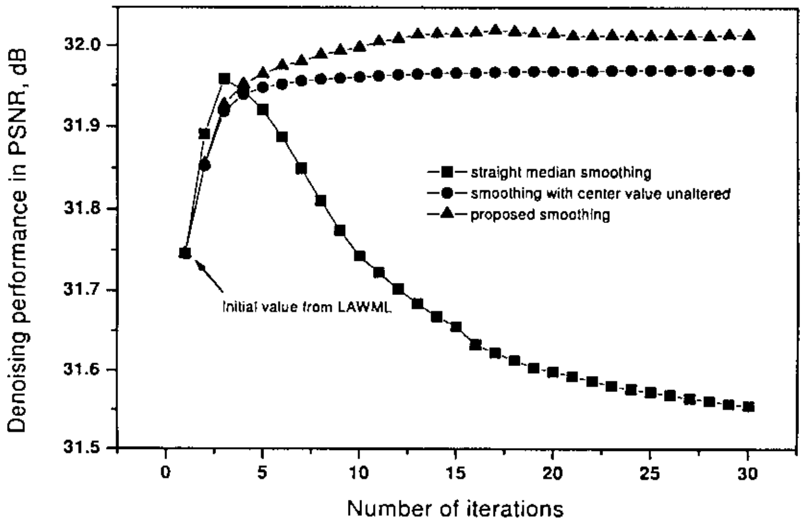


Figure 4.9: Comparing the performances of the straight median filter, ACWMML as in Eqn. (3.16) except that the second term in its second condition is absent, and the proposed method ACWMML for HH_2 coefficients of the image *Lena* with $\sigma_n = 15$ and mask size (5×5).

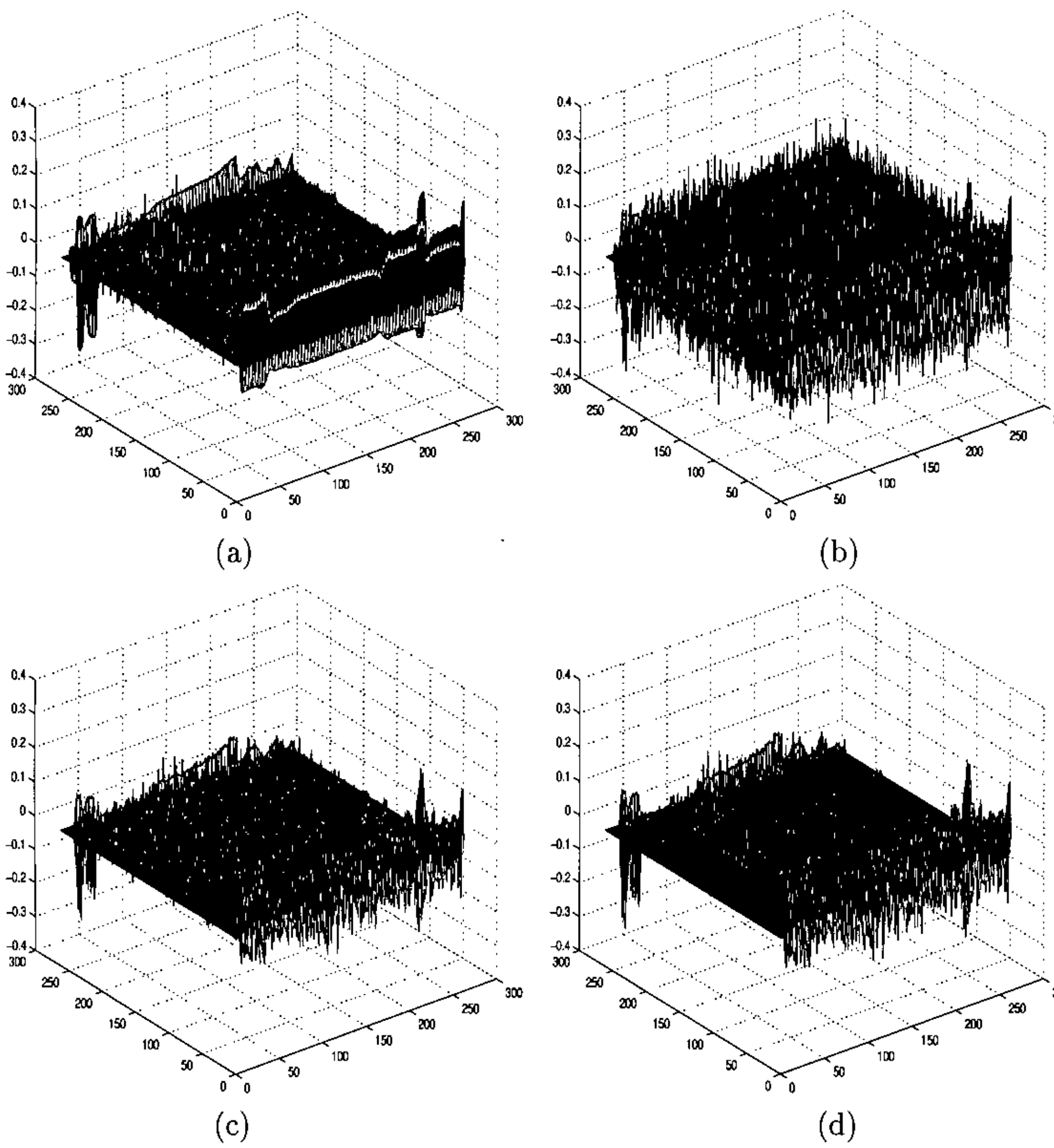


Figure 4.10: Comparing the performances of LAWMAP method with ACWMML method for HL_1 coefficients of the image *Lena* with $\sigma_n = 15$ and mask size (5×5); (a) Original, (b) Noisy observation, (c) Denoising by LAWMAP (d) Denoising by ACWMML.

The last method, which is the proposed ACWMML, shows faster convergence and significant improvement than the other two in terms of PSNR. Here lies the justification of adding the second term in the second condition in Eqn. (3.16). The performance curves shown in Fig. 4.6 shows a slightly different result. Here the proposed method show a little bit less improvement than the straight median filter, but when compare to the presented second type of smoothing it shows a superior performance. This can be explained as – relatively smooth images have very small number of signal coefficients in the HH_1 scale, hence over smoothing would perform better than constrained smoothing. For an image with high details this explanation may not be true. Moreover, all the other scales presented here need a constrained smoothing. Considering the performance in all scales, it can be concluded that the proposed method qualifies to produce the highest overall improvement when comparing these three types of smoothing.

Now we can show a typical study that compares the coefficient domain denoising performance in terms of PSNR using the ACWMML method over the LAWML method [21] and the LAWMAP method [22]. Table 4.2 shows the improvement results for the coefficients in first and second level decomposition, using the ACWMML method, the LAWML method, and the LAWMAP method for the image *Lena* degraded by a noise level of $\sigma_n = 15$ using a mask size of 5×5. It is found that the ACWMML method outperforms the other two in each subband presented here. It is also noticed that the improvement in the first level is much higher than the second level. This is due to the fact that the smoothing process can easily eliminate noise from the high frequency level, where signal portion is insignificant compared to the noise. On the other hand, at the higher decomposition level signal portion compared to noise is relatively high, hence smoothing process has less capability to denoise. The 3–D plot of the denoised wavelet coefficients for HL_1 using the ACWMML and LAWMAP method is shown in Fig. 4.10 for further verification of the improved performance. As can be seen, smoothing of the coefficients is obtained preserving the detail information.

4.2.2 Mandrill

The image *Mandrill* has a large amount of high frequency and oscillating patterns. Hence, the denoising performance of many cited papers are not very attractive.

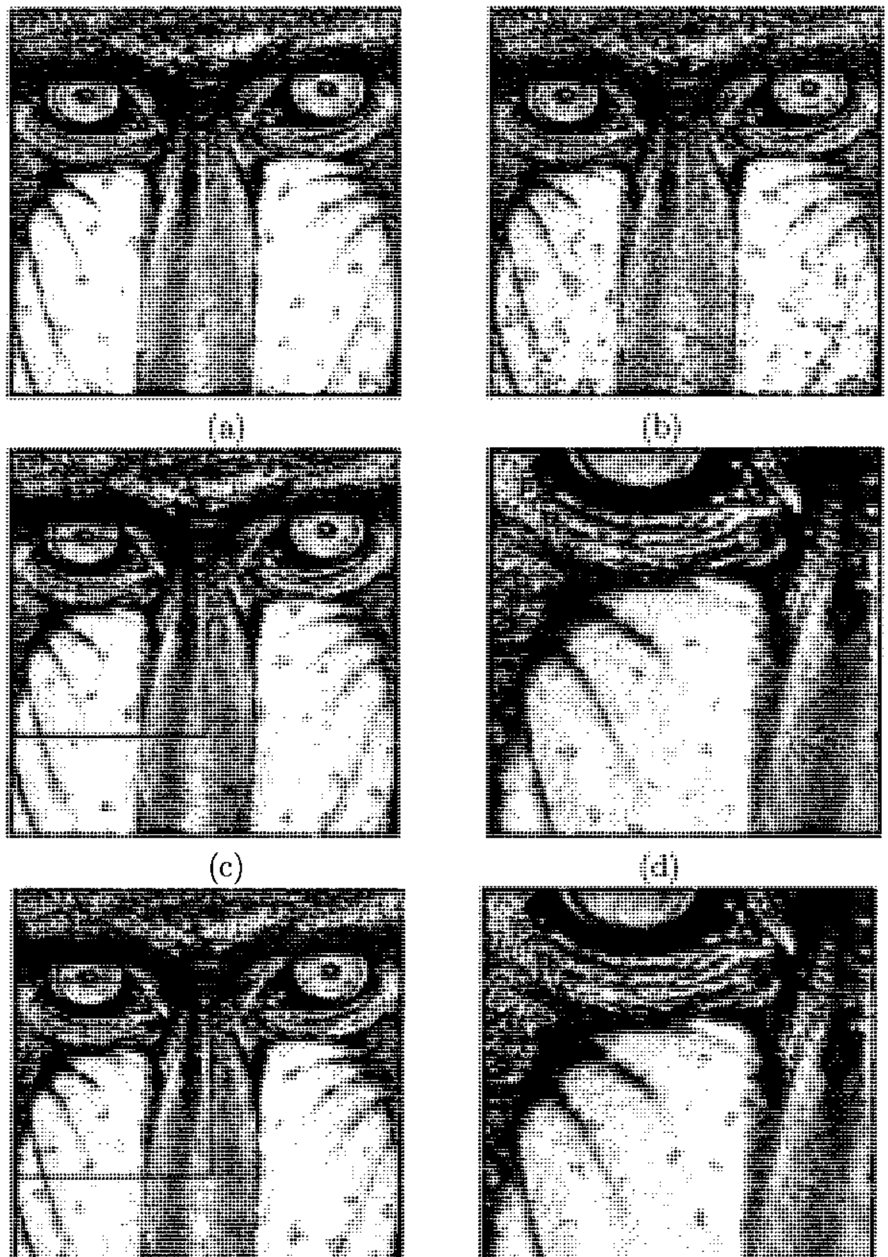


Figure 4.11: Comparing the performance of LAWMAP method with ACWMML method for *Mandrill* with $\sigma_n = 20$ and mask size 3×3; (a) Original, (b) Noisy observation, (c) Denoising by LAWMAP, (d) Zoomed-in section of (c), (e) Denoising by ACWMML, (f) Zoomed-in section of (e).

Table 4.3: Results in PSNR (dB) for several denoising methods for the image *Mandrill.*

Denoising algorithms	Noise standard deviation σ_n			
	10	15	20	25
wiener2	26.51	25.67	24.78	23.89
SAWT	29.68	27.10	25.41	24.19
3×3 LAWML	30.08	27.46	25.73	24.41
3×3 ACWMML	30.26	27.74	26.10	24.86
3×3 LAWMAP	30.08	27.51	25.83	24.61
3×3 ACWMMAP	30.25	27.67	25.97	24.74
5×5 LAWML	30.22	27.68	26.02	24.78
5×5 ACWMML	30.30	27.80	26.18	25.00
5×5 LAWMAP	30.22	27.69	26.04	24.83
5×5 ACWMMAP	30.30	27.78	26.12	24.91
7×7 LAWML	30.24	27.72	26.08	24.86
7×7 ACWMML	30.27	27.77	26.15	24.96
7×7 LAWMAP	30.25	27.72	26.09	24.89
7×7 ACWMMAP	30.28	27.77	26.13	24.94

Table 4.4: Results in PSNR (dB) for several denoising methods for the image *MRI.*

Denoising algorithms	Noise standard deviation σ_n			
	10	15	20	25
wiener2	33.60	31.04	28.91	27.06
SAWT	34.03	32.17	30.97	29.97
3×3 LAWML	34.01	31.61	29.90	28.43
3×3 ACWMML	34.75	32.70	31.27	30.07
3×3 LAWMAP	34.52	32.53	31.23	30.18
3×3 ACWMMAP	34.89	32.89	31.52	30.41
5×5 LAWML	34.53	32.39	30.90	29.61
5×5 ACWMML	34.86	32.89	31.59	30.53
5×5 LAWMAP	34.73	32.76	31.45	30.39
5×5 ACWMMAP	34.89	32.93	31.60	30.54
7×7 LAWML	34.63	32.56	31.17	29.98
7×7 ACWMML	34.75	32.77	31.46	30.37
7×7 LAWMAP	34.73	32.74	31.45	30.39
7×7 ACWMMAP	34.78	32.83	31.53	30.48

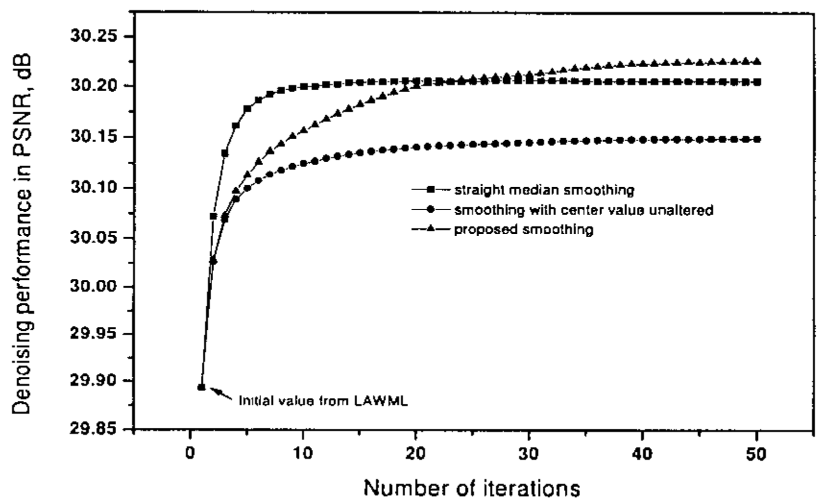


Figure 4.12: Comparing the performances of the straight median filter, ACWMML as in Eqn. (3.16) except that the second term in its second condition is absent, and the proposed method ACWMML for HH_1 coefficients of the image *Mandrill* with $\sigma_n = 15$ and mask size (5×5).

However, it is often chosen as an extreme example of images having large details. Table 4.3 shows the PSNR values of different denoising methods. A typical improvement for denoising the image with $\sigma_n = 20$ by the ACWMML method and the LAWMAP method with a mask size 3×3 is shown in Fig. 4.11. The denoised image using the ACWMML is more smoother than LAWMAP as can be seen from the zoomed-in section. The interesting result is found when our denoising algorithm is applied on the HH_1 coefficients of the image *Mandrill*. Fig. 4.12 shows the denoising performance using three different type of median filtering as mentioned in the previous section. We observe that though the straight median filter performs slightly better than our proposed smoothing when applied to the HH_1 coefficients of the image *Lena*, for the same condition with the image *Mandrill* the straight median filter fails to produce better results than the proposed one. This is because, the image *Mandrill* has signal coefficients in HH_1 level higher than the smooth image *Lena*. This can also be verified from Fig. 3.1, where *Mandrill* has the highest average signal coefficient magnitude in LH_1, LH_1, HH_1 level as compared to the other images presented. Hence, in conclusion, a straight median smoothing has a greater risk of degrading its performance after first few iterations as compared to our proposed smoothing.

4.2.3 MRI

The image *MRI* is an example of the image class that has relatively more flat regions and less amount of sharp transitions. It is a more smooth image when compared to the image *Lena*. This can be verified from Fig. 3.1, where *MRI* has the lowest average signal coefficient magnitude in LH_1, LH_1, HH_1 level as compared to the other images presented. It is chosen here for simulation not only as an example of smooth image but also as an example of the medical imagery. The denoising performance for different methods in PSNR is shown in Table 4.4. The improvement is quite satisfactory, for the mask size of 3×3 and 5×5. Although mask size of 7×7 shows improved PSNR, the smaller window is always preferable for image smoothing by a median filter [9]. The improvement of the ACWMML method over the LAWML method using a mask size 5×5, for denoising this image corrupted by $\sigma_n = 15$ is shown in Fig. 4.13. The improvement in the visual quality can be easily noticed in the zoomed-in section presented here.

4.2.4 Earthspace

As an astronomical image we choose the image *Earthspace*, which has moderate oscillations and sharp variations. The performance in PSNR is shown in Table 4.5. The comparison of the LAWML method with the ACWMMAP method using a mask size 3×3, for denoising this image corrupted by $\sigma_n = 15$ is shown in Fig. 4.14. Although the denoising improvement can be observed in the true-size of the image, zoomed-in sections are presented here to show it more clearly. As can be seen the artifacts remain when the denoising method is LAWML. In contrast, the ACWMMAP method can remove the bloches sufficiently.

4.2.5 Pentagon

The image *Pentagon* is chosen as an example of low altitude RADAR image. It is a synthetic image, since its color version is transformed to grayscale. The image has sharp transitions as in the image *Mandrill*. Table 4.6 shows the PSNR improvement for different methods. Here also consistent improvement is observed.

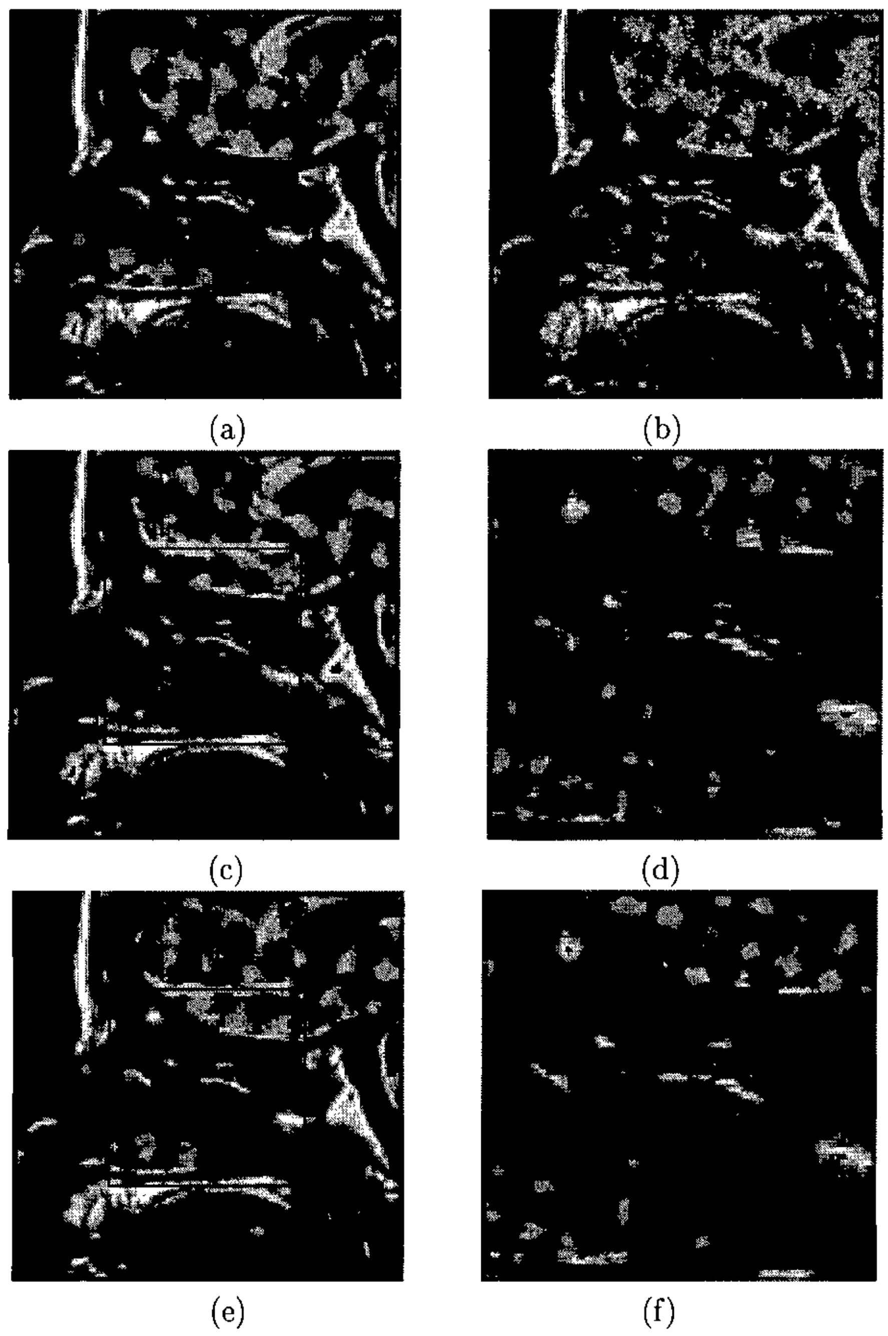


Figure 4.13: Comparing the performance of LAWML method with ACWMML method for *MRI* with $\sigma_n = 15$ and mask size 5×5; (a) Original, (b) Noisy observation, (c) Denoising by LAWML, (d) Zoomed-in section of (c), (e) Denoising by ACWMML, (f) Zoomed-in section of (e).

Table 4.5: Results in PSNR (dB) for several denoising methods for the image *Earthspace*.

Denoising algorithms	Noise standard deviation σ_n			
	10	15	20	25
wiener2	32.88	30.77	28.80	27.13
SAWT	32.94	31.09	29.73	28.61
3×3 LAWML	32.87	30.62	28.94	27.57
3×3 ACWMML	33.38	31.41	29.96	28.81
3×3 LAWMAP	33.23	31.33	30.03	28.95
3×3 ACWMMAP	33.41	31.50	30.18	29.11
5×5 LAWML	33.23	31.16	29.67	28.48
5×5 ACWMML	33.48	31.55	30.22	29.13
5×5 LAWMAP	33.36	31.45	30.15	29.07
5×5 ACWMMAP	33.45	31.54	30.22	29.14
7×7 LAWML	33.30	31.31	29.89	28.75
7×7 ACWMML	33.41	31.48	30.14	29.06
7×7 LAWMAP	33.37	31.45	30.13	29.05
7×7 ACWMMAP	33.42	31.50	30.17	29.08

Table 4.6: Results in PSNR (dB) for several denoising methods for the image *Pentagon*.

Denoising algorithms	Noise standard deviation σ_n			
	10	15	20	25
wiener2	29.12	27.85	26.56	25.34
SAWT	29.88	27.61	26.16	25.19
3×3 LAWML	30.34	28.02	26.43	25.28
3×3 ACWMML	30.50	28.22	26.79	25.77
3×3 LAWMAP	30.36	28.10	26.63	25.63
3×3 ACWMMAP	30.48	28.20	26.69	25.66
5×5 LAWML	30.43	28.19	26.70	25.66
5×5 ACWMML	30.47	28.21	26.82	25.87
5×5 LAWMAP	30.43	28.20	26.74	25.74
5×5 ACWMMAP	30.48	28.25	26.78	25.76
7×7 LAWML	30.38	28.16	26.71	25.72
7×7 ACWMML	30.41	2816	26.77	25.81
7×7 LAWMAP	30.39	28.16	26.72	25.74
7×7 ACWMMAP	30.41	28.20	26.76	25.78

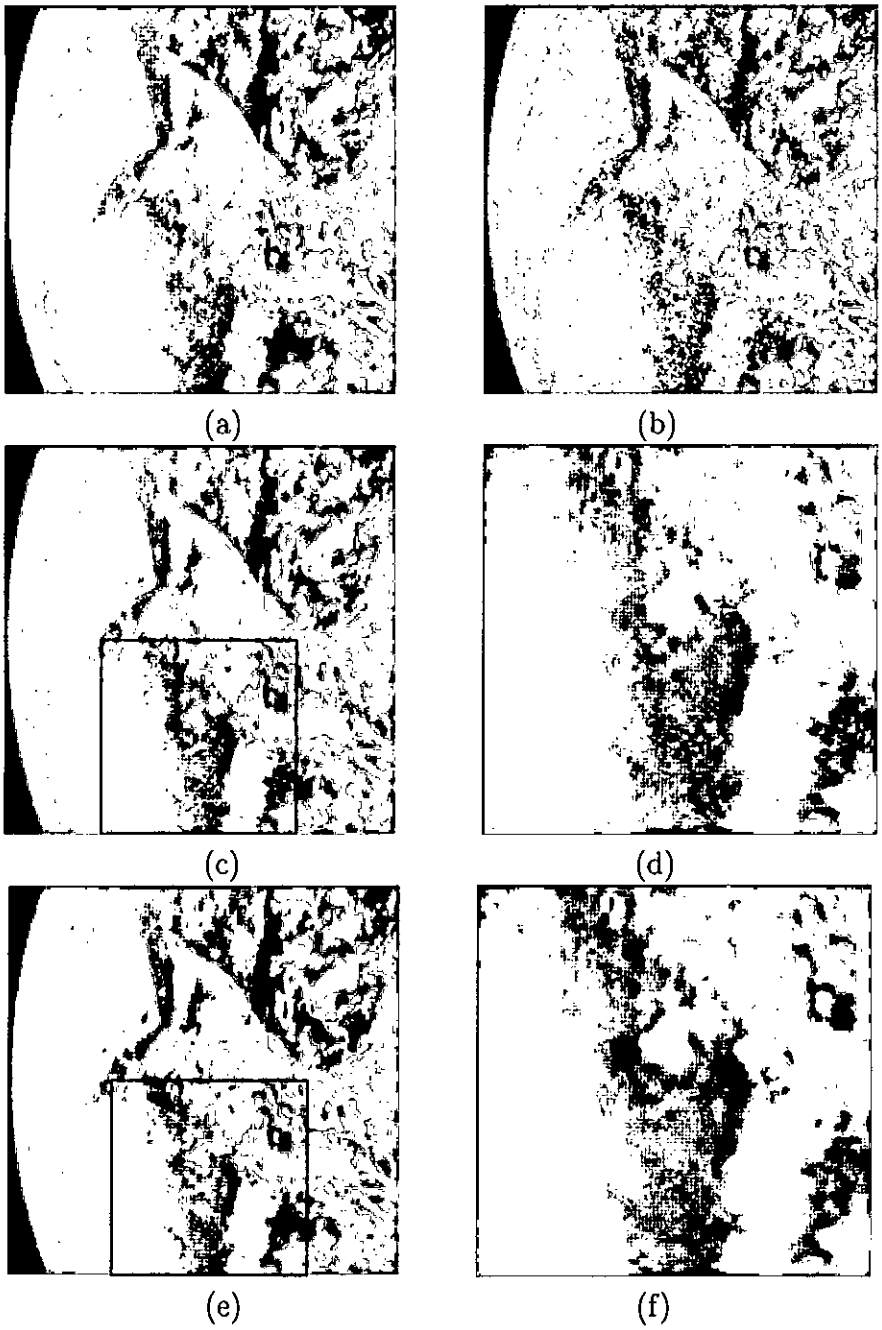


Figure 4.14: Comparing the performance of LAWML method with ACWMMAP method for *Earthspace* with $\sigma_n = 15$ and mask size 3×3; (a) Original, (b) Noisy observation, (c) Denoising by LAWML, (d) Zoomed-in section of (c), (e) Denoising by ACWMMAP, (f) Zoomed-in section of (e).

4.3 Conclusion

In the experimental study, we have also simulated our denoising methods on the images like *Saturn*, *Barbara*, *Peppers*, *Elaine*, *Goodhill* etc. But complete results of only a selected standard image are presented here. So that the presented results from different angles show a more logical impression on the denoising method proposed, we select different images with different classes based on their contents of transitions and flat regions. Detail results for several coefficient domains are presented for the image *Lena*, since it falls in the over lapping region of the two extreme categories. As can be seen, a smoother version of the denoised image is found without degradation of the edge information. The coefficient domain result in the 3-D plot shows a better visualization of smoothness. Considering the image quality, PSNR results in the tabular form, and given results in the coefficients domain, we can conclude that the proposed methods outperform the popular recently reported denoising techniques.

Chapter 5

Conclusion

5.1 Summary

A wavelet domain adaptive center weighted median filter (ACWMF) has been used successfully with the minimum mean square error (MMSE) estimator to denoise image data corrupted by additive white Gaussian noise (AWGN). The success of denoising by the proposed method lies in the fact that the ACWMF provides a better estimate of the image wavelet coefficient variance for the MMSE over some of the recent denoising methods. Basically, the ACWMF re-estimates a coefficient variance from a preliminary estimated one using an iterative smoothing process. The purpose of the smoothing process is to account for the mismatch between the assumed statistical model of image coefficients and the actual one. It has been demonstrated using numerical results that a straight median smoothing degrades the performance drastically after few iterations. As such, a constrained smoothing process has been proposed to ensure a stable denoising performance. To define the scale adaptive constraint, first the overall scales have been categorized into two groups, namely, the signal dominant scales and the noise dominant scales. In this work, an exponentially decaying inter scale model for image wavelet coefficients has been proposed to classify the scales. All the coefficient variances in the signal dominated scales have been kept apart from the smoothing process to avoid any risk of over-smoothing. The preliminary estimates of these coefficient variances determines their accuracy. Except these very few number of coefficient variances, all other variances undergo modification. It has been shown that the extent of modification of a coefficient variance by the proposed smoothing filter depends on its own value as well as its local neighborhood. A straight median smoothing has been performed on the noise dominated regions. On the contrary, in the signal dominated regions, smoothing is performed in such a way that the gradient of the variance field be smoothed without significantly distorting the signal component. For this purpose, a smoothing term from the local neighborhood has been calculated, and then added to the center value within the mask. The regions have been identified by comparing the center value of the smoothing mask with a prescribed threshold.

The simulation results on some common standard images, have shown substantial improvement in the denoising performance by the proposed method over some of the recent methods in terms of peak signal to noise ratio (PSNR). To ensure that the proposed denoising method may be applied to a broad class of images, we have tested it on a typical smooth image, moderately smooth image, and image with high variation. In all cases, the proposed methods have shown improvement not only in the overall MSE but also in the visual quality. The numerical results in the coefficient domain also justify the improvement in different scales as well as the convergence of the iterative process.

5.2 Future Work

The work may be extended in several ways. Smoothing of the variance field of some coarse scales are neglected here. More accurate model of inter and inner scale statistics can be exploited for smoothing of these variance fields. Choosing of the mask size both for preliminary estimation and smoothing process is a crucial one. The larger the mask size the better the assumption of Gaussian statistics. On the other hand, the smaller the mask size the greater the chance of contrast invariant smoothing by the median operator. Moreover, the coarse scales are dominated by the signals having a number of coefficients smaller than their immediate finer scales. Then it is intuitive that the mask size of these coarse scales should be smaller than the detail scales. Hence, adaptation of the mask both in inner and inter scales may provide better performance. Another way is to select a best weighting function for the difference term in the ACWMF, which is added to the center value within the concerned mask. Finally, as the smoothing process presented here is iterative, computational expense cannot be neglected. Without using the same stopping criterion for all scales as described in our presented result, a scale variant stopping criterion may significantly reduce the computational burden.

Bibliography

[1] W. Chen, M. Chen, and J. Zhou, "Adaptively regularized constrained total lesst-squares image restoration," *IEEE Trans. Image Processing*, vol. 9, no. 4, pp. 588-596, 2000.

[2] M. R. Banham, and A. K. Katsaggelos, "Digital image restoration," *IEEE Signal Processing Magazine*, pp. 24-41, 1997.

[3] R. M. Rangayyan, M. Ciuc, and F. Faghih, "Adaptive-neighbourhood filtering of images corrupted by signal-dependent noise," *Appl. Opt.*, vol. 37, no. 20, pp. 4477-4487, 1998.

[4] R. C. Gonzalez, and R. E. Woods, *Digital Image Processing*, Addison-Wesley Publishing Company, second edition, 1993.

[5] J. S. Lee, "Digital image enhancement and noise filtering by use of local statistics," *IEEE Trans. Pattern Anal. Mach. Intell.*, PAMI-2, pp. 165-168, 1980.

[6] J. S. Lee, "Refined filtering of image noise using local statistics," *Comput. Graphics Image Process.*, vol. 15, pp. 380-389, 1981.

[7] S. S. Jiang, and A. A. Sawchuk, "Noise updating repeated Wiener filter and other adaptive noise smoothing filters using local image statistics," *Appl. Opt.*, vol. 25, pp. 2626-2337, 1986.

[8] L. Yin, R. Yang, M.Gabbouj, and Y. Neuvo, "Weighted median filters: A tutorial," *IEEE Trans. Circuits Syst.*, vol. 43, pp. 157-191, 1996.

[9] F. Guichard, and J. M. Morel, "Image iterative smoothing and PDE's," *Lecture notes, School of mathematical problems in image processing*, The Abdus Salam International Centre For Theoretical Physics, September, 2000.

[10] H. H. Arsenault, C. Gendron, and M. Denis, "Transformation of film-grain noise into signal-independent Gaussian noise," *Appl. Opt.*, vol. 23, pp. 845-850, 1984.

[11] J. S. Lim, and H. Nawab, "Techniques for speckle noise removal," *Opt. Eng.*, vol. 21, pp. 472-480, 1981.

[12] G. Strang, and T. Nguyen, *Wavelets and Filter Banks*, Wellesley-Cambridge Press, Wellesley MA, first edition, 1996.

[13] J. C. Goswami, and A. K. Chan, *Fundamentals of Wavelets, Theory, Algorithms, and Application*, John Wiley & Sons, first edition, 1999.

[14] I. Daubechies, *Ten Lectures on Wavelets*, SIAM, Philadelphia PA, first edition, 1992.

[15] S. Mallat, *A wavelet tour of signal processing*, Academic Press, San Diego CA, 1998.

[16] Z. Xiong, M. T. Orchard, and K. Ramchandran, "A comparitive study of DCT and wavelet based coding," *Proc. SPIE*, no. 3164, pp. 271-278, 1997.

[17] D. L. Donoho, "Denoising by soft-thresholding," *IEEE Trans. Info. Theory*, vol. 41, no. 3, pp. 613-627, 1995.

[18] D. L. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *Journal of the American Statistical Assoc.*, vol. 90, no. 432, pp. 1200-1224, 1995.

[19] S. G. Chang, B. Yu, and M. Vettereli, "Spatially adaptive wavelet thresholding with context modeling for image denoising," *IEEE Trans. Image Processing*, vol. 9, no. 9, pp. 1522-1531, 2000.

[20] S. G. Chang, B. Yu, and M. Vettereli, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Trans. Image Processing*, vol. 9, no. 9, pp. 1532-1546, 2000.

[21] M. K. Mihcak, I. Kozintsev, and K. Ramchandran, "Spatially adaptive statistical modeling of wavelet image coefficients and its application to denoising," *Proc. IEEE Int. Conf. Acoustics, speech, and signal processing*, Phoenix, AZ, vol. 6, pp. 3253-3256, 1999.

[22] M. K. Mihcak, I. Kozintsev, K. Ramchandran, and P. Moulin, "Low-complexity image denoising based on statistical modeling of wavelet coefficients," *IEEE Signal Process. Lett.*, vol. 6, no. 12, pp. 300-303, 1999.

[23] H. Zhang, A. Nosralinia, and R. O. Wells, Jr., "Image denoising via wavelet-domain spatially adaptive FIR wiener filtering," *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Istambul, Turkey, vol. 5, pp. 2179-2182, 2000.

[24] Z. Cai, T. H. Cheng, C. Lu, and K. R. Subramanium, "Efficient wavelet-based image denoising algorithm," *Electron. Lett.*, vol. 37, no. 11, pp. 683-685, 2001.

[25] J. Liu, and P. Moulin, "Information-theoritic analysis of interscale and intrascale dependencies between image wavelet coefficients," appeared in *IEEE Trans. Image Processing*, 2001.
(Preprint available at:*http://www.ifp.uiuc.edu//*).

[26] J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. Signal Proc.*, vol. 41, pp. 3445-3462, 1993.

[27] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based statistical signal processing using hidden Markov models," *IEEE Trans. Info. Theory*, vol. 45, pp. 846-862, 1999.

[28] M. B. Martin, "Applications of multiwavlets to image compression," *Master's Thesis*, Virginia Tech., Blacjsburg, Virginia, 1999.

[29] D. C. Munson, "A note on Lena," *IEEE Trans. Image Processing*, vol. 5, no. 1, 1996
(Preprint available at:*http://www-2.cs.cmu.edu/ chuck/lennapg/editor.html*).