WAVELET-BASED REDUCTION OF SPATIAL VIDEO NOISE

Antonio De Stefano

A thesis submitted for the degree of

Doctor of Philosophy

University of Southampton

Faculty of Engineering and Applied Science

Institute of Sound and Vibration Research

September 2000

University of Southampton

Abstract

Faculty of Engineering and Applied Science
Institute of Sound and Vibration Research

Doctor of Philosophy

Wavelet-Based Reduction of Spatial Video Noise

by Antonio De Stefano

Many real word images are often contaminated by noise. Noise reduction techniques aim to improve image quality and can be used to facilitate further image processing. This work proposes an alternative method for spatial, additive, Gaussian noise reduction based on a discrete wavelet transform.

A new undecimated and shift invariant filter bank has been used to decompose the image into components. The basic filters are extrapolated from a biorthogonal wavelet basis. Reconstruction is obtained by simply summing the image components.

The noise reduction on the components is obtained by applying thresholding functions on the pixel values of each component. Each thresholding function is a member of a scheme and is characterised by a number of parameters. The scheme describes the shape of a parameterised family of thresholding functions. The parameters select the member of the family to be applied to each component. A new thresholding scheme, obtained from Bayesian optimal estimator theory, is designed. The parameters for each component are dependent on the level of the contaminating noise and are selected using a preliminary training procedure based on a set of video images. The cost function utilised for the training is a weighted version of mean square error designed to reflect the human visual system.

An estimation of the standard deviation level of the noise is required by the technique. Three techniques using the highest frequency band to estimate the level on all the bands are presented and a combined estimator is used.

The method has been tested on large sets of images and levels of additive, Gaussian, white and coloured noises. The method compares favourably with other wavelet based noise reduction techniques and demonstrates significantly increased noise reduction and visual quality.

Contents

CHAPTER 1: INTRODUCTION **1**

1.1 INTRODUCTION 2
1.2 ELEMENTS OF A DIGITAL IMAGE SYSTEM 3
1.3 IMAGE NOISE REDUCTION 5

CHAPTER 2: IMAGE NOISE REDUCTION USING WAVELET ANALYSIS **7**

2.1 INTRODUCTION 8
2.2 DECOMPOSITION-RECONSTRUCTION 10
2.2.1 Basic filters 11
2.2.2 Filter bank implementation 11
2.3 NOISE REDUCTION OF THE COMPONENTS 14
2.3.1 Shape of the thresholding functions 15
2.3.2 Parameters of the thresholding functions 16

CHAPTER 3: WAVELET ANALYSIS **19**

3.1 INTRODUCTION 20
3.2 FOURIER AND WAVELET TRANSFORMS 21
3.3 WAVELET TRANSFORM 22
3.3.1 Continuous Wavelet Transform (CWT) 22
3.3.2 Wavelet series expansion 24
3.3.3 Discrete Wavelet Transform (DWT) 27
3.3.4 Undecimated wavelet transform 29
3.4 FILTER BANKS 32
3.4.1 Shifting 32
3.4.2 Scaling 34
3.4.3 Perfect reconstruction 35
3.4.4 Multiresolution 39
3.5 WAVELET ANALYSIS FOR IMAGE NOISE REDUCTION 43
3.6 THE SPECIFICS OF THE WAVELET TRANSFORM 47
3.7 CONCLUSIONS 52

CHAPTER 4: THRESHOLDING FUNCTION SHAPE **53**

4.1 INTRODUCTION 54
4.2 FAMILIES OF THRESHOLDING FUNCTIONS 55
4.3 THRESHOLDING FUNCTIONS ON THE COMPONENTS 56
4.4 THRESHOLDING SCHEMES 59
4.4.1 Soft and hard thresholding 59
4.4.2 Optimal Bayesian thresholding schemes 60
4.4.3 Bayesian estimator for particular distributions 62
4.4.3.a Laplacian distribution 63
4.4.3.b Generalised Gaussian distribution 68
4.5 SCHEME IN OUR APPROACH 73
4.6 CONCLUSIONS 78

CHAPTER 5: THRESHOLDING FUNCTION PARAMETERS **79**

5.1 INTRODUCTION 80

5.2 COMPARISON OF THE TRAINING ALGORITHMS PERFORMANCES 81
5.3 COST FUNCTION PARAMETERS 83
5.3.1 Use of Hessian matrix to reduce the number of parameters 84
5.3.2 Reduction of the number of parameters of the thresholding functions 88
5.3.3 Number of decomposition levels 91
5.4 COST FUNCTION VALUES 94
5.4.1 Visual model 95
5.4.2 Weighted Mean Square Error (WMSE) 97
5.5 TRAINING RESULTS 100

CHAPTER 6: NOISE ESTIMATION 108

6.1 INTRODUCTION 109
6.2 NOISE CONTAMINATED IMAGE MODELS 110
6.3 NOISE LEVEL ESTIMATION IN THE HIGH FREQUENCY BAND 113
6.3.2 Noise level estimation using cumulative distribution functions (cdfs) 113
6.3.2 Noise level estimation using moment matching 116
6.3.3 Noise level estimation using linear regression 118
6.4 RESULTS OF THE NOISE LEVEL ESTIMATION 120
6.4.1 Results of noise level estimation using the cdfs 120
6.4.2 Results of noise level estimation using moment matching 123
6.4.3 Results of noise level estimation using linear regression 127
6.4.4 Performance comparison and noise level estimation using a combined estimator 129

CHAPTER 7: RESULTS 137

7.1 INTRODUCTION 138
7.2 INFLUENCE OF THE TRAINING PARAMETERS ON THE ALGORITHM PERFORMANCES 140
7.3 PERFORMANCE EVALUATION FOR MORE GENERAL NOISE CONDITIONS 149
7.4 COMPARISON WITH OTHER WAVELET BASED APPROACHES 155
7.4.1 Decimated and undecimated components 155
7.4.2 Thresholding techniques 155
7.4.3 Performance comparison 156
7.5 VISUAL QUALITY ANALYSIS 164

CHAPTER 8: CONCLUSIONS AND FURTHER WORK 185

8.1 INTRODUCTION 186
8.2 WAVELET DECOMPOSITION 187
8.3 SHAPE OF THE THRESHOLDING FUNCTIONS 189
8.4 PARAMETERS OF THE THRESHOLDING FUNCTION 190
8.5 NOISE ESTIMATION 192
8.6 NOISE REDUCTION RESULTS 194
8.7 FURTHER WORK 197

APPENDIX A:TRAINING ALGORITHMS I

A.1 BFGS QUASI NEWTON ALGORITHM II
A.2 DOWNHILL SIMPLEX (NELDER) ALGORITHM V
A.3 DIFFERENTIAL EVOLUTION ALGORITHM VIII

APPENDIX B OPTIMAL THRESHOLDING SCHEMES X

B.1 LAPLACIAN IMAGE MODEL XI
B.2 GENERALISED GAUSSIAN IMAGE MODEL XVI

APPENDIX C: LIST OF IMAGE EXAMPLES ..... **XX**

C.1 DOCUMENTS XXI

REFERENCES ..... **XXI**

Figures and Tables

FIGURE 1-1. ELEMENTS OF A DIGITAL IMAGE SYSTEM.....3
FIGURE 2-1. NOISE REDUCTION PROBLEM.....8
FIGURE 2-2. THE MAIN STEPS OF A WAVELET BASED NOISE REDUCTION PROCEDURE.....9
FIGURE 2-3. DECOMPOSITION AND RECONSTRUCTION.....11
FIGURE 2-4. SIMPLIFIED DECOMPOSITION AND RECONSTRUCTION SCHEME.....13
FIGURE 2-5. HARD AND SOFT THRESHOLDING SCHEMES.....16
FIGURE 2-6. NEW THRESHOLDING SCHEME.....16
FIGURE 3-1. FILTER BANK DECOMPOSITION.....34
FIGURE 3-2. EFFECT OF SUB-SAMPLING AND UP SAMPLING.....36
FIGURE 3-3. TWO-BAND CODING AND RECONSTRUCTION WITH ORTHONORMAL FILTER BANK.....38
FIGURE 3-4. TWO-BAND CODING AND RECONSTRUCTION WITH BIORTHOGONAL FILTER BANK.....39
FIGURE 3-5. TREE FILTER BANK WITH SUB-SAMPLING.....41
FIGURE 3-6. TREE FILTER BANK WITHOUT SUB-SAMPLING.....42
FIGURE 3-7. TREE FILTER BANK WITHOUT SUB-SAMPLING FOR IMAGE NOISE REDUCTION.....45
FIGURE 3-8. IDEAL FREQUENCY DECOMPOSITION.....46
FIGURE 3-9. FILTER APPROXIMATION.....49
FIGURE 3-10. TREE FILTER BANK USED IN OUR APPROACH.....50
FIGURE 3-11. IMAGE PLATE.....51
FIGURE 3-12. TWO-DIMENSIONAL FREQUENCY DECOMPOSITION OF THE ZONE PLATE.....51
FIGURE 4-1. PDFS OF ORIGINAL AND NOISY IMAGES.....57
FIGURE 4-2. PDFS OF HIGH FREQUENCY COMPONENTS OF AND NOISY IMAGES.....58
FIGURE 4-3. SIGNAL ESTIMATION IN PRESENCE OF ADDITIVE NOISE.....59
FIGURE 4-4. HARD AND SOFT THRESHOLDING SCHEMES.....60
FIGURES 4-5. MAP OPTIMAL SCHEME WITH ORIGINAL SIGNAL HAVING LAPLACIAN DISTRIBUTION, AND EFFECT ON THE PDFS OF USING THIS THRESHOLDING.....65
FIGURE 4-6. LMSE OPTIMAL THRESHOLDING SCHEME FOR A LAPLACIAN SIGNAL IN GAUSSIAN NOISE.....67
FIGURE 4-7. EFFECT ON THE PDFS OF LMSE OPTIMAL THRESHOLDING SCHEME FOR A LAPLACIAN SIGNAL IN GAUSSIAN.....70
FIGURE 4-8. MAP OPTIMAL THRESHOLDING SCHEME WHEN THE IMAGE HAS A GENERALISED GAUSSIAN.....70
FIGURE 4-9. MAP (AND LMSE) OPTIMAL THRESHOLDING SCHEME WHEN THE ORIGINAL COMPONENTS HAVE GAUSSIAN DISTRIBUTION.....71
FIGURE 4-10. EFFECT ON PDFS OF MAP (AND LMSE) OPTIMAL THRESHOLDING WHEN THE ORIGINAL IMAGE HAS A GAUSSIAN.....72
FIGURE 4-11. LMSE OPTIMAL THRESHOLDING SCHEME WHEN THE ORIGINAL IMAGE HAS GENERALISED GAUSSIAN.....72
FIGURE 4-12-A. PDFS VARIABILITY IN THE FREQUENCY RANGE FOR 3 IMAGES.....74
FIGURE 4-12-B. PDFS VARIABILITY IN THE IMAGES FOR 3 FREQUENCY.....75
FIGURE 4-13. A GENERAL THRESHOLDING SCHEME.....76
TABLE 5.1. UNWEIGHTED OPTIMISATION OF AN IMAGE CONTAMINATED BY WHITE NOISE WITH STD=15.....82
TABLE 5.2. NUMBER OF ITERATIONS AND CONVERGENCE TIME REQUIRED BY UNWEIGHTED OPTIMISATION OF AN IMAGE CONTAMINATED BY WHITE NOISE WITH STD=15.....82

TABLE 5.3. PARAMETERS OF THE THRESHOLDING FUNCTIONS FOR COMPLETE SCHEME. 84
TABLE 5.4. ANALYSIS OF THE SIGNIFICANCE OF THE PARAMETERS IN THE TRAINING PROCEDURE USING THE HESSIAN MATRIX. 88
FIGURE 5-1. THRESHOLDING SCHEME AFTER FIRST SIMPLIFICATION. 90
TABLE 5.4. ANALYSIS OF THE SIGNIFICANCE OF THE PARAMETERS IN THE TRAINING PROCEDURE USING THE HESSIAN MATRIX. 92
TABLE 5.5-A. UNWEIGHTED AND WEIGHTED OPTIMISATIONS FOR 3 DECOMPOSITION LEVELS AND 3 NOISE LEVELS. 93
TABLE 5.5-B. NOISE REDUCTION USING UNWEIGHTED AND WEIGHTED OPTIMISATIONS FOR 3 DECOMPOSITION LEVELS AND 3 NOISE LEVELS. 93
FIGURE 5-2. VISIBILITY FUNCTION DEPENDENCE ON THE DISTANCE OF A LINE TO AN EDGE. .96
FIGURE 5-3. MODEL FOR THE FREQUENCY RESPONSE OF THE VISUAL SYSTEM. 97
FIGURE 5-4. BLOCK DIAGRAM DESCRIBING THE WMSE EVALUATION. 98
FIGURE 5-5. THE FREQUENCY RESPONSE OF THE WEIGHTING FILTER. 98
FIGURE 5-6. FREQUENCY RESPONSE DESCRIBING THE VISUAL MODEL ASSUMED IN OUR APPROACH. 99
TABLE 5.6. PARAMETER DEPENDENCE ON THE NOISE LEVELS FOR UNWEIGHTED AND WEIGHTED OPTIMISATIONS. 101
FIGURE 5-7-A. RELATION BETWEEN PARAMETERS AND NOISE LEVELS (MSE). 102
FIGURE 5-7-B. RELATION BETWEEN PARAMETERS AND NOISE LEVELS (WMSE). 103
TABLE 5.7. ANALYSIS OF THE NOISE REALISATION INFLUENCE ON THE PARAMETERS. 104
TABLE 5.8-A. PARAMETER COMPARISON FOR NOISE LEVEL 5. 106
TABLE 5.8-B. PARAMETER COMPARISON FOR NOISE LEVEL 7. 106
TABLE 5.8-C. PARAMETER COMPARISON FOR NOISE LEVEL 9. 106
TABLE 5.8-D. PARAMETER COMPARISON FOR NOISE LEVEL 11. 107
TABLE 5.8-E. PARAMETER COMPARISON FOR NOISE LEVEL 13. 107
TABLE 5.8-F. PARAMETER COMPARISON FOR NOISE LEVEL 15. 107
FIGURE 6-1.GENERAL MODEL OF THE IMAGE SYSTEM. 110
FIGURE 6-2. MODEL OF THE IMAGE SYSTEM FOR ADDITIVE NOISE. 110
FIGURE 6-3. MODEL OF THE IMAGE SYSTEM FOR COLOURED NOISE. 111
FIGURES 6-4. FILTERS USED TO SIMULATE COLOURED NOISE 112
FIGURE 6-5. NOISE SPECTRAL DISTRIBUTION ESTIMATION. 112
FIGURE 6-6. NOISE ESTIMATION USING THE CDFS EVALUATION. 113
FIGURE 6-7. CDFS OF AN IMAGE FOR DIFFERENT NOISE LEVELS. 114
FIGURE 6-8. CDFS OF 13 IMAGES FOR NOISE. 114
TABLE 6.1. BEST GREY LEVEL DISCRIMINATORS FOR DIFFERENT NOISE LEVEL RANGES. 120
TABLE 6.2. CONFIDENCE INTERVAL USING OVERALL BEST GREY LEVEL DISCRIMINATOR. ..121
TABLE 6.3. RESULTS OF NOISE ESTIMATION USING THE CDFS. 121
FIGURE 6-9. NOISE ESTIMATION ON 13 IMAGES FOR 6 NOISE LEVELS WITH THE CDFS 122
FIGURE 6-10. MEAN ± ONE STANDARD DEVIATION OF THE OVERALL ESTIMATION, THEORETICAL NOISE LEVEL; CDFS METHOD 123
TABLE 6.4. RESULTS OF NOISE ESTIMATION USING MOMENT MATCHING. 124
FIGURE 6-11. NOISE ESTIMATION ON 13 IMAGES FOR 6 NOISE LEVELS WITH MOMENT MATCHING 125
FIGURE 6-12. MEAN ± ONE STANDARD DEVIATION OF THE OVERALL ESTIMATION, THEORETICAL NOISE LEVEL; MOMENT MATCHING METHOD. 126
TABLE 6.5. RESULTS OF NOISE ESTIMATION USING LINEAR REGRESSION. 127
FIGURE 6-13. NOISE ESTIMATION ON 13 IMAGES FOR 6 NOISE LEVELS WITH LINEAR REGRESSION. 128

FIGURE 6-14. MEAN ± ONE STANDARD DEVIATION OF THE OVERALL ESTIMATION, THEORETICAL NOISE LEVEL; LINEAR REGRESSION METHOD. 129
TABLE 6.6. PERFORMANCES OF THE THREE METHODS WITH RESPECT TO THE NOISE LEVELS AND OVER THE COMPLETE SET OF IMAGES. 130
TABLE 6.7. PERFORMANCES OF THE THREE METHODS WITH RESPECT TO THE INDIVIDUAL IMAGES AND OVER THE COMPLETE SET OF IMAGES. 131
TABLE 6.8. RESULTS OF NOISE ESTIMATION USING THE COMBINED ESTIMATOR. 132
FIGURE 6-15. NOISE ESTIMATION ON 13 IMAGES FOR 6 NOISE LEVELS WITH COMBINED ESTIMATOR 133
FIGURE 6-16. MEAN ± ONE STANDARD DEVIATION OF THE OVERALL ESTIMATION, THEORETICAL NOISE LEVEL; COMBINED ESTIMATOR METHOD. 134
TABLE 6.9. PERFORMANCES USING THE COMBINED ESTIMATOR WITH RESPECT TO THE NOISE LEVELS AND OVER THE COMPLETE SET OF IMAGES. 134
TABLE 6.10. PERFORMANCES USING THE COMBINED ESTIMATOR WITH RESPECT TO THE INDIVIDUAL IMAGES AND OVER THE COMPLETE SET OF IMAGES. 135
TABLE 6.11. COMPARISON OF THE NOISE ESTIMATION METHODS ON IMAGES NOT INCLUDED IN THE TRAINING SET. 136
FIGURE 7-1. TEST IMAGE. 140
FIGURE 7-2-A. FIRST IMAGE OF THE SET OF SIX IMAGES 141
FIGURE 7-2-B. SECOND IMAGE OF THE SET OF SIX IMAGES 141
FIGURE 7-2-C. THIRD IMAGE OF THE SET OF SIX IMAGES 142
FIGURE 7-2-D. FOURTH IMAGE OF THE SET OF SIX IMAGES 142
FIGURE 7-2-E. FIFTH IMAGE OF THE SET OF SIX IMAGES 143
FIGURE 7-2-F. SIXTH IMAGE OF THE SET OF SIX IMAGES 143
TABLE 7.1-A. MSE DEPENDENCE ON THE TRAINING SETTINGS. 145
TABLE 7.1-B. WMSE DEPENDENCE ON THE TRAINING SETTINGS. 145
FIGURE 7-3-A. RELATIONSHIP BETWEEN MSE REDUCTION AND NUMBER OF PARAMETERS. 146
FIGURE 7-3-B. RELATIONSHIP BETWEEN WMSE REDUCTION AND NUMBER OF PARAMETERS 146
TABLE 7.2. WMSE DEPENDENCE ON IMAGE USED IN THE TRAINING. 147
FIGURE 7-4. WMSE DEPENDENCE ON IMAGE USED IN THE TRAINING. 148
TABLE 7.3. WMSE AND MSE REDUCTIONS (IN PERCENTAGE) FOR 13 IMAGES WITH AWGN HAVING STANDARD DEVIATION 15. 149
TABLE 7.4. AVERAGE OVER 13 IMAGES OF THE PERCENTAGE OF WMSE AND MSE REDUCTION FOR 6 AWGN LEVELS. 150
TABLE 7.5. CONFIDENCE INTERVAL FOR MSE AND WMSE. 151
TABLE 7.6-A. MSE REDUCTIONS (IN PERCENTAGE) FOR 13 IMAGES FOR DIFFERENT TYPES OF NOISE, HAVING STANDARD DEVIATION 15. 152
TABLE 7.6-B. WMSE REDUCTIONS (IN PERCENTAGE) FOR 13 IMAGES FOR DIFFERENT TYPES OF NOISE, HAVING STANDARD DEVIATION 15. 152
TABLE 7.7-A. AVERAGE OVER 13 IMAGES OF MSE REDUCTION FOR VARIOUS NOISE TYPES. 153
TABLE 7.7-B. AVERAGE OVER 13 IMAGES OF WMSE REDUCTION FOR VARIOUS NOISE TYPES. 154
TABLE 7.8-A. MSE REDUCTION (IN PERCENTAGE) FOR 13 IMAGES (DECIMATED DECOMPOSITION). 157
TABLE 7.8-B. WMSE REDUCTION (IN PERCENTAGE) FOR 13 IMAGES (DECIMATED DECOMPOSITION). 157

TABLE 7.9-A.MSE REDUCTION (IN PERCENTAGE) FOR 6 NOISE LEVELS (DECIMATED DECOMPOSITION). 158
TABLE 7.9-B. WMSE REDUCTION (IN PERCENTAGE) FOR 6 NOISE LEVELS (DECIMATED DECOMPOSITION). 159
TABLE 7.10-A. MSE REDUCTION (IN PERCENTAGE) FOR 13 IMAGES (UNDECIMATED DECOMPOSITION). 160
TABLE 7.10-B. WMSE REDUCTION (IN PERCENTAGE) FOR 13 IMAGES (UNDECIMATED DECOMPOSITION). 160
TABLE 7.11-A. MSE REDUCTION (IN PERCENTAGE) FOR 6 NOISE LEVELS (UNDECIMATED DECOMPOSITION). 162
TABLE 7.11-B. WMSE REDUCTION (IN PERCENTAGE) FOR 6 NOISE LEVELS (UNDECIMATED DECOMPOSITION). 163
FIGURE 7-5-A. IMAGE CONTAMINATED BY WHITE NOISE. 166
FIGURE 7-5-B. IMAGE CONTAMINATED BY WHITE NOISE AFTER DE-NOISING. 167
FIGURE 7-6-A. IMAGE CONTAMINATED BY LF NOISE. 168
FIGURE 7-6-B. IMAGE CONTAMINATED BY LF NOISE AFTER DE-NOISING. 169
FIGURE 7-7-A. IMAGE CONTAMINATED BY BF NOISE. 170
FIGURE 7-7-B. IMAGE CONTAMINATED BY BF NOISE AFTER DE-NOISING. 171
FIGURE 7-8-A. IMAGE CONTAMINATED BY HF NOISE. 172
FIGURE 7-8-B. IMAGE CONTAMINATED BY HF NOISE AFTER DE-NOISING. 173
FIGURE 7-9 IMAGE 'LENNA' CONTAMINATED BY AWGN AND AFTER DE-NOISING. 174
FIGURE 7-10. IMAGE 'VERTEBRA' CONTAMINATED BY AWGN AND NON-ADDITIVE NOISE AND AFTER DE-NOISING. 174
FIGURE 7-11-A. IMAGE 'GIRL' ORIGINAL. 175
FIGURE 7-11-B. IMAGE 'GIRL' CONTAMINATED BY AWGN. 175
FIGURE 7-11-C AND 7-9-D. IMAGE 'GIRL' THRESHOLDED WITH UNWEIGHTED AND WEIGHTED PARAMETERS. 176
FIGURE 7-12-A. IMAGE 'PLANE' ORIGINAL (CONTAMINATED BY AWGN). 177
FIGURE 7-12-B. IMAGE 'PLANE' AFTER DE-NOISING USING MSE OPTIMISATION. 177
FIGURE 7-12-C. IMAGE 'PLANE' AFTER DE-NOISING USING WMSE OPTIMISATION. 177
FIGURE 7-13-A. IMAGE DIFFERENCE FOR WMSE OPTIMISATION 178
FIGURE 7-13-B. IMAGE DIFFERENCE FOR MSE OPTIMISATION 178
FIGURE 7-14-A. IMAGE DE-NOISED WITH WMSE OPTIMISATION 179
FIGURE 7-14-B. IMAGE DE-NOISED WITH DECIMATION AND VISUAL THRESHOLDING 181
FIGURE 7-14-C. IMAGE DE-NOISED WITH DECIMATION AND SUREVISUAL THRESHOLDING 180
FIGURE 7-14-D. IMAGE DE-NOISED WITH DECIMATION AND HYBRID THRESHOLDING 181
FIGURE 7-14-E. IMAGE DE-NOISED WITH DECIMATION AND MINIMAX THRESHOLDING 183
FIGURE 7-14-F. IMAGE DE-NOISED WITH DECIMATION AND MAD THRESHOLDING 184

Notations

▪ GENERAL

$\mathbf{T_m}$	Translation by m
Λ	Decimation operator
*	Discrete convolution
erfc( )	Complementary error function
sign( )	Signum function
m or μ	Mean
std or σ	Standard deviation
var or σ^2	Variance
$\Im\{\}$	Fourier Transform
D	Operator such that for every discrete filter h, $\mathbf{D}^i h$ is h with 2^i-1 zeros inserted between every pair of filter coefficients
f_N	Nyquist frequency
$\delta_{a,b}$	Kroneker delta function
$H(z)$	Filter discrete frequency response function
$h(I\Delta x)$	Filter discrete impulse response functions
$\mathbf{L}^2(R)$	Class of functions square integrable on the real line
SNR	Signal to Noise Ratio
Γ	Gamma function

▪ WAVELET ANALYSIS AND DECOMPOSITION FILTERS

$a\ (>0)$	Scaling of the wavelet transform
b	Shifting of the wavelet transform
n	Index defining the set of basic function for non-redundant wavelet transform (related to a and b)
$c_{a,b}$	Coefficients of the wavelet transform (continuous or discrete)
$d_{a,b}$	Scaling coefficients of the wavelet transform (continuous or discrete)
c_n	Coefficients of the wavelet transform with biorthogonal basic functions
d_n	Scaling coefficients of the wavelet with biorthogonal basic functions

$\phi(x)$	Scaling function in the wavelet transform (continuous or discrete)
$\psi(x)$	Basic wavelet
$\Psi(f)$	Spectrum of the basic wavelet
$\{\psi_{a,b}(x)\}$	Set of basis functions
$\langle \psi_{a,b}, \psi_{k,j} \rangle = \delta_{a,k}\delta_{b,j}$	Orthonormal set of basic wavelets
$\psi(x)$ and $\tilde{\psi}(x)$	Dual wavelet basic functions
$\langle \psi_{a,b}, \tilde{\psi}_{k,j} \rangle = \delta_{a,k}\delta_{b,j}$	Biorthogonal sets of basic wavelets
CWT	Continuous Wavelet Transform
DWT	Discrete Wavelet Transform
IDWT	Inverse Discrete Wavelet Transform
UDWT	Undecimated Discrete Wavelet Transform
$H_1(z)$	Discrete frequency response of low-pass filters in the wavelet decomposition used to obtain averages
l_n	Coefficients of the wavelet decomposition related with $H_1(z)$
$H_0(z)$	Discrete frequency response of high-pass filters in the wavelet decomposition used to obtain detail
h_n	Coefficients of the wavelet decomposition related with $H_0(z)$
PR	Perfect reconstruction condition in the wavelet decomposition. No distortion and no aliasing
QMF	Quadrature Mirror Filter
HH, HL, LH and LL	Image components related to the first level of decomposition
LLHH, LLHL, LLLH and LLLL	Image components related to the second level of decomposition (decomposing **LL**)
LLLLHH, LLLLHL, LLLLLH and LLLLLL	Image components related to the third level of decomposition (decomposing **LLLL**)
LF and HF	Filter derived from a biorthogonal set of filters and used in the new decomposition

- **TRAINING AND THRESHOLDING FUNCTION ESTIMATION**

MAD	Mean Absolute Deviation (is used as non-adaptive threshold parameters optimisation criterion)
VISUAL	Non-adaptive visual technique to select the threshold parameters

HYBRID	Adaptive visual technique to select the threshold parameters
SURE	Stein's Unbiased Risk Estimate (is used as adaptive threshold parameters optimisation criterion)
RiskShirk Minimax	Threshold parameters optimisation for visual perception
MSE	Mean Square Error
Unweighted optimisation	Training-based minimisation of the MSE
WMSE	Weighted Mean Square Error
Weighted optimisation	Training-based minimisation of the WMSE
GCV	Generalised Cross Validation (is used as criterion for denoising the components)
Direct search methods	Class of non gradient based multidimensional minimisation algorithms
Simplex	Nelder-Mead algorithm for multidimensional minimisation (included in the direct search methods)
Differential evolution	Algorithm for multidimensional minimisation included in the direct search methods
Quasi-Newton methods	Class of gradient based multidimensional minimisation algorithms
BFGS	Broyden, Fletcher, Goldfarb and Shannon algorithm included in the quasi-Newton class
K	Number of functions (components) to be trained concurrently
l	Number of variables for function
$f()$	Cost function to be minimised in the training
H	Hessian matrix derived from Taylor series
$H_c(f)$	Frequency response of the visual system (*Hunt* [51])
W	Weighting filter used in the training
PDF	Probability Density Function
$u(x)$	Original signal
$v(x)$	Observation
$n(x)$	Contaminating noise
$\hat{u}(x)$	Optimal estimate
$f[v(x)]$	Thresholding function
$\hat{u} = \begin{cases} v-K & \text{if } v > K \\ 0 & \text{if } v = 0 \\ v+K & \text{if } v < K \end{cases}$	Soft-thresholding

$\hat{u} = \begin{cases} v & \text{if } \lvert v \rvert > K \\ 0 & \text{if } \lvert v \rvert \leq K \end{cases}$	Hard-thresholding
K	Threshold level
$p(\ )$	Probability distribution of a signal
$p(u/v)$	Posterior probability density function of u given v
$p(v/u)$	Likelihood function
$p(u)$	Prior
$p(v)$	Evidence
$R(\hat{u}) = E[C(\hat{u},u)]$	Risk function
$C(\hat{u},u)$	Cost function
$E[]$	Statistical expectation
MAP	Maximum a Posteriori
ML	Maximum Likelihood
LMSE	Least Mean Square Error
AWGN	Additive White Gaussian Noise
σ_η	Standard deviation of the noise or the noise level
σ_u	Spectral content of the original signal
β_u	Controls the form of the statistical distribution for generalised Gaussian distribution (β_u=0 correspond to a Gaussian distribution)
$\hat{u}_{\substack{MAP \\ Laplacian}}$	Optimal estimate using MAP criterion for components of the original image having Laplacian distribution
$\hat{u}_{\substack{LMSE \\ Laplacian}}$	Optimal estimate using LMSE criterion for components of the original image having Laplacian distribution
$\hat{u}_{\substack{MAP \\ Gaussian}} = \hat{u}_{\substack{LMSE \\ Gaussian}}$	Optimal estimate using MAP or LMSE criterion for components of the original image having Gaussian distribution
$\hat{u}_{\substack{MAP \\ Gen.Gaussian}}$	Optimal estimate using MAP criterion for components of the original image having Generalised Gaussian distribution
$\hat{u}_{\substack{LMSE \\ Gen.Gaussian}}$	Optimal estimate using LMSE criterion for components of the original image having Generalised Gaussian distribution
K1x, K1y and *K2*	Parameters of the new thresholding scheme

- **NOISE ESTIMATION**

$\eta_1(m,n)$	Additive noise
$\eta_2(m,n)$	Multiplicative noise

$\eta_w(m,n)$	AWGN
$\eta_c(m,n)$	Filtered (coloured) noise
LF, *BF* and *HF*	Filters used to colour the white noise
cdf	Cumulative distribution function
h	Grey level
$m_l(h)$, $\sigma_l(h)$	Mean and standard deviation of the cdf at a grey level h computed across the set of images
$f_{l_1,l_2}(h)$	Function to evaluate the optimal grey level between two noise levels
h_0^*	Optimal grey level between two noise levels
$f_{tot}(h)$	Function to evaluate the optimal grey level over all the noise levels
h_0	Optimal grey level over all the noise levels
K	Number of images
H	Number of noise levels
l_i	Noise level
$M1$, $M2$ and $M3$	Three normalised moments having standard deviation dimensions
$\mathbf{A}$	Matrix of the normalised moments for K images and H noise levels
α_i	Regression coefficients
$\underline{x}$	Matrix of the regression coefficients for K images and H noise levels
σ	True noise level (standard deviation)
$\mathbf{G}$	Matrix of the noise levels for K images and H noise levels
i and j	Indices of noise level and image
μ_i and $\sigma 1_i$	Averages and the standard deviation of the estimation over the 13 images and for each noise level l_i
$\mathbf{Q}1_{tot}$	Index to compare the performances of methods with respect to the noise levels
$e_{i,j}$	Noise level estimated over the image j with noise level l_I
$\sigma 2_j$	Standard deviation of the estimation over the H noise level
$\mathbf{Q}2_{tot}$	Index to compare the performances of the methods with respect to the individual images

Accompanying Material

CD-ROM including

- Sequences of static images contaminated by Additive White Gaussian Noise and sequences de-noised using the method proposed in this thesis.
- Set of images contaminated by additive, Gaussian, white and coloured noises.
- Set of images contaminated by additive, Gaussian, white and coloured noises de-noised using the methods proposed in this thesis.

Acknowledgements

I would like to thank my supervisor, Dr. Paul White, for his guidance in the last years. I would like also to thank the following; Dr. William Collis for the encouragement and for sorting all manner of problems, Martin Weston for his constructive advice in countless discussions, and Maureen Strickland for her help in reading this thesis.

I would like to thank my parents and my brother for their support.

Finally it is my pleasure to acknowledge the financial support provided by Snell & Wilcox Ltd.

Chapter 1

Introduction

1.1 Introduction

In the last decades an impressive effort has been made by the scientific world to investigate and to move forward the research concerning techniques operating on images in digital format. Broadcasting of terrestrial and satellite digital television, live streaming of videos over the internet and restoration of old movies are examples of tasks requiring high speed image processing techniques for coding and enhancement. Research in this field has also taken advantage of the increased speed and power of computers and instrumentation as accommodating large data sets with reasonable speed [1].

One aim of image processing techniques is to operate on signals to reduce disturbances caused by acquisition devices to create an output that represents reality as close by as possible. These disturbances are commonly classified into noise and distortion. Noise reduction algorithms, which are the central theme of this thesis, are often included in the wider class of image enhancement algorithms. The area of application of noise reduction algorithms is not limited to the enhancement of visual quality since their preliminary use may also increase the efficiency of other image processing techniques. Reducing the noise facilitates efficient coding, reconstruction, enhancement, and feature extraction.

The specific area of interest in this thesis is video images. As indicated above, the noise reduction techniques may have one of two aims: to improve the visual quality or to facilitate other processing techniques by discarding useless information. This study explores techniques based on wavelet transforms and proposes an innovative approach that significantly reduces noise and improves visual quality.

1.2 Elements of a digital image system

A system to obtain an image from the real world in digital format in general consists of three stages (figure 1-1): acquisition, processing, and displaying.

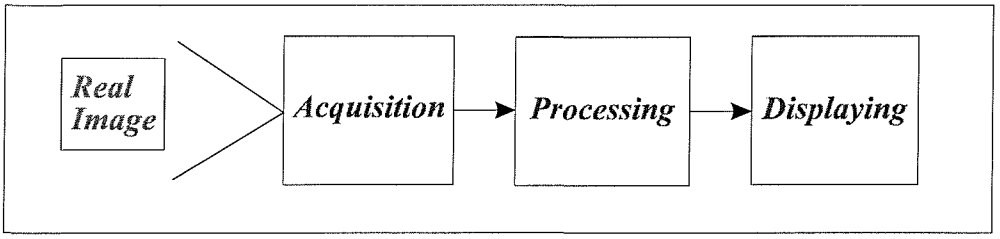


Figure 1-1. Elements of a digital image system.

The acquisition stage has the objective of "measuring" and representing the image in digital format. An image is said to be in digital format, if one (grey level) or more (colour) finite numbers represent a small area of the image (called a pixel). Two dimensional and multidimensional matrices may be used to represent the image in digital format. A conversion into digital format (digitisation) is required when the acquisition element of the equipment represents the image in analogue format. Most modern devices, such as scanners and digital cameras, are able to directly acquire the image in digital format.

When the image has been acquired in digital format one or more operations can be performed in order to transmit it, to improve its quality (defined as required by the application), and in general, to extract information from it. This stage is commonly called processing. Many techniques have been developed for the specific tasks in this stage, for example:

1) Coding and compression algorithms are used to represent the image in compact form for efficient transmission/storage. The Motion Picture Expert Group (MPEG) format is an example of format particularly suitable for video broadcasting and is currently used in digital satellite television [2-4].

2) The aim of enhancement algorithms is to improve the quality of the image as required by the specific application. These can be destined, for example, to increase the visual quality of the image, or to modify the image in order to make specific parts more visible. Restoration of old movies and video quality improvement in the presence of imperfect light conditions are some of the applications of these algorithms [5-7].

3) Reconstruction algorithms are used to obtain images based on several sources of information e.g. a series of images. Some of these techniques are used to build three-dimensional models from a set of images. The algorithms for reconstruction are particularly effective in medical and astronomical applications [8-10].

4) Feature extraction and segmentation algorithms aim to extract areas or sets of parameters characteristic of the image. These algorithms have been demonstrated to be very useful as friendly support for the non-expert user interested only in a limited number of characteristics of the images. The Federal Bureau Investigation uses some of the feature extraction and segmentation algorithms for fingerprint identification [11-13].

The final stage of a digital image system is the displaying. The aim of the displaying stage is to present the processed image to the human eye. There are several display-related pitfalls that should be avoided in order to have correct judgement and thus not degrade image quality.

1.3 Image noise reduction

Image processing algorithms operate on the digital representation of the image and inaccuracies in that representation can make the algorithms less effective. Noise reduction algorithms reduce the disturbance contaminating the image[1]. Reducing the noise avoids coding and compressing useless data, increases the general quality of the image, aids the extraction of specific areas or information from the image, and makes the reconstruction procedure easier.

It is important to underline the difference between noise reduction and enhancement procedures. "*Noise reduction algorithms increase the image quality but aim to keep the original image unchanged*" whilst "*in enhancement the goal is to accentuate certain image features for subsequent analysis*" [10].

Noise reduction is sometimes accomplished simultaneously with the processing stage and consequently uses techniques influenced by the type of processing task required. The application of such algorithms may be limited to the specific processing operation. Video noise reduction algorithms can be classified into temporal and spatial techniques.

a) Temporal techniques utilise a set of noisy images temporally consecutive in order to discern the noise from the original image. Clearly these techniques can only be applied when the processing task is related to the use of sequence of images, e.g. video.

b) Spatial techniques utilise only a single noisy image where the original image is unknown. These techniques must discern the noise from the image based on their different characteristics.

[1] The assumed model for the disturbance (additive noise) contaminating the image is described in section 4.4.

This work presents a new technique for spatial noise reduction on video images which uses wavelet based decomposition. The technique uses new solutions for frequency decomposition and for noise reduction on the components. The technique is shown to produce remarkable reductions in additive white and coloured noise. The next chapter will introduce the reader to the technique, comparing it with other approaches described in the literature and underlining its novelties, whilst chapters 3, 4, 5 and 6, will describe it in more detail. Chapter 7 will present the results and chapter 8 draws conclusions and discusses extensions to the work.

Chapter 2

Image noise reduction using wavelet analysis

2.1 Introduction

The task of image noise reduction techniques is to minimise degradation that has occurred while the image was obtained. Spatial noise reduction algorithms are not trivial because they are required to work in absence of *a priori* information regarding the image. In this work the noise is assumed additive. The general problem of additive noise reduction can be schematised as in figure 2-1.

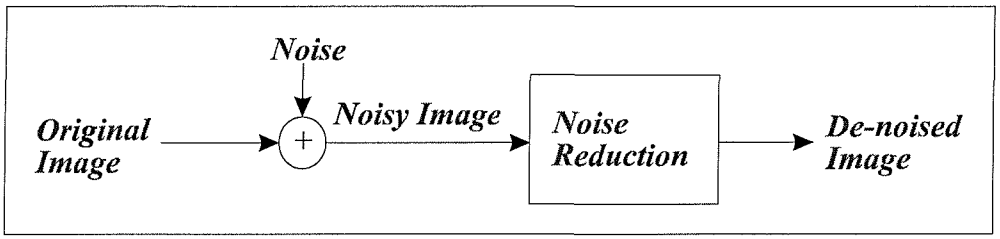


Figure 2-1. Noise reduction problem.

The procedure of noise reduction must identify the characteristics of the noise contaminating the original image in order to obtain an estimate of the original image. The separation of the noise from the image is particularly complicated when the noise and image histograms overlap significantly.

Wavelet analysis is a multiscale signal decomposition method particularly efficient for processing data dominated by transient behaviour or discontinuities [14-19]. This efficiency is a key point in applications where it is a fundamental task to identify and preserve singularities, such as edges in images [20]. Therefore, wavelet decomposition has been proposed as a framework in image noise reduction [21-34] and restoration [35-44] for different classes of images. An image is decomposed into a set of sub-bands, and the information within each sub-band is processed independently of that in the other sub-bands [45-52].

This chapter aims to introduce the reader to the technique of spatial noise reduction using wavelet analysis for signal decomposition. The next sections outline the concepts on which the method is based and the innovative solutions used. Elements characterising wavelet based approaches for noise reduction are:

1. banks of filters used in the decomposition and reconstruction
2. shape of the family of functions utilised for thresholding components
3. criteria employed for selecting one member of the family for each component

The decomposition-reconstruction algorithm and the thresholding scheme are interrelated. Nevertheless in subsequent sections of this chapter these are examined separately leaving the discussion regarding the connections to the following chapters.

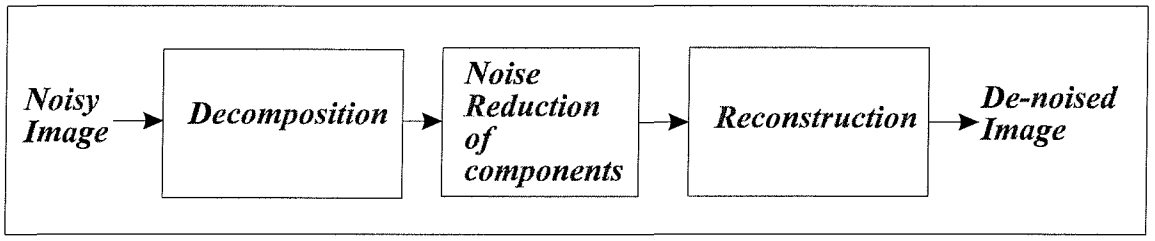


Figure 2-2. The main steps of a wavelet based noise reduction procedure

2.2 Decomposition-reconstruction

Wavelet analysis allows transformation of the image from the spatial domain to a domain related to horizontal and vertical frequency bands. The noise reducer acts so that bands in which the noise is dominant are more heavily attenuated than those which are primarily signal related [53-55]. The transform and the inverse transform are implemented in stages commonly referred to as decomposition and reconstruction (figure 2-3). Decomposition divides the image into the frequency bands, and reconstruction recombines the frequency components [56]. If no intermediate operations are performed, the application of decomposition and reconstruction algorithms should leave the image unchanged.

Discrete Wavelet Transform (DWT) and Inverse Discrete Wavelet Transform (IDWT) are the mathematical instruments used respectively for decomposition and reconstruction. The transforms are implemented by applying shifted and dilated versions of a set of basic filters. The implementation of the filter bank from the set of basic filters may, or may not include a decimation operation and should allow the complete reconstruction of the image without distortion and, in presence of decimation, aliasing. The filters and their implementation as a filter bank will be examined.

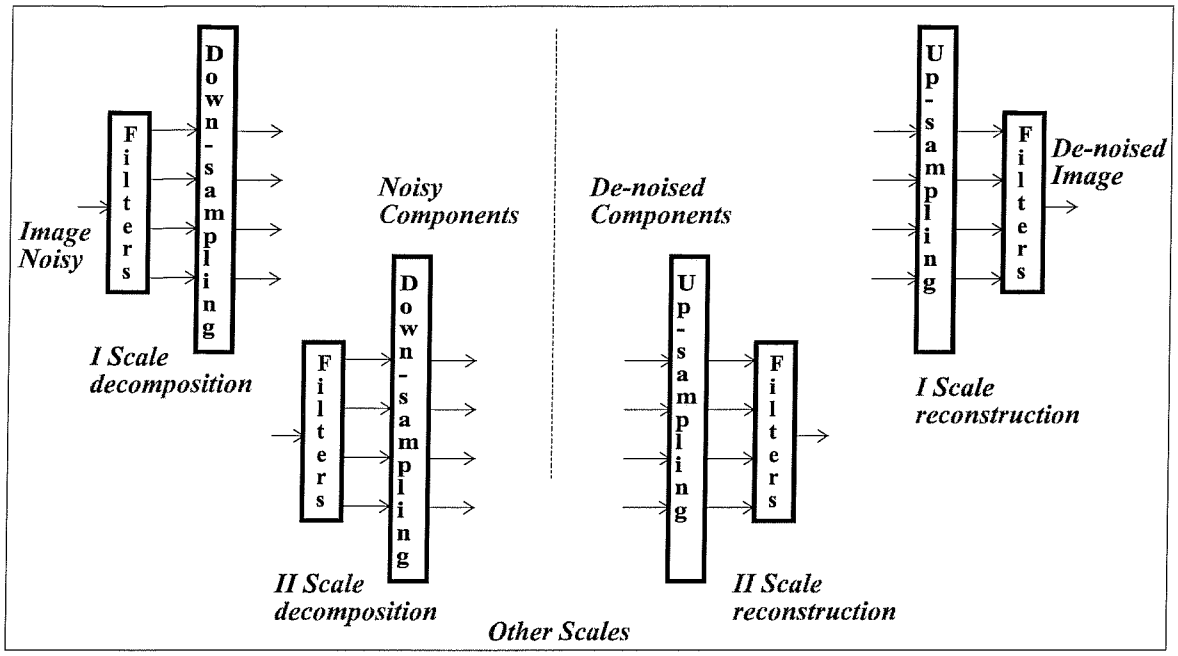


Figure 2-3. Decomposition and reconstruction.

2.2.1 Basic filters

There exists orthogonal [57-59] and biorthogonal [60-62] sets of basic filters able to produce filter banks which achieve perfect reconstruction [63-64]. The differences between these filters are in number of elements of the set and their frequency responses. The selection of the set of basic filters has influence on the frequency decomposition and on the overall complexity of the system. The next chapter details the properties of these sets; here it is sufficient to note that the basic filters used herein constitute a set of biorthogonal filters. From these a non-orthogonal filter bank is implemented which achieves perfect reconstruction. The choice of this set is influenced by the simplification that can be obtained in the implementation of the filter bank.

2.2.2 Filter bank implementation

The filter bank is derived from shifted and dilated versions of the basic filters. The operation of data decimation reduces the data size at each scale in order to remove redundancy and keeps the total data size constant. On the other hand, this operation,

introducing aliasing, produces shift-variance[1] which is the cause of some visual artefacts. The problem of the shift-variance can be solved using redundant transforms. One solution is to use circular shifting of the decimated transform [65-68]. Pyramidal decomposition representing the Fourier domain in terms of polar coordinates instead of Cartesian coordinates can be used to obtain shift-invariant transform of video images [69] and sequences of images [70]. Dual tree filters generating complex coefficients can also be utilised to obtain approximately shift-invariant transforms [71]. These redundant transforms are able to avoid artefacts introduced by shift-variance but the price paid is an increase in the size of the data to be stored.

In this work a shift-invariant undecimated redundant transform is employed. The redundancy cost can in part be offset against the complexity saved by the simplification of the filter bank implementation. The absence of decimation also serves to simplify the condition for perfect reconstruction, which is in general composed of two conditions, one for non-distortion and one for absence of aliasing. The implementation of an undecimated filter bank must only take into account the non-distortion condition. Further, in absence of decimation, the linearity of the system allows the decomposition and reconstruction filtering to be combined before thresholding. The convolution of decomposition and reconstruction filters results in new decomposition filters and the reconstruction is achieved by simply summing the components. The filter frequency response at each scale is implemented using the 'a trous' [71-75] by inserting zeros between the filter coefficients. These filters are applied to each component, each component being the size of the original image. The decomposition-reconstruction method applied here is depicted in figure 2-4.

[1] The demonstration of this is in [63].

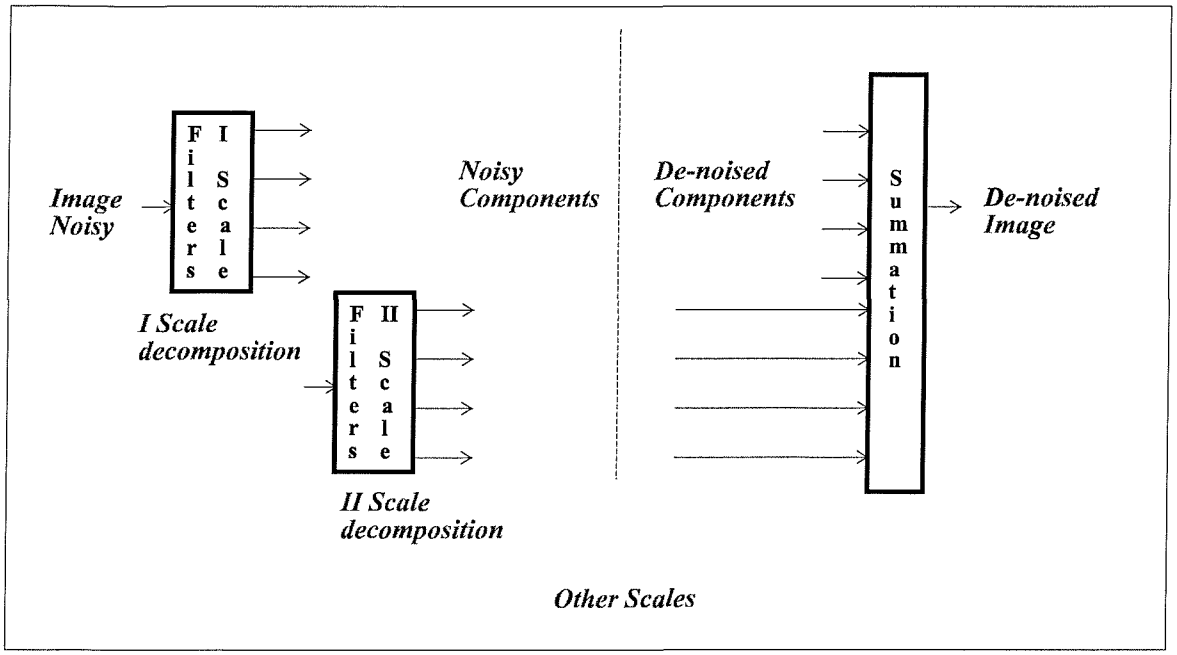


Figure 2-4. Simplified decomposition and reconstruction scheme.

2.3 Noise reduction of the components

The components obtained from the decomposition represent the image in different frequency bands, or equivalently on different detail scales. The original image is in general not uniformly distributed in frequency. Normally there are some bands where the noise is dominant and some where the signal dominates.

Different approaches have been proposed to discern the noise from the original image. The noise identification can be based on the correlation between components at different scales [76-77] making the hypothesis that the noise has a much weaker correlation than the original image. Another method to identify the noise is based on the examination of the local Lipschitz regularity of the components [78-79]. The assumption is that the regularity of the local extrema of the original image is large compared with that of the noise. A method for spatially adaptive restoration of noisy and blurred images is to apply a multiscale Kalman smoothing filter to the components [43-44]. The regularisation parameters of the least squares filters change depending on scale, local SNR, and orientation allowing the spatially adaptive restoration.

Noise identification can also be performed on the basis of an estimation of the probability that the components are affected or unaffected by noise [80]. *A priori* information and components are combined in a Bayesian probabilistic formulation and are implemented as a Markov random field image model.

Finally the most commonly used approach is based on the analysis of the pixel values [65, 81-87]. Thresholding functions are applied to the components and normally attenuate small values, assumed to be related to the noise, whilst retaining large values. In this case, the procedure of noise reduction of the components is dependent on two elements: the generic

shape of the thresholding functions applied, and the parameters specifying the exact values of the thresholding function.

In this work a new thresholding scheme and a new criterion for parameter selection are utilised. These are summarised in next two subsections.

2.3.1 Shape of the thresholding functions

Two of the most common thresholding schemes are termed soft-thresholding and hard-thresholding. Figure 2-5 shows a plot of the output pixel values against input pixel values for these schemes. In the presence of white noise, hard-thresholding yields a smaller mean squared error than soft thresholding but produces visual artefacts. Soft-thresholding provides almost minimax mean square error subject to the condition of similar smoothness between the original image and its estimate [68, 88]. These schemes have also been applied in the presence of speckle [85] and correlated noises [77, 88-89].

The method presented in this thesis uses a new thresholding scheme that is derived using a Least Mean Square Error (LMSE) approach based on a Bayesian procedure. The noise is assumed to have Gaussian distribution and the components are modelled as having either a Laplacian or a generalised Gaussian distribution. The two families of optimal thresholding functions have respectively been obtained analytically and numerically. These families are approximated by the thresholding function scheme shown in figure 2-6. This will be demonstrated in more detail in chapter 4. In the new scheme each function is characterised by three parameters instead of one, like in the hard and soft thresholding. Nevertheless, this scheme includes both hard and soft thresholding schemes.

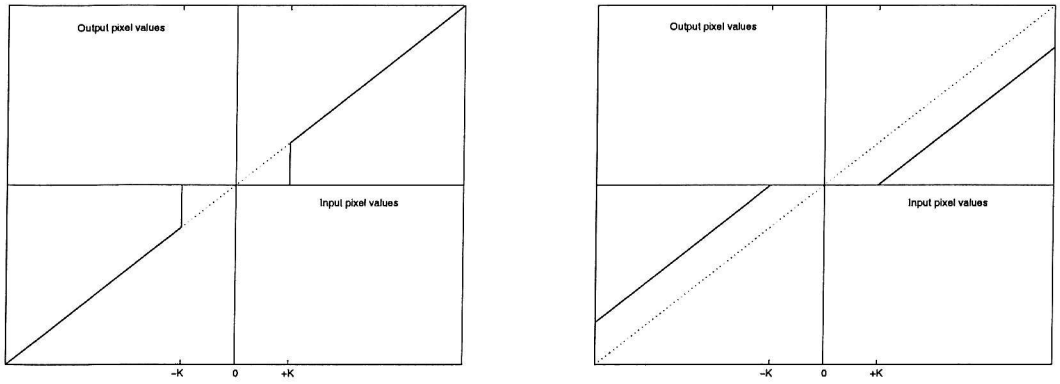


Figure 2-5. Hard and soft thresholding schemes.

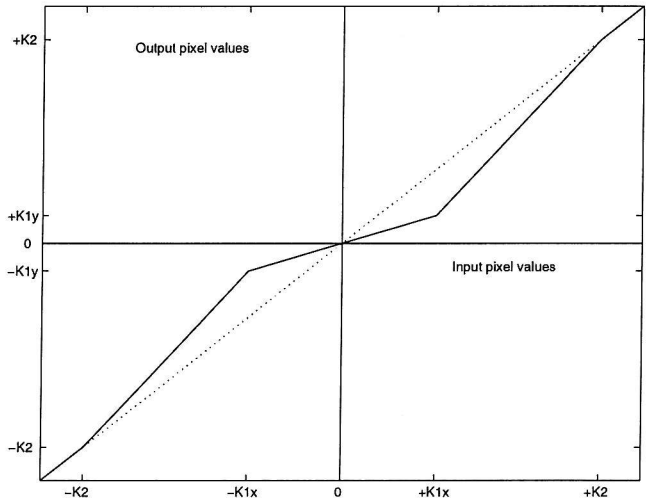


Figure 2-6. New thresholding scheme.

2.3.2 Parameters of the thresholding functions

The hard and soft-thresholding schemes are families where each function is characterised by a single parameter: the thresholding level. Different techniques have been utilised to select this parameter for each image component [90-91]. In general these parameters depend on the level of the noise contaminating the component; therefore the first step is the estimation of the noise level on each components. In the presence of spatially white noise, the level is constant on all the components and the majority of the techniques use the highest frequency component to estimate it. This assumes that the highest frequency is the

one with the lowest signal to noise ratio. The mean absolute deviation (MAD) of the pixel values [75, 83, 85, 88, 92] is the parameter commonly evaluated. In the presence of coloured noise this method is not satisfactory and noise level estimation is required on each component.

The techniques to estimate the thresholding level[2], assuming the noise level on each component to be known, can be classified into two categories. The first category includes techniques where the parameter only depends on the noise level and is constant on all the decomposition scales. In this case only one parameter is selected and the same 'universal thresholding function' has to be applied on all the components. The selection criterion can be based on image size [65-66], or on Minimax optimisation for the visual perception [82-83]. The second category includes techniques where distinct parameters are selected and the thresholding functions are applied to each component. These techniques are adaptive with respect to the noise level and image spectral content. Examples of techniques that fall into this second category are the criteria based on the Stein's unbiased risk estimate (SURE) [82], on the calculation of the median absolute deviation (MAD) of the pixel values for each level [75, 88], and on the interpretation of the thresholding level selection as a multiple hypothesis testing problem [86]. Another method to select the thresholding functions parameters is based on a generalised cross validation (GCV) algorithm [77]. This method is asymptotically optimal and does not require estimation of the noise level.

The criterion used in our approach is based on training the parameters to minimise a weighted version of the mean square error (WMSE) between the de-noised and original images. This index represents a measure of mean square error (MSE) in the frequency bands where the noise is more important. The use of the WMSE as a cost function

[2] Detail concerning these techniques are in section 7.4.2.

produces visual superior image quality compared with the MSE. A modified version of the Nelder algorithm [93-94] has been utilised to minimise the WMSE. This algorithm is based on the use of a Simplex and does not need gradient computation. The dependence between parameters and noise level turns out to be almost linear. The noise level estimation is performed using an estimator combined from three algorithms. In the presence of coloured noise the method is generalised using an estimation of the spectral noise density.

The training has been performed initially on one image and successively on a set of images. The dependence of the parameters on the training image has been studied.

Chapter 3

Wavelet analysis

3.1 Introduction

This chapter aims to introduce the reader to the use of wavelet analysis and filter banks on digital images and to describe their use in this application. The properties of wavelet analysis are initially described for mono-dimensional signals, e.g. time series. Subsequently it is explained how the implementation of filter banks relates to wavelet analysis theory. Particular attention is then given to the case of filter banks for image processing. Finally, the filter bank central to this thesis is presented and its properties are outlined.

3.2 Fourier and wavelet transforms

In signal and image processing, transforms are useful tools to explore signal characteristics. In both cases the use of an alternative domain to the time and space domains allows features to be readily discerned that otherwise might be difficult to identify. Fourier transforms use sinusoidal waves as orthonormal basis functions. For a mono-dimensional time dependent signal, these bases are inappropriate to represent signals having non-zero values only over a short time interval. Transient signals are problematic for Fourier representations because they are based on infinite duration sine waves, and require cancellation over most of the interval in order to create a function non-zero only over a small section. Extending the discussion to the two-dimensional case, natural images can be considered as containing many transient elements since many of their features are highly localised in space. For this reason, the Fourier transform is inappropriate for representing images having a large number of discontinuities. Mathematicians and engineers have explored various transforms having basis functions of limited duration; one of these is known as the wavelet transform.

3.3 Wavelet transform

There are three different but related wavelet transformations:

a) continuous wavelet transform (CWT),

b) wavelet series expansion,

c) discrete wavelet transform (DWT).

It is the DWT that is used for digital image processing. Nevertheless, to describe the DWT's characteristics the basic theory regarding CWT and wavelet series expansion is required. In this section the continuous series expansion and discrete cases are examined, then the filter bank implementation deriving from the wavelet analysis is described.

3.3.1 Continuous Wavelet Transform (CWT)

For a one-dimensional continuous case, wavelet analysis deals with the class of functions $f(x)$ that are square integrable on the real line (also indicated by $f(x) \in \mathbf{L}^2(\boldsymbol{R})$). This can be expressed as

$$\int_{-\infty}^{+\infty} |f(x)|^2 dx < \infty \qquad (3.1)$$

This class can be represented by using a set of basis functions obtained by shifting and translating a single prototype function $\psi(x)$, called the **basic wavelet**, also integrable on the real line ($\psi(x) \in \mathbf{L}^2(\boldsymbol{R})$). A basic wavelet must be a real valued function, whose spectrum $\Psi(f)$ satisfies the properties

$$C_\Psi = \int_{-\infty}^{+\infty} \frac{|\Psi(f)|^2}{|f|} df < \infty$$

$$\Psi(0) = 0 \Rightarrow \int_{-\infty}^{+\infty} \psi(x)dx = 0 \qquad (3.2)$$

$$\Psi(\infty) = 0$$

Translating and scaling the basic wavelet generates a set of basis functions $\{\psi_{a,b}(x)\}$

$$\psi_{a,b}(x) = \frac{1}{\sqrt{a}} \psi\left(\frac{x-b}{a}\right) \qquad (3.3)$$

The coefficients a (>0) and b respectively indicate *scaling* and the *shifting* of the basic function. These two operations are the core of the technique. The basic wavelet $\psi(x)$ is shifted as $\psi(x-b)$. This operation partitions the time axis into bands permitting to be analysed separately. The basic wavelet $\psi(x)$ is scaled as $\psi(x/a)$. This operation partitions the frequency axis non-uniformly allowing a multiresolution analysis. Section 3.4 shows how these two operations are implemented in a filter bank.

The **continuous wavelet transform** (CWT) of the function $f(x)$ with respect to the basic wavelet $\psi(x)$ is defined as

$$W_f(a,b) = \langle f(x), \psi_{a,b}(x) \rangle = \int_{-\infty}^{+\infty} f(x)\psi_{a,b}(x)dx \qquad (3.4)$$

<•,•> represents the inner product. The inverse transform is

$$f(x) = \frac{1}{C_\Psi} \int_0^{+\infty} \int_{-\infty}^{+\infty} W_f(a,b)\psi_{a,b}(x)db\frac{da}{a^2} \qquad (3.5)$$

The CWT is said to be overcomplete, it represents an increase in the dimension of the signal and consequently in complexity of elaboration.

3.3.2 Wavelet series expansion

Now consider the case when:

1. the scaling are factors of 2 (2^a, a integer, *binary scaling*), and
2. the shifts are defined $b/2^a$ (b integer, *dyadic translation*).

A basic wavelet $\psi(x)$ is said to be **orthogonal** if using

$$\psi_{a,b}(x) = 2^{a/2}\psi(2^a x - b) \qquad (3.6)$$

generates a set which is an **orthonormal** basis of $\mathbf{L}^2(\boldsymbol{R})$, this means

$$\left\langle \psi_{a,b}, \psi_{k,j} \right\rangle = \delta_{a,k}\delta_{b,j} \qquad (3.7)$$

and

$$\forall\ f(x) \in \mathbf{L}^2(\mathbf{R}) \Rightarrow f(x) = \sum_{a=-\infty}^{+\infty} \sum_{b=-\infty}^{+\infty} c_{a,b}\psi_{a,b}(x) \qquad (3.8)$$

The coefficients a, k and b, j are integers and determine respectively scaling and translation; $\delta_{a,b}$ is the Kroneker delta function. The coefficients $c_{a,b}$ in the expansion (3.8) are given by

$$c_{a,b} = \left\langle f(x), \psi_{a,b}(x) \right\rangle = 2^{a/2} \int_{-\infty}^{+\infty} f(x)\psi(2^a x - b)dx \qquad (3.9)$$

Equations (3.8) and (3.9) describe a **redundant**[1] **wavelet series expansion** of $f(x)$ with respect to the basic wavelet $\psi(x)$.

To understand the multiscale nature of the wavelet transform, one must introduce the concept of **scaling function** $\phi(x)$. The recursive use of this function permits one to represent a continuous function $f(x)$ at finer and finer scales, where at each scale the coefficients of its representation consist of a weighted sum of shifted and dilated versions of $\phi(x)$ [63, 89].

$$f(x) = \sum_{b=-\infty}^{+\infty} d_{A,b} 2^{A/2} \phi(2^A x - b) + \sum_{a=-\infty}^{A} \sum_{b=-\infty}^{+\infty} c_{a,b} 2^{a/2} \psi(2^a x - b)$$

with

$$c_{a,b} = 2^{a/2} \int_{-\infty}^{+\infty} f(x) \psi(2^a x - b) dx \quad \textit{wavelet coefficients} \tag{3.10}$$

$$d_{a,b} = 2^{a/2} \int_{-\infty}^{+\infty} f(x) \phi(2^a x - b) dx \quad \textit{scaling coefficients}$$

The first equation in (3.10) is the sum of two terms: the approximation of the function at scale A and the refinements at scales $a \leq A$. Wavelet and scaling coefficients can be recovered recursively

$$\begin{aligned} c_{a+1,b} &= \sum_{n=-\infty}^{+\infty} h_{2b-n} d_{a,n} \\ d_{a+1,b} &= \sum_{n=-\infty}^{+\infty} l_{2b-n} d_{a,n} \end{aligned} \tag{3.11}$$

[1] The continuous function $f(x)$ is represented by a doubly infinite sequence, and the transform is overcomplete. Since the basis function extends to infinity in both directions, a complete reconstruction must include all the terms.

As will be seen in subsequent sections, the functions l_n and h_n represent discrete time low-pass and high-pass filters respectively, and a tree filter bank including low-pass, high-pass filters and decimators derives from (3.11).

Now one restricts $f(x)$ and $\psi(x)$ to have compact support. Without loss of generality they may be assumed to be zero outside the interval [0 1] in which case a single index n defines the set.

$$\psi_n(x) = 2^{a/2}\psi(2^a x - b) \qquad (3.12)$$

where

$$\begin{aligned} &\forall\ n\ \ a \text{ is the maximum integer such that } 2^a \le n \\ &\text{and}\quad b\ =\ n\text{-}2^a \end{aligned} \qquad (3.13)$$

In this case the wavelet series expansion is described by the equations

$$\begin{aligned} f(x) &= \sum_{n=0}^{\infty} c_n \psi_n(x) \\ c_n &= \langle f(x), \psi_n(x) \rangle = 2^{a/2} \int_{-\infty}^{+\infty} f(x)\psi(2^a x - b)dx \\ d_n &= \langle f(x), \phi_n(x) \rangle = 2^{a/2} \int_{-\infty}^{+\infty} f(x)\phi(2^a x - b)dx \end{aligned} \qquad (3.14)$$

Equation (3.14) describes a **non-redundant wavelet series expansion**. The transform is no longer overcomplete and the redundancy has been eliminated[2]. The single index n is used to indicate the case of a non-redundant transform. When a finite number of elements of this series is able to approximate the function, the series can be truncated without introducing an appreciable approximation error.

A problem using orthonormal basis with compact support is the lack of symmetry of the basic wavelet function. This can be solved using two **biorthogonal** sets of basic wavelets $\{\psi_{a,b}(x)\}$ and $\{\tilde{\psi}_{k,j}(x)\}$, to decompose and reconstruct the signal, derived from two dual wavelet basic functions $\psi(x)$ and $\tilde{\psi}(x)$.

$$\left\langle \psi_{a,b}, \tilde{\psi}_{k,j} \right\rangle = \delta_{a,k}\delta_{b,j} \qquad (3.15)$$

This can be used to produce two symmetric sets of wavelet basic functions with compact support. In this case equations (3.14) becomes

$$\begin{aligned} c_n &= \left\langle f(x), \psi_n(x) \right\rangle = 2^{a/2} \int_{-\infty}^{+\infty} f(x)\psi(2^a x - b)dx \\ d_n &= \left\langle f(x), \tilde{\psi}_n(x) \right\rangle = 2^{a/2} \int_{-\infty}^{+\infty} f(x)\tilde{\psi}(2^a x - b)dx \end{aligned} \qquad (3.16)$$

for decomposition, and

$$f(x) = \sum_{n=0}^{+\infty} c_n \tilde{\psi}_n(x) = \sum_{n=0}^{+\infty} d_n \psi_n(x) \qquad (3.17)$$

for reconstruction.

3.3.3 Discrete Wavelet Transform (DWT)

The DWT is used to represent a uniformly sampled discrete function with N points $f(i)$ (i = 0...(N - 1)). When the set of basic wavelets is orthonormal the expression for DWT can be directly derived from equation (3.10)

[2] The function $f(x)$ is approximated by a single sum between 0 and ∞.

$$f(i) = \sum_{a=-\infty}^{+\infty} \sum_{b=-\infty}^{+\infty} c_{a,b} 2^{a/2} \psi(2^a i - b)$$

or

$$f(i) = \sum_{b=-\infty}^{+\infty} d_{A,b} 2^{A/2} \phi(2^A i - b) + \sum_{a=-\infty}^{A} \sum_{b=-\infty}^{+\infty} c_{a,b} 2^{a/2} \psi(2^a i - b) \qquad (3.18)$$

with

$$c_{a,b} = 2^{a/2} \sum_{i=0}^{N-1} f(i)\psi(2^a i - b) \quad \textit{wavelet coefficients}$$

$$d_{a,b} = 2^{a/2} \sum_{i=0}^{N-1} f(i)\phi(2^a i - b) \quad \textit{scaling coefficients}$$

If $f(i)$ and $\psi(i)$ are zero outside the interval [0 1], then the single index n can define the set as in (3.12), and the coefficients a and b are constrained by (3.13). In this case the transform will be non-redundant.

When the set of basis functions is biorthogonal the expression for the DWT is derived from (3.16) and (3.17)

$$\begin{aligned} c_n &= 2^{a/2} \sum_{i=0}^{N-1} f(i)\psi(2^a i - b) \\ d_n &= 2^{a/2} \sum_{i=0}^{N-1} f(i)\tilde{\psi}(2^a i - b) \end{aligned} \qquad (3.19)$$

for decomposition, and

$$f(i) = \sum_{n=0}^{+\infty} c_n \tilde{\psi}_n(i) = \sum_{n=0}^{+\infty} d_n \psi_n(i) \qquad (3.20)$$

for reconstruction. This section has shown how wavelet analysis theory allows the definition of a discrete, non-redundant, invertible transforms based on orthonormal and biorthogonal sets of basis functions. These, under specific hypothesis, can exactly recover the original signal.

3.3.4 Undecimated wavelet transform

The most relevant problem when applying the DWT is the absence of shift invariance. Equations (3.11) show that the coefficients are calculated for each scale of decomposition using a low-passed and scaled version of the coefficients at the previous level of decomposition[3]. The scaling operation is implemented in the filter bank by means of a sub-sampling (decimation) operator, which introduces shift-variance. A solution is to define an undecimated version of the discrete wavelet transform. It has been demonstrated [74] that one can define the undecimated version of the wavelet transform from the decimated version. An alternative form for (3.11) to express the decimated wavelet transform is

$$\begin{aligned} d_{a+1} &= \Lambda(l * d_a) \\ c_{a+1} &= \Lambda(h * d_a) \end{aligned} \qquad (3.21)$$

where Λ is the decimation operator, * is the discrete convolution, d_a is the input of the filter at the level a, c_a is the wavelet coefficient at the level a, h is the high-pass filter, and l is the low-pass filter used in (3.11). It is convenient to use the notation

$$[c^i]_j = c_{2^a, b2^a} \qquad (3.22)$$

to indicate the wavelet coefficients of the signal $f(i)$, with time dilatation 2^a and time translation $b2^a$. Let T_m be the operation of translation by m, so that

$$(T_m d)_j \equiv d_{j-m} \qquad (3.23)$$

It is possible to demonstrate [74] that the translation of the input signal (d^0) by $2^i m$ affects the wavelet coefficients according to

[3] Observe the presence of the factor 2 in the index of low and high pass digital filters

$$\left[c^i(T_{2^i m}d^0)\right]_j = \left[c^i(d^0)\right]_{j-m} \qquad (3.24)$$

In this equality d^0 is considered as the argument of c^i in order to include the dependency of the wavelet coefficient at level i on the input of the filter at level 0 (that is the original signal). The **undecimated discrete wavelet transform** is defined as

$$\tilde{c}_{i,j} \equiv \left[\tilde{c}^i(d^0)\right]_j \equiv \left[c^i(T_{-j}d^0)\right]_0 \qquad (3.25)$$

Using (3.24) and (3.25) results in

$$\tilde{c}_{i,2^i j} = c_{i,j} \qquad (3.26)$$

Sampling the undecimated discrete wavelet transform every 2^i samples produces the decimated discrete wavelet transform. The shift-invariance of the transform is demonstrated by

$$\begin{aligned}\left[\tilde{c}^i(T_m d^0)\right]_j &= \left[c^i(T_{-j}T_m d^0)\right]_0 \\ &= \left[c^i(T_{m-j}d^0)\right]_0 \\ &= \left[\tilde{c}^i(d^0)\right]_{m-j} = T_m\left[\tilde{c}^i(d^0)\right]_j\end{aligned} \qquad (3.27)$$

Defining the operator D such that, for every discrete filter h, $D^i h$ is h with 2^i-1 zeros inserted between every pair of filter coefficients, one can rewrite (3.21) in terms of the undecimated form as

$$\begin{aligned}d_{a+1} &= \left(D^i l\right) * d_a \\ c_{a+1} &= \left(D^i h\right) * d_a\end{aligned} \qquad (3.28)$$

In this case the transform will be redundant and every d_a has the same size as d_0 ($f(i)$). This equation will be used in implementing the undecimated tree filter bank.

3.4 Filter banks

To understand the relation between wavelet theory and filter banks, one must consider how the three main concepts related to wavelet analysis (shifting, scaling and multiresolution) are accomplished using filter banks [10]. Consider a signal, *f(x),* having the same properties as the signal considered in the case of non-redundant DWT. This signal is further constrained so that it is:

a) band-limited,

$$F(f) = 0 \quad \text{if } |f| > f_{\max} \tag{3.29}$$

b) sampled in a uniform way to form the discrete signal

$$\begin{aligned} &f(i) \quad i = 0,\ldots,(N-1) \\ &f_{\max} < f_N \,(\text{Nyquist frequency}) \end{aligned} \tag{3.30}$$

3.4.1 Shifting

The frequency axis can be partitioned into disjoint M intervals using a set of M discrete band-pass filters $\{H_j(z)\}$ ($j = 1\ldots M$). If the signal $f(i)$ is presented to each of the filters in parallel (figure 3-1), this effectively decomposes it and allows better identification of its spectral components.

The set of band-pass filters $\{H_j(z)\}$should be constructed so that its elements sum to 1 so that the signal components $g_j(k)$ will sum to form the original signal.

$$\sum_{j=1}^{M} H_j(z) = 1 \Rightarrow \sum_{j=1}^{M} g_j(k) = f(i) \tag{3.31}$$

The filter outputs can be written as:

$$g_j(k) = \sum_{i=0}^{N-1} f(i) h_j(k-i) \qquad (3.32)$$

Where $h_j(i)$ are the filter impulse response functions. Assuming the $H_j(z)$ are real and even, this convolution can be regarded as the inner product between the f and a shifted version of h_j.

The filter bank decomposition of a discrete signal $f(i)$ can be considered as the DWT of this signal using, as basis functions, the set made up of shifted versions of the impulse response function. Therefore, the sets of functions $\{g_j(k)\}$ and $\{h_j(i)\}$ can be respectively regarded as coefficients $c_{a,b}$ and functions $\psi_{a,b}(i)$ of a wavelet transformation.

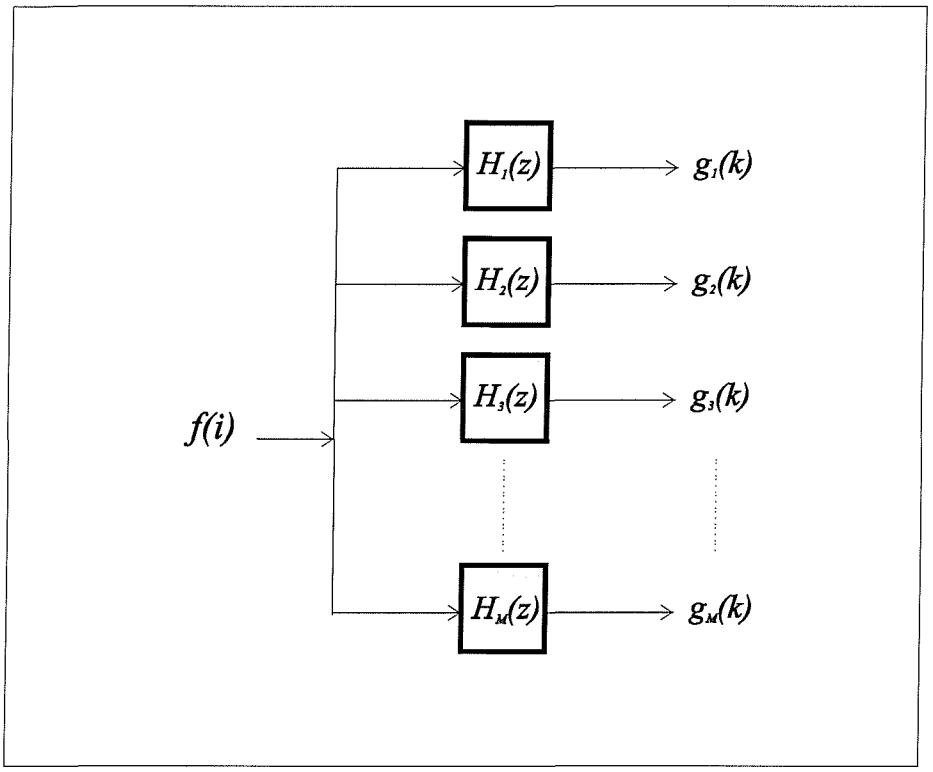


Figure 3-1. Filter bank decomposition.

3.4.2 Scaling

Consider now the case when $M = 2$. It is possible to decompose a signal into two band limited components and represent them without redundancy, in such a way that it is possible to reconstruct the original signal without error. Filtering the signal with half band high-pass and low-pass filters having discrete impulse responses $h_0(i)$ and $h_1(i)$, yields the discrete signals $g_0(k)$ and $g_1(k)$. Then

$$f(i) = g_0(k) + g_1(k) = h_0(i) * f(i) + h_1(i) * f(i) \qquad (3.33)$$

assuming

$$H_0(z) + H_1(z) = 1 \qquad (3.34)$$

The signals $g_0(k)$ and $g_1(k)$ can be sampled with sample spacing as large as $2\Delta x$ without introducing aliasing in the $g_0(k)$ case, and introducing a non-destructive (the original information can be fully recovered in reconstruction) aliasing in the $g_1(k)$ case [63]. This operation is called decimation or sub-sampling. Therefore, it is possible to recover the signal $f(i)$ from the sub-sampled version $g_0(k)$ and $g_1(k)$. There is an analogy between this filter bank and the wavelet transform expression in (3.11). Low-pass filters, having discrete frequency response $H_1(z)$, are used to obtain averages and their coefficients are related with the coefficients of l_n. High-pass filters, having discrete frequency response $H_0(z)$, are used to obtain detail and their coefficients are related with the coefficients h_n.

3.4.3 Perfect reconstruction

There are limitations on the choice of the low and high-pass filters. The condition of reconstruction without error has a fundamental importance in wavelet based filter bank design. There are four conditions describing the different levels of reconstruction reliability, only the most restrictive of which produces the complete recovery of the original signal (perfect reconstruction) and is considered here. In discrete frequency, the effect of sub-sampling and up-sampling is described in figure 3-2 and by the expression (3.35) [63].

$$\begin{aligned} X_1(z) &= A(z)X(z) \\ Y_1(z) &= \frac{1}{2}[X_1(z^{1/2}) + X_1(-z^{1/2})] = \frac{1}{2}[A(z^{1/2})X(z^{1/2}) + A(-z^{1/2})X(-z^{1/2})] \\ Y(z) &= Y_1(z^2) = \frac{1}{2}[A(z)X(z) + A(-z)X(-z)] \end{aligned} \quad (3.35)$$

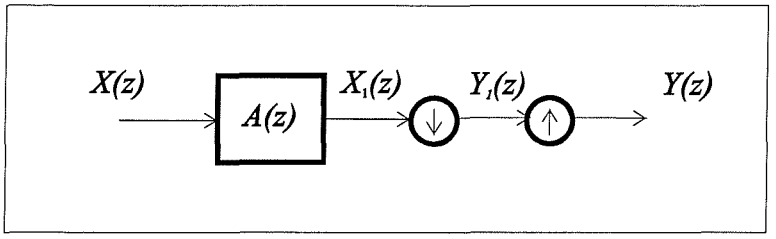


Figure 3-2. Effect of sub-sampling and up sampling.

For this the conditions for perfect reconstruction in the presence and absence of sub/up-sampling and for orthonormal and biorthogonal filter banks can be deduced.

a) In the discrete frequency domain in the absence of sub-sampling and up-sampling for an orthonormal filter bank (figure 3-3-a) the condition for perfect reconstruction is given by

$$\hat{F}(z) = G_0(z)H_0(z) + G_1(z)H_1(z) = \frac{1}{2}\{F(z)H_0(z)H_0(z) + F(z)H_1(z)H_1(z)\} =$$
$$= \frac{1}{2}\{F(z)[H^2{}_0(z) + H^2{}_1(z)]\} \qquad (3.36)$$
$$\begin{cases} \hat{F}(z) = F(z)z^{-p} \\ \text{Perfect reconstruction with a delay of } p \end{cases} \Rightarrow H_0^2(z) + H_1^2(z) = 2z^{-p} \text{ (no distortion)}$$

b) In the presence of sub-sampling and up-sampling for an orthonormal filter bank (figure 3-3-b) the condition for perfect reconstruction is given by

$$\hat{F}(z) = G_0{}'(z)H_0(z) + G_1{}'(z)H_1(z) =$$
$$= \frac{1}{2}H_0(z)[H_0(z)F(z) + H_0(-z)F(-z)] + \frac{1}{2}H_1(z)[H_1(z)F(z) + H_1(-z)F(-z)] =$$
$$= \frac{1}{2}\{F(z)[H^2{}_0(z) + H^2{}_1(z)] + F(-z)[H_0(z)H_0(-z) + H_1(z)H_1(-z)]\}$$
$$\begin{cases} \hat{F}(z) = F(z)z^{-p} \\ \text{Perfect reconstruction with a delay of } p \end{cases} \Rightarrow \begin{cases} H_0^2(z) + H_1^2(z) = 2z^{-p} & \text{(No distortion)} \\ H_0(z)H_0(-z) + H_1(z)H_1(-z) = 0 \text{ (No aliasing)} \end{cases}$$
(3.37)

In the case of biorthogonal filter bank, the corresponding filters in the analysis and synthesis banks are not anymore equal.

c) In the absence of sub-sampling and up-sampling for biorthogonal filter bank (figure 3-4-a) the condition for perfect reconstruction is given by

$$\hat{F}(z) = G_0(z)\tilde{H}_0(z) + G_1(z)\tilde{H}_1(z) = \frac{1}{2}\left\{F(z)H_0(z)\tilde{H}_0(z) + F(z)H_1(z)\tilde{H}_1(z)\right\} =$$

$$= \frac{1}{2}\left\{F(z)[H_0(z)\tilde{H}_0(z) + H_1(z)\tilde{H}_1(z)]\right\}$$

$$\begin{cases} \hat{F}(z) = F(z)z^{-p} \\ \text{Perfect reconstruction with a delay of } p \end{cases} \Rightarrow H_0(z)\tilde{H}_0(z) + H_1(z)\tilde{H}_1(z) = 2z^{-p} \text{ (no distortion)} \quad (3.38)$$

If we choose

$$\begin{aligned} \tilde{H}_0(z) &= H_1(-z) \\ \tilde{H}_1(z) &= -H_0(-z) \end{aligned} \quad (3.39)$$

and defining the product filter

$$P_0(z) = H_0(z)\tilde{H}_0(z) \quad (3.40)$$

the condition described in (3.38) becomes

$$\begin{cases} \hat{F}(z) = F(z)z^{-p} \\ \text{Perfect reconstruction with a delay of } p \end{cases} \Rightarrow P_0(z) - P_0(-z) = 2z^{-p} \text{ (no distortion)} \quad (3.41)$$

d) In the presence of sub-sampling and up-sampling for biorthogonal filter bank (figure 3-4-b) the condition for perfect reconstruction is given by

$$\hat{F}(z) = G_0'(z)H_0(z) + G_1'(z)H_1(z) =$$
$$= \frac{1}{2}H_0(z)[H_0(z)F(z) + H_0(-z)F(-z)] + \frac{1}{2}H_1(z)[H_1(z)F(z) + H_1(-z)F(-z)] =$$
$$= \frac{1}{2}\{F(z)[H^2{}_0(z) + H^2{}_1(z)] + F(-z)[H_0(z)H_0(-z) + H_1(z)H_1(-z)]\}$$
$$\begin{cases} \hat{F}(z) = F(z)z^{-p} \\ \text{Perfect reconstruction with a delay of } p \end{cases} \Rightarrow \begin{cases} H_0^2(z) + H_1^2(z) = 2z^{-p} & \text{(No distortion)} \\ H_0(z)H_0(-z) + H_1(z)H_1(-z) = 0 & \text{(No aliasing)} \end{cases}$$
(3.42)

Because conditions (3.39) are removing the aliasing the perfect reconstruction condition can be expressed also in this case as (3.41).

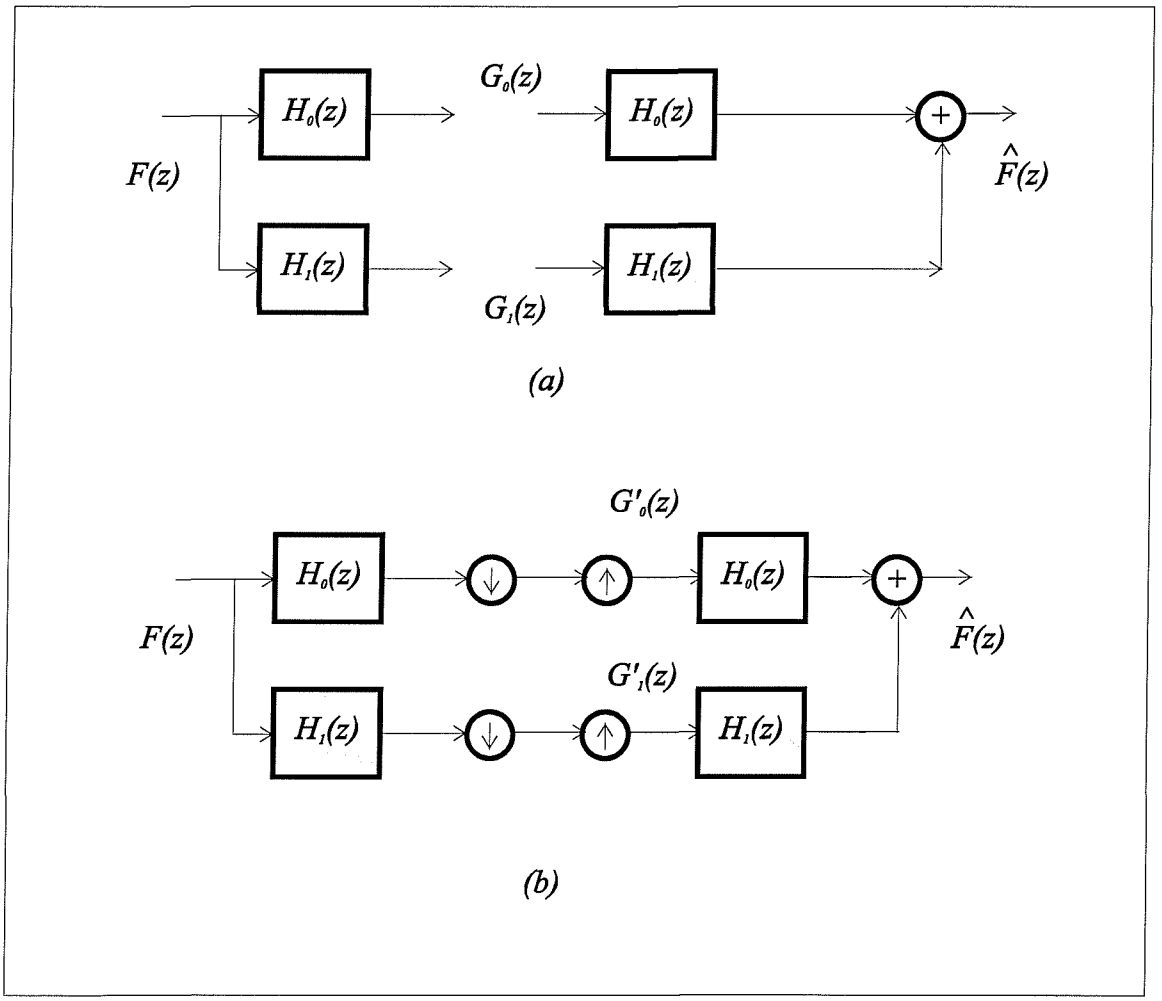


Figure 3-3. Two-band coding and reconstruction with orthonormal filter bank.

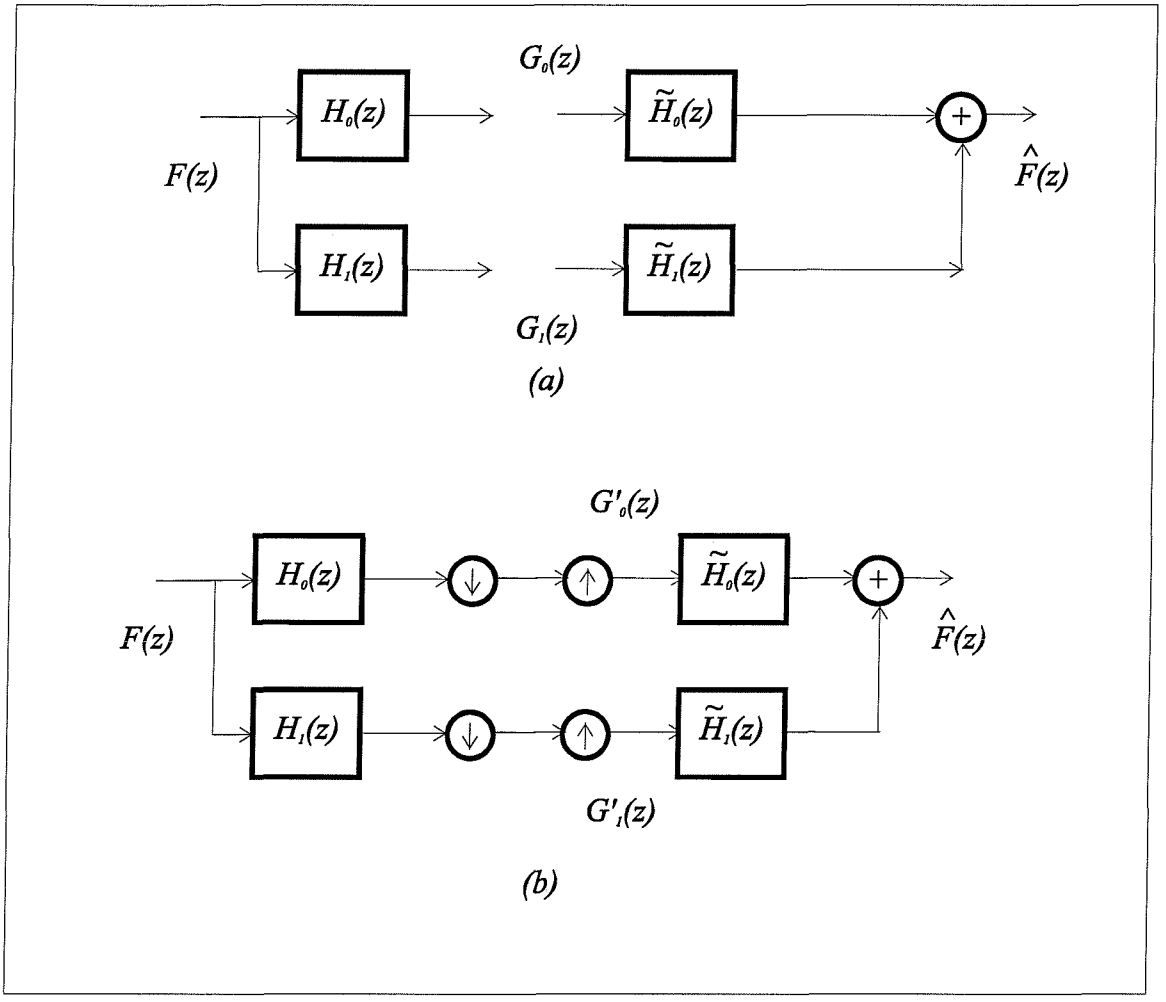


Figure 3-4. Two-band coding and reconstruction with biorthogonal filter bank.

3.4.4 Multiresolution

Figures 3.3 and 3.4 show single level filter banks in the presence and absence of sub-sampling. Consider the problem of implementing a recursive tree filter bank decomposing the signal *f(i)* in different levels of resolution. This type of filter bank represents the implementation of the recursive expressions seen for wavelet analysis. When the signal is sub-sampled, its size is halved at each level to keep unchanged the total amount of data allowing the absence of redundancy. The filters applied on the different levels are always the same. Figure 3-5 represents the tree filter biorthogonal bank in the presence of sub-sampling.

To implement a tree filter bank in the absence of sub-sampling the definition of undecimated wavelet transform (equation (3.28)) is used. This is implemented by inserting 2^a -1 zeros between the filter coefficients at every decomposition level a. This tree filter bank is redundant because at each level of decomposition the size of the signal is unchanged. On the other hand, the absence of decimation makes easier the design of the basic filters. Figure 3-6 represents a tree filter biorthogonal bank in the absence of sub-sampling. In this case the filters change on each level of decomposition and the discrete frequency responses are related by

$$\forall \text{decomposition level } a \begin{cases} H_{a1}(z) = [h_{1,1}, \underbrace{0...0}_{2^a-1}, h_{2,1}, \underbrace{0...0}_{2^a-1}, ..., h_{N_f-1,1}, \underbrace{0...0}_{2^a-1}, h_{N_f,1}] \\ H_{a0}(z) = [h_{1,0}, \underbrace{0...0}_{2^a-1}, h_{2,0}, \underbrace{0...0}_{2^a-1}, ..., h_{N_f-1,0}, \underbrace{0...0}_{2^a-1}, h_{N_f,0}] \\ \tilde{H}_{a1}(z) = [\tilde{h}_{1,1}, \underbrace{0...0}_{2^a-1}, \tilde{h}_{2,1}, \underbrace{0...0}_{2^a-1}, ..., \tilde{h}_{N_f-1,1}, \underbrace{0...0}_{2^a-1}, \tilde{h}_{N_f,1}] \\ \tilde{H}_{a0}(z) = [\tilde{h}_{1,0}, \underbrace{0...0}_{2^a-1}, \tilde{h}_{2,0}, \underbrace{0...0}_{2^a-1}, ..., \tilde{h}_{N_f-1,0}, \underbrace{0...0}_{2^a-1}, \tilde{h}_{N_f,0}] \end{cases} \quad (3.43)$$

with

$$H_{01}(z) = [h_{1,1}, h_{2,1}, ..., h_{N_f-1,1}, h_{N_f,1}]$$

$$H_{00}(z) = [h_{1,0}, h_{2,0}, ..., h_{N_f-1,0}, h_{N_f,0}]$$

$$\tilde{H}_{01}(z) = [\tilde{h}_{1,1}, \tilde{h}_{2,1}, ..., \tilde{h}_{N_f-1,1}, \tilde{h}_{N_f,1}]$$

$$\tilde{H}_{00}(z) = [\tilde{h}_{1,0}, \tilde{h}_{2,0}, ..., \tilde{h}_{N_f-1,0}, \tilde{h}_{N_f,0}]$$

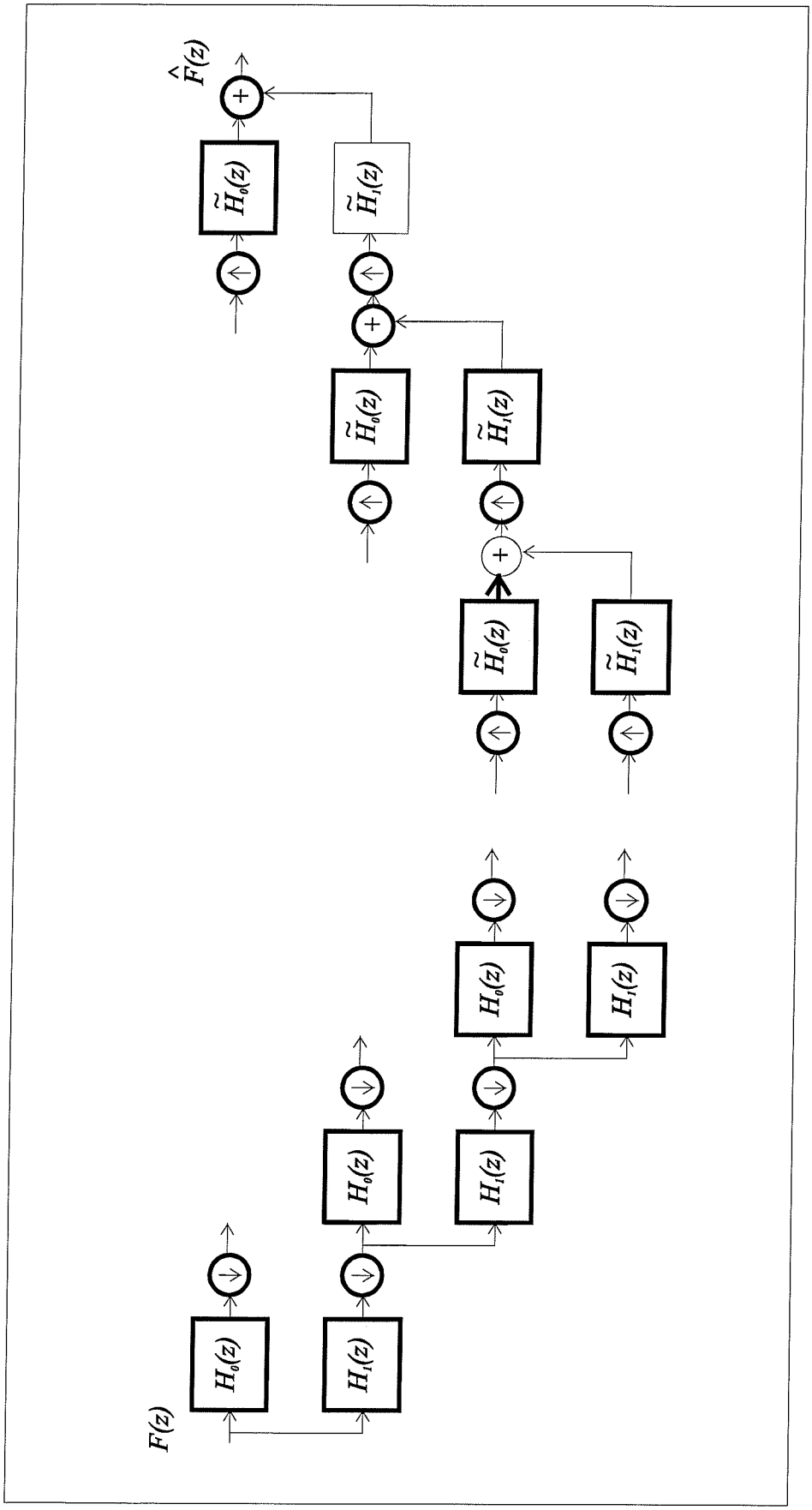


Figure 3-5. Tree filter bank with sub-sampling.

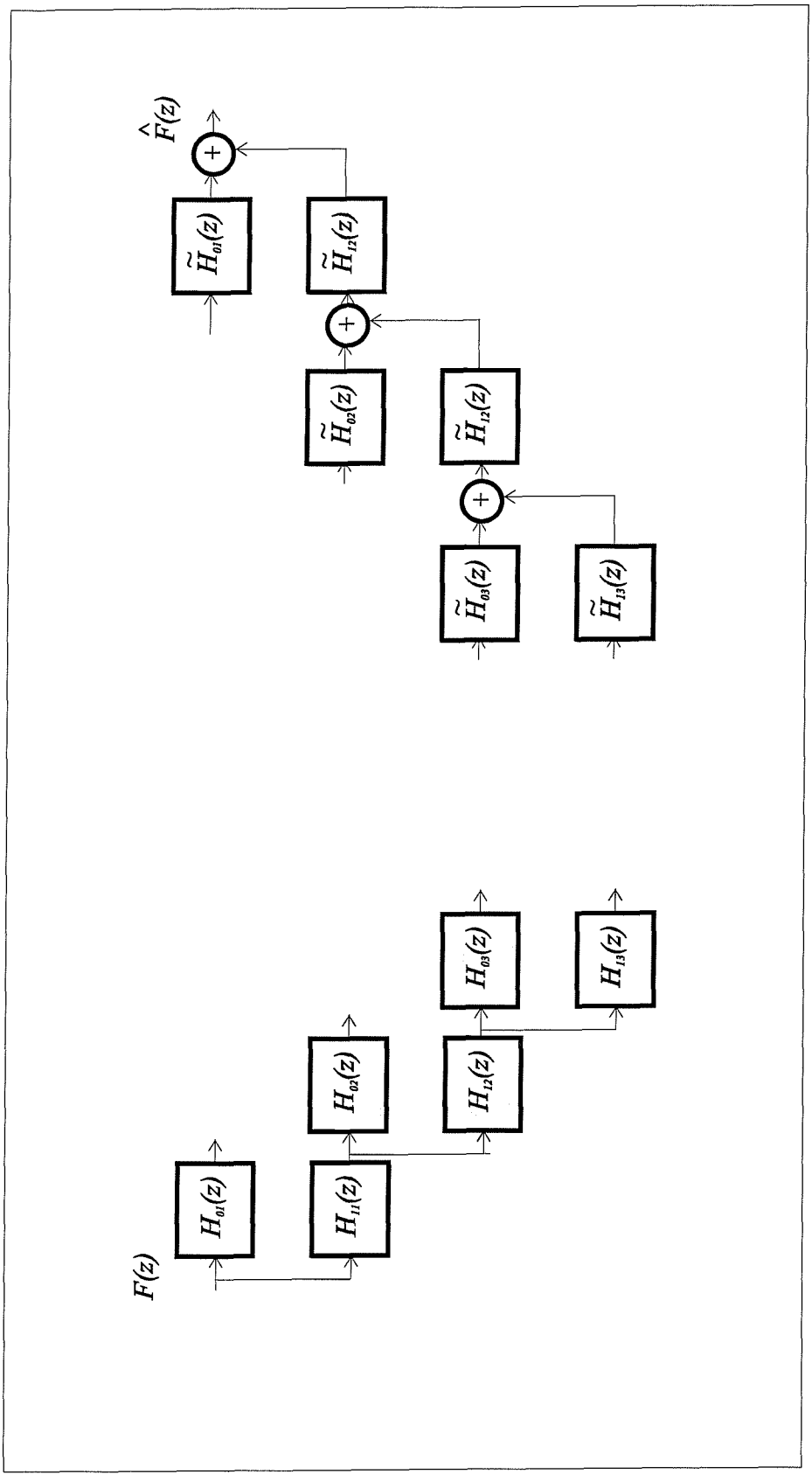


Figure 3-6. Tree filter bank without sub-sampling.

3.5 Wavelet analysis for image noise reduction

There has been a significant trend in recent research towards the use of multiscale approaches, such as wavelet analysis, for image processing. In particular, the use of the Discrete Wavelet Transform (DWT) has been shown to be useful for noise reduction. Most of the image energy is concentrated in a limited number of spectral components, whereas the noise energy is generally spread over the entire spectrum. Applying noise reduction techniques only on particular components can significantly reduce the noise energy whilst the signal energy remains relatively unaffected. Here only the case when the noise reduction is performed on the individual pixel values of the decomposed signal will be considered. The general procedure to reduce the noise using DWT approach consists of three steps:

a) the image is transformed into the wavelet domain by filtering and subsampling at each level (analysis),

b) a non-linear function is applied to the transformed coefficients in order to reduce the noise on each component, and

c) the inverse transform is applied to recover the image by filtering and upsampling at each level (synthesis).

The filter banks described for a one-dimensional signal in the previous section can be generalised for two-dimensional signals. Two-dimensional filter banks derived from separable two-dimensional wavelet bases are implemented by alternating the application of the $H_0(z)$ and $H_1(z)$ filters on rows (using $H_{0h}(z)$ and $H_{1h}(z)$) and columns (using $H_{0v}(z)$ and $H_{1v}(z)$) of the image. The analysis and synthesis filter banks must satisfy the perfect reconstruction condition. Figure 3-7 describes the undecimated biorthogonal filter bank for

three levels of decomposition. The filter bank is characterised by 24 filters, but the design is based only on one filter.

a) $H_{01v}(z), H_{11v}(z), H_{02v}(z), H_{12v}(z), H_{03v}(z), H_{13v}(z),$
$\tilde{H}_{01v}(z), \tilde{H}_{11v}(z), \tilde{H}_{02v}(z), \tilde{H}_{12v}(z), \tilde{H}_{03v}(z), \tilde{H}_{13v}(z),$

are the transposed versions[4] of

$H_{01h}(z), H_{11h}(z), H_{02h}(z), H_{12h}(z), H_{03h}(z), H_{13h}(z),$
$\tilde{H}_{01h}(z), \tilde{H}_{11h}(z), \tilde{H}_{02h}(z), \tilde{H}_{12h}(z), \tilde{H}_{03h}(z), \tilde{H}_{13h}(z),$

b) $H_{02v}(z), H_{12v}(z), H_{03v}(z), H_{13v}(z), \tilde{H}_{02v}(z), \tilde{H}_{12v}(z), \tilde{H}_{03v}(z), \tilde{H}_{13v}(z)$

are obtained by inserting zeros in

$H_{01v}(z), H_{11v}(z), \tilde{H}_{01v}(z), \tilde{H}_{11v}(z)$

c) $\tilde{H}_{01v}(z) = H_{11v}(-z)$ and $\tilde{H}_{11v}(z) = -H_{01v}(-z)$ *anti-aliasing condition*

d) $H_{01}(z)H_{11}(-z) - H_{01}(-z)H_{11}(z) = 2z^{-p}$ *anti distortion condition*

Finally the only filter to design is $H_{01}(z)$ (or $H_{11}(z)$, or even better the half-band product filter $H_{01}(z)\, H_{11}(-z)$).

Figure 3-8 describes the frequency decomposition obtained from this filter bank. At every level of decomposition, frequency bands are halved in the horizontal and in the vertical direction but the total size of the image components is unchanged. To understand the numbering system used for components in this figure, one should realise that the component 4 (LL) is decomposed in the components 5 (LLHH), 6 (LLHL), 7 (LLLH), and 8 (LLLL). The component 8 (LLLL) is decomposed again in the components 9 (LLLLHH), 10 (LLLLHL), 11 (LLLLLH), and 12 (LLLLLL).

[4] This means that they are the same filters applied on the image rotated by 90 degrees.

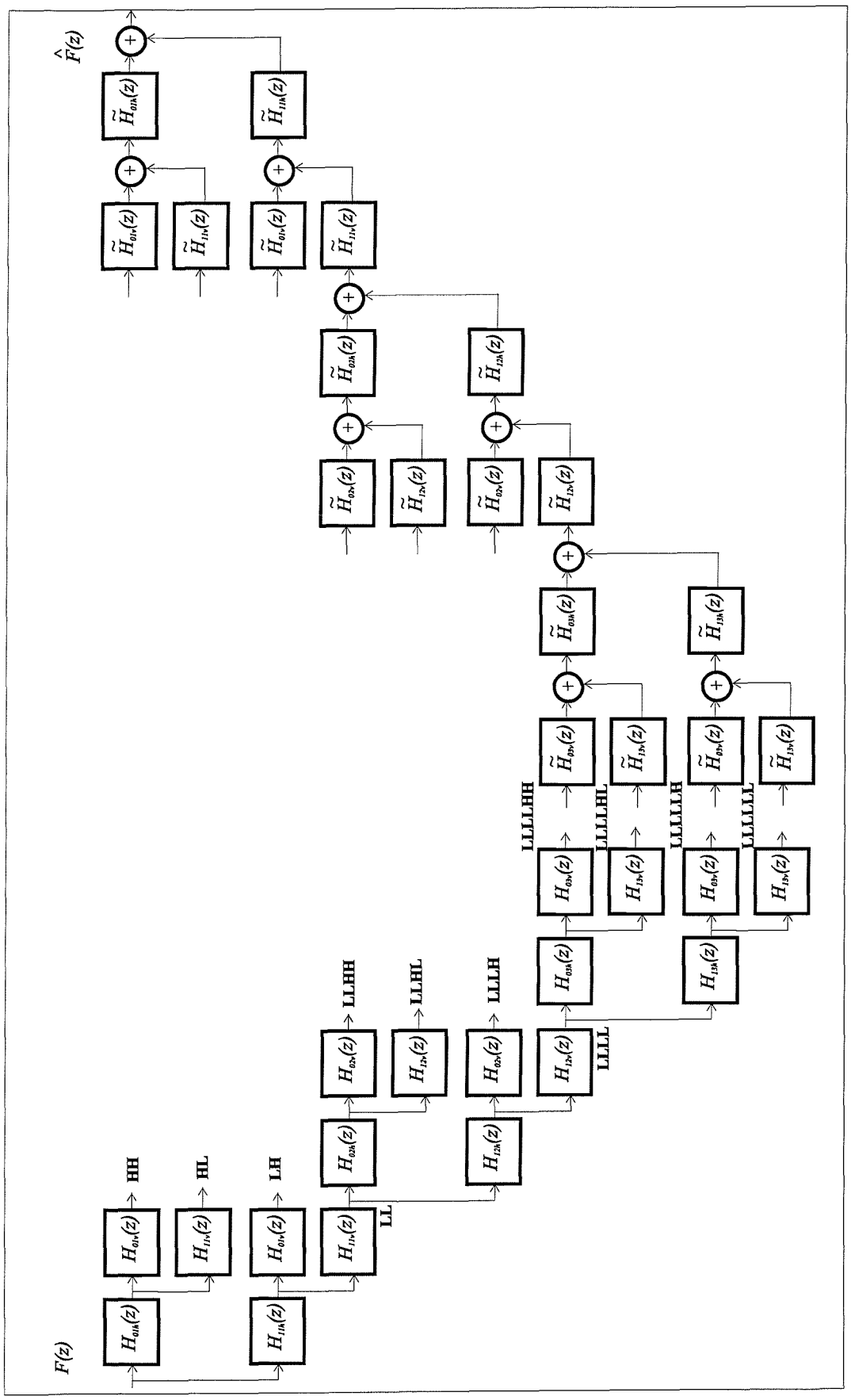


Figure 3-7. Tree filter bank without sub-sampling for image noise reduction.

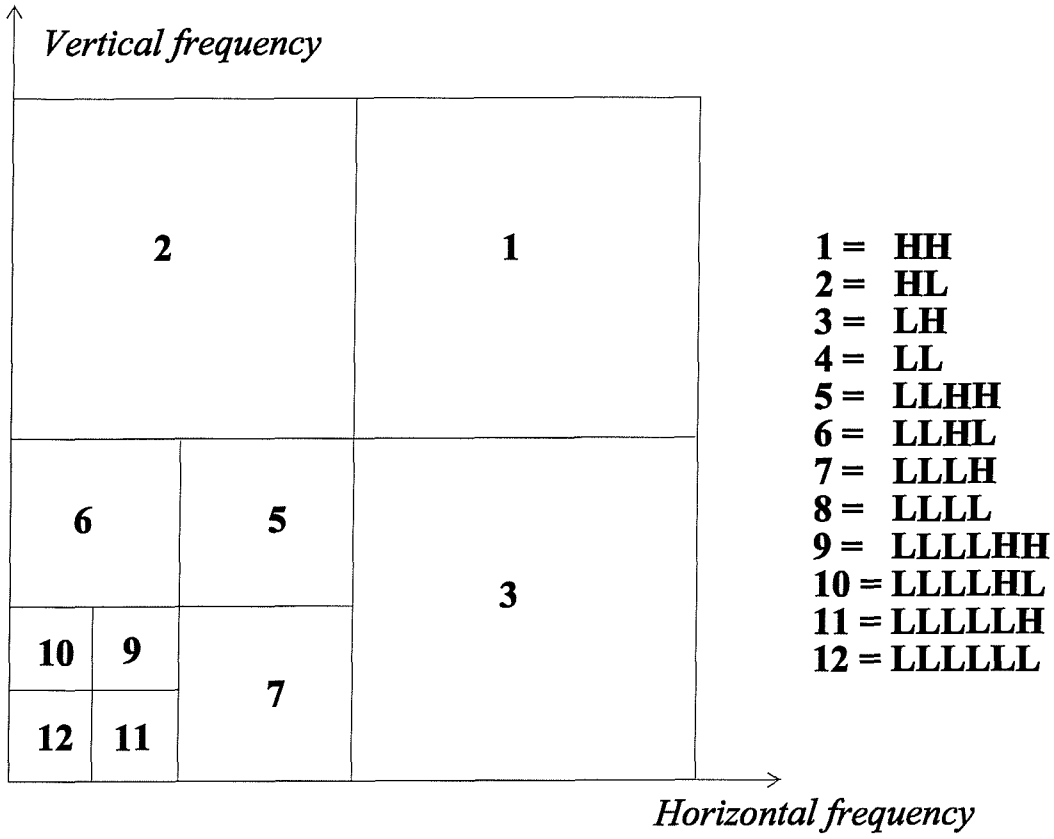


Figure 3-8. Ideal frequency decomposition.

3.6 The specifics of the wavelet transform

The filter bank used here is derived from the undecimated biorthogonal tree filter bank seen in figure 3.8. For a single level tree filter bank, the decomposition is defined by the biorthogonal set of filters[5]

$$H_0 = \frac{[-1,\ 2,\ 6,\ 2,\ -1]}{8} \qquad H_1 = \frac{[1,\ -2,\ 1]}{8}$$
$$\tilde{H}_0 = \frac{[1,\ 2,\ 1]}{4} \qquad \tilde{H}_1 = \frac{[1,\ 2,\ -6,\ 2,\ 1]}{4} \tag{3.44}$$

The absence of decimation allows the combination of the filter H_0 and H_1 respectively with $\tilde{H}_0$ and $\tilde{H}_1$ before the noise reduction procedure. The combined filters are:

$$LF1 = H_0 * \tilde{H}_0 = \frac{[-1,\ 0,\ 9,\ 16,\ 9,\ 0,\ -1]}{32}$$
$$HF1 = H_1 * \tilde{H}_1 = \frac{[1,\ 0,\ -9,\ 16,\ -9,\ 0,\ 1]}{32} \tag{3.45}$$

The image reconstruction procedure simply consists of a summation of the components. The filter in (3.45) can be approximated by

$$LF = \frac{[1,\ 2,\ 1]}{4}$$
$$HF = \frac{[-1,\ 2,\ -1]}{4} \tag{3.46}$$

Figures 3-9 compare the frequency responses of the filters in (3.45) and (3.46).

[5] This set is demonstrated in [63] to satisfy the perfect reconstruction condition.

The pair of filters, (3.46), allow perfect reconstruction, since they satisfy (3.36), and are an attractive solution in practice because of their simplicity. The use of the filters in (3.46) for the decomposition described in this thesis follows. The first level of decomposition is performed by applying *HF* and *LF* in the horizontal and vertical directions. The decomposition filters for the successive levels of decomposition are obtained, from the definition of undecimated wavelet transform, inserting 2^a - 1 (where a is the decomposition level) zeros between each filter coefficient. Reconstruction only requires the addition of individual components. The absence of decimation in this filter bank ensures the shift-invariance and that each component is overcomplete and has same size as that of the original image. Figure 3-10 describes the filter bank used. The applied filters are described by

$$h_{11h} = \frac{[1,\ 2,\ 1]}{4} \quad h_{01h} = \frac{[-1,\ 2,\ -1]}{4} \quad h_{11v} = h_{11h}^{\ T} \quad h_{01v} = h_{01h}^{\ T}$$

$$h_{12h} = \frac{[1,\ 0,\ 2,\ 0,1]}{4} \quad h_{02h} = \frac{[-1,\ 0,2,\ 0,\ -1]}{4} \quad h_{12v} = h_{12h}^{\ T} \quad h_{02v} = h_{02h}^{\ T}$$

$$h_{13h} = \frac{[1,\ 0,0,0,\ 2,\ 0,0,0,1]}{4} \quad h_{03h} = \frac{[-1,\ 0,0,0,2,\ 0,0,0,\ -1]}{4} \quad h_{13v} = h_{13h}^{\ T} \quad h_{03v} = h_{03h}^{\ T}$$

(3.47)

The filters used are non-orthogonal, therefore the components are correlated and this will influence the training procedure. The filters applied in the lowest decomposition level (LLLLHH, LLLLHL, and LLLLLH) have some non-zero areas outside the lowest frequency band. This means that the components produced by filtering the original image with these filters have a small influence on the higher frequency components of the image. A method to visualise this is to decompose an image where the spatial position is related to the frequency (a zone plate). A circular zone plate is shown in figure 3-11, and figure 3-12

shows the two-dimensional frequency decomposition produced by the filter bank. Figure 3-12 emphasises the presence of non-zero areas on each component outside the related frequency band.

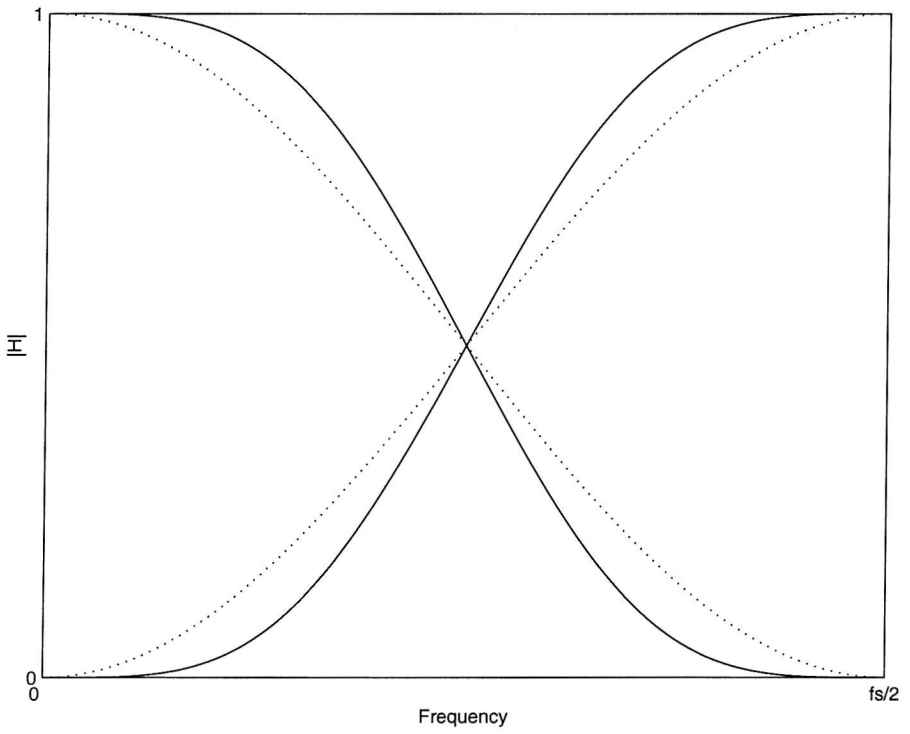


Figure 3-9. Filter approximation. Continuous lines are the frequency responses of the filters described by (3.45) and dotted lines are the frequency responses of the filters described by (3.46).

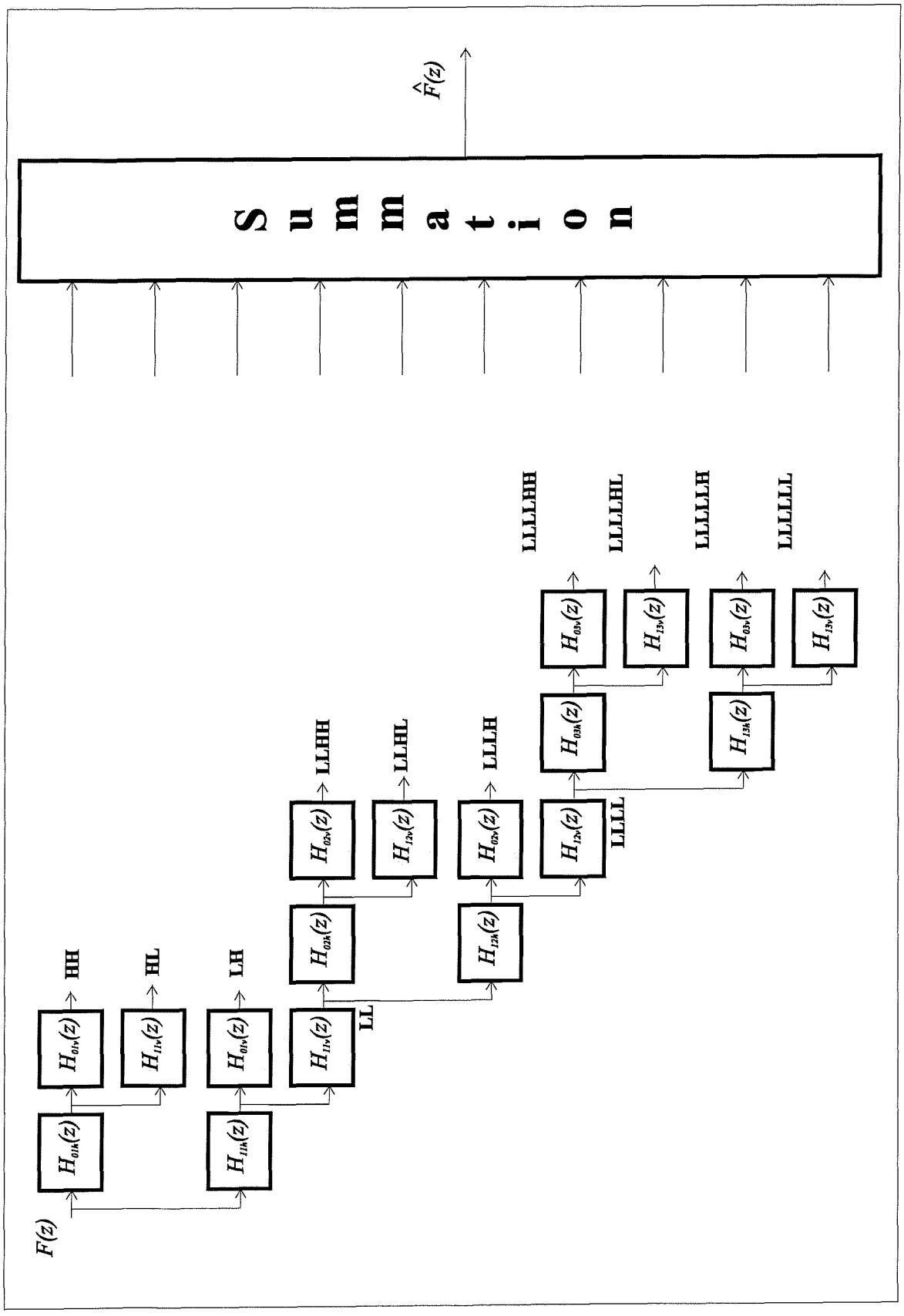


Figure 3-10. Tree filter bank used in our approach.

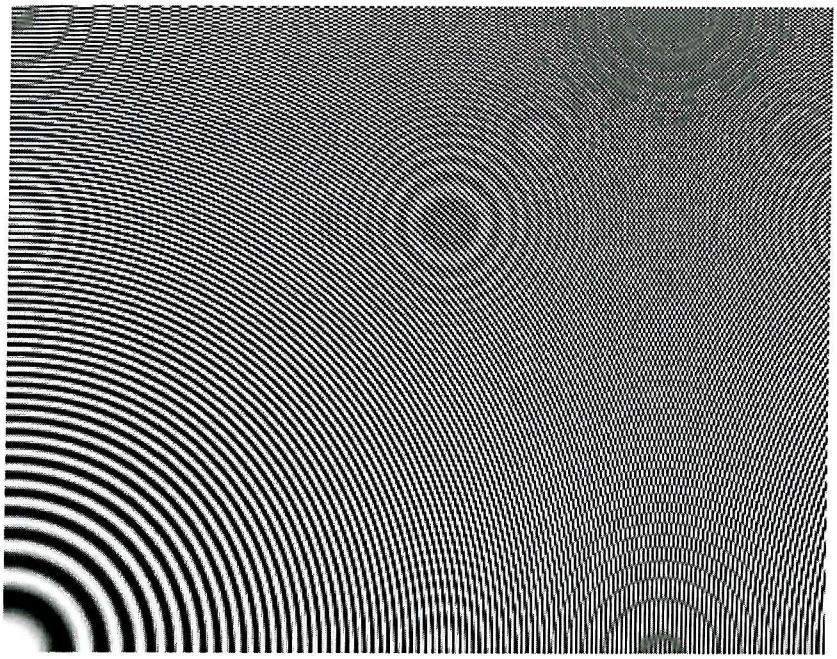


Figure 3-11. Image Plate.

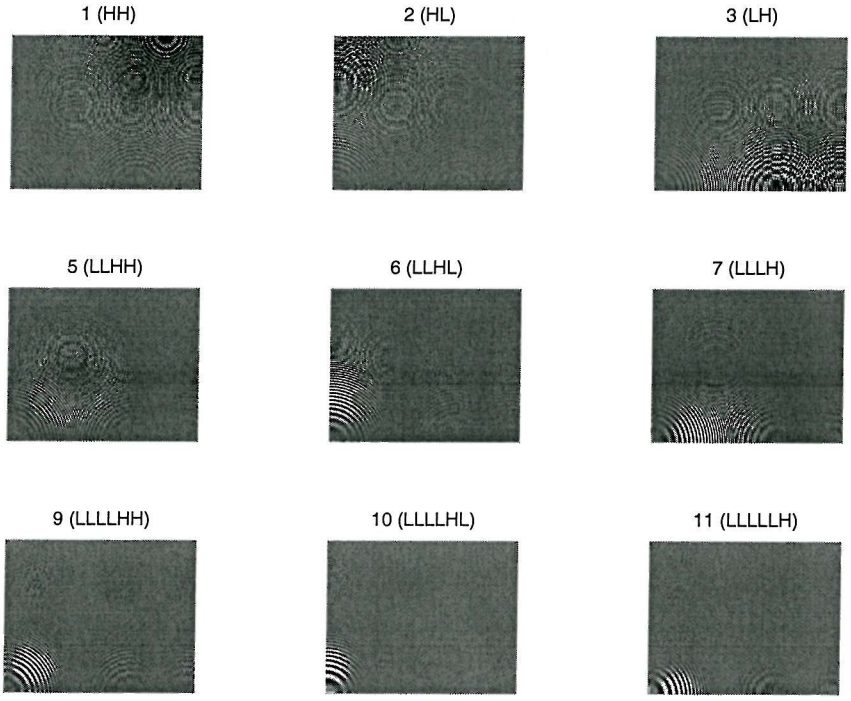


Figure 3-12. Two-dimensional frequency decomposition of the zone plate.

3.7 Conclusions

Wavelet analysis provides a method to define a transform in a suitable domain for image noise reduction. Wavelet analysis theory naturally leads to the implementation of filter banks. This chapter has described an implementation of a tree filter bank derived from a biorthogonal set of basis functions. This filter bank has the appearance of a scheme for sub-band coding and it does produce perfect reconstruction. The components obtained from the decomposition are undecimated and consequently have the same size as the original image. From the point of view of the transmission, the disadvantage of the increase of data is partially compensated by the constant size of the components because this permits the use of uniform transmission bandwidth. Another drawback of this solution is the increase of required computations due to the redundant scheme. This is partially compensated by the absence of filtering in reconstruction and by simplicity of the filters used in decomposition. Because of the absence of aliasing, these filters are constrained only by the non-distortion condition.

Chapter 4

Thresholding function shape

4.1 Introduction

The previous chapter described the tree filter bank used in order to decompose an image into components. This and subsequent chapters examine the procedure of noise reduction on the components derived from the decomposition. The noise reduction approach adopted is based on the analysis of the pixel values of the components and on the application of thresholding functions to them. Two elements characterise the thresholding functions: shape and parameters. This chapter explores the origin of the family of functions used, describing the shape of the family of thresholding functions to be applied. The next chapter will estimate the parameters defining a member of the family for every component. In this chapter initially optimal schemes are derived using Bayesian estimation in the Maximum a Posteriori (MAP), Maximum Likelihood (ML), and Least Mean Square Error (LMSE) senses. The components of the original (noise free) image are assumed to have either a Laplacian or a generalised Gaussian homogeneous probability distribution. The noise is assumed to be additive and its components are assumed to have Gaussian probability distribution. Finally a parameterised family of piecewise linear functions is defined that approximates the optimal families.

4.2 Families of thresholding functions

The components derived from the noisy image decomposition represent the energetic partition on frequency bands of the image contaminated by noise. These components, in every band, are dependent on both the components of the original image and on the noise. In general the energy of the original images is very unevenly distributed over the entire frequency range. Conversely the noise energy normally covers the entire range (although not necessarily uniformly). There generally exist frequency bands which predominantly contain noise. Eliminating or reducing the components related to these bands produces reduction of the noise energy with only a small reduction of the energy of the original image energy. The noise reduction procedure consists of applying non-linearities, called thresholding functions, to the components. It is convenient to define a parameterised family of functions (sometime called a thresholding scheme) including all the possible functions that can be applied to the components. The identification of the precise thresholding function to be applied on each component consists of the selection of one or more parameters.

4.3 Thresholding functions on the components

To understand the advantage of applying different thresholding functions on different components, consider initially the case when a thresholding function is directly applied on the complete image without decomposition. In this case, a given image grey level $u \in [-L\ L]$ is mapped into another grey level $v \in [-L\ L]$ according to a function

$$v = f(u) \qquad (4.1)$$

Applying the thresholding directly to the image results in little noise reduction and introduces significant image distortion. The cause of this can be seen from observing figure 4.1. Applying a thresholding function to an image results in the reduction of the small values of its histogram. In figure 4-1 the normalised histograms (probability density functions (PDFs)) of original and noisy images are compared for 13 images[1]. To discern noise from original image using the complete non decomposed images is complicated. Reducing the small values in the PDFs will remove a small amount of noise and will also reduce the energy of the original image.

Conversely, when the image is decomposed, these methods have been productively applied on components. Using the same set of 13 images, figures 4-2 compares the PDFs of the original and noisy components related to the high frequency band. For this component the noise contamination is clearly discernible for most of the images and thresholding can be an effective method to reduce noise without unduly distorting the original image.

[1] The set of 13 images has been provided by BBC and describes the class of video images. The images belonging to this set are shown in appendix C.

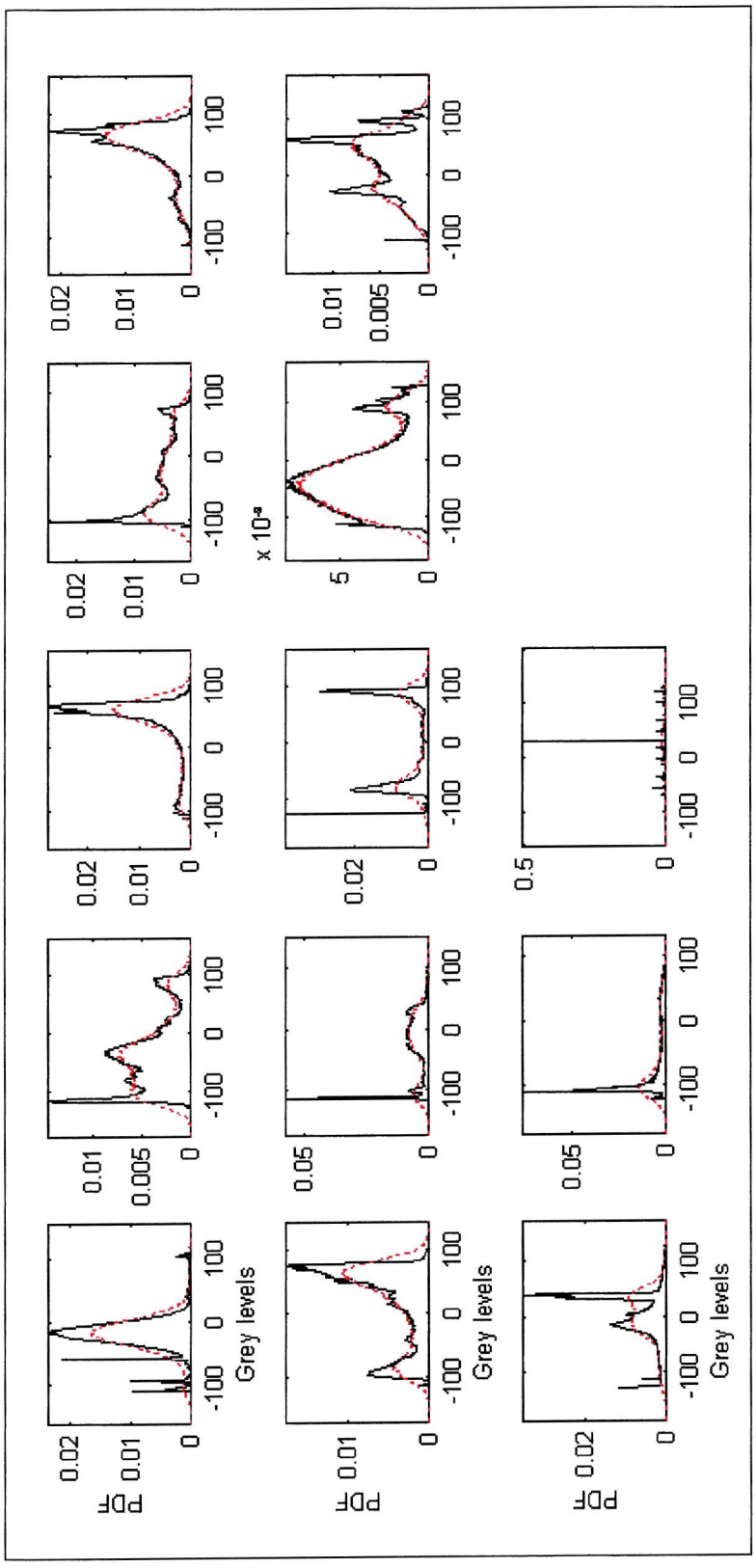


Figure 4-1. PDFs of original (black continuous lines) and noisy (red dotted lines) images.

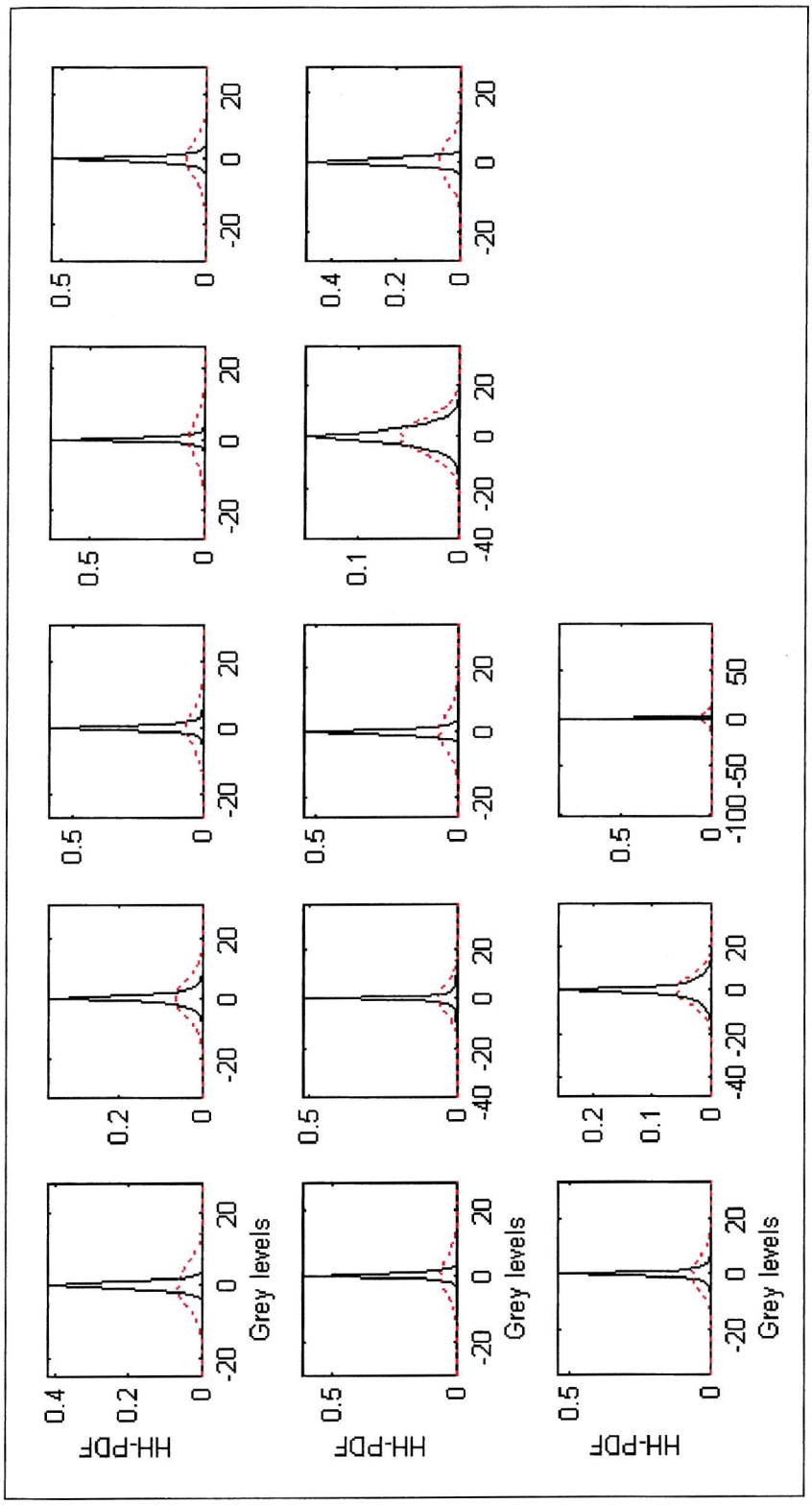


Figure 4-2. PDFs of high frequency components of original (black continuous lines) and noisy (red dotted lines) images.

4.4 Thresholding schemes

The shape of the thresholding function critically affects the efficiency of the noise reduction algorithm. Initially consider the case of a signal contaminated by additive Gaussian noise. Given an original signal $u(x)$, and an observation of this $v(x)$, after it has been contaminated by the noise $n(x)$, one seeks to find the optimal estimate $\hat{u}(x)$ of the original signal. This can be formulated as follows

$$v(x) = u(x) + n(x) \qquad (4.2)$$

The thresholding scheme will have the form

$$\hat{u}(x) = f[v(x)] \qquad (4.3)$$

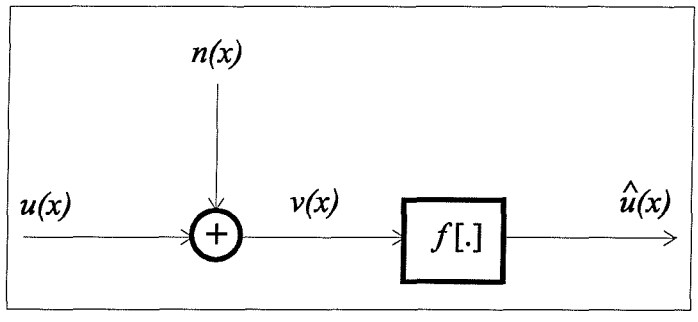


Figure 4-3. Signal estimation in presence of additive noise.

The objective is to force the estimate $\hat{u}(x)$ to approximate $u(x)$ in some optimal fashion.

4.4.1 Soft and hard thresholding

The two predominant thresholding schemes used in the literature [65, 81-87] are the soft-thresholding and the hard-thresholding. They are described by

$$\hat{u} = \begin{cases} v & \text{if } |v| > K \\ 0 & \text{if } |v| \leq K \end{cases} \quad \text{Hard - thresholding}$$

$$\hat{u} = \begin{cases} v - K & \text{if } v > K \\ 0 & \text{if } -K \leq v \leq K \\ v + K & \text{if } v < -K \end{cases} \quad \text{Soft - thresholding} \tag{4.4}$$

These thresholding functions are depicted in figure 4-4

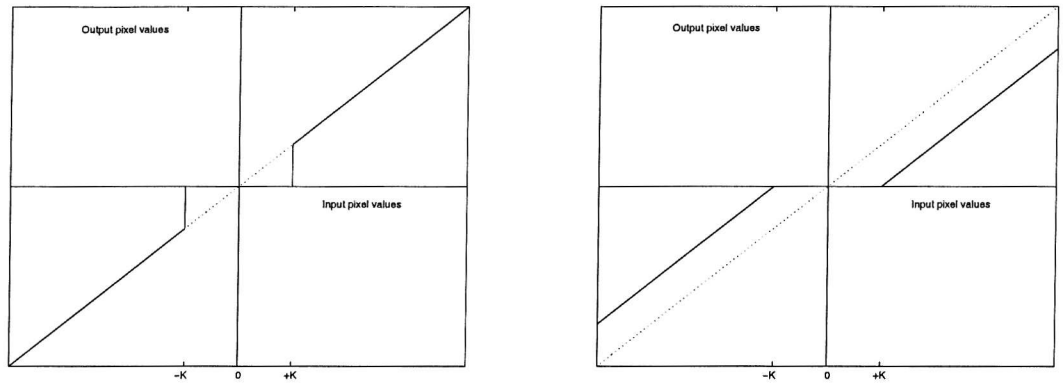


Figure 4-4. Hard and soft thresholding schemes.

The parameter K is called threshold level, and it strongly influences the performance of the thresholding scheme. To select an appropriate value for K one requires knowledge of the noise level (noise standard deviation). The selection of threshold levels, or more generally set of parameters defining the thresholding functions, is the subject of the next chapter. This chapter concentrates on defining the shape of the thresholding functions.

4.4.2 Optimal Bayesian thresholding schemes

Assuming the original signal has a specific distribution and using Bayesian theory it is possible to develop expressions for the optimal scheme (4.3) to apply in order to reduce noise having Gaussian distribution.

If $p(u)$, $p(v)$, and $p(n)$ are respectively the probability distributions of original signal u, observed signal v, and noise n, the Bayes rule states that

$$p(u/v) = \frac{p(v/u)p(u)}{p(v)} \qquad (4.5)$$

- $p(u/v)$ is the **posterior** probability density function of u given v
- $p(v/u)$ is the **likelihood** function
- $p(u)$ is the **prior,** and
- $p(v)$ is the **evidence**.

Bayesian [107] estimation aims to obtain an estimate $\hat{u}$ of u from the observation v by minimising a **risk function** with respect to $\hat{u}$.

$$R(\hat{u}) = E[C(\hat{u},u)] = \iint C(\hat{u},u)p(u/v)p(v)dvdu \qquad (4.6)$$

Given that the evidence, $p(v)$, is constant for a given observation, then the general expression of the Bayesian estimate is

$$\hat{u}_{BAYES} = Arg\ \underset{\hat{u}}{\text{Min}}\left(\int C(\hat{u},u)p(u/v)du\right) \qquad (4.7)$$

$C(\hat{u},u)$ is the **cost function** characterising the estimator. Examples are:

1. Maximum a Posteriori Estimation (**MAP**). In this case

$$\begin{aligned} C(\hat{u},u) &= 1-\delta(\hat{u},u) \\ \hat{u}_{MAP} &= Arg\ \underset{\hat{u}}{\text{Max}}\left(p(\hat{u}/v)\right) \end{aligned} \qquad (4.8)$$

which is equivalent to finding the maximum of $p(\hat{u}/v)$, i.e. solving

$$\frac{\partial p(\hat{u}/v)}{\partial \hat{u}} = 0$$
$$\text{or} \qquad (4.9)$$
$$\frac{\partial \log\{p(\hat{u}/v)\}}{\partial \hat{u}} = 0$$

2. Maximum Likelihood (**ML**). In this case

$$C(\hat{u},u) = 1 - \delta(\hat{u},u) \text{ and } p(u) = const.$$
$$\hat{u}_{ML} = Arg \underset{\hat{u}}{\text{Min}}\left(p(v/\hat{u})\right) \qquad (4.10)$$

which is equivalent to maximising $p(v/\hat{u})$, i.e. solving

$$\frac{\partial p(v/\hat{u})}{\partial \hat{u}} = 0$$
$$\text{or} \qquad (4.11)$$
$$\frac{\partial \log\{p(v/\hat{u})\}}{\partial \hat{u}} = 0$$

3. Least Mean Squares Error (**LMSE**). In this case

$$C(\hat{u},u) = [(\hat{u}-u)(\hat{u}-u)^T]$$
$$\hat{u}_{LMSE} = \int up(u/v)du \qquad (4.12)$$

This is the conditional mean of the posterior probability. The quality index used in noise reduction is the mean square error or its weighted version. Consequently this estimator has a particular relevance in this analysis.

4.4.3 Bayesian estimator for particular distributions

Assume that *n(x)* is Additive White Gaussian Noise (AWGN), with zero mean, having a distribution

$$p(n) = \frac{1}{\sigma_\eta \sqrt{2\pi}} e^{-\frac{n^2}{2\sigma_\eta^2}} \qquad (4.13)$$

where σ_η is the standard deviation of the noise or the noise level. The following details the solution to these optimisations for various assumed distributions for the image.

4.4.3.a Laplacian distribution

The image distribution is given by

$$p(u) = \frac{\sqrt{2}}{\sigma_u} e^{-\frac{\sqrt{2}}{\sigma_u}|u|} \qquad (4.14)$$

where σ_u is the spectral content of the original signal[2]. Since the noise is Gaussian, *p(v/u)* can be expressed as

$$p(v/u) = \frac{1}{\sigma_\eta \sqrt{2\pi}} e^{-\frac{(v-u)^2}{2\sigma_\eta^2}} \qquad (4.15)$$

The optimal MAP estimator can be obtained by solving (4.9). In Appendix B it is shown that the optimal scheme is given by

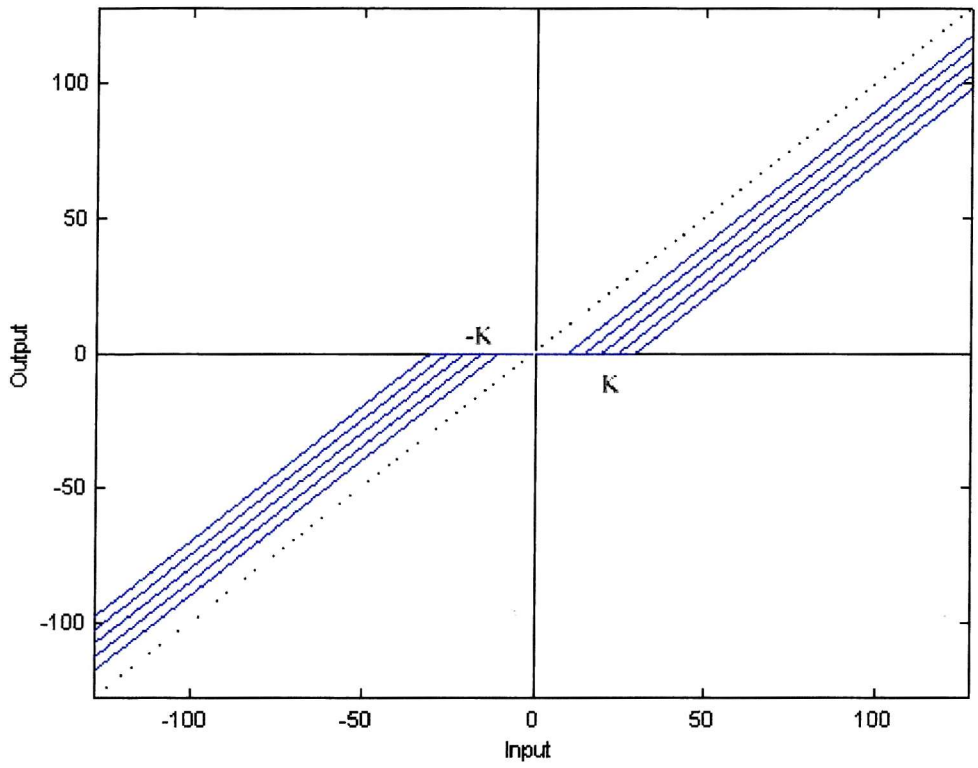
$$\hat{u}_{\substack{MAP \\ Laplacian}} = \begin{cases} v - \dfrac{\sigma_\eta^2 \sqrt{2}}{\sigma_u} & \text{if } v > 0 \\ 0 & \text{if } v = 0 \\ v + \dfrac{\sigma_\eta^2 \sqrt{2}}{\sigma_u} & \text{if } v < 0 \end{cases} \qquad (4.16)$$

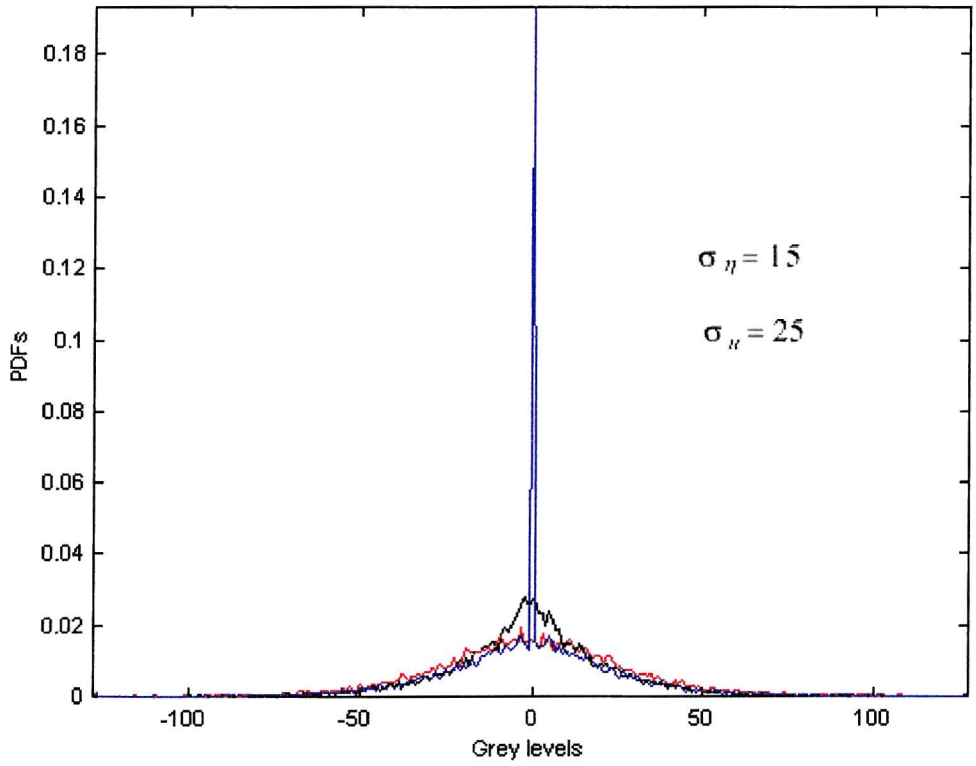
When the original signal has a Laplacian distribution the MAP optimal scheme has the shape of the soft thresholding scheme with threshold level

[2] In compression wavelet coefficients are often assumed Laplacian

$$K = \frac{\sigma_\eta^2 \sqrt{2}}{\sigma_u} \qquad (4.17)$$

This demonstrates that when the original signal has Laplacian distribution the soft-thresholding scheme is the optimal scheme in the MAP sense. Figure 4-5 describes this scheme and the effect on the PDFs of the scheme on a synthetic Laplacian image with σ_u=25 and σ_η=15.





Figures 4-5. MAP optimal scheme with original signal having Laplacian distribution, and effect on the PDFs of using this thresholding (black is the original, red is the noisy, and blue is the thresholded).

The optimal ML estimator can be obtained by solving (4.11). In appendix B it is demonstrated that the optimal scheme is given by

$$\hat{u}_{ML} = v \qquad (4.18)$$

Since such a scheme leaves the image unaltered it is of no use for a noise reduction.

The optimal LMSE estimator can be found by solving (4.12). The posterior distribution is computed using (4.5). The evidence is computed as

$$p(v) = \int_{-\infty}^{+\infty} p(u-v)p(u)du \qquad (4.19)$$

$p(u)$ is given by (4.13). Appendix B shows that the resulting scheme is

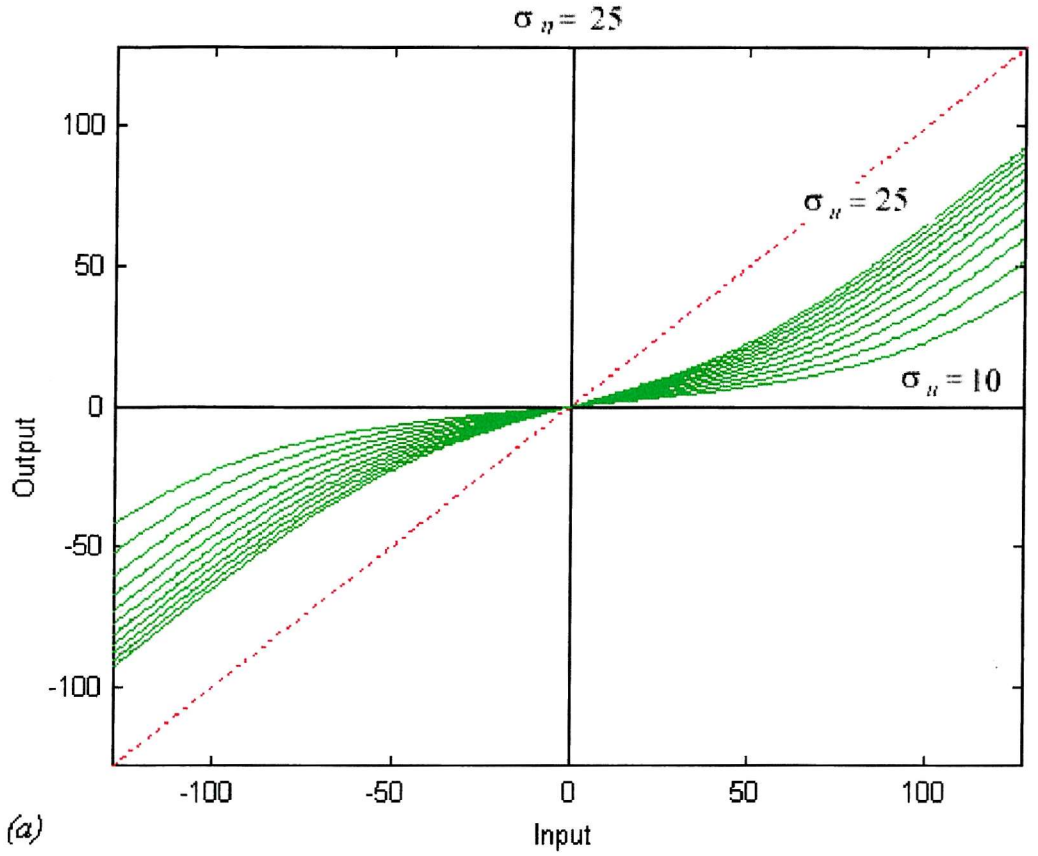
$$\hat{u}_{\substack{LMSE \\ Laplacian}} = \sigma_\eta \sqrt{2} \frac{\left\{ \left(\frac{\sigma_\eta}{\sigma_u} + \frac{v}{\sigma_\eta \sqrt{2}} \right) e^{\frac{\sqrt{2}}{\sigma_u} v} \, erfc\left(\frac{\sigma_\eta}{\sigma_u} + \frac{v}{\sigma_\eta \sqrt{2}} \right) - \left(\frac{\sigma_\eta}{\sigma_u} - \frac{v}{\sigma_\eta \sqrt{2}} \right) e^{-\frac{\sqrt{2}}{\sigma_u} v} \, erfc\left(\frac{\sigma_\eta}{\sigma_u} - \frac{v}{\sigma_\eta \sqrt{2}} \right) \right\}}{\left\{ e^{\frac{\sqrt{2}}{\sigma_u} v} \, erfc\left(\frac{\sigma_\eta}{\sigma_u} + \frac{v}{\sigma_\eta \sqrt{2}} \right) + e^{-\frac{\sqrt{2}}{\sigma_u} v} \, erfc\left(\frac{\sigma_\eta}{\sigma_u} - \frac{v}{\sigma_\eta \sqrt{2}} \right) \right\}} \qquad (4.20)$$

The LMSE optimal scheme depends on two parameters:

a) signal level σ_u, and

b) noise level σ_η.

Figure 4-6 describes the LMSE optimal scheme under these conditions. Figure 4-7 describes the effect on the PDFs of applying the LMSE optimal scheme on synthetic Laplacian image with σ_u=25 and σ_η=15.



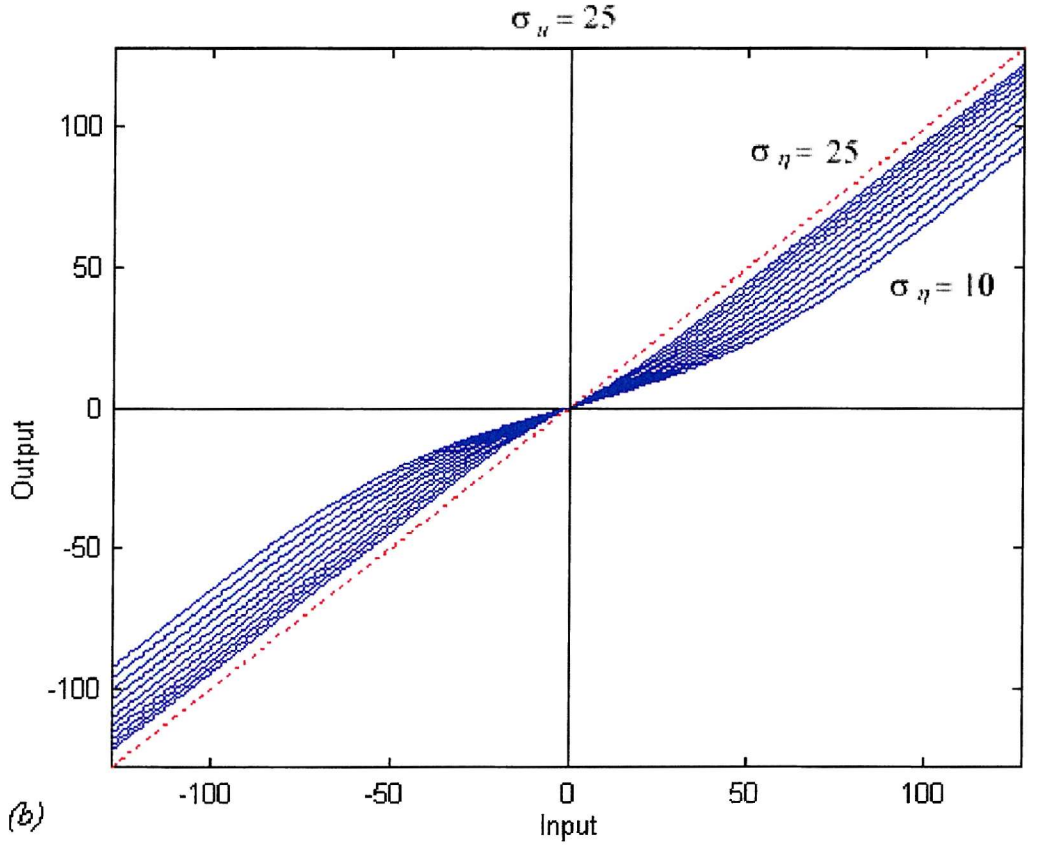


Figure 4-6. LMSE optimal thresholding scheme for a Laplacian signal in Gaussian noise; a)σ_η fixed (25), σ_u varies; b) σ_u fixed (25), σ_η varies.

4.4.3.b Generalised Gaussian distribution

Consider now the case when the signal $u(x)$ is assumed to have generalised Gaussian distribution, with a PDF of the form [107]:

$$p(u,\sigma_u,\beta_u)=\frac{\omega(\beta_u)}{\sigma_u}e^{-c(\beta_u)\left|\frac{u}{\sigma_u}\right|^{\frac{2}{1+\beta_u}}} \qquad -\infty<u<\infty$$

$$\omega(\beta_u)=\frac{\left\{\Gamma\left[\frac{3}{2}(1+\beta_u)\right]\right\}^{\frac{1}{2}}}{(1+\beta_u)\left\{\Gamma\left[\frac{1}{2}(1+\beta_u)\right]\right\}^{\frac{3}{2}}} \qquad c(\beta_u)=\left\{\frac{\Gamma\left[\frac{3}{2}(1+\beta_u)\right]}{\Gamma\left[\frac{1}{2}(1+\beta_u)\right]}\right\}^{\frac{1}{1+\beta_u}} \qquad -1<\beta_u<1 \tag{4.21}$$

In this expression β_u controls the form of the statistical distribution, in which β_u=0 corresponds to a Gaussian distribution. The noise is assumed to be described by (4.12). The optimal MAP estimator can be obtained by solving (4.9). Appendix B shows that the optimal scheme is given by the solution of the equation

$$\frac{v-\hat{u}}{\sigma_\eta^2}-\frac{2c(\beta_u)}{(1+\beta_u)\sigma_u^{\frac{2}{1+\beta_u}}}sign(\hat{u})\left|\hat{u}\right|^{\left(\frac{1-\beta_u}{1+\beta_u}\right)}=0 \tag{4.22}$$

Figure 4-8 describes the MAP optimal scheme when the original signal has generalised Gaussian distribution for various choices of β_u.

The equation (4.22) when β_u=0 (Gaussian distribution) becomes

$$\hat{u}_{\substack{MAP\\Gaussian}}=\frac{\sigma_u^2}{\sigma_u^2+\sigma_\eta^2}v=\frac{1}{1+\frac{1}{SNR}}v$$

where (4.23)

$$SNR=\frac{\sigma_u^2}{\sigma_\eta^2}$$

Figure 4-9 describes the MAP optimal scheme with the original image having a Gaussian distribution for various *SNR*s. Figure 4-10 describes the effect on the PDFs of applying the MAP optimal scheme for a synthetic Gaussian image with *SNR*=2.7.

The optimal estimator in ML sense is the same than in the Laplacian case because the prior distribution has no influence.

Appendix B shows that the LMSE optimal scheme when the original signal has a generalised Gaussian distribution is given by the solution of (4.12) and can be expressed in integral form as:

$$\widehat{u}(\beta_u)_{\substack{LMSE \\ Gener.Gaussian}} = \frac{\int_{-\infty}^{\infty} u e^{\frac{-(v-u)^2}{2\sigma_\eta^2}} e^{-c(\beta_u)\left|\frac{u}{\sigma_u}\right|^{\frac{2}{1+\beta_u}}} du}{\int_{-\infty}^{\infty} e^{\frac{-(v-u)^2}{2\sigma_\eta^2}} e^{-c(\beta_u)\left|\frac{u}{\sigma_u}\right|^{\frac{2}{1+\beta_u}}} du} \qquad (4.24)$$

This scheme does not have a closed form solution but it can be solved numerically to estimate the LMSE optimal scheme. This scheme depends on three parameters, β_u, σ_u and σ_η. Figure 4-11 describes this scheme when the two parameters σ_u and σ_η are fixed and the model index β_u varies.

It can be seen that equation (4.24) when β_u=0 (Gaussian distribution) produces (4.23) again, so that, when the input signal has a Gaussian distribution, the optimal LMSE and MAP estimators are coincident.

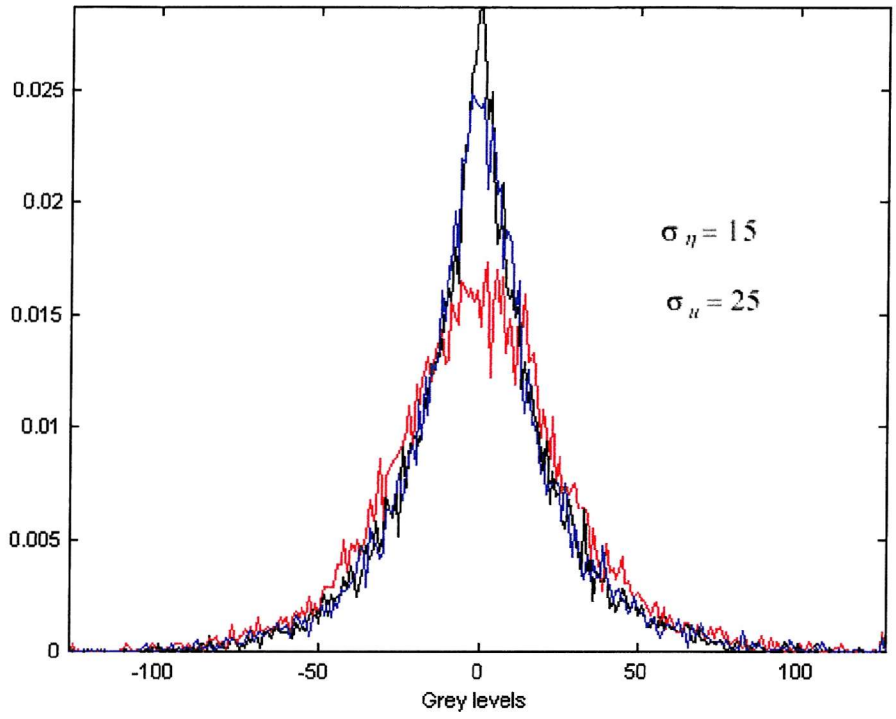


Figure 4-7. Effect on the PDFs of LMSE optimal thresholding scheme for a Laplacian signal in Gaussian noise (black is the original, red is the noisy, and blue is the thresholded).

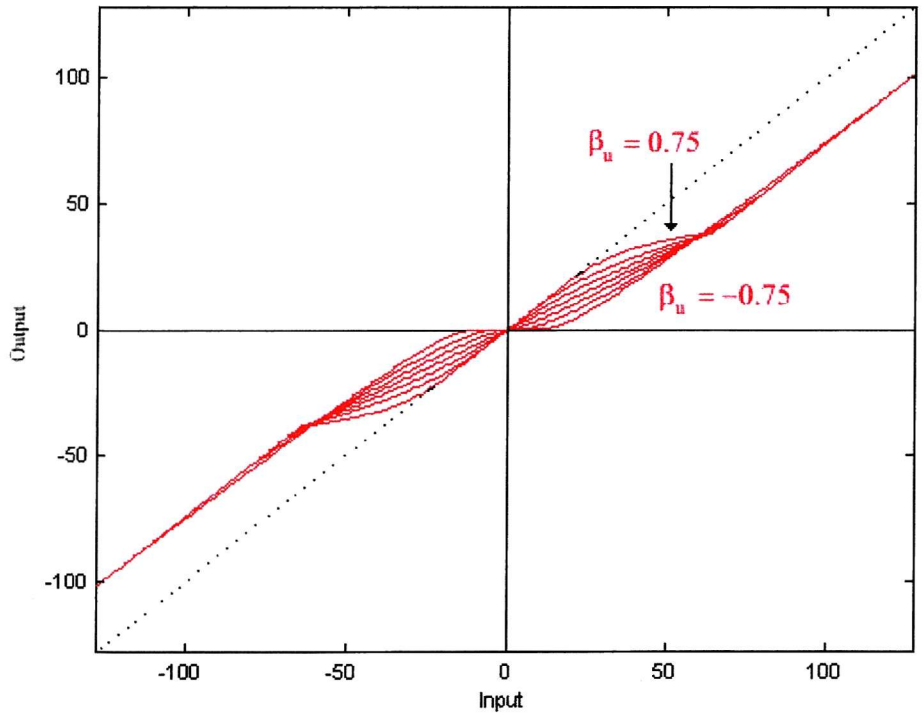


Figure 4-8. MAP optimal thresholding scheme when the image has a generalised Gaussian distribution (noise level fixed σ_η=20 and signal level σ_u=25 are fixed).

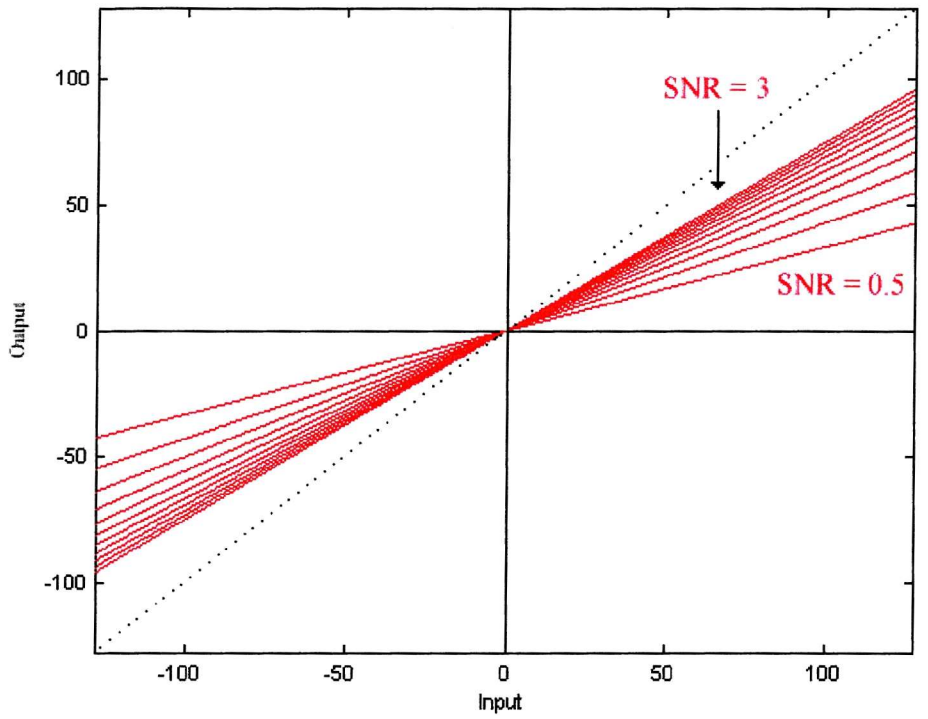


Figure 4-9. MAP (and LMSE) optimal thresholding scheme when the original components have a Gaussian distribution.

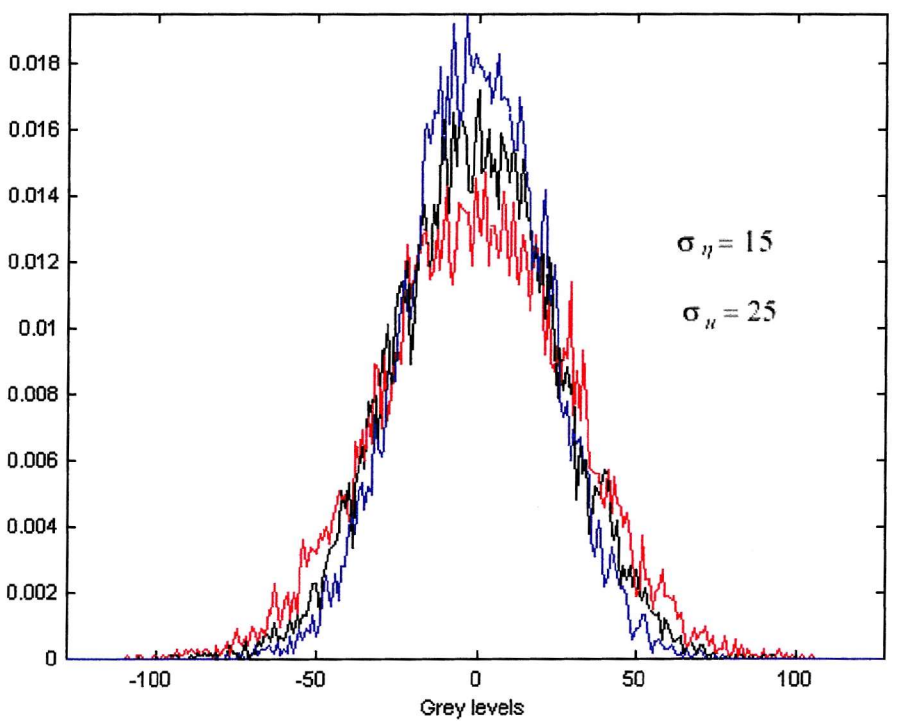


Figure 4-10. Effect on PDFs of MAP (and LMSE) optimal thresholding when the original image has a Gaussian distribution (black is the original, red is the noisy, and blue is the thresholded).

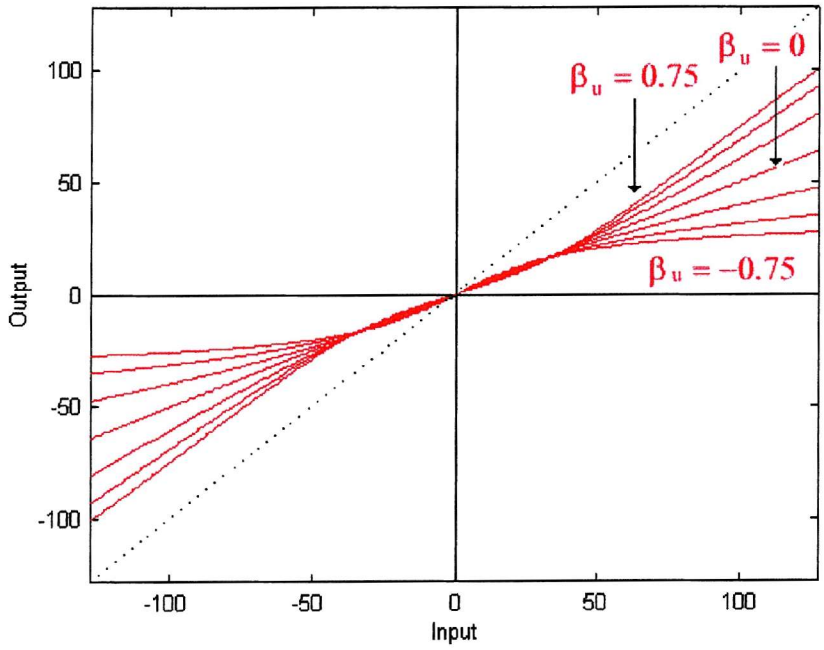


Figure 4-11. LMSE optimal thresholding scheme when the original image has generalised Gaussian distribution (noise level σ_η=15 and signal level σ_u=15 are fixed).

4.5 Scheme in our approach

The previous section has presented several schemes that could be applied to the components obtained from the decomposition of the noisy image in order to reduce noise. These were

a) Soft-thresholding and hard thresholding schemes (equations (4.4) and figures 4-4),

b) MAP optimal scheme assuming a Laplacian original signal distribution (equation (4.16) and figure 4-5),

c) LMSE optimal scheme assuming a Laplacian original image distribution (equation (4.20) and figure 4-6),

d) MAP or LMSE scheme assuming a Gaussian original image distribution (equation (4.23) and figure 4-9),

e) MAP optimal scheme assuming a generalised Gaussian original image distribution, (equation (4.22) and figure 4-8),

f) LMSE optimal scheme assuming a generalised Gaussian original image distribution, (equation (4.24) and figure 4-11).

To analyse the efficiency on the noisy components of these schemes is useful to identify which statistical distribution is suitable to model the image components. Considering the set of 13 training images, for every image the statistical distribution of one of the components obtained from the decomposition depends on:

1) Image frequency distribution,

2) Range of frequency corresponding to that component.

Figure 4-12-a shows the distributions of three components for three images. Figure 4-12-b shows the distribution of the same components for three images.

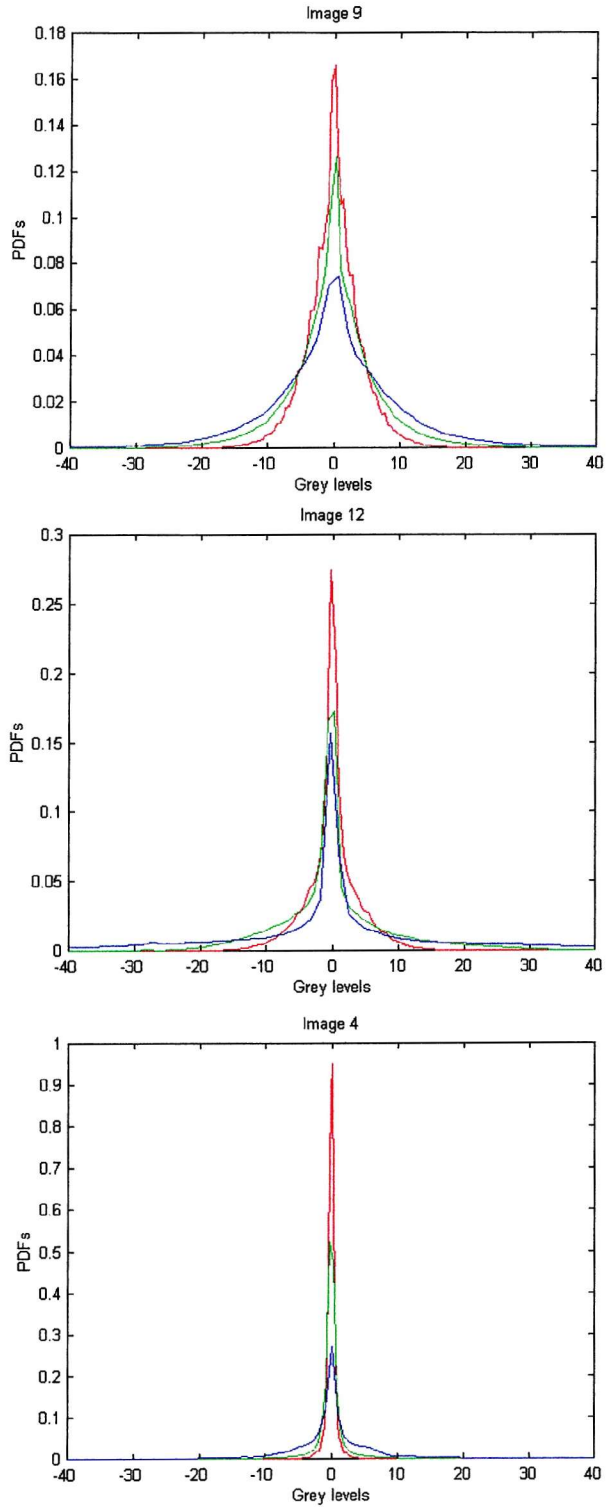


Figure 4-12-a. PDFs variability in the frequency range for 3 images (red is HH, green is HL, and blue is LH).

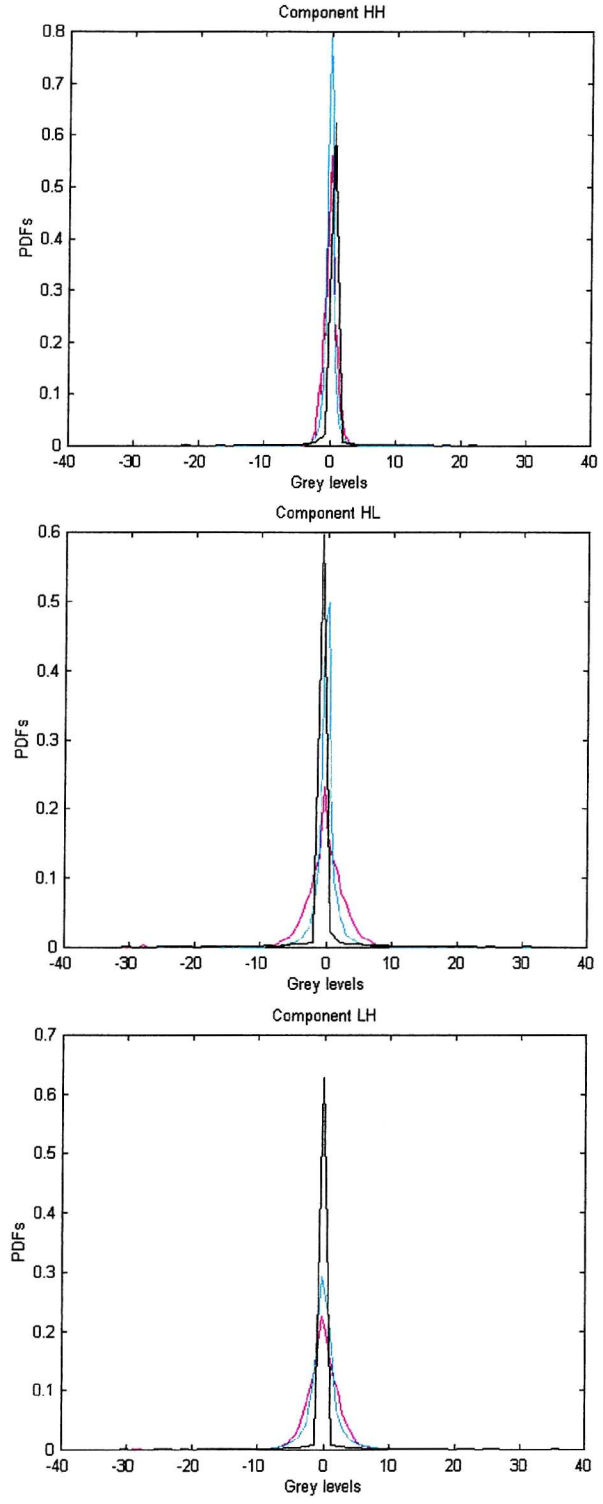


Figure 4-12-b. PDFs variability in the images for 3 frequency ranges (magenta is image 1, blue is image 6, and black is image 13).

From this one can deduce that the components obtained from the decomposition could not be modelled using only a single distribution. A compromise solution seems to be to define

a scheme including (or approximating) the described schemes. One such scheme is shown in figure 4-13.

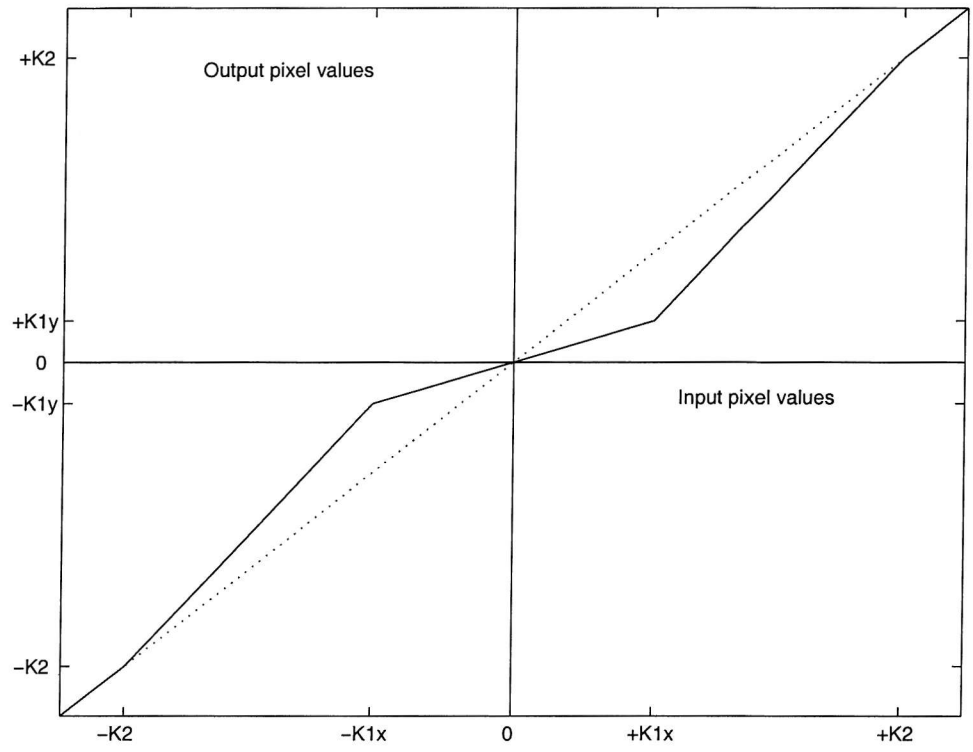


Figure 4-13. A general thresholding scheme.

The analytical expression of this scheme is

$$\hat{u} = \begin{cases} v & \text{if } |v| > K2 \\ \dfrac{K1y}{K1x} v & \text{if } K1x \le |v| \le K2 \\ (v - sign(v)K1x)\dfrac{K2 - K1y}{K2 - K1x} + sign(v)K1y & \text{if } |v| \le K1x \end{cases} \qquad (4.25)$$

By suitable choice of the three parameters it is possible to approximate the optimal schemes previously derived.

- Scheme a) is approximated when $lim\ K1y \to 0$, $K1y \to K1x$.

- Scheme b) is approximated when $K1x \neq K1y$, $K1y = 0$, $K2 \rightarrow \infty$.
- Scheme d) is approximated when $K1x \rightarrow \infty$ $K1y \neq 0$.
- Selecting appropriate $K1x$, $K1y$, and $K2$, it is possible to have a piece-wise linear approximation of schemes in c), e), and f).

4.6 Conclusions

This chapter has analysed the problem of the identification of the shape of the thresholding functions to apply on the components obtained from the decomposition. Each component is treated as a stationary (homogeneous) process and the thresholds are applied globally and not locally in the spatial domain. To identify the shape of these functions in a general form it is useful to define a scheme or a parameterised family of functions. Two classical schemes used in the literature were considered and other optimal schemes were derived from Bayesian estimation theory. Two main criteria for optimal estimation have been considered: MAP and LMSE. The noise has been assumed to be AWGN. The optimal schemes have been described when the original component is assumed to have either a Laplacian or a generalised Gaussian distribution. In particular the case of Gaussian distribution has been considered. Nevertheless it is difficult to model the totality of the component using a single statistical distribution. For this reason a new scheme which approximates (or even includes) both classical and optimal schemes was derived. The selection of suitable parameters for the scheme is the subject of the next chapter.

Chapter 5

Thresholding function parameters

5.1 Introduction

The shape of a set of near optimal thresholding functions has been described in the previous chapter. This shape defines a parameterised family of thresholding functions. This chapter considers the problem of selecting the parameters of the thresholding functions to achieve the best results. This selection can be seen as an optimisation of some cost function related to the image quality and depending on a finite number of variables (the thresholding function's parameters). A training procedure to perform this optimisation is used. MSE and WMSE are assumed as cost functions. The training is performed initially on a test image and then extended to a set of images. The parameters resulting from the training define the set of thresholding functions for the components and depend on the level of noise contaminating the image. Consequently a noise estimation procedure is required. The techniques to estimate the noise level will be examined in next chapter.

The performances of the algorithms tested for the cost function minimisation are compared in the next section. The detailed description of the algorithms is given in Appendix A. The parameters utilised in the cost function are presented in section 5.3. The cost function values used in the training are described in section 5.4. The results of the training are listed and commented on in section 5.5.

5.2 Comparison of the training algorithms performances

Two significant categories of optimisation methods are: gradient based and direct search methods. Three algorithms belonging to these categories have been considered in this thesis.

The cost function to be minimised is non-linear and the hypothesis of differentiability cannot always be justified. Direct search methods are characterised by a strategy that generates variation of the parameter vector and decision criteria to determine whether or not the newly derived parameters should be accepted. Consequently the hypothesis of differentiability is not required and this makes the methods in this category seem more suitable than gradient based approaches for this application. In the presence of multimodal cost functions, the main problem with direct search methods is ensuring the algorithm converges to the global, rather than to a local, minimum. In this section a quasi-Newton gradient based algorithm [94] and two direct search algorithms, Simplex [93] and differential evolution [95], are compared.

In order to compare the performance of the minimisation algorithms a noise free test image has been contaminated by additive white Gaussian noise and, after the decomposition, the parameters of the thresholding functions applied on the components have been optimised to minimise the noise. In table 5.1 the sets of parameters derived using the three algorithms are listed. The MSE has been used as a cost function (unweighted optimisation) and the image is contaminated by white noise having a standard deviation level of 15. The size of the parameter sets depends on the number of decomposition levels used. Considering 3, 2 and 1 levels, sets composed respectively of 11, 8 and 5 parameters are utilised. The assumption related to these sizes and the notation of the parameters will be explained in

detail in section 5.3. Table 5.2 shows the number of iterations and time required by the three algorithms to converge. The training has been performed using a PC Pentium 133 MHz.

The differences between the MSE reductions using the three algorithms are not significant (of the order of 0.1 %). Table 5.1 shows a strong similarity between the sets of parameters obtained for the different minimisation techniques. Nevertheless table 5.2 shows that the convergence speed of the Simplex algorithm is superior compared with the other two algorithms. The Simplex algorithm needs a smaller number of iterations compared with the differential evolution algorithm and a smaller time for single iteration compared with the quasi-Newton algorithm. For this reason the Simplex algorithm has been preferred to the other algorithms and the results in the remainder of the chapter are obtained using this technique.

	K1y	K2	K1x HH	K1x HL	K1x LH	K1x LL HH	K1x LL HL	K1x LL LH	K1x LL LL HH	K1x LL LL HL	K1x LL LL LH
5 parameters											
Simplex	0.05	6.89	20.51	9.90	7.23						
Quasi-Newton	0.05	6.89	20.61	9.90	7.23						
Differ-Evolution	0.04	6.89	20.67	9.97	7.23						
8 parameters											
Simplex	0	3.89	37.60	10.21	7.51	1.96	2.15	2.38			
Quasi-Newton	0	3.89	20.67	10.21	7.51	1.96	2.15	2.38			
Differ-Evolution	0.01	3.76	25.45	10.19	7.49	1.97	2.15	2.37			
11 parameters											
Simplex	0.01	3.67	51.87	10.39	7.83	1.99	2.04	2.39	0.92	0.98	0.97
Quasi-Newton	0.01	3.90	20.70	10.19	7.51	1.94	1.99	2.17	0.87	0.90	1.04
Differ-Evolution	0.01	3.81	32.45	10.24	7.79	1.95	2.01	2.23	0.9	0.89	1.10

Table 5.1. Unweighted optimisation of an image contaminated by white noise with std=15.

	5 parameters		8 parameters		11 parameters	
	No. Iterat.	**Cnv. Time**	**No. Iterat.**	**Cnv. Time**	**No. Iterat.**	**Cnv. Time**
Simplex	540	2hrs 15min	800	3hrs 20min	1100	4hrs 35min
Quasi-Newton	475	2hrs 15min	795	3hrs 45min	1130	5hrs 20min
Differ-Evolution	775	2hrs 35min	1050	3hrs 30min	1575	5hrs 15min

Table 5.2. Number of iterations and convergence time required by unweighted optimisation of an image contaminated by white noise with std=15

5.3 Cost function parameters

The filters used in the decomposition are non-orthogonal resulting in an overcomplete representation. As a consequence the minimisation of the noise on each component is not equivalent to minimising the noise on the full image. To minimise the noise on one component requires minimisation of a function of l parameters. Consequently to minimise the overall noise by minimising the noise on the K components individually requires the minimisation of K separate functions each of l variables. On the other hand, to minimise the noise on the K components concurrently requires the minimisation of one function of $(K \mathbf{x} l)$ variables. The non-orthogonal nature of the decomposition forces one to optimise over all the parameters simultaneously leading to an optimisation task over a relatively large number of parameters. Table 5.3 details the 27 parameters resulting from minimisation of the noise in a concurrent scheme. There are 27 parameters since 3 levels of decomposition result in 9 components to be processed and each component is thresholded using a function described by 3 parameters (see (4.25)). Results are presented for two cost functions and three levels of noise.

This procedure is expensive in terms of computational time and the optimisation results are sensitive to the initial conditions. From table 5.3 it can be seen that some of the parameters are negative which leads to counter-intuitive thresholding functions. This may be an indication that the minimisation routine has not converged to the global minimum. To avoid this, it is prudent to consider simplifying the threshold functions in order to reduce the number of parameters involved in the optimisation.

Each thresholding function is characterised by l parameters and is applied to one of the K image components. The number of thresholding function parameters depends on the

thresholding scheme and the number of components depends on the number of decomposition levels used. The total number of parameters of the cost function is due to these two factors. In the next sub-sections these are analysed separately.

	Unweighted optimisation			**Weighted optimisation**		
	Noise level 4	**Noise level 5**	**Noise level 6**	**Noise level 4**	**Noise level 5**	**Noise level 6**
K1x (HH)	2.82	3.75	4.69	5.56	4.15	5.20
K1y (HH)	0.15	0.11	-0.26	0.02	-0.08	-0.18
K2 (HH)	24.72	28.44	33.75	14.95	16.21	19.99
K1x (HL)	2.01	2.64	3.00	3.71	2.85	3.34
K1y (HL)	0	0	0	0	0.03	-0.01
K2 (HL)	7.37	8.98	11.25	11.09	9.57	11.80
K1x (LH)	1.38	1.66	2.29	03.19	2.37	3.99
K1y (LH)	0.17	0.04	0.23	0.71	0.56	1.51
K2 (LH)	8.05	10.15	11.13	12.76	9.28	10.00
K1x (LLHH)	0.63	0.92	1.05	0.98	0.86	1.07
K1y (LLHH)	0	0.01	0	0	-0.07	0
K2 (LLHH)	1.62	1.85	2.27	2.32	1.97	2.29
K1x (LLHL)	0.43	0.69	0.95	0.95	0.59	0.86
K1y (LLHL)	0.01	0	0	0	-0.04	0
K2 (LLHL)	2.84	2.68	2.90	2.88	3.64	2.93
K1x (LLLH)	0.71	0.87	1.10	1.06	1.15	1.12
K1y (LLLH)	0.03	0.02	0	0	0.31	0
K2 (LLLH)	2.50	3.01	3.41	2.99	2.85	3.24
K1x (LLLLHH)	0.26	0.33	0.33	0.65	0.16	0.39
K1y (LLLLHH)	0.02	0.038	0	-0.05	-0.18	0
K2 (LLLLHH)	0.66	1.20	1.70	0	1.36	1.60
K1y (LLLLHL)	0.19	0.37	0.36	0.30	0.24	0.35
K1x (LLLLHL)	0.01	0.01	0.05	-0.01	0.07	-0.01
K2 (LLLLHL)	1.21	1.24	1.60	1.96	2.18	1.57
K1x (LLLLLH)	0.16	0.26	0.03	0.07	0.45	0.01
K1y (LLLLLH)	0.01	0.02	-0.15	0.03	0	-0.18
K2 (LLLLLH)	0.91	1.61	6.01	1.49	0.85	2.43

Table 5.3. Parameters of the thresholding functions for complete scheme.

The notation used to name the parameters is composed of two parts: the first part (*K*1*x*, *K*1*y* or *K*2) indicates the parameter of the thresholding function, the second (HH, HL, LH, etc.) refers to the component where the thresholding function has been applied (see figure 3-8).

5.3.1 Use of the Hessian matrix to reduce the number of parameters

Simplifying the thresholding schemes it is important to identify which parameters of the thresholding scheme are the most important. This section considers the use of the local

characteristics of the cost function surface to assess the importance of each parameter. The method employed here examines the Hessian matrix of the cost function evaluated at the optimum [109].

Consider the Taylor series expansion of $f(h_1, h_2, ..., h_n)$ about its minimum point $(h_1^*, h_2^*, ..., h_n^*)$

$$f(h_1, h_2, ..., h_n) = f(h_1^*, h_2^*, ..., h_n^*) + \sum_i (h_i - h_i^*) \frac{\partial f(h_1^*, h_2^*, ..., h_n^*)}{\partial h_i} + \\ + \frac{1}{2} \sum_i \sum_j (h_i - h_j^*)(h_j - h_j^*) \frac{\partial^2 f(h_1^*, h_2^*, ..., h_n^*)}{\partial h_i \partial h_j} + ... \quad (5.1)$$

If this is truncated to only include terms up to second order terms and it is noted that since h_i^* is the minimum then

$$f(h_1, h_2, ..., h_n) \approx f(h_1^*, h_2^*, ..., h_n^*) + \frac{1}{2} (\underline{h} - \underline{h}^*)^T H (\underline{h} - \underline{h}^*) \quad (5.2)$$

where $\underline{h} = (h_1, h_2, ..., h_n)$ and the matrix H is the Hessian matrix defined as

$$H = \begin{bmatrix} \frac{\partial^2 f(h_1 ... h_n)}{\partial h_1^2} & \frac{\partial^2 f(h_1 ... h_n)}{\partial h_2 \partial h_1} & \cdot & \frac{\partial^2 f(h_1 ... h_n)}{\partial h_n \partial h_1} \\ \frac{\partial^2 f(h_1 ... h_n)}{\partial h_1 \partial h_2} & \frac{\partial^2 (h_1 ... h_n)}{\partial h_2^2} & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \frac{\partial^2 f(h_1 ... h_n)}{\partial h_1 \partial h_n} & \cdot & \cdot & \frac{\partial^2 f(h_1 ... h_n)}{\partial h_n^2} \end{bmatrix} \quad (5.3)$$

Defining $\Delta \underline{h} = \underline{h} - \underline{h}^*$, then the change in the cost function resulting from a change in parameters of $\Delta \underline{h}$ is approximated by

$$\frac{1}{2}\Delta\underline{h}^T H\Delta\underline{h} \qquad (5.4)$$

To see the significance of the i^{th} parameter, one can consider the following optimisation.

Minimise

$$f(h_1,h_2,\ldots,h_n) \quad (5.5)$$

subject to

$$\underline{u}^T\underline{h} = 0 \qquad (5.6)$$

where

$$u_k = \begin{cases} 0 & k \neq i \\ 1 & k = i \end{cases} \qquad (5.7)$$

The change in value of $f(\ )$ for this constrained optimisation is a measure of the importance of the i^{th} parameter. Using the quadratic approximation for $f(\ )$ then the solution to the constrained approximation is given by using method of Lagrange multipliers, as follows.

Consider

$$\begin{aligned} \Psi &= \frac{1}{2}(\underline{h}-\underline{h}^*)^T H(\underline{h}-\underline{h}^*) + \lambda(\underline{u}^T\underline{h}) \\ \frac{\partial\Psi}{\partial\underline{h}} &= H\underline{h} - H\underline{h}^* + \lambda\underline{u} = 0 \qquad (5.8) \\ \frac{\partial\Psi}{\partial\lambda} &= \underline{u}^T\underline{h} = 0 \end{aligned}$$

this implies

$$\underline{h} = \underline{h}^* - \lambda H^{-1}\underline{u} \qquad (5.9)$$

Pre-multiplying by $\underline{u}^T$ gives

$$\underline{u}^T \underline{h} = 0 = \underline{u}^T \underline{h}^* - \lambda \underline{u}^T H^{-1} \underline{u} \qquad (5.10)$$

and then

$$\lambda = \frac{\underline{u}^T \underline{h}^*}{\underline{u}^T H^{-1} \underline{u}} \qquad (5.11)$$

Equation (5.9) can be rewritten

$$\underline{h} = \underline{h}^* - \frac{H^{-1} \underline{u}\underline{u}^T \underline{h}^*}{\underline{u}^T H^{-1} \underline{u}} \qquad (5.12)$$

Substituting into the cost function (5.8)

$$\begin{aligned} \Psi &= \frac{(H^{-1}\underline{u}\underline{u}^T \underline{h}^*)^T H (H^{-1}\underline{u}\underline{u}^T \underline{h}^*)}{(\underline{u}^T H^{-1} \underline{u})^2} = \\ &= \frac{\underline{h}^{*T} \underline{u} (\underline{u}^T H^{-1} \underline{u}) \underline{u}^T \underline{h}^*}{(\underline{u}^T H^{-1} \underline{u})^2} = \\ &= \frac{\underline{h}^{*T} \underline{u}\underline{u}^T \underline{h}^*}{(\underline{u}^T H^{-1} \underline{u})} = \frac{(\underline{h}^{*T} \underline{u})^2}{(\underline{u}^T H^{-1} \underline{u})} \end{aligned} \qquad (5.13)$$

Using the definition of $\underline{u}$ one obtains

$$\Psi = \frac{(\underline{h}_i^*)^2}{(H^{-1})_{i,i}} \qquad (5.14)$$

This is the change in the cost function, $f(\ )$, if the i^{th} parameter is set to zero based on a local quadratic approximation. Thus the parameter for which (5.14) is smallest is the parameter which is of least significance. The influence of each parameter on the results, i.e.

the influence of each variable in minimising a multidimensional function, is investigated by examination of the Hessian matrix. The Hessian matrix has been computed using finite differences:

$$\frac{\partial^2 f(x_1,.....,x_N)}{\partial x_i \partial x_j} = \begin{cases} \dfrac{f(x_1,...,x_i-\Delta x_i,...x_N)-2f(x_1,...,x_i,...x_N)+f(x_1,...,x_i+\Delta x_i,...x_N)}{\Delta x_i^2} & if\ i=j \\ \dfrac{f(x_1,...,x_i-\Delta x_i,x_j-\Delta x_j,...x_N)+f(x_1,...,x_i+\Delta x_i,x_j+\Delta x_j,...x_N)-f(x_1,...,x_i+\Delta x_i,x_j-\Delta x_j,...x_N)+}{\Delta x_i \Delta x_j} & \\ \dfrac{-f(x_1,...,x_i-\Delta x_i,x_j+\Delta x_j,...x_N)}{\Delta x_i \Delta x_j} & if\ i \neq j \end{cases} \quad (5.15)$$

where $\Delta x_i = 1\%$ of x_i

Table 5.4 shows the results of the procedure to identify the significance of the parameters. The results show that assuming quadratic behaviour of the cost function the parameters related to the first level of decomposition are the least significant and the parameters constant for all the components (*K*1*y* and *K*2) cannot be neglected. This justifies the choice of thresholding scheme done in section 5.3.2.

	Noise level 5	**Noise level 10**	**Noise level 15**
Least significant	K1xLH	K1xLH	K1xHL
2nd	K1xHL	K1xHL	K1xLH
3rd	K1xHH	K1xHH	K1xHH
4th	K1xLLLLLH	K1xLLLLHL	K1xLLLLHL
5th	K1xLLLLHL	K1xLLLLLH	K1xLLLLHH
6th	K1xLLLLHH	K1xLLLLHH	K1xLLHH
7th	K1y	K1xLLHH	K1y
8th	K1xLLHH	K1y	K1xLLLH
9th	K1xLLLH	K1xLLLH	K1xLLLLLH
10th	K1xLLHL	K1xLLHL	K1xLLHL
Most significant	K2	K2	K2

Table 5.4. Analysis of the significance of the parameters in the training procedure using the Hessian matrix.

5.3.2 Reduction of the number of parameters of the thresholding functions

The task of this sub-section is to discuss several options for reducing the number of parameters involved in defining each thresholding function. According to the scheme

described in the previous chapter each function depends on 3 parameters *K1x, K1y,* and *K2*. The initial function to be minimised is of the form

$$f(K1x_1, K1y_1, K2_1, ...K1x_i, K1y_i, K2_i, ...) \quad (5.16)$$
$$i = 1.....K \text{ is the components index}$$

The first simplification is to assume *K1y* and *K2* are linearly dependent on *K1x* with coefficients of proportionality that are constant for all the components of the same image i.e. assuming

$$K1y_i = K1y * K1x_i$$
$$K2_i = K2 * K1x_i \quad (5.17)$$
$$i = 1.....K \text{ is the component index}$$

The function to be minimised has the form

$$f(K1y, K2, K1x_1, ...K1x_i, ...) \quad (5.18)$$
$$i = 1.....K \text{ is the component index}$$

The analytical form of this thresholding scheme is,

$$v = \begin{cases} u & \text{if } |u| > K2 * K1x \\ K1y * u & \text{if } K1x \le |u| \le K2 * K1x \\ (u - sign(u)K1x)\dfrac{K2 - K1y}{K2 - 1} + sign(u)K1y * K1x & \text{if } |u| \le K1x \end{cases} \quad (5.19)$$

The family of thresholding functions is showed in figure 5-1.

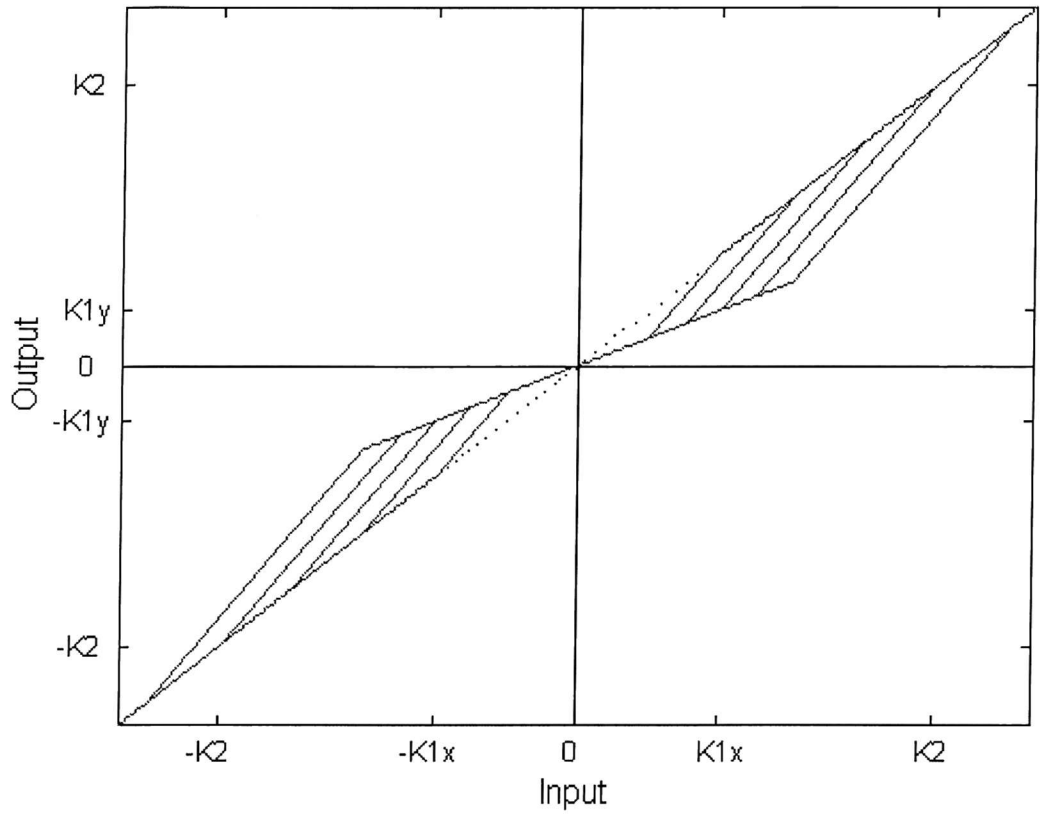


Figure 5-1. Thresholding scheme after first simplification.

The assumption related to the first simplification seems reasonable when this scheme is compared with the optimal thresholding schemes derived in the previous chapter (figure 4-13). If l is the number of components resulting from the decomposition, the first simplification reduces the number of parameters from $3l$ to $l+2$.

Further simplifications could be made assuming

- $K1y = 0$
- $K1y = 0$ and $K2 = 3$[1].

These simplifications reduce the number of variables from $3l$ respectively to $l+1$ and l.

These assumptions have not been considered in this chapter because they reduce slightly the number of parameters but also produce a grosser approximation of the family described

in figure 4-13 and consequently of the families derived theoretically in the previous chapter. In the remainder of the chapter the scheme related to the first simplification (l+2 variables of the cost function) has been used because it represents an acceptable compromise between number of cost function variables and similarity with the optimal schemes.

5.3.3 Number of decomposition levels

In the previous sub-section several options for reducing the number of variables defining the thresholding functions have been proposed. In all cases the complexity of the training procedure is roughly proportional to the number of levels of decomposition. Hence one way of reducing the number of parameters is to reduce the number of decomposition levels but this reduces the ability of the method to analyse limited frequency bands. The best choice is based on a compromise between the amount of frequency decomposition needed to reduce the noise on the image, and computational load. In this work 3, 2, and 1 levels of decomposition are considered, corresponding to 10, 7, and 4 components. Furthermore, no manipulations are performed on the component in the lowest frequency band because it is hard to distinguish noise and image in this component. Hence the number of components to be thresholded is 9, 6, and 3 and the number of variables of the cost function is 11, 8, and 5. Table 5.5-a lists the parameters for 3, 2, and 1 decomposition levels for weighted[2] and unweighted MSE and for 3 levels of noise. Table 5.5-b presents the MSE and WMSE (weighted MSE) levels applying these sets of parameters.

[1] The choice of this value facilitates the hardware implementation of the technique.
[2] Details of the weighting are given in section 5.4.2.

Unweighted optimisation	**K1y**	**K2**	**K1x HH**	**K1x HL**	**K1x LH**	**K1x LL HH**	**K1x LL HL**	**K1x LL LH**	**K1x LL LL HH**	**K1x LL LL HL**	**K1x LL LL LH**
5 parameters											
Noise level 5	0.05	3.67	4.17	2.89	2.33						
Noise level 10	0.02	3.99	10.94	5.96	4.79						
Noise level 15	0.05	6.89	20.51	9.90	7.23						
8 parameters											
Noise level 5	0.08	3.74	4.27	2.72	2.14	0.68	0.78	0.90			
Noise level 10	0	3.73	11.91	5.62	4.49	1.29	1.55	1.72			
Noise level 15	0	3.89	37.60	10.21	7.51	1.96	2.15	2.38			
11 parameters											
Noise level 5	0.09	3.62	4.31	2.68	2.22	0.73	0.73	0.87	0.32	0.35	0.28
Noise level 10	0.01	3.60	10.99	5.74	4.59	1.39	1.43	1.59	0.61	0.59	0.78
Noise level 15	0.01	3.67	51.87	10.39	7.83	1.99	2.04	2.39	0.92	0.98	0.97
Weighted optimisation	**K1y**	**K2**	**K1x HH**	**K1x HL**	**K1x LH**	**K1x LL HH**	**K1x LL HL**	**K1x LL LH**	**K1x LL LL HH**	**K1x LL LL HL**	**K1x LL LL LH**
5 parameters											
Noise level 5	0.36	2.57	8.82	2.95	2.97						
Noise level 10	0.09	6.30	13.02	4.01	4.08						
Noise level 15	0.05	14.98	21.00	7.96	6.79						
8 parameters											
Noise level 5	0.24	2.82	12.76	1.81	2.06	0.94	1.15	1.18			
Noise level 10	0.54	2.19	164.7	0.14	4.88	1.81	2.08	2.13			
Noise level 15	0.68	2.52	547.2	209.6	54.82	2.79	2.89	2.79			
11 parameters											
Noise level 5	0.08	4.30	4.60	1.08	1.11	1.13	1.24	1.26	0.40	0.44	0.46
Noise level 10	0.34	3.32	10.25	3.85	4.21	1.82	1.96	1.89	0.65	0.63	0.72
Noise level 15	0.47	2.67	72.66	7.29	8.09	2.81	2.71	2.72	0.88	0.89	1.00

Table 5.5-a. Unweighted and weighted optimisations for 3 decomposition levels and 3 noise levels.

MSE				
Unweighted optimisation	**0 parameters**	**5 parameters.**	**8 parameters**	**11 parameters**
Noise level 5	25	14.1	13.5	13.4
Noise level 10	99.2.	43.2	40.1	39.7
Noise level 15	226.7	80.1	73.0	72.2
Weighted optimisation				
Noise level 5	25	15.1	14.7	14.7
Noise level 10	99.2	55.8	51.2	51.4
Noise level 15	226.7	123	102.1	101.2
WMSE				
Weighted optimisation				
Noise level 5	2.5	2.3	2.1	2
Noise level 10	10.1	8.8	7.2	7.1
Noise level 15	23.1	19.1	15.3	14.7
Unweighted optimisation				
Noise level 5	2.5	2.3	2.2	2.1
Noise level 10	10.1	9	8	7.8
Noise level 15	23.1	19.3	17.1	16.4

Table 5.5-b. Noise reduction using weighted and unweighted optimisations for 3 decomposition levels and 3 noise levels.

In this case the notation used to name the parameters is composed of two parts only for $K1x$ while $K1y$ and $K2$ are constant for all the components (see equations (5.17) and (5.18)). Table 5.5-a shows that the parameters related to a level of decomposition depend on the number of decomposition levels applied. This confirms the need of the concurrent optimisation of all the components across the different decomposition levels. The table also emphasises that the larger differences between the parameters are obtained using the two cost functions especially at the first level of decomposition (columns 4, 5 and 6). The reason for this will be explained in section 5.4 where the cost functions are described in detail. Table 5.5-b demonstrated a significant part of the noise reduction is performed on the second and third level of decomposition. The selection of the number of decomposition levels applied depends on the spectral characteristics of the image considered. In the rest of this chapter, 3 levels of decomposition are utilised.

5.4 Cost function values

The primary goal of this work is to reduce the noise contaminating an image, but meantime it is important that the visual quality of the image should not be significantly degraded. Thus the cost function used in the training procedure should be a measure of noise reduction and of the visual quality of the image. Two indices have been considered as cost functions in this work.

The first index is the mean square error (MSE), defined as:

$$\mathrm{MSE} = \frac{\sum_{m,n}(u(m,n) - \hat{u}(m,n))^2}{N} \qquad (5.20)$$

where N is the number of pixels in the original image $u(m, n)$, and $\hat{u}(m,n)$ is the estimated image after the noise reduction technique has been applied to a version of $u(m, n)$ contaminated by additive noise.

$$\begin{aligned} v(m,n) &= u(m,n) + n(m,n) \\ \hat{u}(m,n) &= f[v(m,n)] \end{aligned} \qquad (5.21)$$

where $f[\ ]$ is the thresholding operation. This is the simplest of all measures of image quality but is not always sufficient to judge the performance of a technique. Assuming independence between image and noise and N large the MSE can be rewritten as

$$\mathrm{MSE} \approx \frac{\sum_{m,n}(v(m,n) - \hat{v}(m,n))^2}{N} + \frac{\sum_{m,n}(n(m,n) - \hat{n}(m,n))^2}{N} \qquad (5.22)$$

Distortion

Noise reduction

The percentage of reduction of this index measures the difference between the percentage of noise reduced and the percentage of distortion introduced by processing. Consequently the percentage of noise reduced by minimising the MSE may be large but may introduce unacceptable distortions.

The second measure of image quality is a modified form of MSE, adjusted to partially reflect the human sensitivity to high frequency distortions. To evaluate the significance of any distortion, a model of the human visual system is necessary. The determination of a model of the human visual system is made difficult by its inaccessible and distributed nature. Experiments demonstrate that the human eye is particularly sensitive to distortion in high frequency bands. So that if a small amount of high frequency distortion is introduced then the perceived image quality may be reduced, even in the presence of noise reduction. The relationship between the WMSE used herein and visual models presented in literature is explored later.

5.4.1 Visual model

To describe the human visual system is a very complex problem. Experiments have been performed to investigate some of the eye properties. This sub-section describes two concepts emphasised by these experiments in order to justify the procedure utilised in the approach used herein.

The first concept is that the first perceptible distortion produced by filtering occurs at the major edges of the image. From this, one can assume that the threshold perception of distortion could be studied by analysing edge perception in the filtered images. A model describing this perception has been used to design filters for noise reduction [96, 97, 98]. The visibility of lines as a function of the distance to an edge can be expressed by:

$$v(x) = 1 - a^x \qquad (5.23)$$

In this expression x is the distance to an edge, and a is a parameter related to the ratio of luminance on high and low levels sides (figure 5.2).

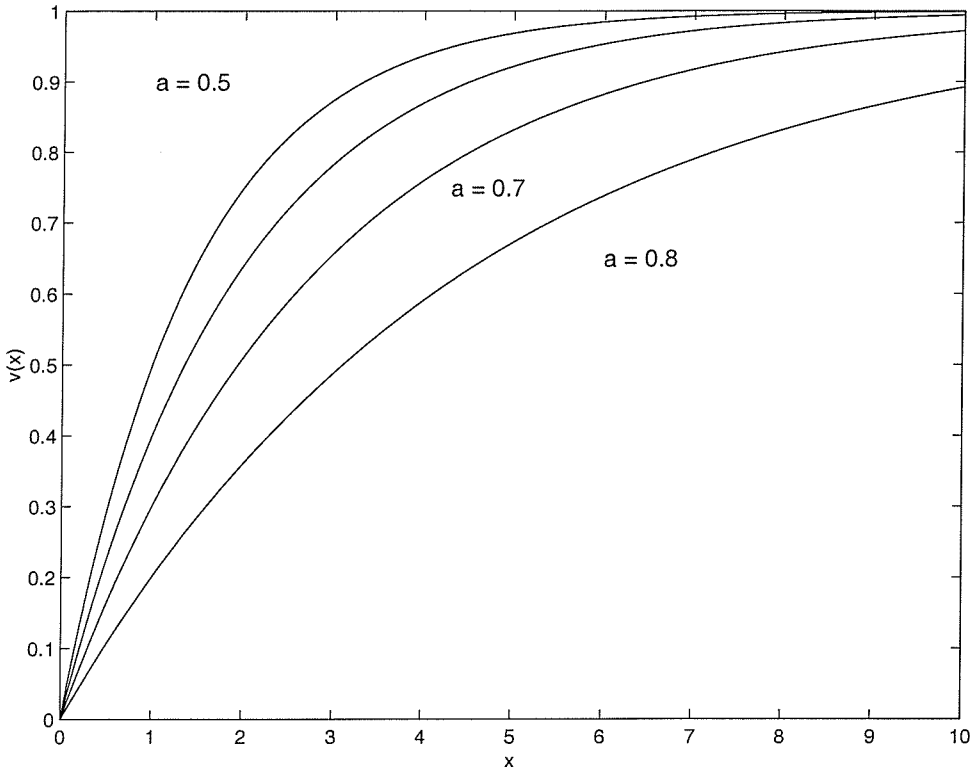


Figure 5-2. Visibility function dependence on the distance of a line to an edge.

The second concept is that there is an upper limit to the spatial frequency that the eye can perceive at a given viewing distance. Therefore the eye behaves as a low-pass filter. Moreover the visual system weights the middle range frequencies more heavily than either the low or the high frequencies. An approximation of the response of the visual system [99] as a spatial frequency filter can be described by the function

$$H_c(f)^2 = (1 + 50f^2)e^{-(2f)^2} \quad \text{for } 0 < f < 1 \ \left[\frac{\text{cycles}}{\text{Unit distance}}\right] \qquad (5.24)$$

which is shown in figure 5-3 where the frequency axis is normalised so that 1 cycle per unit distance corresponds to the frequency cutoff of the visual system.

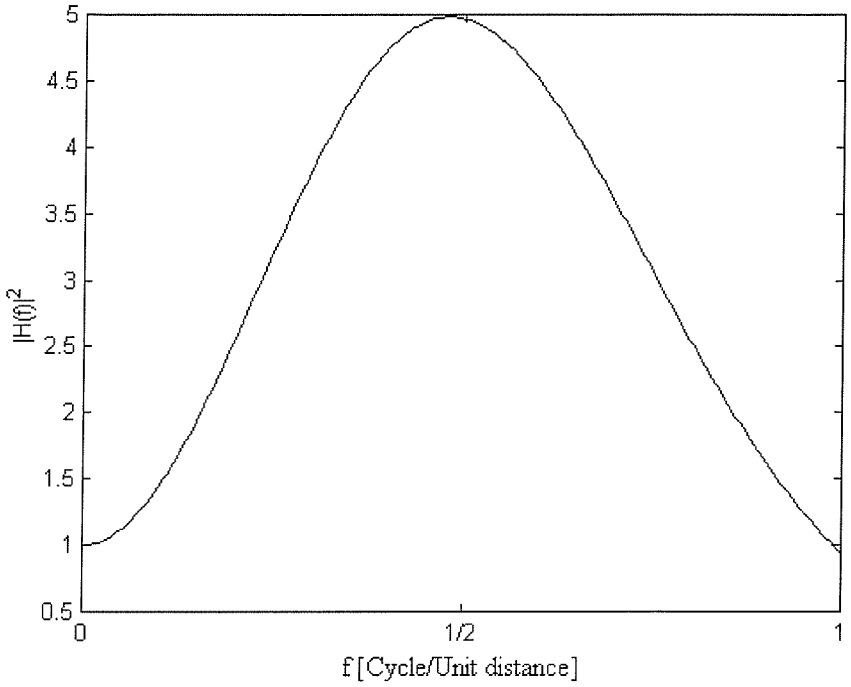


Figure 5-3. Model for the frequency response of the visual system.

5.4.2 Weighted Mean Square Error (WMSE)

A method to evaluate the image visual quality based on MSE is to weight differently the error components in different frequency bands. The human eye is particularly sensitive to edges and discontinuities. To preserve the discontinuities less noise reduction should be performed in the high frequency bands. In order to better determine the visual quality of an image, the error is weighted using a low-pass filter retaining the low frequency content and attenuating the high frequency elements. The limit of this procedure is the assumption of the eye perception as linear whereas in fact it is non-linear. Figure 5-4 describes the block

diagram of the method to measure the WMSE, and figure 5-5 shows the frequency response of the low-pass weighting filter used[3].

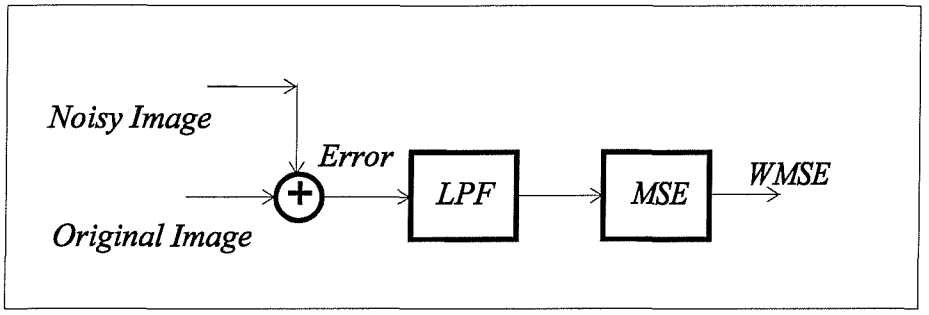


Figure 5-4. Block diagram describing the WMSE *evaluation.*

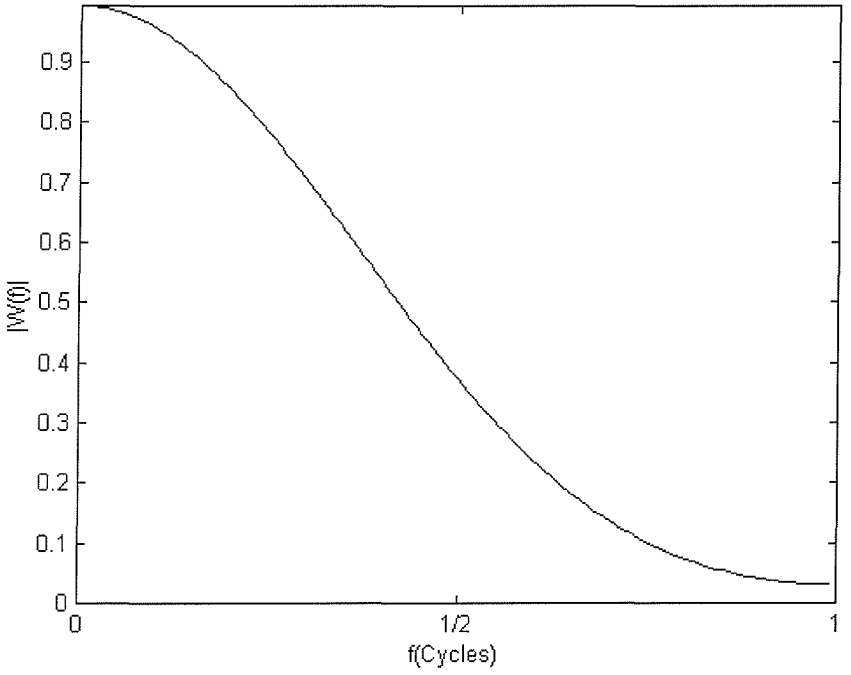


Figure 5-5. The frequency response of the weighting filter.

The weighting filter serves to limit the highest frequency influencing the measured image quality. The de-noising scheme used here imposes a low frequency limit by the fact that the very lowest frequency component is never subjected to the thresholding function. The visibility function extracted from these two filters has the characteristics of a band pass

[3] This FIR filter with impulse response W=[0.0347 0.2406 0.4495 0.2406 0.0347], has been suggested in a private communication with M. Weston.

filter. The frequency response of this filter is shown in figure 5-6. One can observe the similarity with the frequency response obtained from the visual model, see figure 5-3. It is left until later chapters to illustrate that minimising WMSE does indeed result in improving visual quality of the final image.

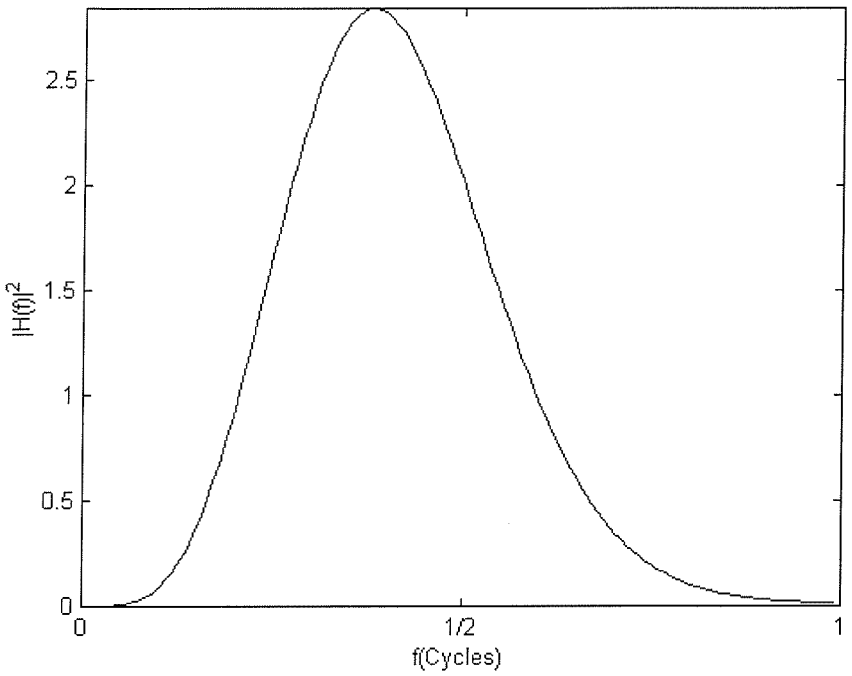


Figure 5-6. Frequency response describing the visual model assumed in the method presented in this thesis.

5.5 Training results

Section 5.2 compares the performances of the algorithms used to minimise the cost function, section 5.3 describes the parameters of the cost function and section 5.4 describes the form of the cost function. In this section the results of the training procedure in these conditions are compared. Additive White Gaussian Noise (AWGN) is assumed and further that the standard deviation of the noise is known (procedures for noise estimation are described in next chapter).

The influence of the noise level (standard deviation) on the threshold parameters is explored via a series of simulations. Six different levels of noise standard deviation in the range [5, 15] were added to an image. The range of noise levels was chosen to cover the range expected in real applications. Table 5.6 shows the parameters obtained for six of the noise levels utilising MSE and WMSE as the cost functions. Figure 5-7 plots the dependence of these parameters[4] on the noise level emphasising the approximately linear dependence between the threshold parameters and the noise level.

[4] The 9 parameters respectively related to the 9 frequency bands are considered. The other 2 parameters related to all the band have an almost constant shape.

Unweighted optimisation	**K1y**	**K2**	**K1x HH**	**K1x HL**	**K1x LH**	**K1x LL HH**	**K1x LL HL**	**K1x LL LH**	**K1x LL LL HH**	**K1x LL LL HL**	**K1x LL LL LH**
Noise level 5	0.09	3.62	4.31	2.68	2.22	0.73	0.73	0.87	0.32	0.35	0.28
Noise level 7	0.07	3.47	5.99	3.40	3.25	0.94	0.99	1.11	0.43	0.49	0.51
Noise level 9	0	3.65	7.09	4.46	4.25	1.17	1.31	1.23	0.57	0.58	0.76
Noise level 11	0.01	3.61	15.42	6.98	5.06	1.29	1.52	1.54	0.65	0.69	0.81
Noise level 13	0.02	3.54	25.64	8.70	6.47	1.57	1.79	1.97	0.78	0.79	0.92
Noise level 15	0.01	3.67	51.87	10.39	7.83	1.99	2.04	2.39	0.92	0.98	0.97
Weighted optimisation											
Noise level 5	0.08	4.30	4.60	1.08	1.11	1.13	1.24	1.26	0.40	0.44	0.46
Noise level 7	0.24	2.30	12.18	2.01	2.59	1.19	1.45	1.48	0.48	0.49	0.54
Noise level 9	0.33	2.33	14.05	3.13	3.74	1.29	1.80	1.82	0.62	0.60	0.67
Noise level 11	0.41	2.51	33.67	4.91	4.87	1.46	2.07	2.11	0.69	0.68	0.78
Noise level 13	0.46	2.54	52.01	5.23	6.62	1.79	2.40	2.44	0.79	0.79	0.99
Noise level 15	0.47	2.67	72.66	7.29	8.09	2.81	2.71	2.72	0.88	0.89	1.00

Table 5.6. Parameter dependence on the noise levels for unweighted and weighted optimisations.

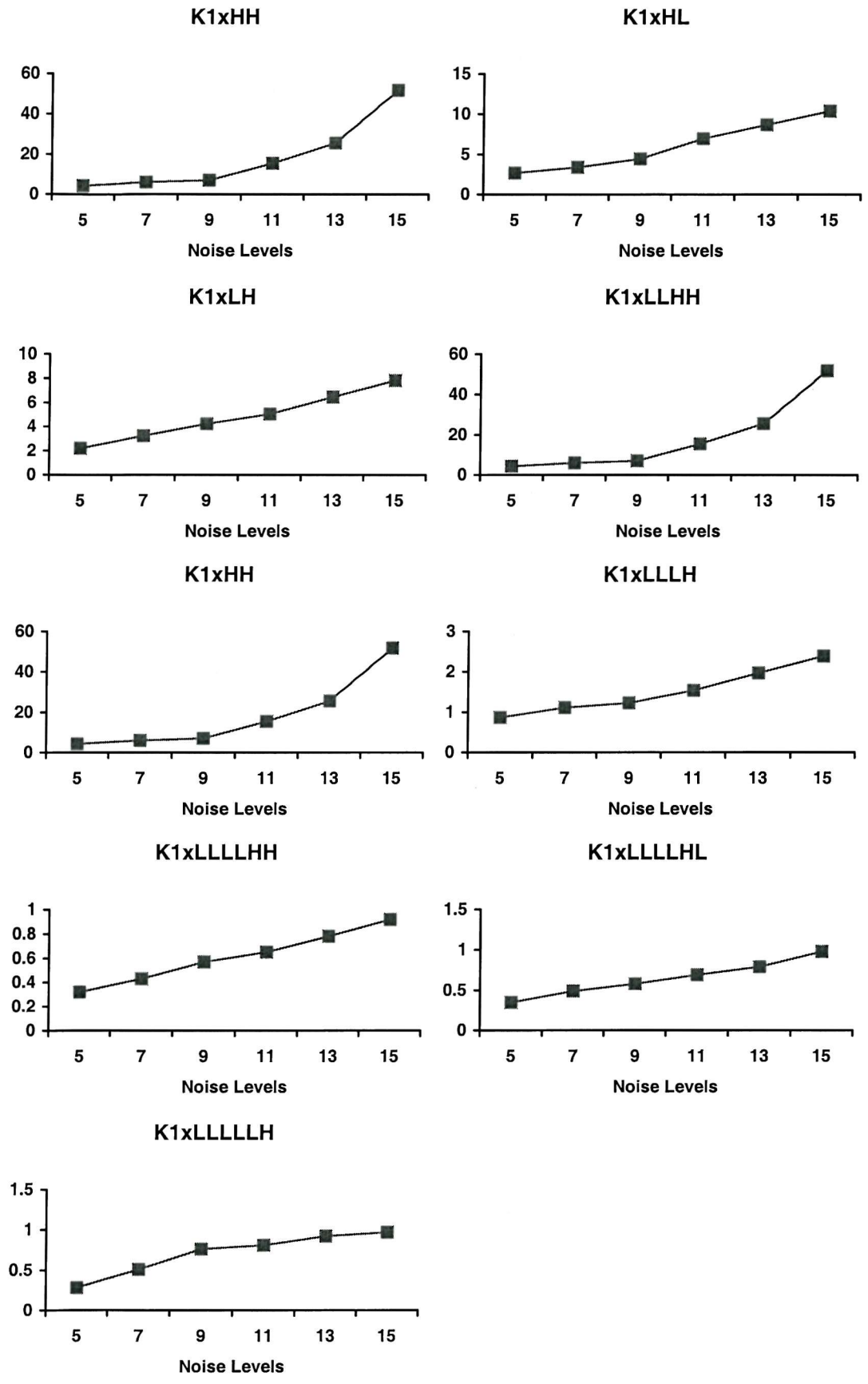


Figure 5-7-a. Relation between parameters and noise levels (unweighted optimisation, 11 parameters, and Nelder algorithm).

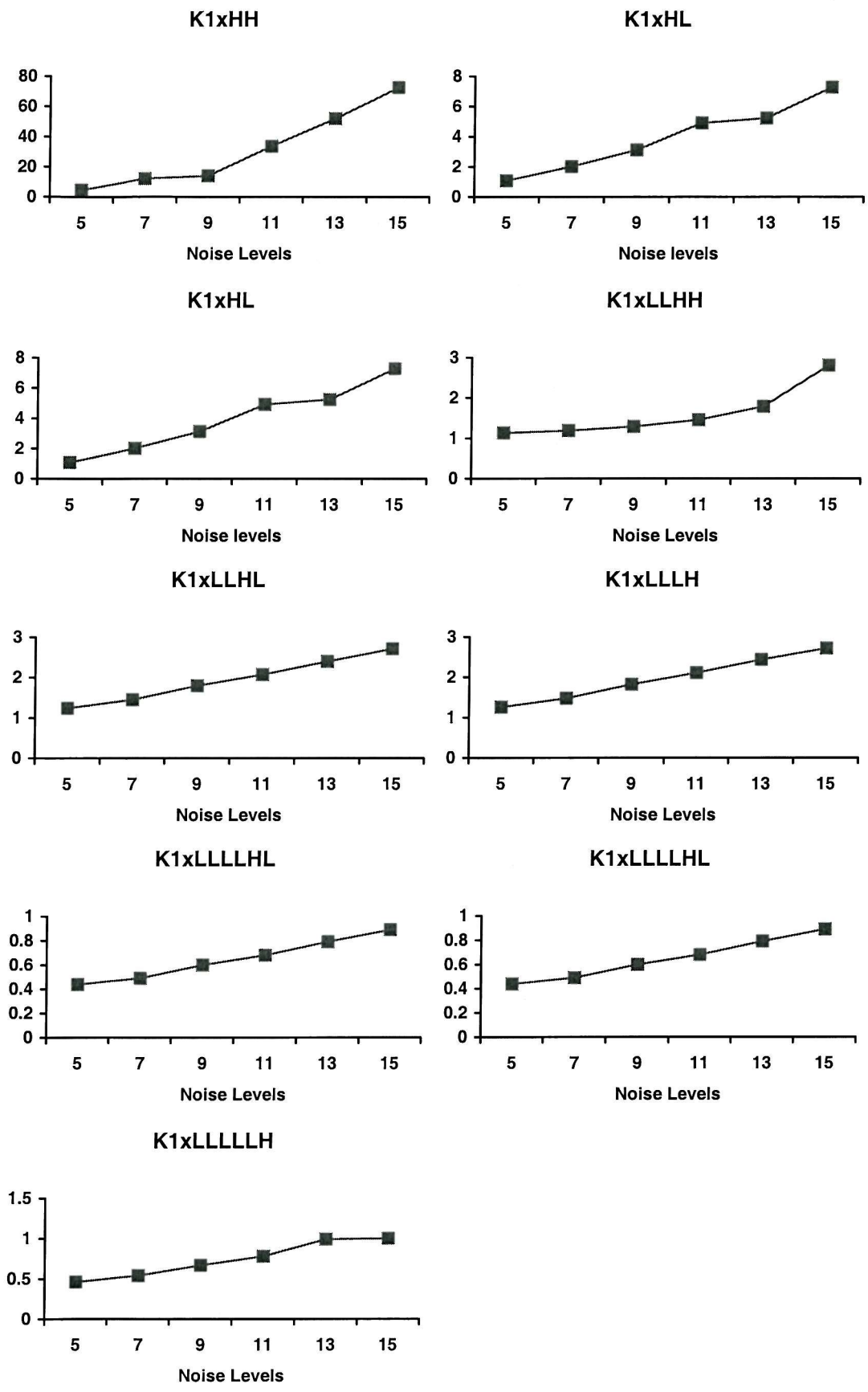


Figure 5-7-b. Relation between parameters and noise levels (weighted optimisation, 11 parameters, and Nelder algorithm).

To assess the robustness of these results with regard to the realisation of the noise, threshold parameters were computed for five different noise realisations (for each of the six noise levels) minimising the WMSE. The means and standard deviations of the parameters are computed. Table 5.7 lists these for each parameter and uses the ratio (normalised standard deviation)

$$\mathbf{var\%} = \frac{\mathbf{standard\ deviation}}{\mathbf{mean}} \qquad (5.25)$$

as an index of significance of the noise realisation on each parameter. This table demonstrates that the parameters related to the first level of decomposition are least robust.

		K1y	K2	K1x HH	K1x HL	K1x LH	K1x LL HH	K1x LL HL	K1x LL LH	K1x LL LL HH	K1x LL LL HL	K1x LL LL LH
Noise level 5	**Mean**	0.01	1.87	12.94	0.10	0.09	1.16	1.28	1.29	0.41	0.47	0.43
	Std	0.03	0.12	4.81	0.06	0.06	0.05	0.08	0.04	0.02	0.02	0.02
	Var %	27.63	6.67	37.18	59.58	70.30	4.88	6.52	3.64	5.94	5.60	6.77
Noise level 7	**Mean**	0.24	2.11	13.84	0.11	0.10	1.39	1.56	1.50	0.49	0.51	0.55
	Std	0.03	0.13	5.57	0.08	0.11	0.07	0.09	0.04	0.03	0.05	0.05
	Var %	13.27	6.44	40.28	72.24	118.5	5.32	5.96	2.92	7.09	11.70	10.16
Noise level 9	**Mean**	0.37	2.41	17.78	0.12	0.11	1.76	1.78	1.75	0.55	0.63	0.60
	Std	0.02	0.07	7.15	0.07	0.031	0.03	0.04	0.07	0.01	0.02	0.03
	Var %	6.54	2.92	40.26	55.64	27.73	1.85	2.31	4.52	2.66	4.12	5.65
Noise level 11	**Mean**	0.43	2.59	46.88	0.13	0.11	2.04	2.06	2.05	0.66	0.69	0.72
	Std	0.02	0.08	9.40	0.04	0.05	0.04	0.03	0.05	0.04	0	0.03
	Var %	5.06	3.40	20.05	34.49	48.49	2.32	1.77	2.85	6.88	1.28	5.22
Noise level 13	**Mean**	0.49	2.60	52.53	0.13	0.12	2.42	2.44	2.35	0.75	0.83	0.85
	Std	0.03	0.15	11.91	0.07	0.08	0.08	0.12	0.08	0.03	0.05	0.03
	Var %	7.85	5.94	22.67	54.93	67.16	3.61	5.11	3.79	5.20	6.09	3.75
Noise level 15	**Mean**	0.53	2.68	80.67	0.14	0.13	2.82	2.65	2.57	0.86	0.90	1.00
	Std	0.02	0.13	13.97	0.09	0.11	0.02	0.06	0.05	0.06	0.06	0.07
	Var %	3.84	4.88	17.32	66.48	84.07	1.01	2.26	2.28	7.94	7.34	7.02

Table 5.7. Analysis of the noise realisation on the parameters.

This work concerns video images and this is a category that can include images having disparate spectral distributions. Therefore the generalisation of the parameters for such a large class is not a trivial task. A set of seven images having various spectral distributions was chosen as example images. The optimal thresholding parameters for those images were

computed for six noise levels. This subset was used because it represents a diverse image set. The relation between the size of each parameter and the energetic content of the related frequency component was investigated. The relation between confidence intervals for the parameters and the corresponding variance of the performance was explored. The customised Nelder method was used to minimise the WMSE, six noise levels were considered, and three levels of decomposition were used. Tables 5.8 show the parameters obtained by training on seven different images, demonstrating, unsurprisingly, that the parameters differ depending on the training image used, i.e. the spectral content of the image influences the results. However in chapter 7 it will be shown that the performance of the algorithm is not critically dependent upon using the parameters obtained by training on that image.

Noise level 5	K1y	K2	K1x HH	K1x HL	K1x LH	K1x LL HH	K1x LL HL	K1x LL LH	K1x LL LL HH	K1x LL LL HL	K1x LL LL LH
Test	0.08	4.30	4.60	1.08	1.11	1.13	1.24	1.26	0.40	0.44	0.46
Graph	0.58	2.76	11.4	3.21	1.17	0.69	0.42	0.83	0.21	0.26	0.24
Girl	0.20	2.13	9.53	0.02	0.04	0.85	0.81	0.69	0.25	0.28	0.24
Interview	0.04	2.14	23.42	0.01	0	1.17	1.24	1.09	0.43	0.51	0.37
Tree	0.01	1.89	11.39	0	0.40	1.12	1.25	1.27	0.41	0.44	0.45
Text	0.01	1.88	11.39	0	0.38	1.14	1.26	1.26	0.40	0.43	0.45
Synthetic	0.05	1.86	11.39	0	0.39	1.15	1.25	1.27	0.40	0.43	0.47

Table 5.8-a. Parameter comparison for noise level 5.

Noise level 7	K1y	K2	K1x HH	K1x HL	K1x LH	K1x LL HH	K1x LL HL	K1x LL LH	K1x LL LL HH	K1x LL LL HL	K1x LL LL LH
Test	0.24	2.30	12.18	2.01	2.59	1.19	1.45	1.48	0.48	0.49	0.54
Graph	0.49	3.05	11.84	6.65	2.61	1.59	0.86	2.25	0.41	0.53	0.37
Girl	0.28	2.15	10.82	4.94	0.03	1.38	1.19	1.61	0.52	0.63	0.52
Interview	0.09	2.40	24.04	3.57	0	1.49	1.46	1.70	0.60	0.75	0.55
Tree	0.25	2.29	12.19	0	0.59	1.31	1.44	1.47	0.47	0.50	0.53
Text	0.25	2.29	12.18	0.01	0.59	1.32	1.45	1.48	0.50	0.49	0.53
Synthetic	0.25	2.31	12.18	0	0.58	1.32	1.44	1.48	0.49	0.49	0.54

Table 5.8-b. Parameter comparison for noise level 7.

Noise level 9	K1y	K2	K1x HH	K1x HL	K1x LH	K1x LL HH	K1x LL HL	K1x LL LH	K1x LL LL HH	K1x LL LL HL	K1x LL LL LH
Test	0.33	2.33	14.05	3.13	3.74	1.29	1.80	1.82	0.62	0.60	0.67
Graph	0.53	3.34	13.62	7.96	6.71	1.98	1.91	1.83	0.57	0.66	0.86
Girl	0.35	1.84	13.39	7.62	7.67	1.71	2.02	1.67	0.78	0	0.86
Interview	0.21	2.04	26.14	6.21	6.06	1.88	2.39	2.08	0.85	0.98	0.85
Tree	0.34	2.33	14.06	0.01	0.74	1.64	1.79	1.83	0.63	0.60	0.68
Text	0.30	2.33	14.06	0.01	0.75	1.66	1.81	1.84	0.62	0.62	0.67
Synthetic	0.33	2.35	14.07	0	0.74	1.66	1.81	1.82	0.62	0.60	0.67

Table 5.8-c. Parameter comparison for noise level 9.

Noise level 11	K1y	K2	K1x HH	K1x HL	K1x LH	K1x LL HH	K1x LL HL	K1x LL LH	K1x LL LL HH	K1x LL LL HL	K1x LL LL LH
Test	0.41	2.51	33.67	4.91	4.87	1.46	2.07	2.11	0.69	0.68	0.78
Graph	0.45	2.51	17.86	9.87	8.81	2.51	2.45	2.81	0.82	0.74	0.95
Girl	0.30	1.81	11.56	9.69	10.37	2.17	2.33	2.15	0.90	1.14	0.94
Interview	0.19	1.96	34.56	8.59	7.93	2.38	2.74	2.53	0.99	1.34	1.05
Tree	0.41	2.52	33.66	0.01	0.88	2.01	2.06	2.12	0.69	0.70	0.77
Text	0.41	2.52	33.68	0.01	0.88	1.98	2.06	2.12	0.69	0.67	0.79
Synthetic	0.42	2.50	33.67	0.01	0.85	1.99	2.05	2.11	0.70	0.70	0.78

Table 5.8-d. Parameter comparison for noise level 11.

Noise level 13	K1y	K2	K1x HH	K1x HL	K1x LH	K1x LL HH	K1x LL HL	K1x LL LH	K1x LL LL HH	K1x LL LL HL	K1x LL LL LH
Test	0.46	2.54	52.01	5.23	6.62	1.79	2.40	2.44	0.79	0.79	0.99
Graph	0.39	2.65	21.67	11.23	10.84	3.11	3.08	2.90	0.87	0.86	1.46
Girl	0.23	1.92	19.87	11.54	12.07	2.41	2.63	2.38	1.00	1.23	1.10
Interview	0.16	1.91	53.93	10.45	9.5	2.80	3.19	3.08	1.14	1.58	1.24
Tree	0.44	2.54	52.09	0.02	0.91	2.41	2.40	2.45	0.78	0.80	0.90
Text	0.45	2.54	52.11	0.01	0.89	2.40	2.41	2.46	0.81	0.78	0.91
Synthetic	0.47	2.53	52.1	0.01	0.92	2.41	2.40	2.44	0.78	0.78	0.89

Table 5.8-e. Parameter comparison for noise level 13.

Noise level 15	K1y	K2	K1x HH	K1x HL	K1x LH	K1x LL HH	K1x LL HL	K1x LL LH	K1x LL LL HH	K1x LL LL HL	K1x LL LL LH
Test	0.47	2.67	72.66	7.29	8.09	2.81	2.71	2.72	0.88	0.89	1.00
Graph	0.35	2.08	25.46	13.33	13.07	3.80	4.25	3.47	1.09	0.97	1.77
Girl	0.19	1.95	27.27	13.11	14.47	2.97	3.01	2.74	1.10	1.44	1.18
Interview	0.13	1.77	59.37	13.23	11.62	3.32	3.65	3.60	1.35	1.89	1.52
Tree	0.51	2.61	71.85	0	1.09	2.82	2.69	2.7	0.88	0.89	1.00
Text	0.50	2.59	71.85	0	1.10	2.80	2.71	2.72	0.88	0.88	0.99
Synthetic	0.49	2.59	71.84	0.01	1.11	2.82	2.71	2.71	0.87	0.88	1.00

Table 5.8-f. Parameter comparison for noise level 15.

Chapter 6

Noise estimation

6.1 Introduction

Noise level estimation, i.e. estimation of the noise standard deviation level, is needed on each component to select the appropriate set of thresholding function parameters. Images are non-stationary, two-dimensional processes, and may in general be corrupted by additive, impulse, or signal dependent noise. This work assumes that the noise is additive, Gaussian and either white or coloured.

This chapter initially describes noise contaminated image models and specifies the assumptions made herein. Subsequently in section 6.3, three methods for noise level estimation are presented, and the results of these estimation procedures are listed and commented on in the last section.

6.2 Noise contaminated image models

The general model of a noise contaminated image system can be described by [7, 13, 100, 101, 102]:

$$v(m,n) = [u(m,n) + \eta_1(m,n)]\eta_2(m,n) \qquad (6.1)$$

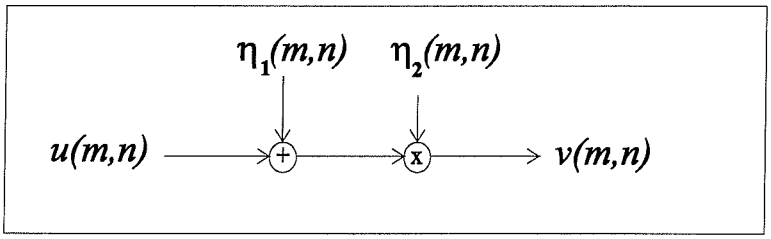


Figure 6-1.General model of the image system.

where $v(m,n)$ is the observed image, $u(m,n)$ is the original image, $\eta_1(m,n)$ is an additive noise, and $\eta_2(m,n)$ is a multiplicative noise. This study initially considers zero mean Additive White Gaussian Noise (AWGN). The less restrictive hypothesis of coloured Gaussian noise is considered later.

In the AWGN case the image model is:

$$v(m,n) = u(m,n) + \eta_w(m,n) \qquad (6.2)$$

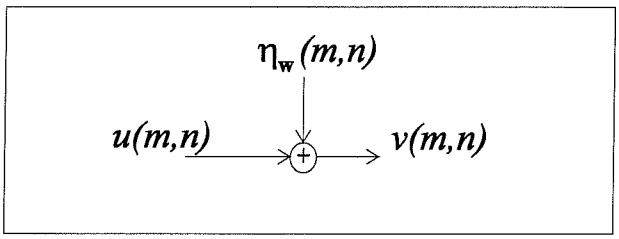


Figure 6-2. Model of the image system for additive noise.

Where $\eta_w(m,n)$ is AWGN and it is characterised by a single parameter: its standard deviation (the mean is zero by assumption). In the following section three methods for estimating the standard deviation of the noise contaminating the high frequency band are described. The high frequency band[1] is used because the image is in general smallest in this band and therefore one can better discern the noise. Under the AWGN assumption knowledge of the noise standard deviation in one frequency band allows the computation of the overall noise variance.

When the less restrictive hypothesis of coloured Gaussian noise is assumed then the image model is:

$$v(m,n) = u(m,n) + \eta_c(m,n) \qquad (6.3)$$

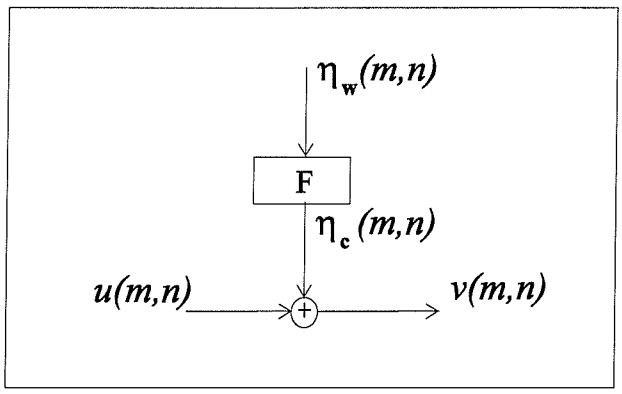
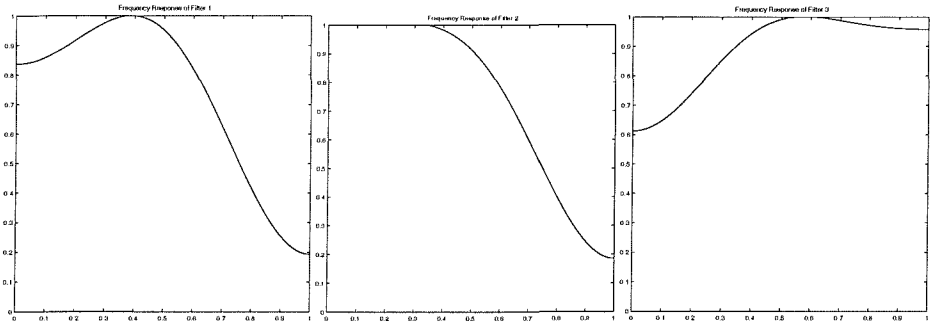


Figure 6-3. Model of the image system for coloured noise.

Where $\eta_w(m,n)$ is AWGN and $\eta_c(m,n)$ is the filtered (coloured) noise and it is characterised by its spectral distribution. In tests, three spectral distributions for the noise have been used; the frequency responses of the three filters used to generate the coloration are shown in figures 6-4. These filters have been designed in order to generate a range of spectral distributions for the noise.

[1] This corresponds to the coarsest scale in the wavelet decomposition



Figures 6-4. Filters used to simulate coloured noise: a) band-pass, b) low and c) high-pass.

The methods described in the following section are not specifically designed for coloured noise scenarios where the noise standard deviation should at least be estimated in each band individually. The spectral distribution of the noise may be known a priori or may be deduced using two consecutive video frames (see figure 6-5). In this figure **Nw1** and **Nw2** are two white noise realisations; **Imn1** and **Imn2** are the two frames contaminated with coloured noises derived by filtering **Nw1** and **Nw2**; **ImAv** is the image obtained averaging **Imn1** and **Imn2**. This latter procedure allows one to compute the ratios of the standard deviation of the noise in any frequency band.

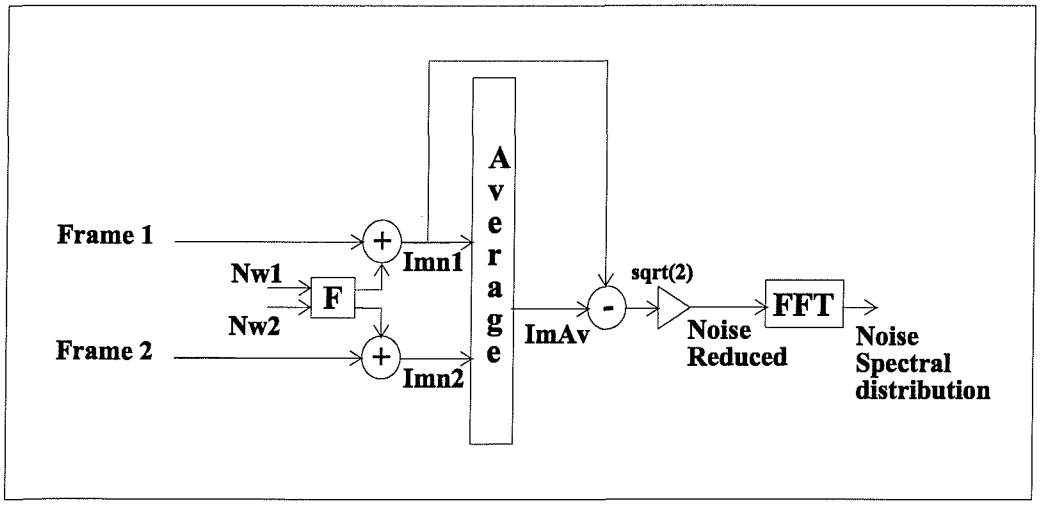


Figure 6-5. Noise spectral distribution estimation.

6.3 Noise level estimation in the high frequency band

In this section, three methods for noise level estimation are considered. These methods can be applied directly assuming AWGN or as part of the procedure for coloured noise estimation. Where convenient it shall be assumed that the component under consideration represents the highest frequency band as this is the one which is used when the noise is assumed to be white.

6.3.2 Noise level estimation using cumulative distribution functions (cdfs)

One method to estimate the noise levels is based on the Cumulative Distribution Functions (cdfs) [110, 111]. The cdf is computed for the squared, low passed pixel values (see figure 6-6). This operation computes local estimates of the local energy. The cdf value for a value h is defined as the proportion of pixels for which the local energy is less than h.

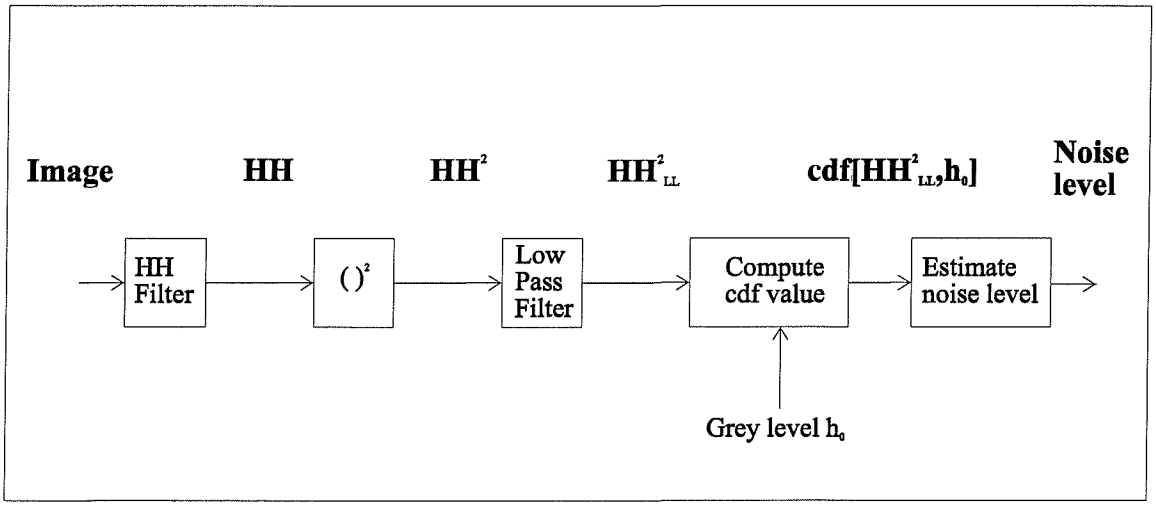


Figure 6-6. Noise estimation using the cdfs evaluation.

For a given point h the cdf value depends on the noise level. The cdfs of one image for six different noise levels are shown in figure 6-7. However the form of this relationship depends on the image and on the value of h. The cdfs of 13 different images for the same noise level are shown in figure 6-8.

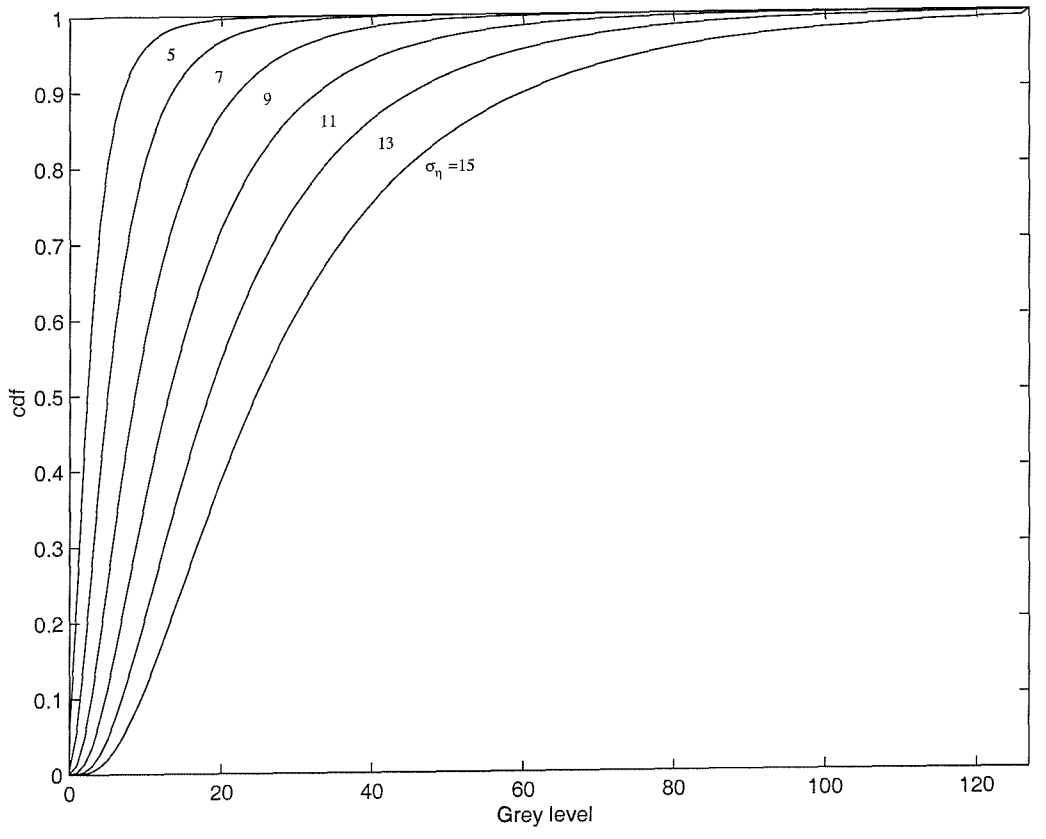


Figure 6-7. Cdfs of an image for different noise levels.

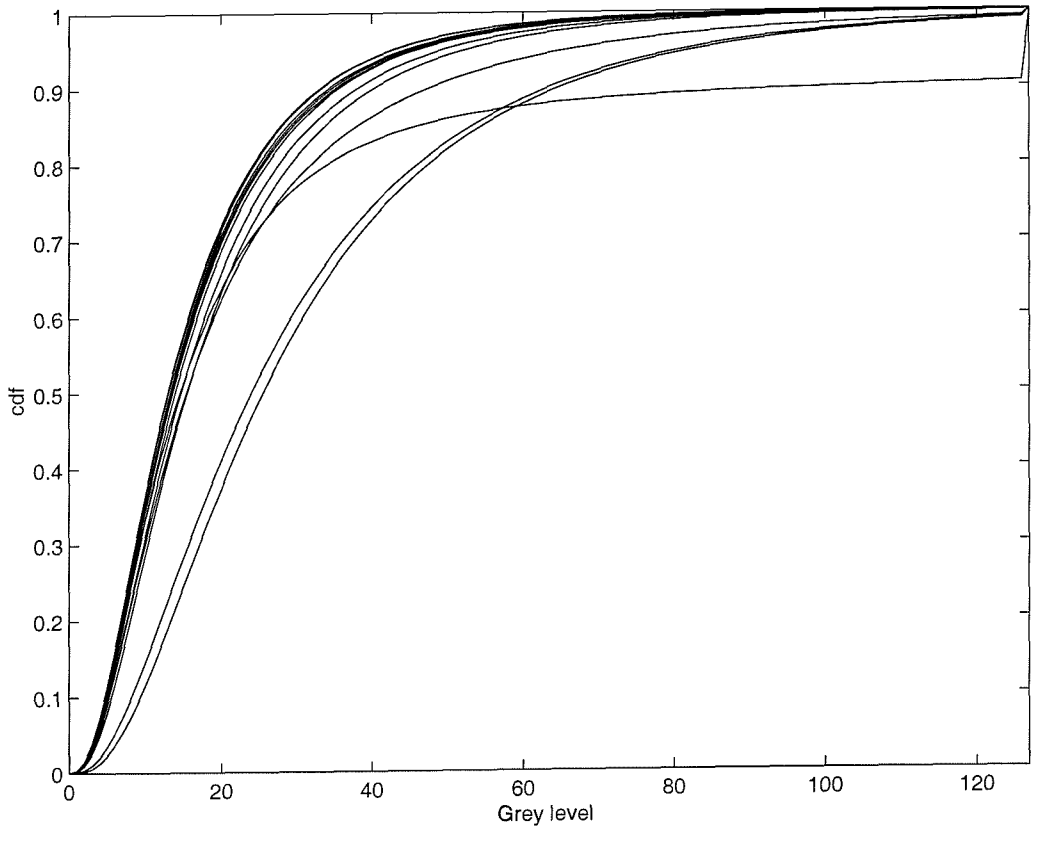


Figure 6-8. Cdfs of 13 images contaminated with noise std level 15.

Consider the problem of determining the grey level h_0 for which the image cdf permits the best estimate of the noise level over all the training set.

Initially the optimal grey level for discrimination between pairs of noise levels, $h_0^*(l_1, l_2)$, is determined using the function

$$f_{l_1,l_2}(h) = \frac{\left|m_{l_1}(h) - m_{l_2}(h)\right|}{\sqrt{\sigma_{l_1}(h) + \sigma_{l_2}(h)}} \qquad (6.4)$$
$$h_0^*(l_1, l_2) = arg\max_h(f_{l_1,l_2}(h))$$

where l_1 and l_2 are the two noise levels, $m_l(h)$ and $\sigma_l(h)$, are the mean and the standard deviation of the cdf at a grey level h computed across the set of images.

The optimal grey level, h_0, over the set of noise levels is determined using the function:

$$f_{tot}(h) = \sum_{i,j} f_{l_i,l_j}(h) \qquad i < j$$
$$h_0 = arg\max_h(f_{tot}(h)) \qquad (6.5)$$

where the summation is taken across all the noise levels considered. To determine h_0, 13 images at 6 noise levels were used. The noise level is estimated by evaluating the cdf at h_0 and comparing this with the range of cdf values obtained by training over the sets of images and noise levels.

This method assumes that the set of 13 images is representative and that the noise level is in the range considered. The main advantage of this method is that it uses information from grey levels where the image has least influence, i.e. planar areas in the image. Moreover, the image components are not assumed to have any particular statistical distribution.

6.3.2 Noise level estimation using moment matching

A second method for estimating the noise level is to assume a statistical distribution for the noise free image component [110, 111]. For example one may assume that the component has a zero mean Laplacian distribution[2] with standard deviation σ_u :

$$p(u) = \frac{\sqrt{2}}{\sigma_u} e^{-\frac{\sqrt{2}}{\sigma_u}|u|} \quad (6.6)$$

The noise component is assumed to have a zero mean Gaussian distribution with standard deviation σ_η :

$$p(n) = \frac{1}{\sigma_\eta \sqrt{2\pi}} e^{-\frac{n^2}{2\sigma_\eta^2}} = p(v-u) = \frac{1}{\sigma_\eta \sqrt{2\pi}} e^{-\frac{(v-u)^2}{2\sigma_\eta^2}} \quad (6.7)$$

Assuming independence between the original image and noise, it is possible to obtain an expression for the distribution of the noisy image component as:

$$p(v) = \int_{-\infty}^{\infty} p(v-u)p(u)du \quad (6.8)$$

Substituting (6.6) and (6.7) in (6.8) and considering the symmetry of both the distributions, it is possible to solve the integral[3], with the result:

$$p(v) = \frac{e^{\frac{\sigma_\eta^2}{\sigma_u^2}}}{\sigma_u \sqrt{2}} \left\{ e^{\frac{\sqrt{2}}{\sigma_u} v} erfc\left(\frac{\sigma_\eta}{\sigma_u} + \frac{v}{\sigma_\eta \sqrt{2}} \right) + e^{-\frac{\sqrt{2}}{\sigma_u} v} erfc\left(\frac{\sigma_\eta}{\sigma_u} - \frac{v}{\sigma_\eta \sqrt{2}} \right) \right\} \quad (6.9)$$

[2] This assumption is realistic in particular for the HH component
[3] The solution is detailed in appendix B.

One can use the method of matching moments [10, 12] to fit the model to the data. One can estimate the variance $E\{v^2\}$, and the absolute moment $E\{|v|\}$ from a component of the noisy image. It is possible to compare these with the corresponding theoretical expressions evaluated from the distribution $p(v)$. Thus the noise levels can be estimated by solving the resulting system of two equations, for σ_u and σ_η:

$$E\{v^2\} = \int_{-\infty}^{\infty} v^2 p(v)dv$$
$$E\{|v|\} = \int_{-\infty}^{\infty} |v| p(v)dv \qquad (6.10)$$

The approach can be generalised when the component of the original image has a statistical distribution which is a member of the standardised normal distribution family. In this case the equivalent of (6.6) has the form:

$$p(u, \sigma_u, \beta_u) = \frac{\omega(\beta_u)}{\sigma_u} e^{-c(\beta_u)\left|\frac{u}{\sigma_u}\right|^{\frac{2}{1+\beta_u}}} \qquad -\infty < u < \infty$$
$$\omega(\beta_u) = \frac{\left\{\Gamma\left[\frac{3}{2}(1+\beta_u)\right]\right\}^{\frac{1}{2}}}{(1+\beta_u)\left\{\Gamma\left[\frac{1}{2}(1+\beta_u)\right]\right\}^{\frac{3}{2}}} \qquad \sigma_u > 0 \qquad (6.11)$$
$$c(\beta_u) = \left\{\frac{\Gamma\left[\frac{3}{2}(1+\beta_u)\right]}{\Gamma\left[\frac{1}{2}(1+\beta_u)\right]}\right\}^{\frac{1}{1+\beta_u}} \qquad -1 < \beta_u < 1$$

see (4.21) for detail. In this case the evaluation of the integral (6.8) in closed form is not trivial. Nevertheless it is possible to use numerical techniques to evaluate the noise level

for each value of β_u, and then for each model assumed. This system defines the link between the statistical model assumed and the estimated noise level. Nevertheless this is a system of two equations and three unknowns and cannot be used to estimate the noise level when the statistical distribution of the original component is unknown, i.e. when β_u is unknown. In subsection 6.3.3 the use of other moments is considered to overcome this problem.

6.3.3 Noise level estimation using linear regression

The final method of noise level estimation is once again based on estimating the image moments [110, 111]. In this case one avoids making explicit assumptions about statistical distribution of the noise free image. In this scheme it is assumed that there exists a linear relationship between normalised moments and the noise level. Linear regression is then used to compute the coefficients of the unknown linear relationship based on a set of images at various noise levels. Three normalised moments are used in this regression

$$\begin{aligned} M1 &= E\{|v|\} \\ M2 &= \frac{E\{v^2\}}{M1} \\ M3 &= \frac{E\{v^4\}}{(M2)(M1)^2} \end{aligned} \qquad (6.12)$$

Note that the three regressors $M1$, $M2$ and $M3$ have the same dimensions as the standard deviation of v. The assumed model is

$$\alpha_1 M1 + \alpha_2 M2 + \alpha_3 M3 = \sigma \qquad (6.13)$$

where α_i are the regression coefficients and σ is the true noise level (standard deviation). Given a set of training examples from the K images at H noise levels, one solves the system of equations

$$\mathbf{A}\underline{x} = \mathbf{G} \qquad (6.14)$$

where

$$\mathbf{A} = \begin{pmatrix} M1_{1,1} & M2_{1,1} & M3_{1,1} \\ . & . & . \\ M1_{1,H}. & M2_{1,H}. & M3_{1,H} \\ M1_{2,1} & M2_{2,1} & M3_{2,1} \\ . & . & . \\ M1_{2,H}. & M2_{2,H} & M2_{2,H} \\ . & . & . \\ . & . & . \\ . & . & . \\ . & . & . \\ M1_{K,1} & M2_{K,1} & M3_{K,1} \\ . & . & . \\ M1_{K,H} & M2_{K,H} & M3_{K,H} \end{pmatrix} \quad \underline{x} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} \quad \mathbf{G} = \begin{pmatrix} \sigma_1 \\ \sigma_2 \\ . \\ \sigma_H \\ \sigma_1 \\ \sigma_2 \\ . \\ \sigma_H \\ . \\ \sigma_1 \\ \sigma_2 \\ . \\ \sigma_H \end{pmatrix} \qquad (6.15)$$

in which $M1_{i,j}$, $M2_{i,j}$ and $M3_{i,j}$ are the normalised moments computed from the i^{th} image at the j^{th} noise level. Assuming $K\,H>3$ then the least squares regression is given by

$$\underline{x} = (\mathbf{A}^t\mathbf{A})^{-1}\mathbf{A}^t\mathbf{G} \qquad (6.15)$$

Clearly one could extend this method to use other regressions and/or non linear regression schemes.

6.4 Results of the noise level estimation

Three methods for the estimation of the noise level have been compared along with a combined algorithm. In order to judge the performances of these methods, a set of 13 images has been considered; the images have been contaminated with 6 different levels of AWGN creating a training set of 78 images.

6.4.1 Results of noise level estimation using the cdfs

The first method for noise level estimation is based on the evaluation of the cumulative density function at a particular grey level. A method for computing the optimal level for discriminating between two noise levels is described in section 6.3.1. This level is where the cdf can best discern noise levels. The grey level is selected to optimise the results. The first column of table 6.1 lists the grey levels which are the best discriminators for pairs of noise levels. The remaining columns list the means and the standard deviations of the 13 cdfs for the two noise levels.

Noise levels k - k+1	Noise discriminating grey level	Value of the cdfs mean for noise level k	Value of the cdfs mean for noise level k +1	Value of the cdfs std for noise level k	Value of the cdfs std for noise level k+1
Noise levels 5-7	5	0.5720	0.3105	0.2004	0.1107
Noise levels 7-9	6	0.3974	0.1995	0.1349	0.0683
Noise levels 9-11	10	0.4345	0.2585	0.1240	0.0750
Noise levels 11-13	14	0.4242	0.2715	0.1051	0.0677
Noise levels 13-15	21	0.4801	0.3418	0.0935	0.0690

Table 6.1. Best grey level discriminators for different noise level ranges.

A single overall grey level has been selected to discriminate over the whole range of noise levels using equation (6.5). The best grey level computed for the whole image set is 8. Table 6.2 shows the means and the standard deviations for the 13 images at grey level 8.

Noise levels	Mean of the cdfs value in the best discriminating grey level	Std of the cdfs value in the best discriminating grey level
Noise level 5	0.7443	0.2167
Noise level 7	0.5394	0.1650
Noise level 9	0.3230	0.1011
Noise level 11	0.1699	0.0532
Noise level 13	0.0830	0.0256
Noise level 15	0.0392	0.0121

Table 6.2. Confidence interval using overall best grey level discriminator.

Table 6.4 shows the results of this complete noise level estimation over all the 78 images. In figure 6-9 the results of the estimation are depicted for the 13 images. Figure 6-10 shows the confidence intervals for the method, based on the means and the standard deviations listed in table 6.3.

	Noise level 5	Noise level 7	Noise level 9	Noise level 11	Noise level 13	Noise level 15
Image 1	2.2440	6.0440	8.5229	10.6444	12.7071	14.7620
Image 2	5.4190	7.3563	9.3522	11.3144	13.2737	14.4396
Image 3	3.3043	6.3517	8.6347	10.7784	12.7698	14.7509
Image 4	1.9722	5.8165	8.2757	10.4455	12.5387	14.6345
Image 5	3.0292	6.2144	8.5470	10.6408	12.7632	14.6935
Image 6	1.9293	5.7728	8.2833	10.4469	12.5308	14.5648
Image 7	5.0228	6.9942	8.9700	10.9785	12.9917	14.9677
Image 8	2.4850	6.0254	8.4108	10.5867	12.6308	14.6765
Image 9	9.8879	11.0017	12.6051	13.9580	14.4048	15.9362
Image 10	2.2378	5.9622	8.3711	10.5457	12.5848	14.5871
Image 11	4.7093	6.9263	8.9547	10.9380	12.9750	14.9468
Image 12	8.7155	10.0722	11.7562	13.4849	13.2620	15.3065
Image 13	3.9211	6.4078	8.5904	10.6658	12.7506	14.7737
Average	4.2213	6.9958	9.1749	11.2211	12.9028	14.8492
Std	2.5519	1.6538	1.3795	1.2464	0.4018	0.3931

Table 6.3. Results of noise estimation using the cdfs.

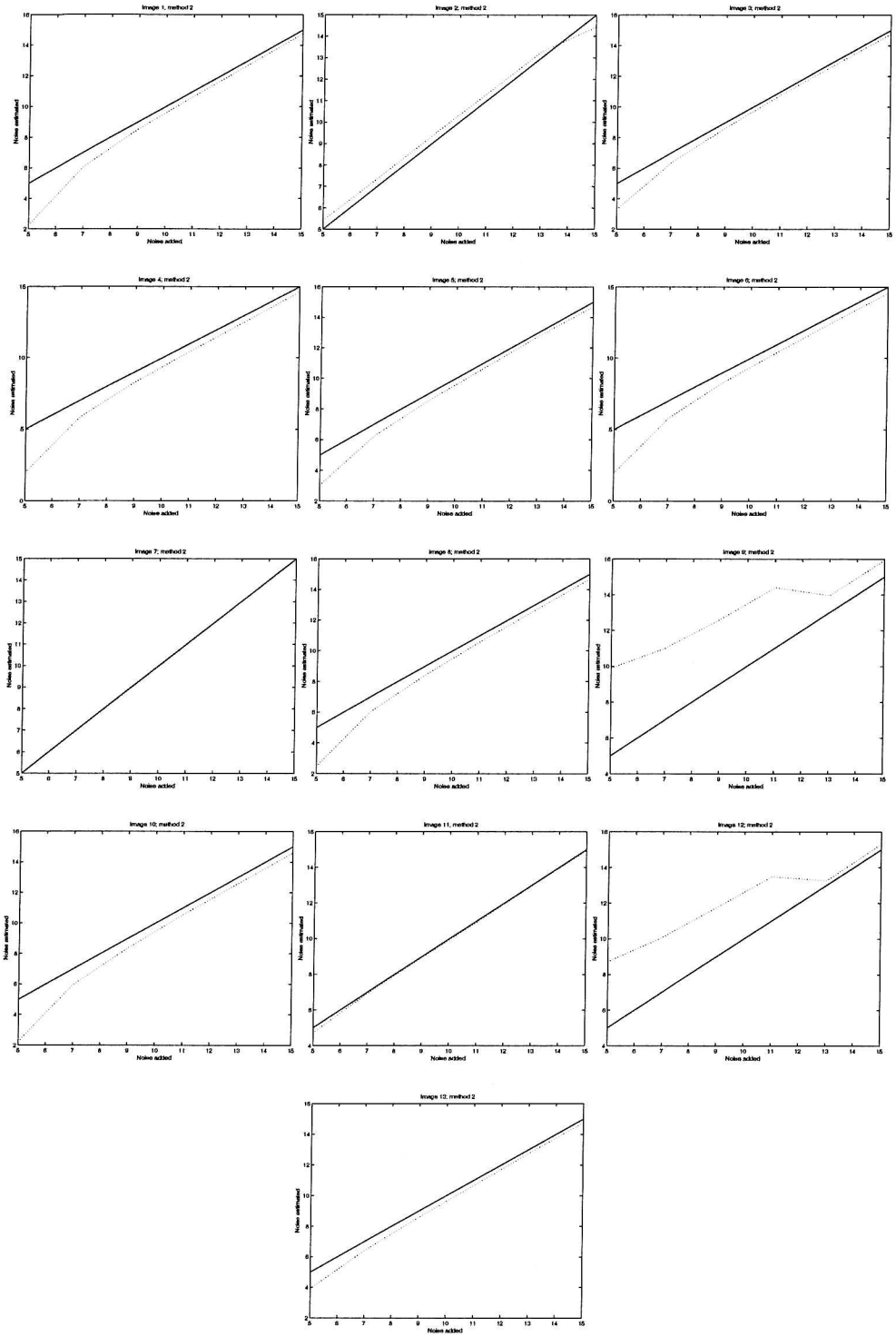


Figure 6-9. Noise estimation on 13 images for 6 noise levels with the cdfs. The continuous line is the standard deviation of the noise introduced and the dotted line is the estimate of the noise level.

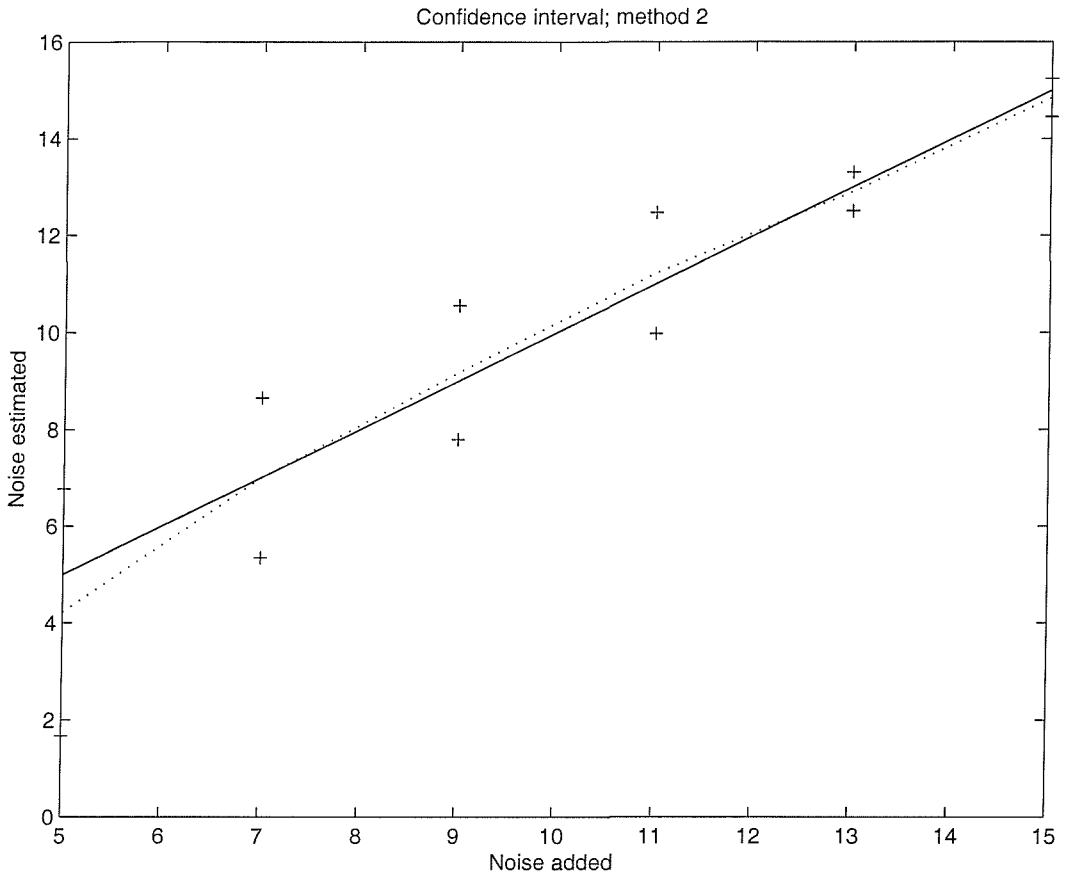


Figure 6-10. Mean (dotted line) ± one standard deviation (+ marks) of the overall estimation, theoretical noise level (continuous line). Cdfs method.

6.4.2 Results of noise level estimation using moment matching

The second method to estimate the noise level, based on the method of matching moments, does not need a priori information but is based on the hypothesis that the highest frequency component in the image has a prescribed distribution; here a Laplacian distribution is assumed. The moments chosen for matching are the first absolute moment and the variance. The method was then applied to the 78 images to estimate the noise level. Table 6.4 shows the results of these trials. The last two lines of the table indicate the average and standard deviation of the estimates computed over each image. The precision of the estimate for each image is depicted in figure 6-11, the confidence intervals of the estimates are illustrated in figure 6-12.

	Noise level 5	**Noise level 7**	**Noise level 9**	**Noise level 11**	**Noise level 13**	**Noise level 15**
Image 1	5.2013	7.2694	9.1013	10.9674	12.9027	14.7023
Image 2	2.8428	5.5582	7.9521	10.1012	12.3460	14.4396
Image 3	3.4872	5.9613	8.2249	10.4253	12.5388	14.3896
Image 4	4.1473	6.5895	8.5812	10.5792	12.8318	15.0879
Image 5	1.6135	5.0445	7.6744	9.8699	12.0782	14.2641
Image 6	4.2636	6.4677	8.5440	10.4837	12.9867	14.9589
Image 7	4.6973	1.3955	5.8718	8.6760	11.0853	13.3639
Image 8	1.8106	5.3398	7.7521	10.0738	12.1033	14.4289
Image 9	7.3575	8.8441	10.4749	12.2460	13.9580	15.9362
Image 10	3.3116	5.8921	8.1085	10.4486	12.4147	14.4152
Image 11	2.0230	5.2571	7.7222	9.9718	12.2276	14.2323
Image 12	5.5926	7.4604	9.3333	11.3060	13.2620	15.3065
Image 13	10.0984	12.7865	14.4631	16.0786	18.3421	20.0096
Average	4.3421	6.4512	8.7541	10.8636	13.0059	15.0412
Std	2.3800	2.5677	2.0200	1.7719	1.7431	1.6152

Table 6.4. *Results of noise estimation using moment matching.*

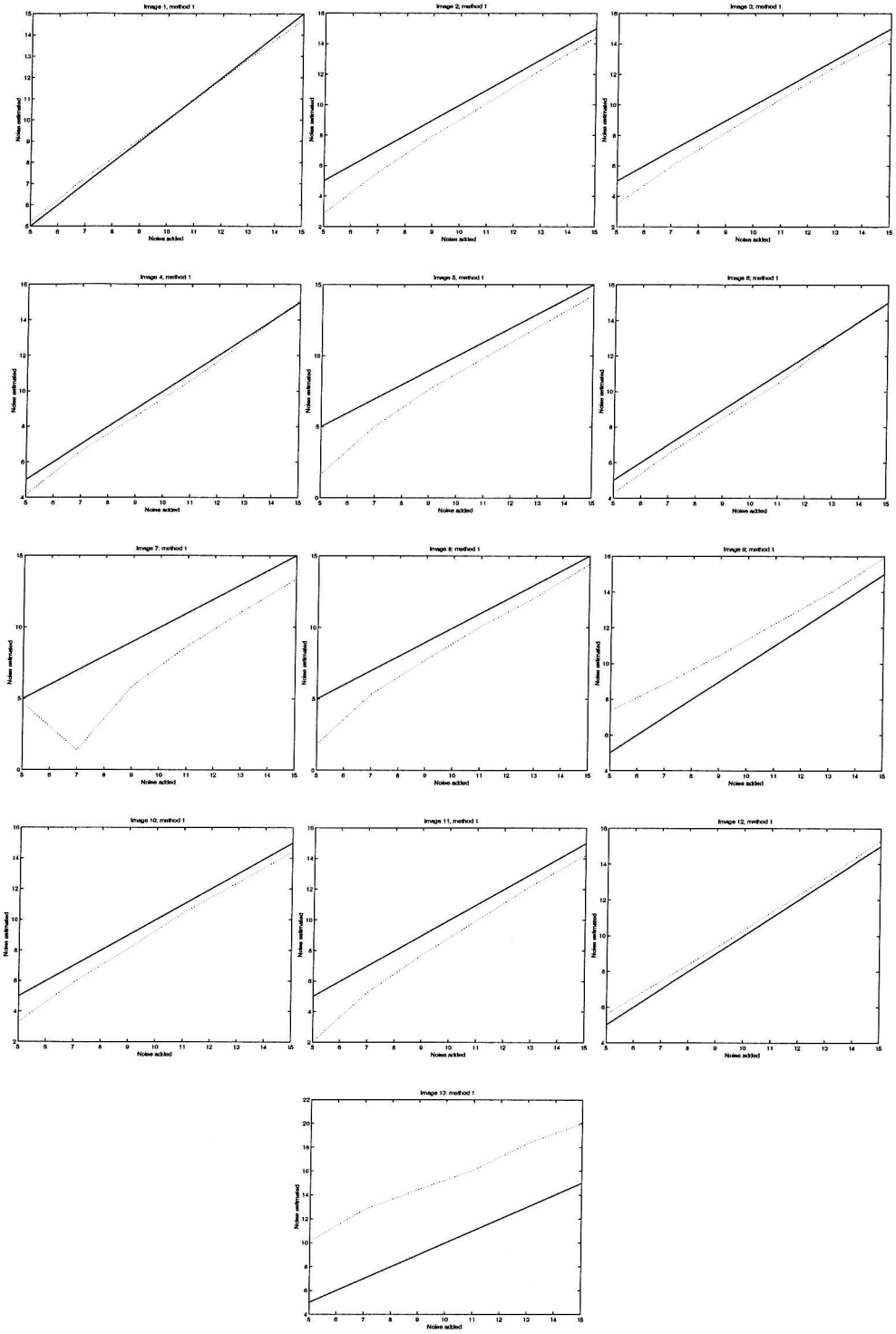


Figure 6-11. Noise estimation on 13 images for 6 noise levels with the moment matching. The continuous line is the standard deviation of the noise introduced and the dotted line is the estimate of the noise level.

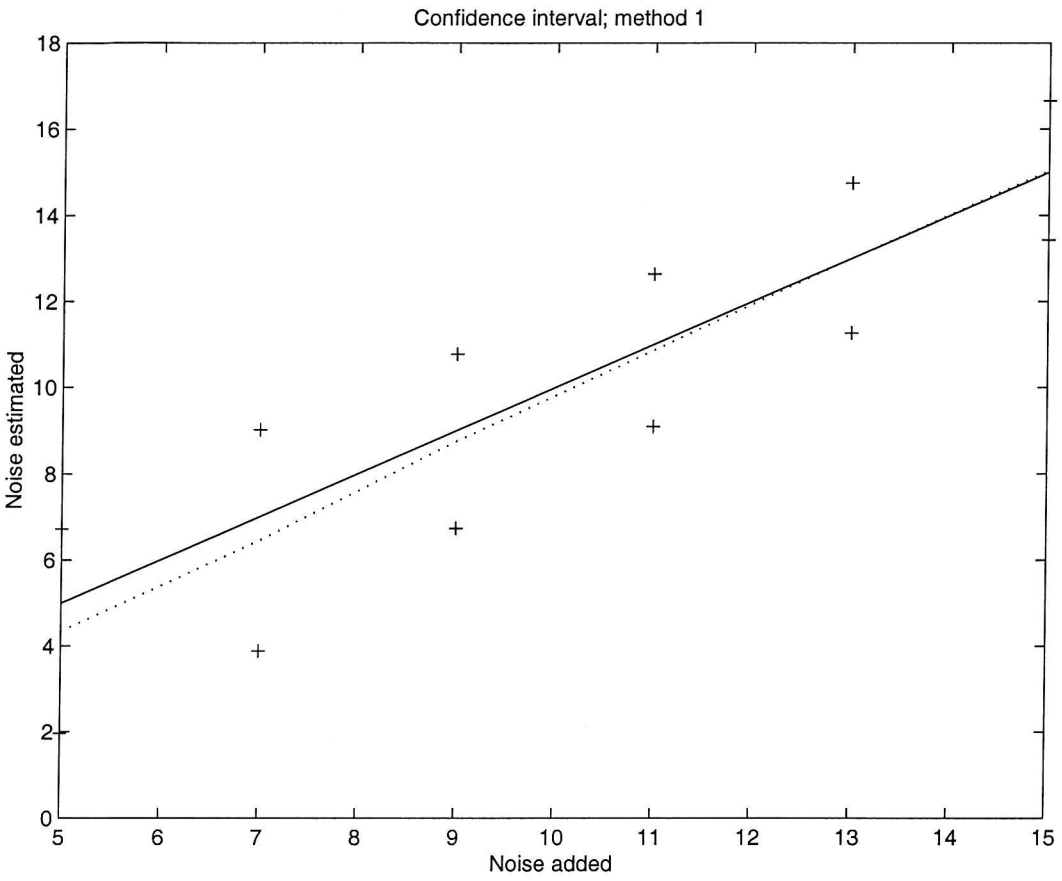


Figure 6-12. *Mean (dotted line) ± one standard deviation (+ marks) of the overall estimation, theoretical noise level (continuous line). Moment matching method.*

6.4.3 Results of noise level estimation using linear regression

This method is based on a linear regression applied onto a set of normalised moments. The full training set is used to generate an overdetermined system of 78 equations. Table 6.5 shows the results applying this method to the complete training data. The last two rows of the table show the mean and standard deviation of the estimates describing the overall performance of the method. Figure 6-13 shows the results of the estimates for the 13 images. Figure 6-14 shows the corresponding confidence interval.

	Noise level 5	**Noise level 7**	**Noise level 9**	**Noise level 11**	**Noise level 13**	**Noise level 15**
Image 1	4.9610	6.5661	8.2703	9.9522	11.7014	13.3561
Image 2	5.8633	7.3789	8.9898	10.6231	12.2932	13.9885
Image 3	5.0657	6.7194	8.4147	10.1510	11.8612	13.4815
Image 4	4.6870	6.4360	8.1266	9.8500	11.6165	13.3596
Image 5	4.9309	6.6318	8.3761	10.0574	11.8348	13.5300
Image 6	4.7159	6.4042	8.1332	9.8557	11.5837	13.3395
Image 7	5.7453	7.4559	9.1359	10.8505	12.5704	14.1921
Image 8	4.7760	6.5315	8.2532	10.0185	11.7470	13.4619
Image 9	9.7065	10.7464	11.9210	13.2750	14.6526	16.1291
Image 10	4.7857	6.5021	8.1950	9.9839	11.6797	13.3952
Image 11	5.3995	7.0515	8.7151	10.3712	12.0687	13.7574
Image 12	8.8637	10.0721	11.4235	12.8251	14.2949	15.8255
Image 13	4.5612	6.5403	8.4841	10.4101	12.4488	14.3321
Average	5.6970	7.3105	8.9568	10.6326	12.3348	14.0114
Std	1.6508	1.4239	1.2494	1.1178	1.0036	0.9344

Table 6.5. Results of noise estimation using linear regression.

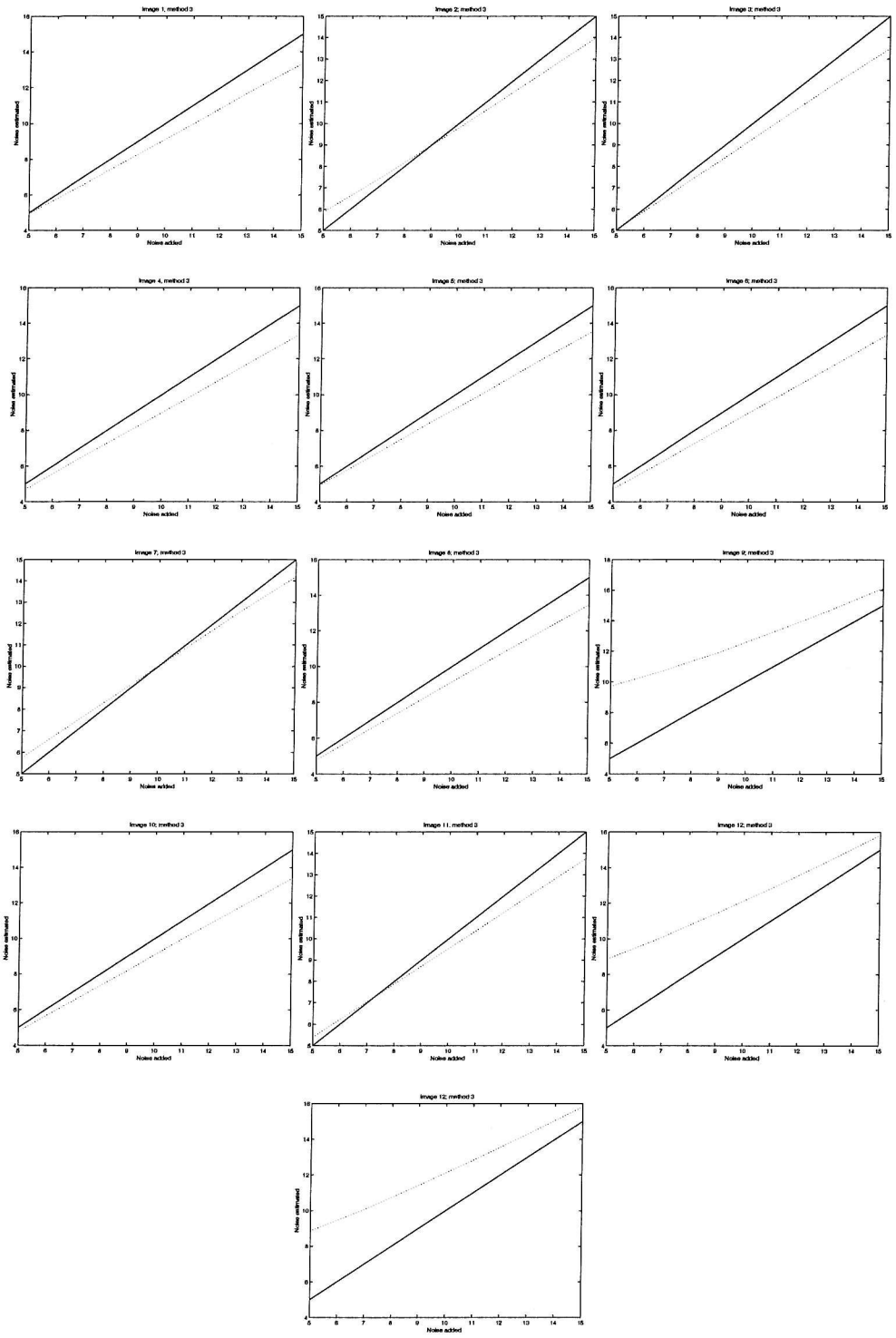


Figure 6-13. Noise estimation on 13 images for 6 noise levels with the linear regression. The continuous line is the standard deviation of the noise introduced and the dotted line is the estimate of the noise level.

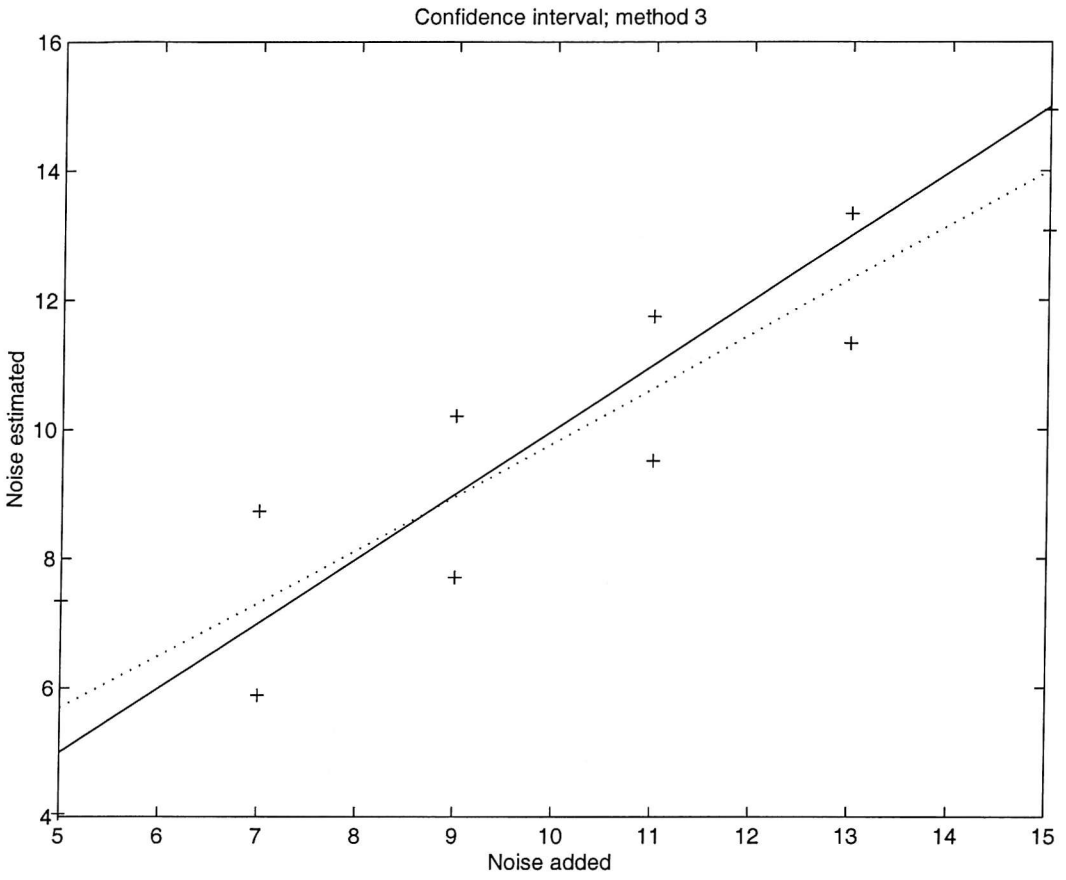


Figure 6-14. Mean (dotted line) ± one standard deviation (+ marks) of the overall estimation, theoretical noise level (continuous line). Linear regression method.

6.4.4 Performance comparison and noise level estimation using a combined estimator

To compare the performances of the three methods with respect to the noise level the following function has been used

$$\begin{aligned} \mathbf{Q1}_i &= |\mu_i - l_i| + \sigma 1_i \\ \mathbf{Q1}_{tot} &= \sum_{i=1}^{H} [Q1_i] \end{aligned} \qquad (6.16)$$

where μ_i and $\sigma 1_i$ are the averages and the standard deviation of the estimates over the 13 images and for each noise level l_i, and H is the number of noise levels. Table 6.6 compares the performances over the complete set and with respect to each noise level.

$Q1_i$	**Noise level 5**	**Noise level 7**	**Noise level 9**	**Noise level 11**	**Noise level 13**	**Noise level 15**	$Q1_{tot}$
Method 1	3.33	1.66	1.55	1.47	0.50	0.55	9.07
Method 2	3.04	3.12	2.27	1.97	1.74	1.66	13.80
Method 3	2.32	1.73	1.30	1.49	1.67	1.92	10.44

Table 6.6. Performances of the three methods with respect to the noise levels and over the complete set of images.

This table shows that in general the first method produces the best estimation. For low levels of noise the third method performs better than the other methods while at high noise levels the first method has a superior performance.

To compare the performances of the three methods with respect to the individual images the following function has been used

$$\mathbf{Q2}_j = \sum_{i=1}^{H} \left| e_{i,j} - l_i \right| + \sigma 2_j$$
$$\mathbf{Q2}_{tot} = \sum_{j=1}^{K} \left[Q2_j \right] \qquad (6.16)$$

where $e_{i,j}$ is the noise level estimated over the image j with noise level l_i, $\sigma 2_j$ is the standard deviation of the estimates over the H noise levels, and K is the number of images included in the set. Table 6.7 compares the performances over the complete set and with respect to each individual image.

$\mathbf{Q2}_j$	**Method 1**	**Method 2**	**Method 3**
Image 1	6.05	1.11	5.78
Image 2	2.38	7.35	3.72
Image 3	3.99	5.36	4.98
Image 4	7.33	2.63	6.43
Image 5	4.77	10.44	5.16
Image 6	7.49	2.59	6.48
Image 7	0.13	16.70	3.01
Image 8	6.03	9.44	5.70
Image 9	19.33	9.37	17.77
Image 10	6.63	5.85	5.97
Image 11	0.65	9.41	3.98
Image 12	14.06	2.39	14.44
Image 13	3.21	32.07	3.31
$\mathbf{Q2}_{tot}$	82.05	114.71	86.73

Table 6.7. Performances of the three methods with respect to the individual images and over the complete set of images.

This table shows again that, in general, the first method produces the best estimates, producing the best performances on six images (2, 3, 5, 7, 11 and 13). The second method is superior on the remaining images (1, 4, 6, 9, 10 and 12), and the third method achieves the best performance on only one image (8).

The results in tables 6.6 and 6.7 demonstrate that the best single algorithm is method 1 but both methods 2 and 3 should not be discarded because there are conditions when their performances are superior.

A final approach to noise level estimation is to use a weighted average of all three methods discussed. The estimation is based on the average between results from the first and third methods, but the first method is substituted by the second when the image contains low levels of noise and consequently the cdf evaluation is not efficient to discern the noise. Table 6.8 and figures 6-15 and 6-16 describe the result formed by averaging the results of the previous methods. Tables 6.9 and 6.10 show that the combined estimator produces performances superior in terms of $\mathbf{Q1}_{tot}$ and $\mathbf{Q2}_{tot}$.

	Noise level 5	**Noise level 7**	**Noise level 9**	**Noise level 11**	**Noise level 13**	**Noise level 15**
Image 1	3.6025	6.3050	8.3966	10.2983	12.2042	14.0590
Image 2	5.6412	7.3676	9.1710	10.9688	12.7834	14.2140
Image 3	4.1850	6.5355	8.5247	10.4647	12.3155	14.1162
Image 4	3.3296	6.1263	8.2012	10.1478	12.0776	13.9970
Image 5	3.9800	6.4231	8.4616	10.3491	12.2990	14.1117
Image 6	3.3226	6.0885	8.2083	10.1513	12.0572	13.9521
Image 7	5.3840	7.2250	9.0530	10.9145	12.7810	14.5799
Image 8	3.6305	6.2784	8.3320	10.3026	12.1889	14.0692
Image 9	9.7972	10.8740	12.2630	13.8399	14.3053	16.0326
Image 10	3.5117	6.2322	8.2831	10.2648	12.1322	13.9912
Image 11	5.0544	6.9889	8.8349	10.6546	12.5219	14.3521
Image 12	8.7896	10.0722	11.5899	13.1550	13.7785	15.5660
Image 13	4.2411	6.4740	8.5372	10.5380	12.5997	14.5529
Average	4.9592	7.1531	9.0659	10.9269	12.6188	14.4303
Std	2.0785	1.5368	1.3127	1.1782	0.6861	0.6471

Table 6.8. Results of noise estimation using the combined estimator.

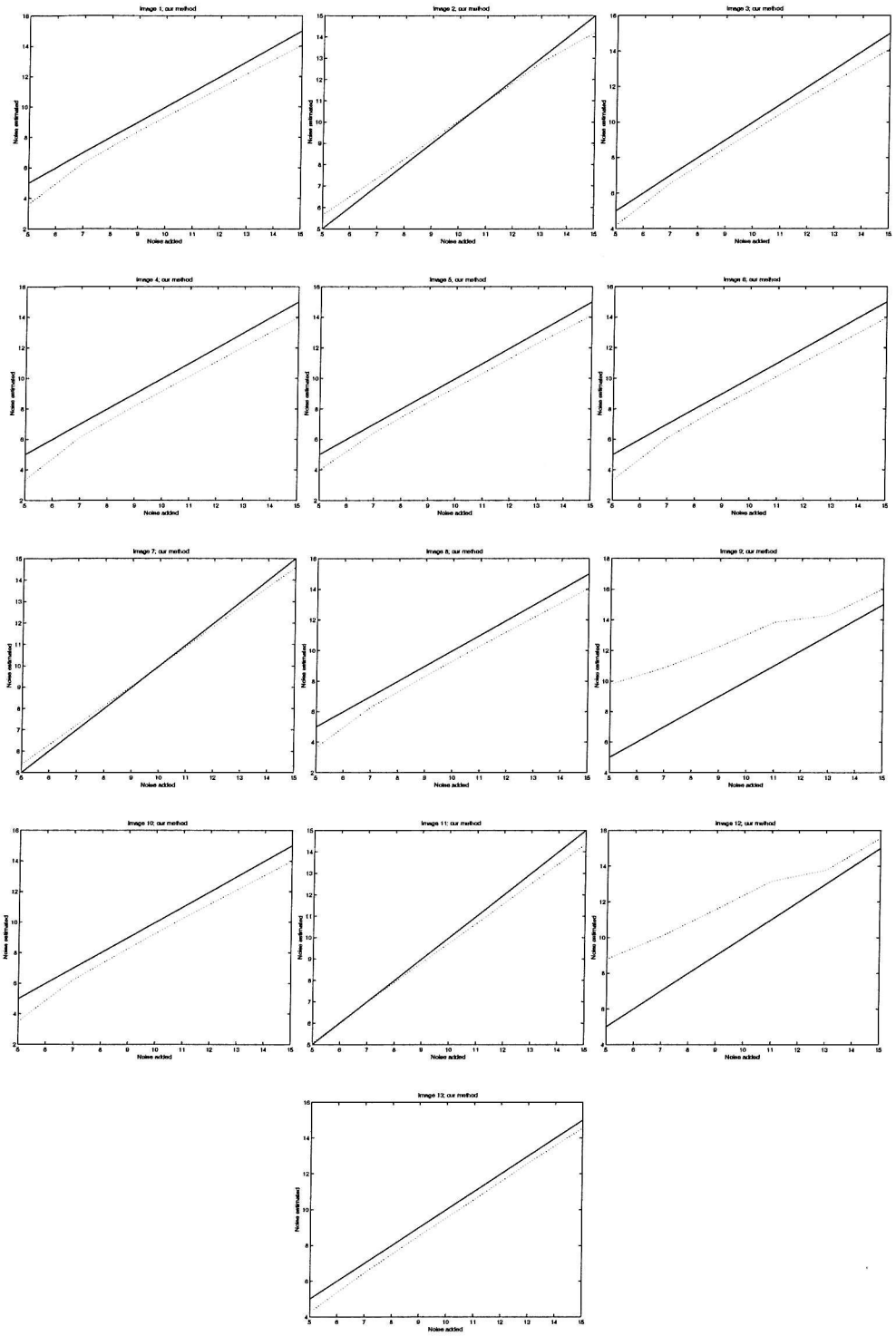


Figure 6-15. Noise estimation on 13 images for 6 noise levels with the combined estimator. The continuous line is the standard deviation of the noise introduced and the dotted line is the estimate of the noise level.

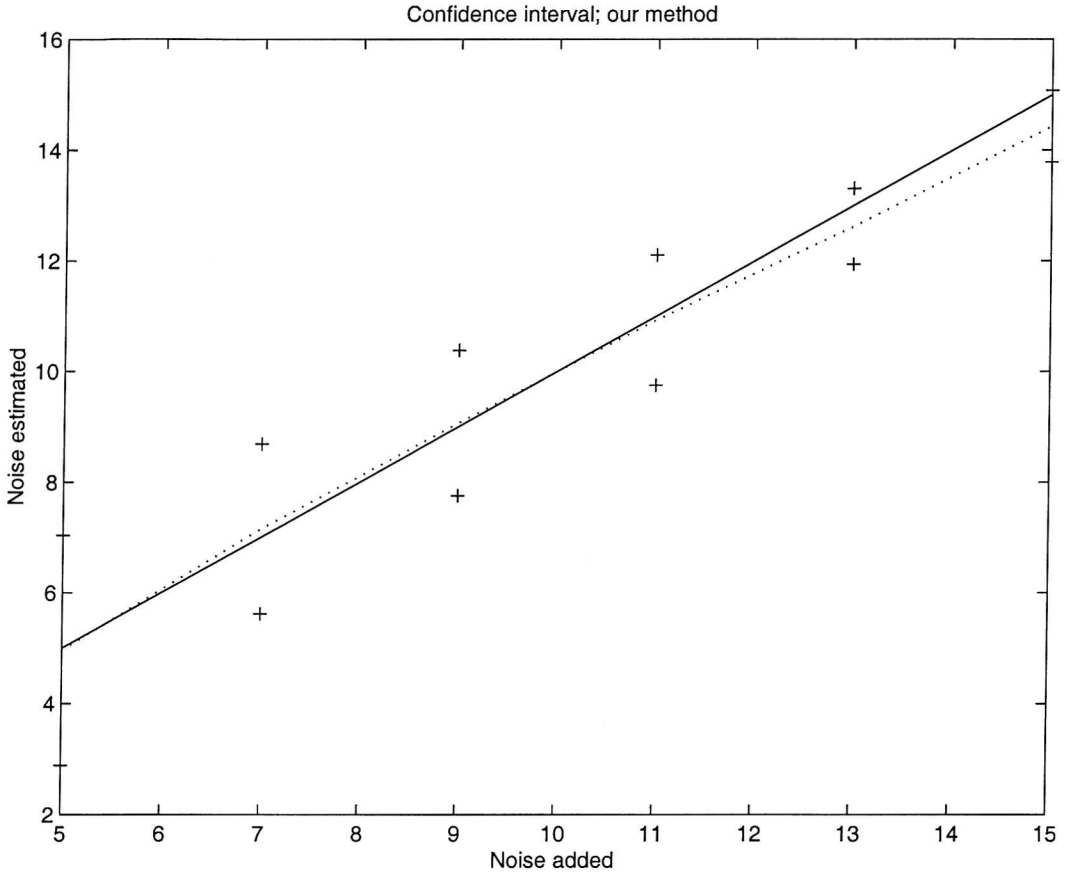


Figure 6-16. Mean (dotted line) ± one standard deviation (+ marks) of the overall estimation, theoretical noise level (continuous line). Combined estimator method.

$Q1_i$	Noise level 5	Noise level 7	Noise level 9	Noise level 11	Noise level 13	Noise level 15	$Q1_{tot}$
Combined estimator	2.12	1.69	1.38	1.25	1.07	1.22	8.72

Table 6.9. Performances using the combined estimator with respect to the noise levels and over the complete set of images.

$Q2_j$	Combined estimator
Image 1	5.42
Image 2	2.50
Image 3	4.04
Image 4	6.45
Image 5	4.56
Image 6	6.55
Image 7	1.54
Image 8	5.46
Image 9	18.57
Image 10	5.88
Image 11	1.95
Image 12	14.23
Image 13	3.19
$Q2_{tot}$	80.34

Table 6.10. Performances using the combined estimator with respect to the individual images and over the complete set of images.

The results described relate to the same set of images as used in the training procedures. For a more rigorous evaluation of the performance, it is convenient to test the algorithms on images outside this set. The noise level has been estimated on the other three images (see appendix C): 'Lenna', 'plate', and 'spine'[4]. Table 6.11 compares the performances of the four methods. It is possible to see that for 'Lenna', the first method produces the best estimates and for low levels of noise the third method has the best performance. Nevertheless the combined estimator continues to produce the best overall performance in terms of $\mathbf{Q}1_{tot}$ and $\mathbf{Q}2_{tot}$.

[4] This is a videofluoroscopic image already contaminated by noise whose distribution is unknown.

		Noise level 5	**Noise level 7**	**Noise level 9**	**Noise level 11**	**Noise level 13**	**Noise level 15**	$Q1_{tot}$	$Q2_j$
Method 1	**'Plate'**	3.06	6.33	8.67	10.80	12.83	14.97		4.04
	'Lenna'	5.40	7.21	9.23	10.91	13.12	14.84		1.32
	'Spine'	7.90	9.00	10.60	12.18	14.00	12.59		11.83
	$Q1_i$	2.87	1.87	1.50	1.06	0.93	2.21	10.44	
	$Q2_{tot}$								17.19
Method 2	**'Plate'**	2.13	5.34	7.90	9.76	12.32	14.54		8.87
	'Lenna'	5.62	7.50	9.41	11.35	13.35	15.44		2.77
	'Spine'	10.11	12.12	13.02	15.21	17.12	18.12		26.46
	$Q1_i$	4.96	4.78	3.74	3.91	3.79	2.90	24.08	
	$Q2_{tot}$								38.10
Method 3	**'Plate'**	4.98	6.75	8.44	10.12	11.82	13.63		4.79
	'Lenna'	5.71	7.38	9.17	10.85	12.60	14.31		2.74
	'Spine'	7.63	8.89	10.31	11.66	13.21	14.81		7.85
	$Q1_i$	2.47	1.77	1.25	0.89	1.16	1.34	8.88	
	$Q2_{tot}$								15.38
Combined estimator	**'Plate'**	4.02	6.54	8.56	10.46	12.33	14.30		3.99
	'Lenna'	5.55	7.29	9.20	10.88	12.86	14.57		1.90
	'Spine'	7.46	8.95	10.45	11.92	13.61	13.70		9.46
	$Q1_i$	2.36	1.82	1.37	0.84	0.71	1.26	8.36	
	$Q2_{tot}$								15.36

Table 6.11. Comparison of the noise estimation methods on images not included in the training set.

Chapter 7

Results

7.1 Introduction

This chapter presents the results of the reduction algorithm described in this thesis and compares these results with similar noise reduction techniques based on the wavelet decompositions and applying thresholding functions to the components.

The degree of decomposition, the procedure for noise reduction on the components, the choice of the image and the contaminating noise have an influence on the efficiency of each technique. The procedure for noise reduction on the components depends on the characteristics of the applied thresholding functions. These are characterised by their shape and by the values of the controlling parameters. The shape is described by the selected thresholding scheme; the parameters are optimised using a training procedure characterised by a minimisation algorithm, number of variables of the cost function, value of the cost function to be minimised and images used in the training. To evaluate and generalise the results, the procedure has been tested for different images contaminated by different noise spectra, noise levels and noise realisations. The factors that should be considered examining the results can be summarised as:

1) Number of decomposition levels used (3, 2 and 1 levels).

2) Shape of the applied thresholding functions (complete and with Ky=0).

3) Minimisation algorithm used for training the thresholding parameters values: (Simplex and Quasi Newton)

4) Form of the cost function (MSE and WMSE)

5) Images used in the training (test image and set of 6 images)

6) Original image to which the technique is applied (set of 13 images and 3 external images).

7) Level of contaminating noise (6 standard deviation levels between 5 and 15).

8) Noise realisation (5 different generations).

9) Spectrum of the noise (white, LF, BF, and HF).

The next section considers the dependence of the results on the parameters of the training procedure. In section 7.3 the results are then extended to consider the dependence on the original image and noise characteristics.

Other noise reduction methods based on wavelet decompositions can use decimated [64, 65, 69, 71, 82, 83] or undecimated [66, 67, 68, 75] components, and the thresholding functions can be selected using fixed or decomposition level adaptive criteria [77]. In section 7.4 the results using conventional wavelet methods and those described herein are compared.

Finally, in section 7.5, images are presented to allow one to compare the visual quality of results. A limited number of the significant images are shown in this section. A more complete set of images is included in the CD-ROM provided with this thesis.

7.2 Influence of the training parameters on the algorithm performance

This section examines how the performance of the algorithm depends on the training procedure. The standard deviation level of the noise is assumed known. The indices of quality considered are the MSE and WMSE. Training and performance evaluation is initially conducted only on the test image (figure 7-1) with AWGN contaminating noise. Later in this section training is performed on a set of 6 images (figures 7-2) and the performance evaluated over these images. The generalisation of these results is the object of the next section where a set of 13 images and different noise types are considered.

Figure 7-1. Test image.

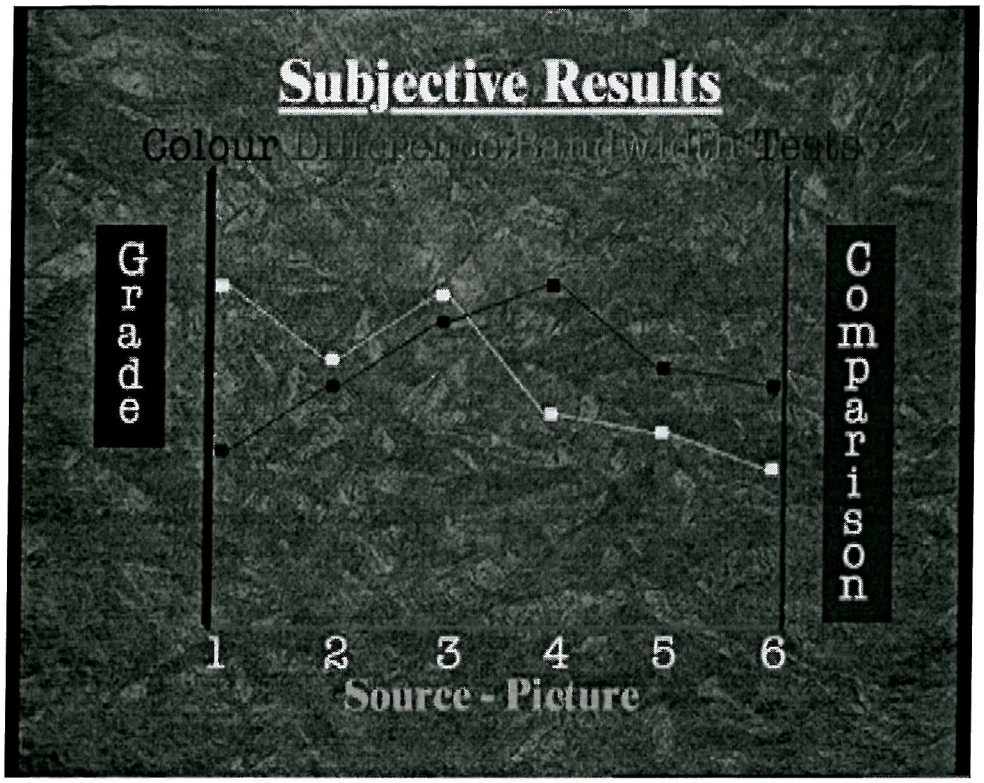


Figure 7-2-a. First image of the set composed of six images ('graph').

Figure 7-2-b. Second image of the set composed of six images ('girl').

Figure 7-2-c. Third image of the set composed of six images ('interview').

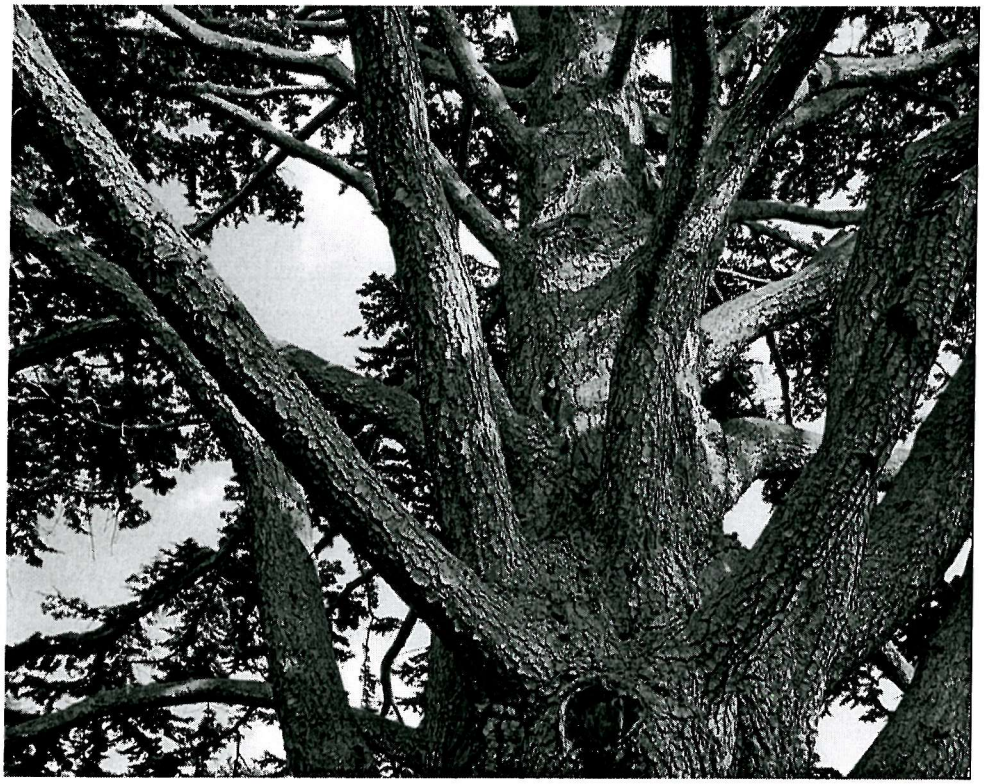


Figure 7-2-d. Fourth image of the set composed of six images ('tree').

Figure 7-2-e. Fifth image of the set composed of six images ('text').

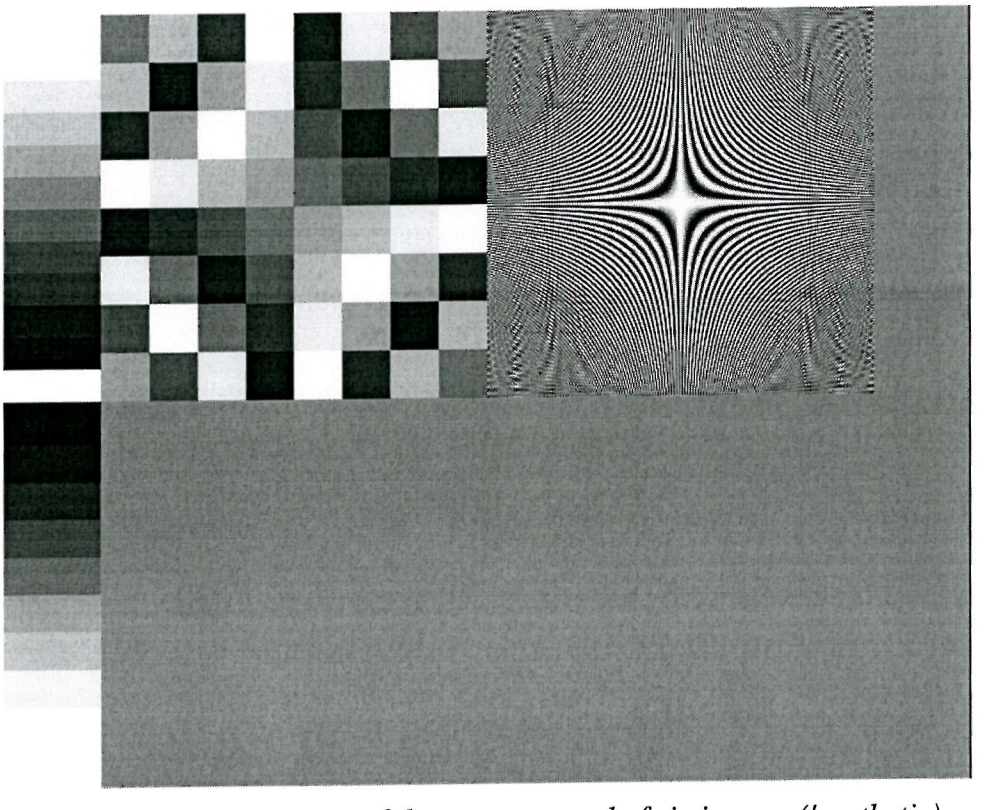


Figure 7-2-f. Sixth image of the set composed of six images ('synthetic).

Tables 7.1 compare the MSE (7.1-a) and WMSE (7.1-b) reductions on the test image contaminated by 3 levels of AWGN using the parameters obtained by training the test image using different settings. Results are compared using:

a) Nelder and Quasi-Newton minimisation algorithms;

b) Unweighted and weighted training;

c) 3 degrees of frequency decomposition: 1 level (4 or 5 cost function parameters), 2 levels (7 or 8 parameters), and 3 levels (10 or 11 parameters);

d) 2 shapes for the thresholding function: complete (11, 8, or 5 parameters) and assuming Ky=0 (10, 7 and 4 parameters).

In tables 7.1 and in figures 7-3 the case when no thresholding functions are applied is indicated with 0 cost function parameters.

	MSE						
Unweighted optimisation, Nelder Method	**0 param.**	**4 param.**	**5 param.**	**7 param.**	**8 param.**	**10 param.**	**11 param.**
Noise level 5	25	14.1	14.1	13.6	13.5	13.5	13.4
Noise level 10	99.2.	43.2	43.2	40.1	40.1	39.7	39.7
Noise level 15	226.7	80.5	80.1	73.0	73.0	72.2	72.2
Unweighted optimisation, Q-Newton. method	**0 param.**	**4 param.**	**5 param.**	**7 param.**	**8 param.**	**10 param.**	**11 param.**
Noise level 5	25	14.1	14.1	13.6	13.5	13.5	13.4
Noise level 10	99.2	43.2	43.2	40.1	40.0	39.7	39.6
Noise level 15	226.7	80.5	80.1	73	73.0	71.8	71.8
Weighted optimisation, Nelder Method	**0 param.**	**4 param.**	**5 param.**	**7 param.**	**8 param.**	**10 param.**	**11 param.**
Noise level 5	25	14.8	15.1	14.6	14.7	14.6	14.7
Noise level 10	99.2	44.5	45.8	56	51.2	56.3	51.4
Noise level 15	226.7	130.5	123	119.7	102.1	119.2	101.2
Weighted optimisation, Q-Newton. Method	**0 param.**	**4 param.**	**5 param.**	**7 param.**	**8 param.**	**10 param.**	**11 param.**
Noise level 5	25	14.3	15.1	14.1	14.5	14.1	14.6
Noise level 10	99.2	44.2	45.9	55.9	51	55.1	50.5
Noise level 15	226.7	130.7	123.2	119.8	102.2	117.9	100.8

Table 7.1-a. MSE dependence on the training settings.

	WMSE						
Weighted optimisation, Nelder method	**0 param.**	**4 param.**	**5 param.**	**7 param.**	**8 param.**	**10 param.**	**11 param.**
Noise level 5	2.5	2.3	2.3	2.1	2.1	2.1	2
Noise level 10	10.1	8.8	8.8	7.6	7.2	7.2	7.1
Noise level 15	23.1	19.1	19.1	16.7	15.3	15.3	14.7
Weighted optimisation, Q-Newton. Method	**0 param.**	**4 param.**	**5 param.**	**7 param.**	**8 param.**	**10 param.**	**11 param.**
Noise level 5	2.5	2.3	2.3	2.1	2.1	2.1	2
Noise level 10	10.1	8.8	8.8	7.6	7.3	7.3	7
Noise level 15	23.1	19.1	19.1	16.7	15.3	15.3	14.7
Unweighted optimisation, Nelder method	**0 param.**	**4 param.**	**5 param.**	**7 param.**	**8 param.**	**10 param.**	**11 param.**
Noise level 5	2.5	2.3	2.3	2.2	2.2	2.2	2.1
Noise level 10	10.1	9.9	9	8	8	7.8	7.8
Noise level 15	23.1	19.2	19.3	17.1	17.1	16.5	16.4
Unweighted optimisation, Q-Newton. Method	**0 param.**	**4 param.**	**5 param.**	**7 param.**	**8 param.**	**10 param.**	**11 param.**
Noise level 5	2.5	2.3	2.3	2.2	2.2	2.2	2.1
Noise level 10	10.1	9.9	9.0	8.0	8.0	7.7	7.7
Noise level 15	23.1	19.2	19.3	17.1	17.1	16.2	16.2

Table 7.1-b. WMSE dependence on the training settings.

The results in tables 7.1 are obtained using the same image for testing and training. As expected, one obtains the best MSE reduction in the presence of unweighted minimisation and the best WMSE reduction in the presence of weighted minimisation. The two minimisation algorithms produce similar performances but in the following the Nelder algorithm has been preferred for reasons of computational efficiency (see section 5.2). Figures 7-3 show the MSE (7-3-a) and WMSE (7-3-b) reductions dependence on number of cost function parameters respectively for unweighted and weighted training.

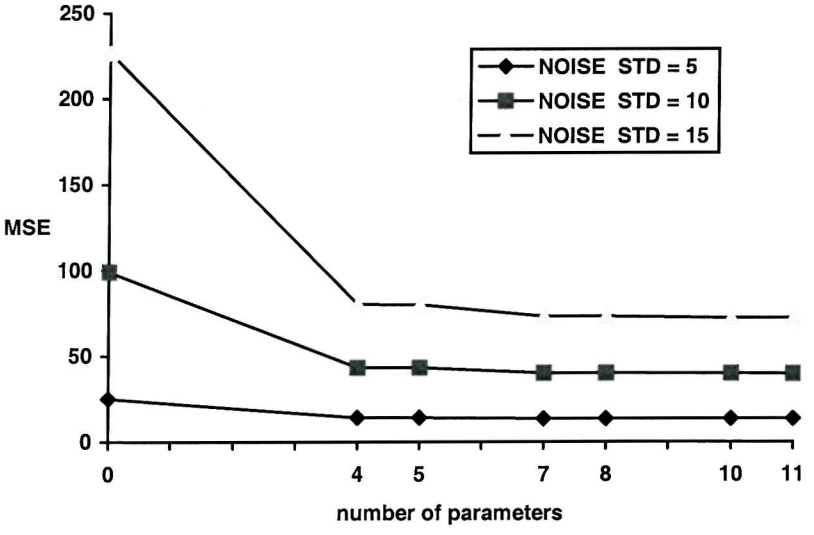


Figure 7-3-a. Relationship between MSE and number of threshold parameters.

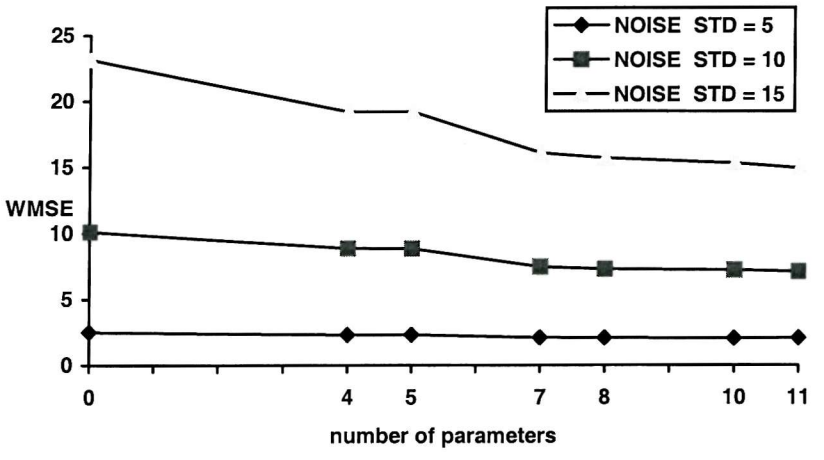


Figure 7-3-b. Relationship between WMSE and number of threshold parameters

These figures demonstrate that the number of degrees of decomposition required by the method depends on the amount of contaminating noise. When the standard deviation of the noise is 15 the largest part of the reduction is achieved with the first 4-5 parameters, whereas when the noise level is 5, more degrees of decomposition (and more parameters) are required in order to have similar MSE and WMSE reductions.

The results showed in tables 7.1 and in figures 7-3 refer to the case when the training and testing are performed on the same single image (figure 7-1). Table 7.2 and figure 7-4 report the results achieved performing training and testing on a set of 6 images (figures 7-2). Each of the six images was processed 7 times. The thresholding parameters obtained by training on each image individually were applied to every image. A final set of parameters was obtained by averaging the six parameter sets. The noise standard deviation is 15 and the reduction in WMSE is measured. This provides information about the dependence of the performance on the spectral content of the image. Comparing these results one can see that using the parameters related to that image produce an WMSE reduction only slightly superior to that obtained when the parameters related to other images are used. The average set of parameters yields results which are reasonably consistent across the image set.

The results also show that the method seems less effective on particular images. Nevertheless even on these images the method dramatically increases the visual quality.

WMSE % REDUCTION		TRAINED						Average Parameters
		Im 1	Im 6	Im 7	Im 9	Im 12	Im 13	
T E S T E D	Im 1	46.6	45.8	42.7	41.4	40.5	25.1	40
	Im 6	50.2	51.3	48.9	44.1	44	32.3	46.5
	Im 7	57.3	59.2	61.2	50.2	50.6	55.7	54.5
	Im 9	23.3	19.3	16.4	29.3	27.4	13.5	26.3
	Im 12	21.2	16.9	17.8	27.2	28.6	20.1	27.4
	Im 13	63.7	67.6	73.5	57.5	59.7	81.2	65.7

Table 7.2. WMSE dependence on image used in the training.

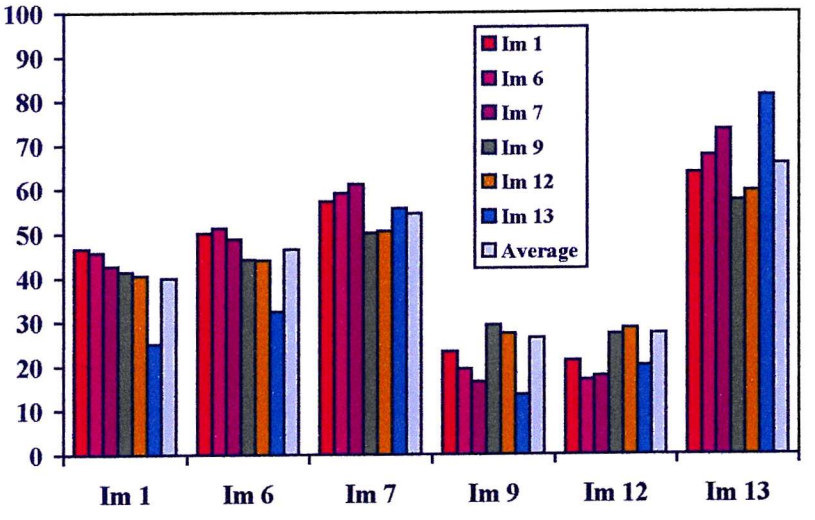
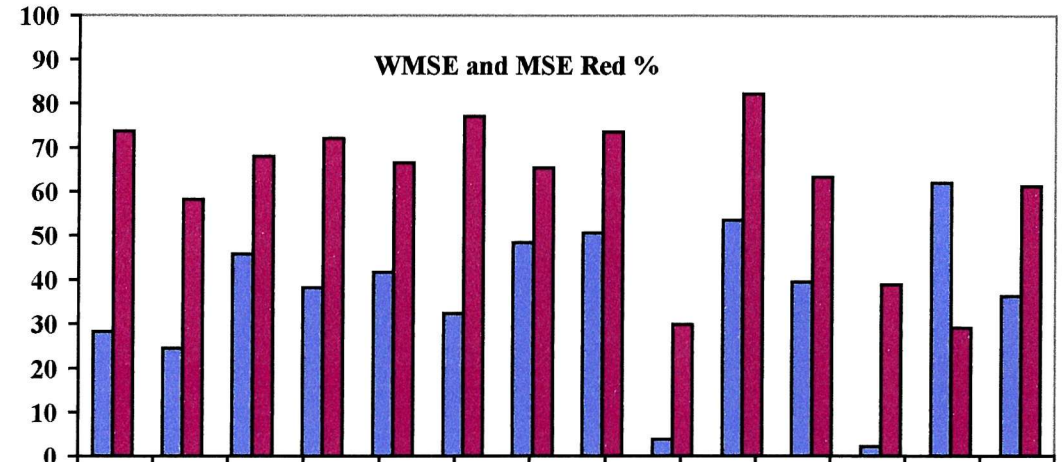


Figure 7-4. WMSE dependence on image used in the training.

7.3 Performance evaluation for more general noise conditions

This section explores the performance of the method for a wider range of conditions. The algorithm uses parameters obtained by training on the test image (figure 7-1), 3 levels of decomposition and the complete scheme for the thresholding function (11 parameters). The performance on the full set of 13 images is considered, with a variety of noise spectra at a range of noise levels.

In table 7.3, WMSE and MSE reduction are shown as percentages for 13 images contaminated by AWGN with a standard deviation of 15. The last column shows the average percentage of noise reduction on the 13 images. The parameters obtained by training on the test image are utilised.



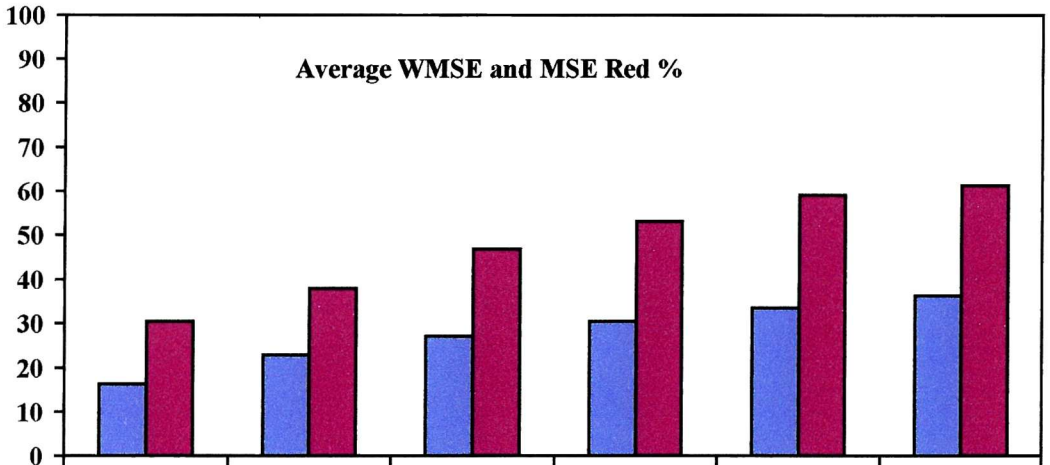
	Im 1	Im 2	Im 3	Im 4	Im 5	Im 6	Im 7	Im 9	Im 10	Im 11	Im 12	Im 13	Aver.
WMSE optim.	28.3	24.5	45.9	38.2	41.7	32.4	48.5	3.8	53.6	39.5	2.2	62.1	36.3
MSE optim.	73.8	58.2	68.1	72.1	66.6	77.2	65.5	29.9	82.3	63.4	38.9	29.2	61.4

Table 7.3. WMSE and MSE reductions (in percentage) for 13 images with AWGN having standard deviation of 15.

It is important to note that the set of images includes two images (9 and 12) which have very different characteristics from the other images. The noise reducing technique is less

effective on these images because of their large high frequency content. If these images are omitted the average reductions rise to 66.3 % and 42.3 % for the MSE and WMSE respectively.

In table 7.4, the average MSE and WMSE reductions computed over 13 images are shown for six levels of AWGN. The percentage reduction increases with the standard deviation of the contaminating AWGN. This is to be expected because these are percentages of reduction. The percentage of MSE and WMSE reduction needed to obtain the same visual quality increases with the standard deviation of the AWGN.



	Noise level 5	Noise level 7	Noise level 9	Noise level 11	Noise level 13	Noise level 15
WMSE optim.	16.3	22.9	27.1	30.5	33.6	36.3
MSE optim.	30.4	37.9	46.8	53.1	59.1	61.4

Table 7.4. Average WMSE and MSE reduction for 6 AWGN levels.

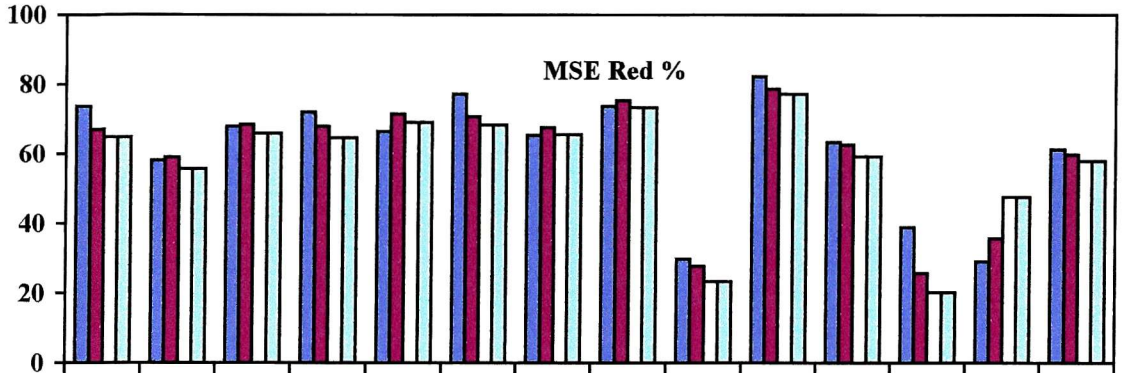
The results presented so far have used computer generated AWGN. The following simulation explores the robustness of the results to the realisation of the noise sequence. Table 5.7[1] describes the dependence of the parameters on the noise realisation in terms of confidence intervals for the parameters. Using these intervals, the dependence of MSE and

WMSE reductions on the noise realisation are shown in tables 7.5. MSE and WMSE are evaluated using the sets of parameters obtained by training for each noise realisation. The results demonstrate that the numerical realisation of the noise has a very small influence on the performance.

	Original MSE	**MSE confidence interval**	**Original WMSE**	**WMSE confidence interval**
Noise level 5	25.1	[14.5, 14.7]	2.5	[2.06, 2,08]
Noise level 7	49	[28.1, 28.3]	5	[3.82, 3.83]
Noise level 9	81	[42.4 42.5]	8.4	[5.99, 6.01]
Noise level 11	121.8	[59.6, 59.9]	12.4	[8.59, 8.60]
Noise level 13	168.7	[78.2, 79.1]	17.1	[11.36, 11.40]
Noise level 15	223.8	[99.3, 99.9]	23.3	[14.85, 14.91]

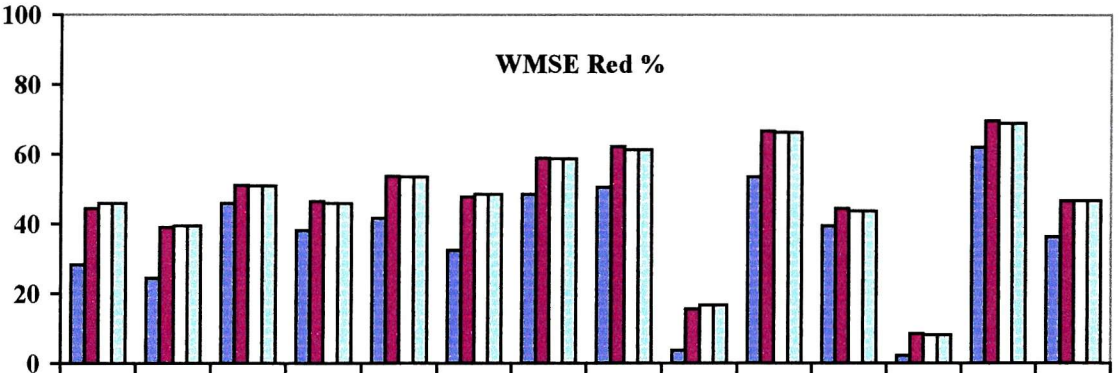
Table 7.5. Confidence intervals for MSE and WMSE.

So far only results considering AWGN have been presented. The performance of the algorithm will now be assessed when coloured noise is added. Three types of coloured additive Gaussian noise are considered. These are obtained filtering white noise with three filters having low-pass, band-pass and high-pass frequency characteristics. In this case the procedure for spectral estimation (see section 6.2) allows the estimation of the standard deviation of the noise in each band. The four types of noise have been normalised so as to produce the same total noise power. Tables 7.6 compare the MSE and WMSE reductions for white and coloured noises contaminating the 13 images. The standard deviation of the AWGN is 15.



	Im 1	Im 2	Im 3	Im 4	Im 5	Im 6	Im 7	Im 8	Im 9	Im 10	Im 11	Im 12	Im 13	Aver.
White	73.8	58.2	68.1	72.1	66.6	77.2	65.5	73.7	29.9	82.3	63.4	38.9	29.2	61.4
LPF	67.1	59.1	68.5	68	71.5	70.7	67.6	75.4	27.8	78.8	62.6	25.8	35.8	59.9
BPF	64.8	55.8	65.9	64.7	69.1	68.3	65.6	73.3	23.4	77.2	59.3	20.2	47.6	58
HPF	64.8	55.8	65.9	64.7	69.1	68.3	65.6	73.3	23.4	77.2	59.3	20.2	47.6	58

Table 7.6-a. MSE reductions (in percentage) for 13 images for different types of noise, with standard deviation 15.



	Im 1	Im 2	Im 3	Im 4	Im 5	Im 6	Im 7	Im 8	Im 9	Im 10	Im 11	Im 12	Im 13	Aver.
White	28.3	24.5	45.9	38.2	41.7	32.4	48.5	50.6	3.8	53.6	39.5	2.2	62.1	36.3
LPF	44.4	39.2	51.1	46.6	53.8	47.8	58.9	62.2	15.6	66.8	44.4	8.6	69.6	46.8
BPF	46	39.5	50.9	46	53.5	48.6	58.8	61.3	16.7	66.3	43.8	8.2	68.9	46.8
HPF	46	39.5	50.9	46	53.5	48.6	58.8	61.3	16.7	66.3	43.8	8.2	68.9	46.8

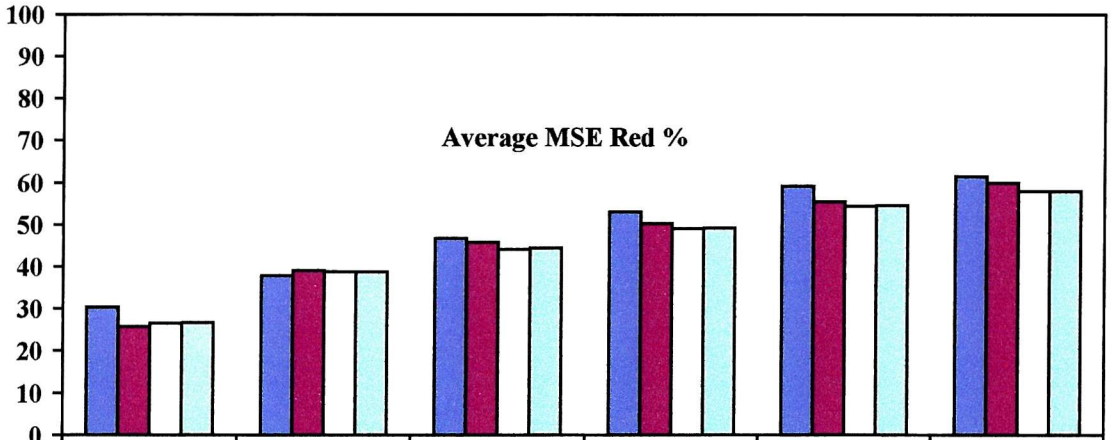
Table 7.6-b. WMSE (in percentage) reductions for 13 images for different types of noise, with standard deviation 15.

The percentage of MSE reduction is broadly comparable for all types of noise. Examining the images individually, in the majority of the cases the reduction for coloured noises is

superior to that in the white noise case. Nevertheless the average MSE reduction for the white noise case is superior to that in the case of coloured noise.

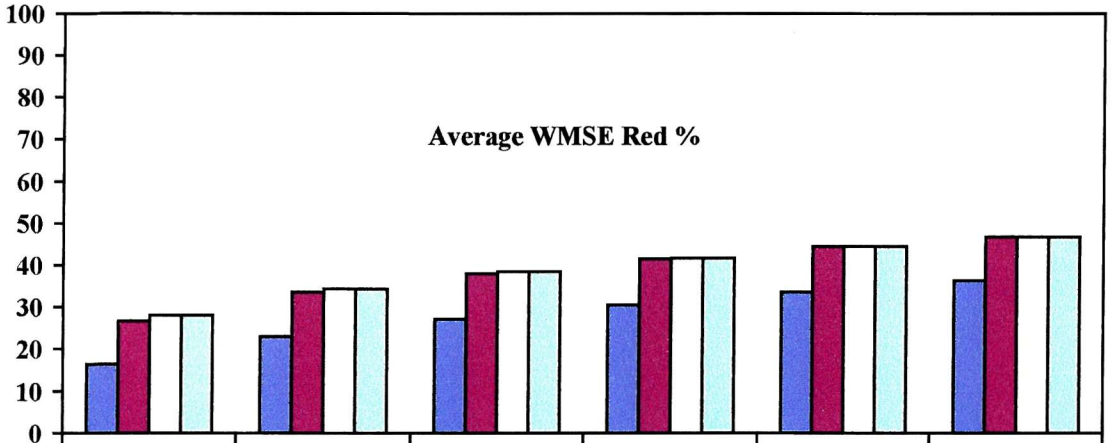
In the case of coloured noise the average and individual reductions of WMSE are always superior to the white noise case.

Tables 7.7 compare the average MSE and WMSE reduction for a range of noise levels and follow the trends seen in tables 7.6.



	Noise level 5	Noise level 7	Noise level 9	Noise level 11	Noise level 13	Noise level 15
White	30.4	37.9	46.8	53.1	59.1	61.4
LPF	25.8	39.1	45.8	50.3	55.5	59.9
BPF	26.5	38.8	44.2	49	54.4	58
HPF	26.6	38.7	44.5	49.2	54.6	58

Table 7.7-a. Average over 13 images of MSE reduction for various noise types.



	Noise level 5	Noise level 7	Noise level 9	Noise level 11	Noise level 13	Noise level 15
White	16.3	22.9	27.1	30.5	33.6	36.3
LPF	26.6	33.5	38	41.5	44.4	46.8
BPF	28	34.3	38.5	41.7	44.4	46.8
HPF	28	34.3	38.5	41.7	44.4	46.8

Table 7.7-b. Average over 13 images of WMSE reduction for various noise types.

7.4 Comparison with other wavelet based approaches

This section compares the performance of the method presented in this thesis with the performance of other wavelet based approaches. The MATLAB toolbox WAVELAB has been used to measure the performance using various decompositions and thresholding criteria. Initially in this section the characteristics of the decompositions and the thresholding functions of existing algorithms are summarised. The performance is compared using the same set of images and identical noises.

7.4.1 Decimated and undecimated components

The frequency decomposition is characterised by the digital filters used in the filter banks and by the presence (or absence) of down and up sampling operations (also called decimation and interpolation). The decomposition employed thus far produces undecimated components, so no sub and up sampling operations are used, allowing one to select very simple digital filters. The algorithm used for comparative purposes employs the same decomposition in which the components are decimated. This decomposition is based on a periodized and orthogonal wavelet 2-D transform utilising QMF filters (Coiflet 2) [63].

7.4.2 Thresholding techniques

The thresholding functions are characterised by the scheme applied and by the parameters used in the thresholding functions. The algorithms used in the comparison assume that the image is preliminarily normalised with respect to the standard deviation of the noise. Thresholding algorithms are:

a) VISUAL. Soft-thresholding is applied and the parameters are the same for all the components and depend on the image size [65, 66].

b) SURE. Hard-thresholding is applied and the parameters are level dependent and selected using principle of Stein's Unbiased Risk Estimate [82].

c) HYBRID. Soft threshold is applied and the parameters are level dependent and selected according to the methods used in a) or in b).

d) MINIMAX. Hard thresholding is applied, the parameters are the same for all the components and are selected from a fixed vector on the basis of the image size [82, 83].

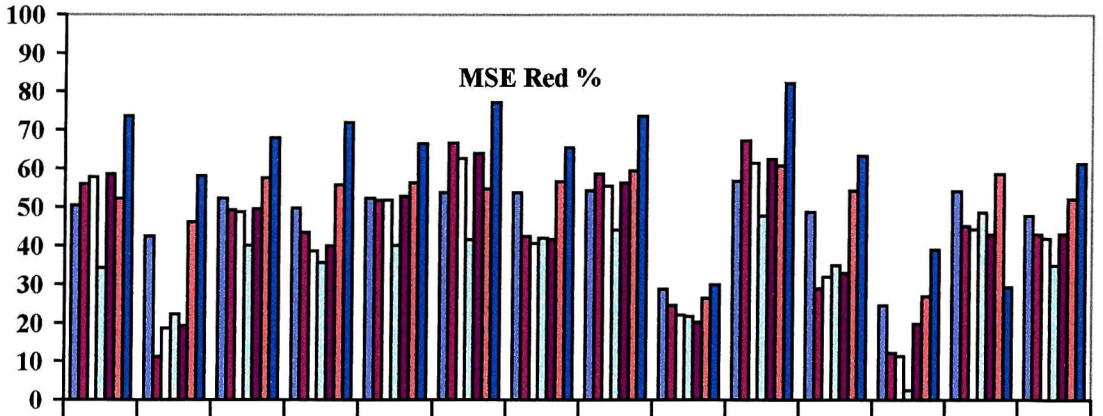
e) MAD. Soft-thresholding is applied, the parameters are level dependent and selected using the median absolute deviation (MAD) of the pixel values for each level [75, 88].

7.4.3 Performance comparison

The performances of these algorithms are summarised for the 13 images in terms of MSE and WMSE reductions; 6 levels of AWGN are considered.

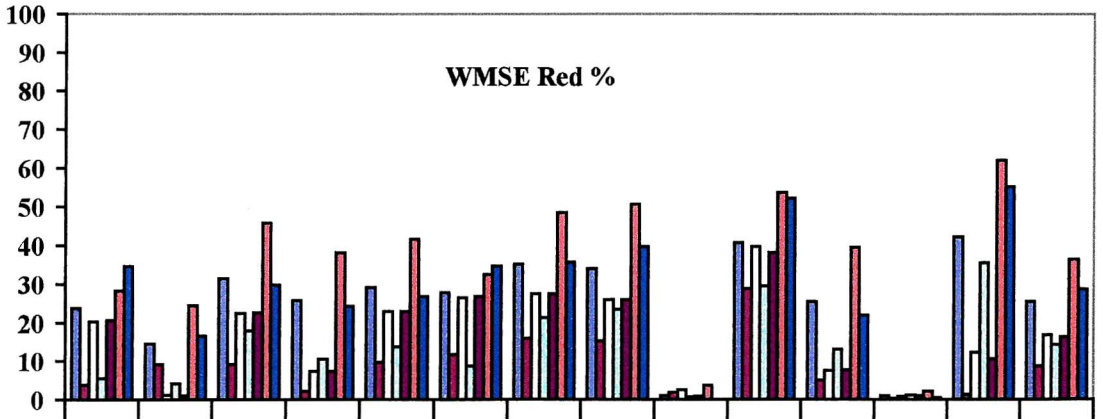
Tables 7.8 and 7.9 compare the methods discussed in this thesis (undecimated decomposition and thresholding criterion based on MSE and WMSE optimisations on a single image) with wavelet based methods using the decimated decomposition described in section 7.4.1 and the 5 thresholding criteria described in section 7.4.2.

Tables 7.8 compare MSE (7.8-a) and WMSE (7.8-b) reductions for 13 images and for noise level 15.



	Im 1	Im 2	Im 3	Im 4	Im 5	Im 6	Im 7	Im 8	Im 9	Im 10	Im 11	Im 12	Im 13	Aver.
Visual	50.5	42.4	52.4	49.8	52.3	53.9	53.9	54.4	28.7	56.8	48.9	24.6	54.2	47.9
SURE	56	11.3	49.4	43.5	51.9	66.7	42.5	58.7	24.6	67.3	29	12.3	45.1	42.9
Hybrid	57.8	18.6	48.8	38.7	51.8	62.7	40.7	55.5	22.1	61.6	31.9	11.4	44.3	41.9
Minimax	34.3	22.2	40.2	35.6	40.1	41.7	42	44.1	21.7	47.9	34.9	2.6	48.6	35
MAD	58.7	19.3	49.6	39.9	52.8	64.1	41.6	56.4	20.2	62.5	32.9	19.7	42.9	43.1
WMSE optim.	52.4	46.1	57.7	55.9	56.4	54.8	56.6	59.5	26.4	60.8	54.4	26.9	58.7	52.2
MSE optim.	73.8	58.2	68.1	72.1	66.6	77.2	65.5	73.7	29.9	82.3	63.4	38.9	29.2	61.4

Table 7.8-a. MSE reduction (in percentage) for 13 images (decimated decomposition).

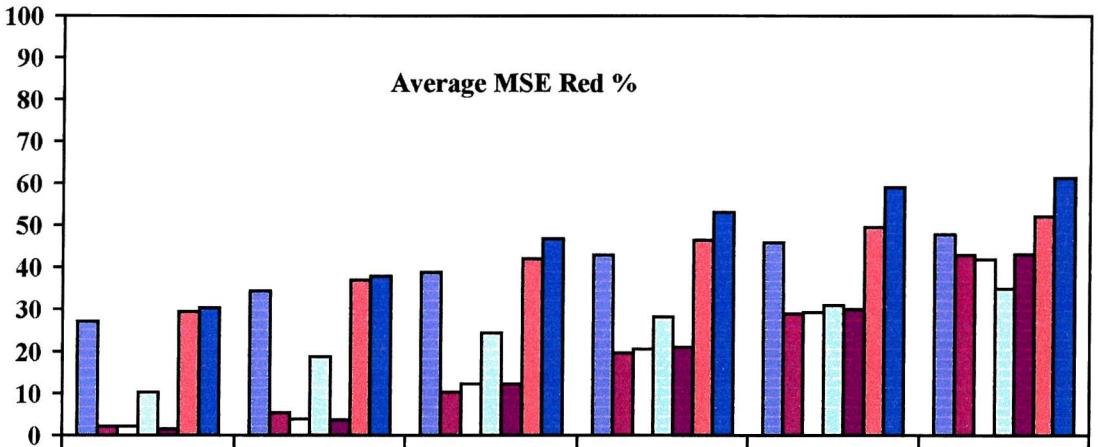


	Im 1	Im 2	Im 3	Im 4	Im 5	Im 6	Im 7	Im 8	Im 9	Im 10	Im 11	Im 12	Im 13	Aver.
Visual	23.7	14.5	31.5	25.8	29.1	27.8	35.1	33.9	1.1	40.7	25.4	1.1	42.1	25.5
SURE	3.9	9.2	9.2	2.2	9.8	11.8	15.9	15.2	1.9	28.8	5.1	0.3	1.3	8.8
Hybrid	20.3	1.2	22.5	7.4	22.9	26.5	27.4	26	2.6	39.6	7.5	0.8	12.3	16.7
Minimax	5.6	4.3	18	10.6	13.7	8.8	21.2	23.4	0.7	29.4	13	1.2	35.4	14.3
MAD	20.6	1.1	22.6	7.4	23	26.7	27.4	25.9	0.9	38.2	7.7	1	10.5	16.3
WMSE optim.	28.3	24.5	45.9	38.2	41.7	32.4	48.5	50.6	3.8	53.6	39.5	2.2	62.1	36.3
MSE optim.	34.7	16.6	29.8	24.2	26.7	34.7	35.7	39.6	0.1	52.2	21.9	0.5	55.1	28.6

Table 7.8-b. WMSE reduction (in percentage) for 13 images (decimated decomposition).

These tables demonstrate that both unweighted and weighted optimisation produce superior averaged MSE and WMSE reductions (see last columns) compared with the decimated frequency decomposition with all 5 thresholding criteria. Note that MSE optimisation produces the largest MSE reduction for 12 out of the 13 images[2], while the WMSE optimisation produces the best WMSE reduction 11 out of the 13 images[3].

Tables 7.9 show the average MSE and WMSE reductions for six noise levels.



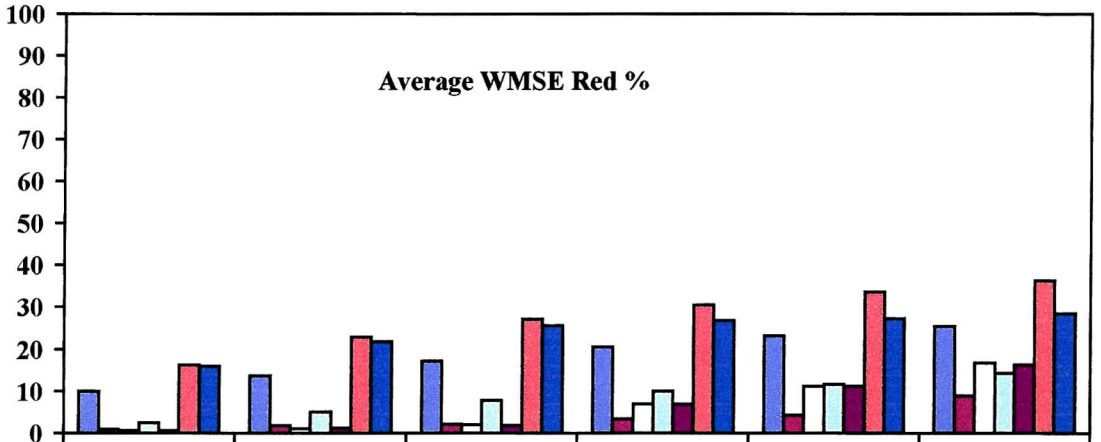
	Noise level 5	Noise level 7	Noise level 9	Noise level 11	Noise level 13	Noise level 15
Visual	27.1	34.3	38.9	42.9	45.9	47.9
SURE	2.1	5.3	10.3	19.6	29	42.9
Hybrid	2.1	3.8	12.2	20.6	29.3	41.9
Minimax	10.2	18.7	24.4	28.2	31	35
MAD	1.5	3.7	12.3	21	30	43.1
WMSE optim.	29.4	37	42	46.5	49.6	52.2
MSE optim.	30.4	37.9	46.8	53.1	59.1	61.4

Table 7.9-a. MSE reduction (in percentage) for 6 noise levels (decimated decomposition).

From table 7.9-a one can see that the best alternative algorithm for reducing MSE at all noise levels is the visual scheme. The performance advantage offered by the algorithm discussed in this thesis reduces at lower noise levels.

[2] The remaining image is the synthetic image (figure 7-2-f) which is a very atypical image. It can be observed that for this image the weighted (WMSE) optimisation produces MSE reduction clearly superior than the unweighted (MSE) optimisation.

[3] For the remaining two images (figures 7-2-a and 7-2-b) the unweighted (MSE) optimisation performs best.

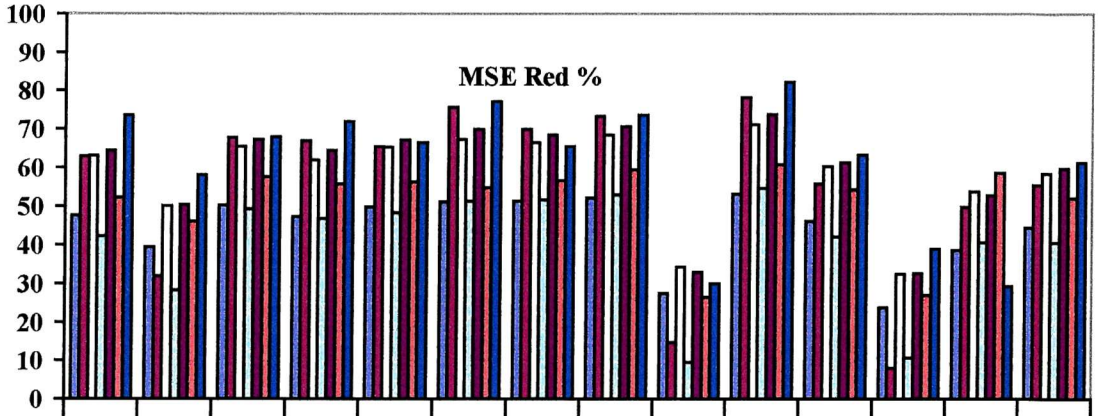


	Noise level 5	Noise level 7	Noise level 9	Noise level 11	Noise level 13	Noise level 15
Visual	9.9	13.7	17.2	20.5	23.1	25.5
SURE	0.9	1.8	2.1	3.4	4.2	8.8
Hybrid	0.5	1.1	1.9	6.9	11.1	16.7
Minimax	2.4	5.1	7.8	9.9	11.7	14.3
MAD	0.6	1.2	1.8	6.9	11.1	16.3
WMSE optim.	16.3	22.9	27.1	30.5	33.6	36.3
MSE optim.	16	21.7	25.6	26.9	27.3	28.6

Table 7.9-b. WMSE reduction (in percentage) for 6 noise levels (decimated decomposition).

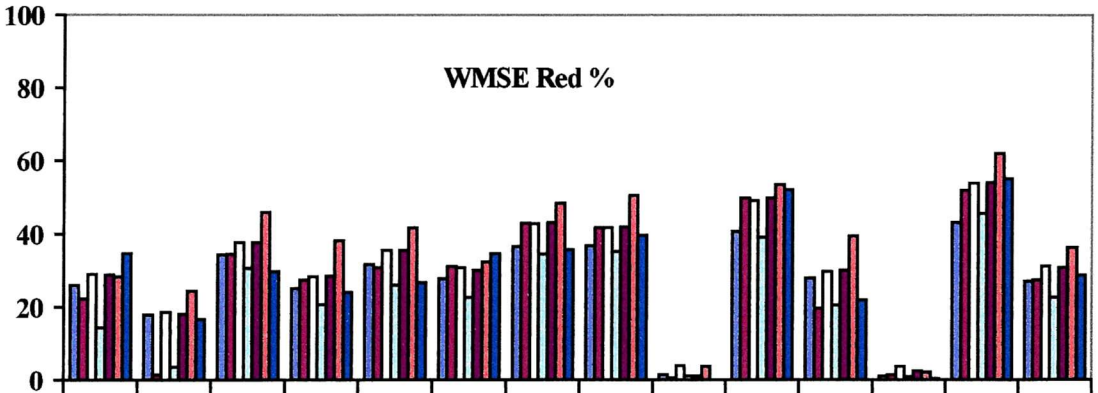
From table 7.9-b one can see that in terms of WMSE the visual algorithm once again exhibits the best performance of all the alternative schemes at all noise levels. The difference in performance between visual and WMSE optimisation algorithm remains around 10% for all noise levels considered.

Tables 7.10 and 7.11 detail the performance of all the wavelet based methods using the *undecimated* decomposition described in this thesis and the 5 different thresholding criteria described in section 7.4.2. Tables 7.10 compare MSE and WMSE reduction for 13 images at a noise level of 15.



	Im 1	Im 2	Im 3	Im 4	Im 5	Im 6	Im 7	Im 8	Im 9	Im 10	Im 11	Im 12	Im 13	Aver.
Visual	47.7	39.5	50.3	47.4	49.8	51.1	51.3	52.1	27.5	53.2	46.1	23.8	38.6	44.5
SURE	63.1	32	67.9	67	65.6	75.8	70.1	73.4	14.7	78.2	55.8	8	49.9	55.5
Hybrid	63.2	50.1	65.6	62.1	65.4	67.3	66.5	68.6	34.3	71.2	60.3	32.4	53.9	58.5
Minimax	42.3	28.3	49.3	46.8	48.4	51.3	51.6	53	9.5	54.6	42.2	10.7	40.6	40.7
MAD	64.6	50.5	67.4	64.6	67.2	70.1	68.6	70.8	33	73.9	61.4	32.6	52.9	59.8
WMSE optim.	52.4	46.1	57.7	55.9	56.4	54.8	56.6	59.5	26.4	60.8	54.4	26.9	58.7	52.2
MSE optim.	73.8	58.2	68.1	72.1	66.6	77.2	65.5	73.7	29.9	82.3	63.4	38.9	29.2	61.4

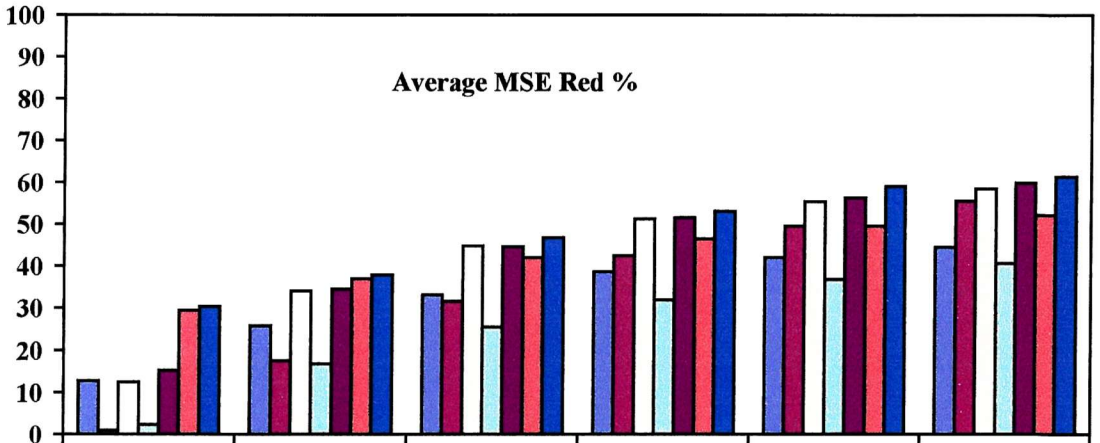
Table 7.10-a. MSE reduction (in percentage) for 13 images (undecimated decomposition).



	Im 1	Im 2	Im 3	Im 4	Im 5	Im 6	Im 7	Im 8	Im 9	Im 10	Im 11	Im 12	Im 13	Aver.
Visual	26.1	17.9	34.4	25.2	31.7	27.8	36.7	36.8	1.6	40.6	28.1	1.2	43.2	27
SURE	22.4	1.6	34.5	27.5	30.9	31.2	43	41.8	0.6	49.9	19.7	1.5	52	27.4
Hybrid	29.1	18.7	37.6	28.3	35.6	30.8	42.8	41.8	4	49.3	29.7	3.7	53.9	31.2
Minimax	14.3	3.6	30.7	20.7	26.1	22.7	34.5	35.2	1.1	39.1	20.6	0.9	45.7	22.7
MAD	28.9	18	37.7	28.5	35.6	30	43.2	42	1.2	49.8	30	2.6	54.1	30.9
WMSE optim.	28.3	24.5	45.9	38.2	41.7	32.4	48.5	50.6	3.8	53.6	39.5	2.2	62.1	36.3
MSE optim.	34.7	16.6	29.8	24.2	26.7	34.7	35.7	39.6	0.1	52.2	21.9	0.5	55.1	28.6

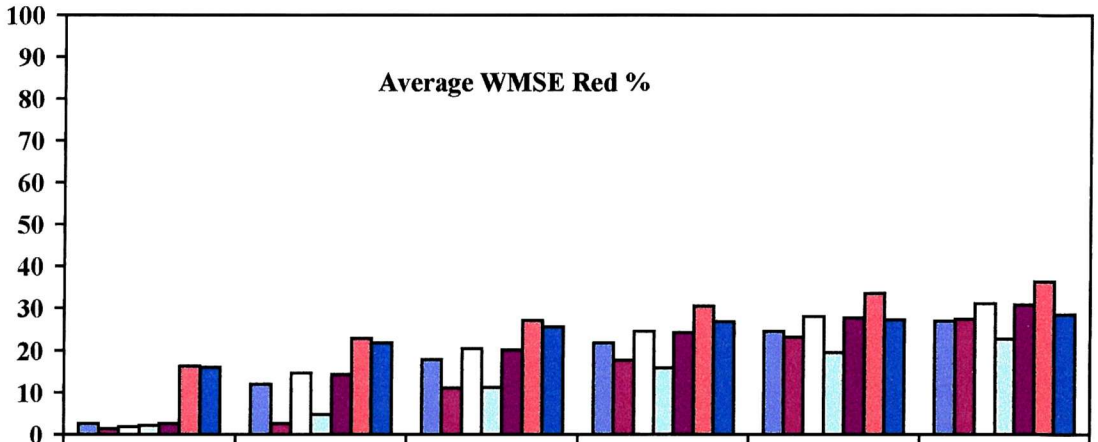
Table 7.10-b. WMSE reduction (in percentage) for 13 images (undecimated decomposition).

Tables 7.10 show that, for noise standard deviation 15, using the undecimated decomposition the MSE and WMSE reductions are much closer for all the thresholding criteria. Nevertheless the thresholding criteria based on the MSE optimisation produces the largest average MSE reduction and is the best performing algorithm for 11 out of the 13 images. As far as the WMSE is concerned, the WMSE optimisation produces the best average performance and is the best algorithm for 9 out of the 13 images. The average results are extended for 6 standard deviation levels of the noise in tables 7.11. The results in table 7.11-a illustrate that the MSE optimisation algorithm produces the best performance at all the noise levels with the MAD or Hybrid algorithms achieving the second best performances. The results in table 7.11-b illustrate that the WMSE optimisation algorithm achieves the best performance at all the noise levels. The performance gain being more significant than in the MSE case.



	Noise level 5	Noise level 7	Noise level 9	Noise level 11	Noise level 13	Noise level 15
Visual	12.7	25.8	33.1	38.6	42.1	44.5
SURE	0.9	17.4	31.6	42.5	49.6	55.5
Hybrid	12.4	34.1	44.8	51.3	55.4	58.5
Minimax	2.3	16.7	25.4	31.9	36.8	40.7
MAD	15.1	34.6	44.7	51.6	56.4	59.8
WMSE optim.	29.5	37	42	46.5	49.6	52.2
MSE optim.	30.4	37.9	46.8	53.1	59.1	61.4

Table 7.11-a. MSE reduction (in percentage) for 6 noise levels (undecimated decomposition).



	Noise level 5	Noise level 7	Noise level 9	Noise level 11	Noise level 13	Noise level 15
Visual	2.6	11.9	17.8	21.8	24.6	27
SURE	1.4	2.6	11	17.7	23.2	27.4
Hybrid	1.8	14.5	20.4	24.6	28.1	31.2
Minimax	2.1	4.7	11.1	15.8	19.5	22.7
MAD	2.6	14.3	20.1	24.3	27.7	30.9
WMSE optim.	16.3	22.9	27.1	30.5	33.6	36.3
MSE optim.	16	21.7	25.6	26.9	27.3	28.6

Table 7.11-b. WMSE reduction (in percentage) for 6 noise levels (undecimated decomposition).

7.5 Visual quality analysis

The evaluation of the visual quality of an image is not a trivial task. This problem is compounded when the results are represented on the printed page via a printer. In this section some example images are presented and in Appendix C there is list of images included in the CD-ROM accompanying the thesis.

The image selected for the presentation is number 3 of the set. Figure 7-5-a shows the image contaminated by AWGN having standard deviation of 15, whilst figure 7-5-b shows the de-noised image using the algorithm based on WMSE optimisation. Figures 7-6, 7-7 and 7-8 show corresponding results for images contaminated by LF, BF and HF noises.

To illustrate the ability of the method to generalise, the method was tested on two other images not included in the training set. Image 'Lenna' is contaminated by AWGN with standard deviation of 15, and image 'vertebra' is contaminated by a combination of AWGN and non-additive noise due to the videofluoroscopic instrumentation employed to acquire the image[4]. Figures 7-9 and 7-10 show these results. In this case the noise standard deviation level is unknown and it is estimated using the combined estimator described in chapter 6.

Two other images are presented to allow one to appreciate the visual differences between results obtained for MSE and WMSE optimisations. The images 'girl' and 'plane' are shown in figures 7-11 and 7-12. The image 'girl' has been artificially contaminated by AWGN while the image 'plane' was already contaminated by typical television noise. A comparison between the two optimisations shows that the visual quality of the images obtained using the WMSE procedure is superior. It is not trivial to discern the differences between the two

de-noised images, nevertheless the images de-noised using the WMSE optimisation can be seen to preserve more detail than the images de-noised using the MSE optimisation. In this case the noise standard deviation level is unknown and it is estimated using the combined estimator described in chapter 6.

A method to visualise ability to preserve detail and the amount of artefacts introduced by the de-noising procedure is to analyse the difference between the noise free image and the de-noised image. Figures 7-13 show the images differences when the test image is contaminated by AWGN with standard deviation level 12 and then de-noised using WMSE and MSE optimisations. The amount of features of the noise free image discernible from the difference images is a measure of the ability of the method to preserve the detail and the presence of 'blocks' is a measure of the distortion introduced. The WMSE optimisation clearly preserves more detail than the MSE optimisation and the 'blocks' are hardly discernible.

Another approach to test the visual performance of a method is to use a static sequence of images contaminated by a different realisation of AWGN. When the sequence is viewed as a video sequence the presence of noise is accentuated by the fact that it changes from frame to frame whereas the image remains stationary. Examples of such sequences are given in the CD-ROM.

In order to illustrate the visual quality of the method with respect to alternative wavelet based approaches figure 7-14 shows an image processed with the WMSE optimisation 5 methods based on decimated decomposition and classical thresholding techniques. The noise free image is contaminated by AWGN with standard deviation 15. The blocking

[4] This signal dependent type of noise has statistical characteristics not completely understood and its study is beyond the scope of this work

artefacts are less evident in the image processed with the WMSE optimisation than in the images processed with the classical wavelet based methods.

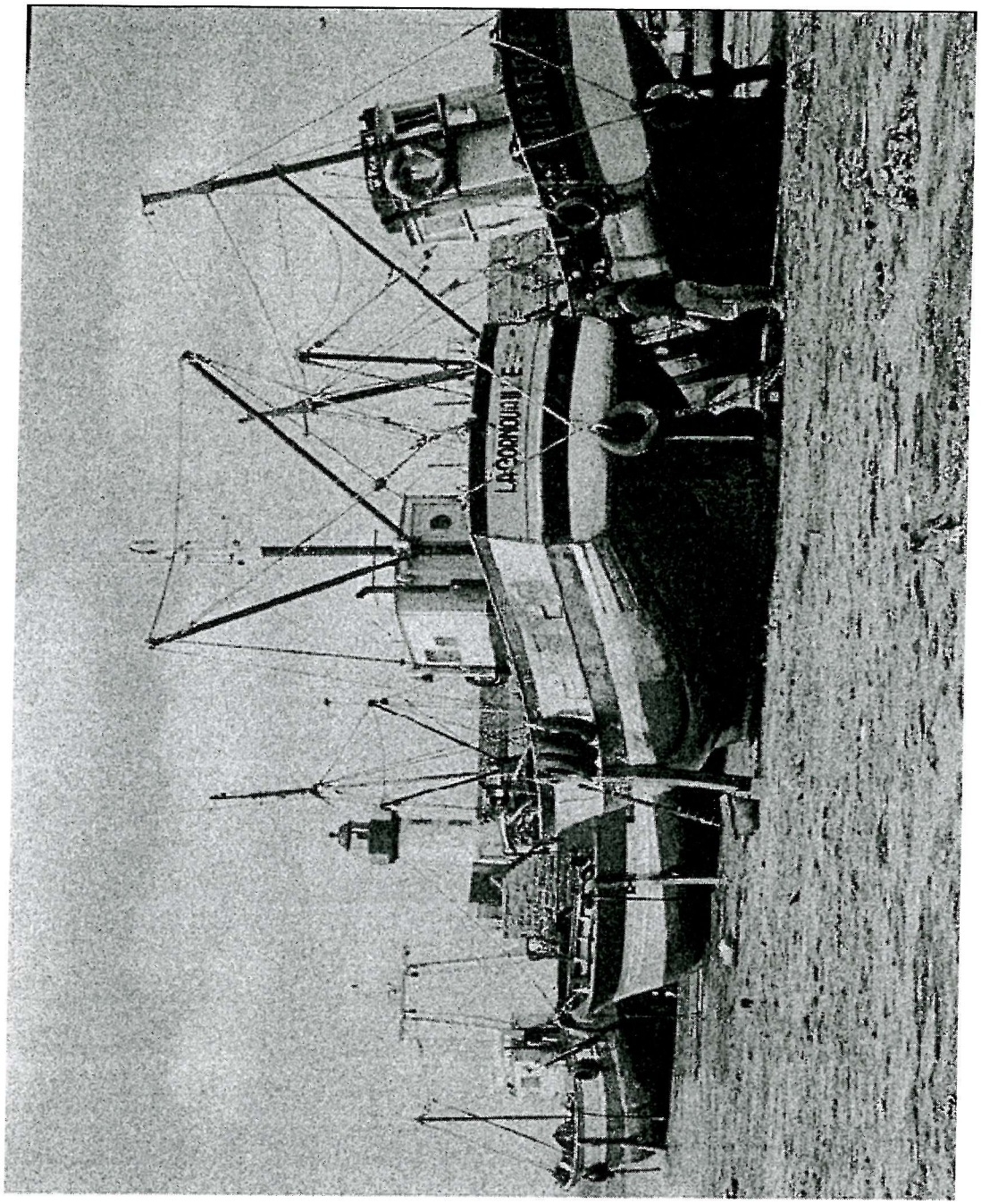


Figure 7-5-a. Image contaminated by white noise.

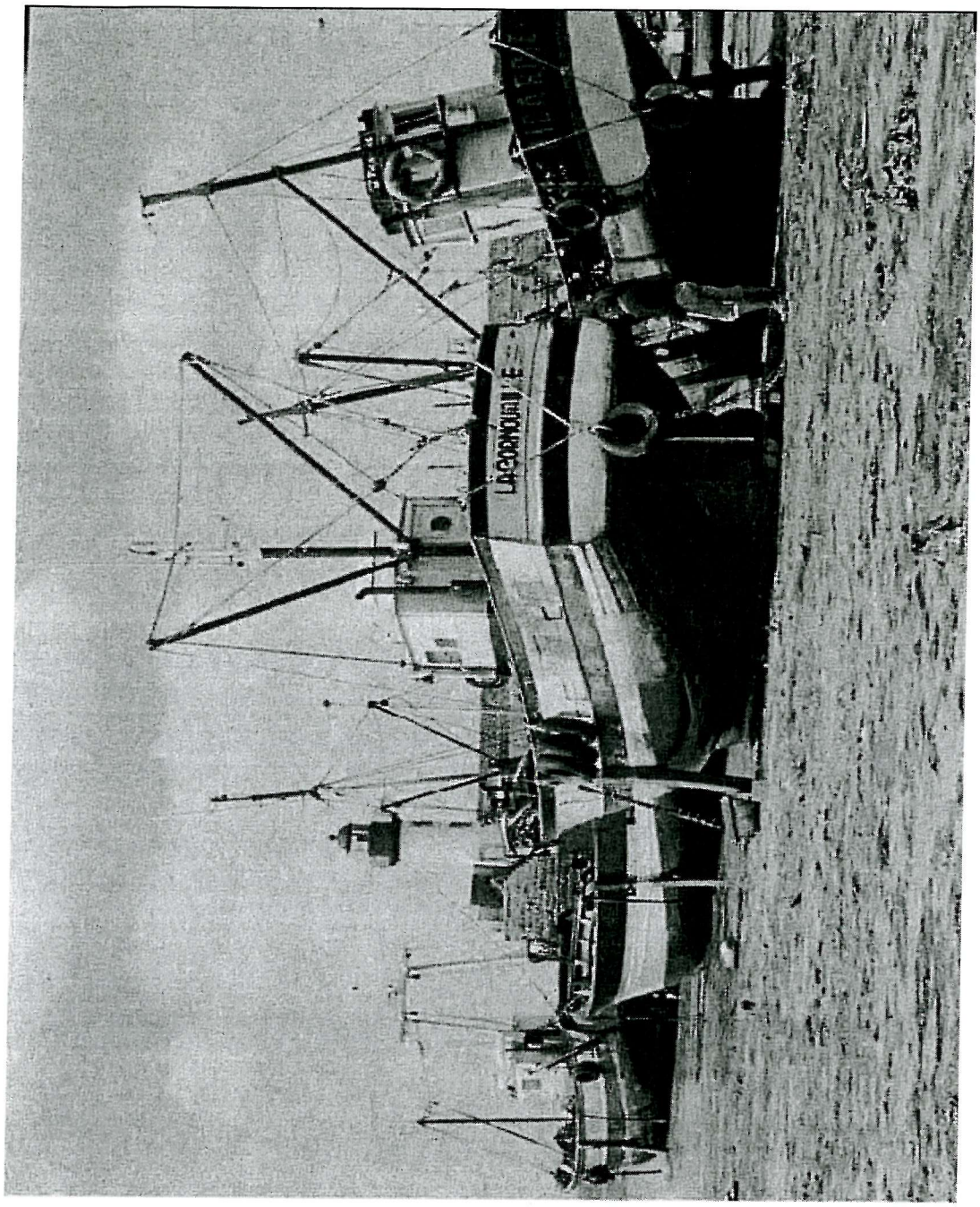


Figure 7-5-b. Image contaminated by white noise after de-noising.

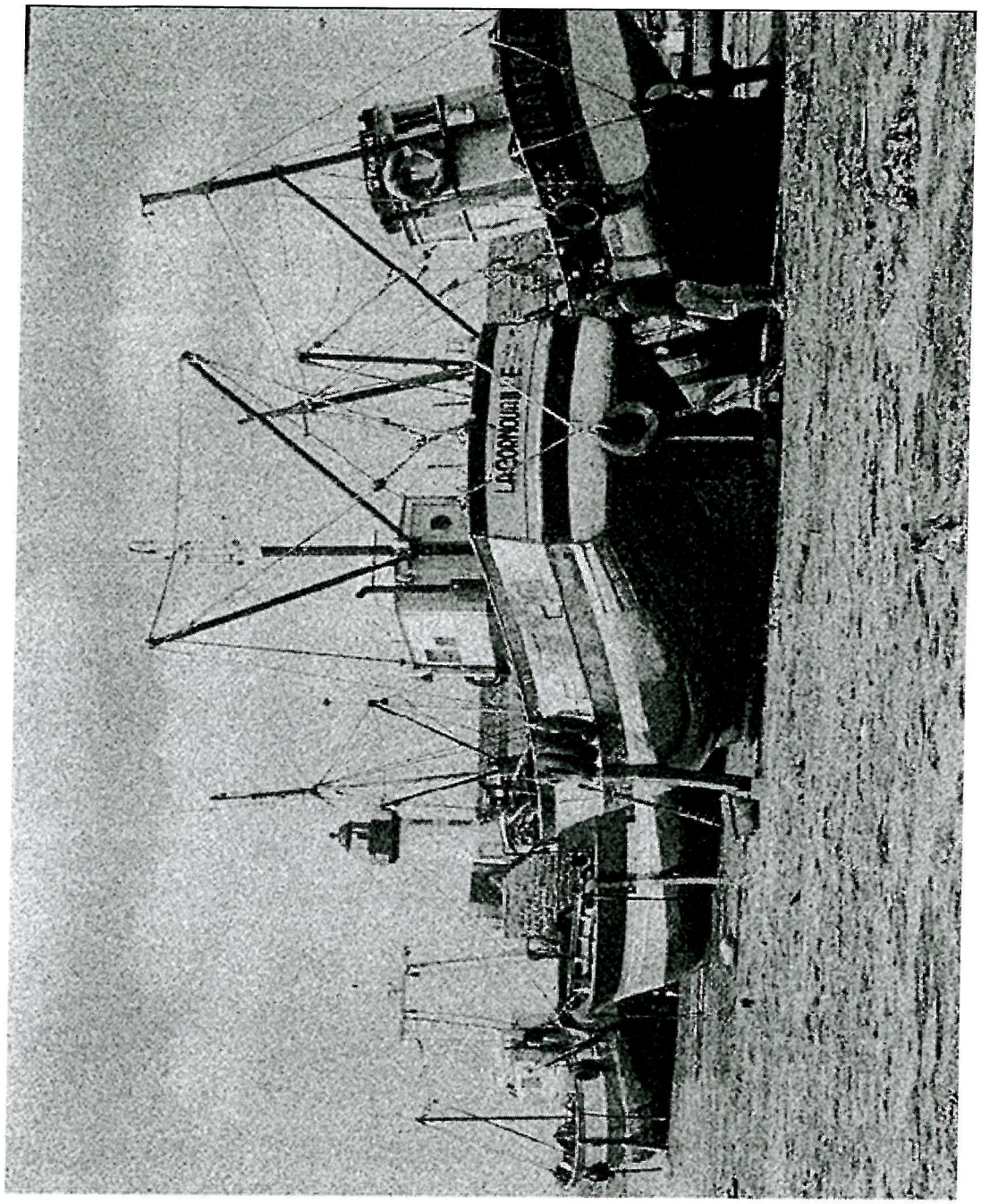


Figure 7-6-a. Image contaminated by LF noise.

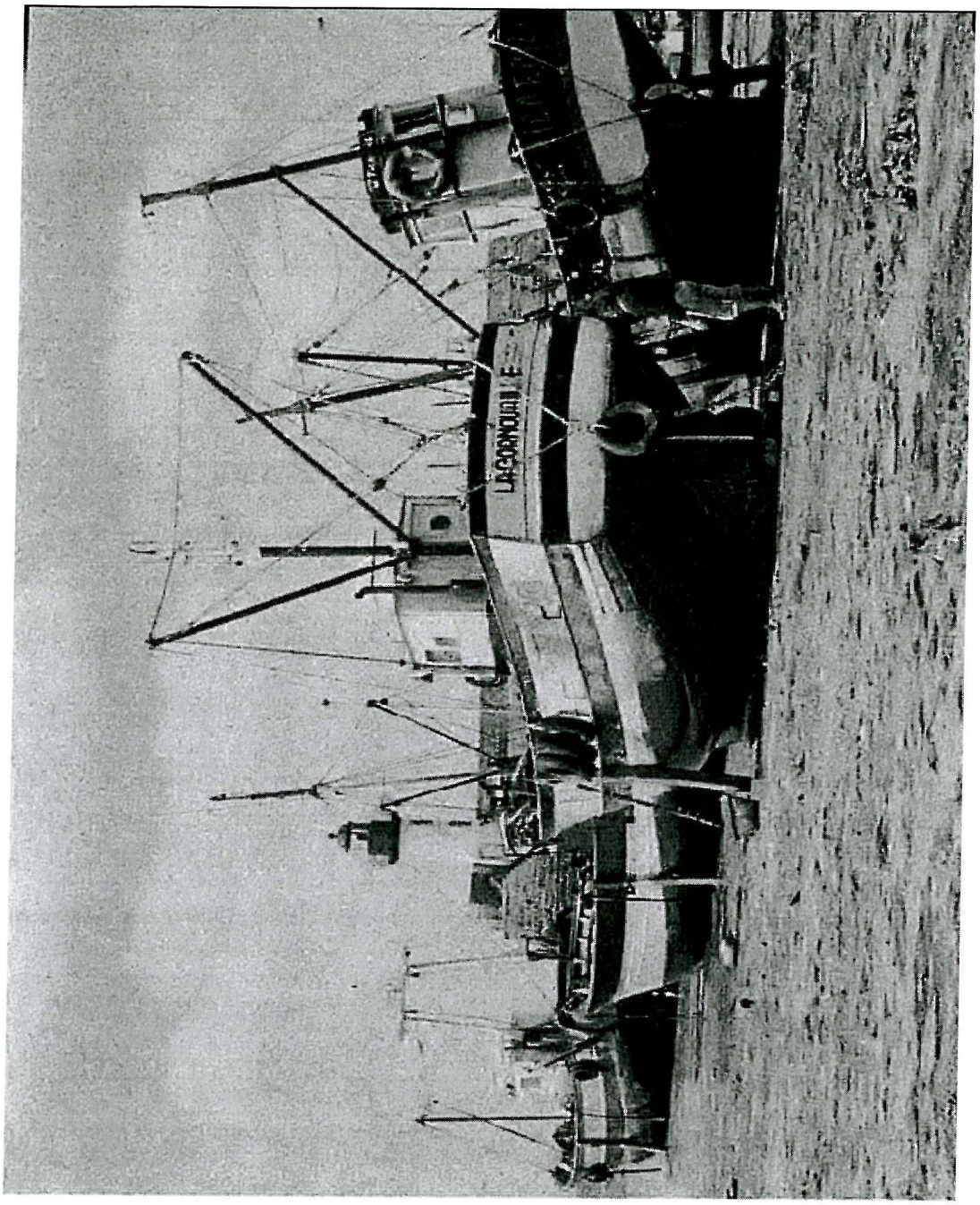


Figure 7-6-b. Image contaminated by LF noise after de-noising.

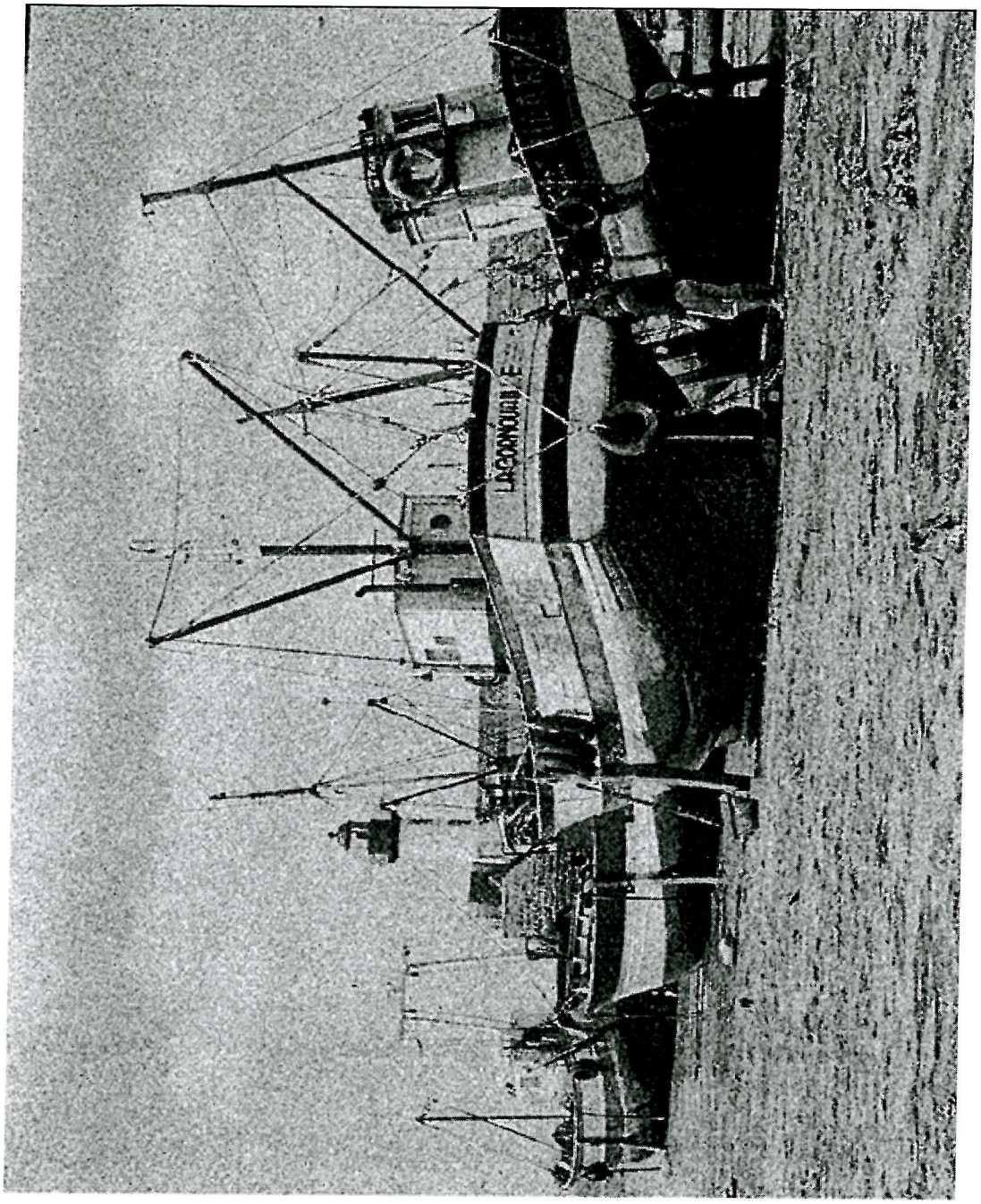


Figure 7-7-a. Image contaminated by BF noise.

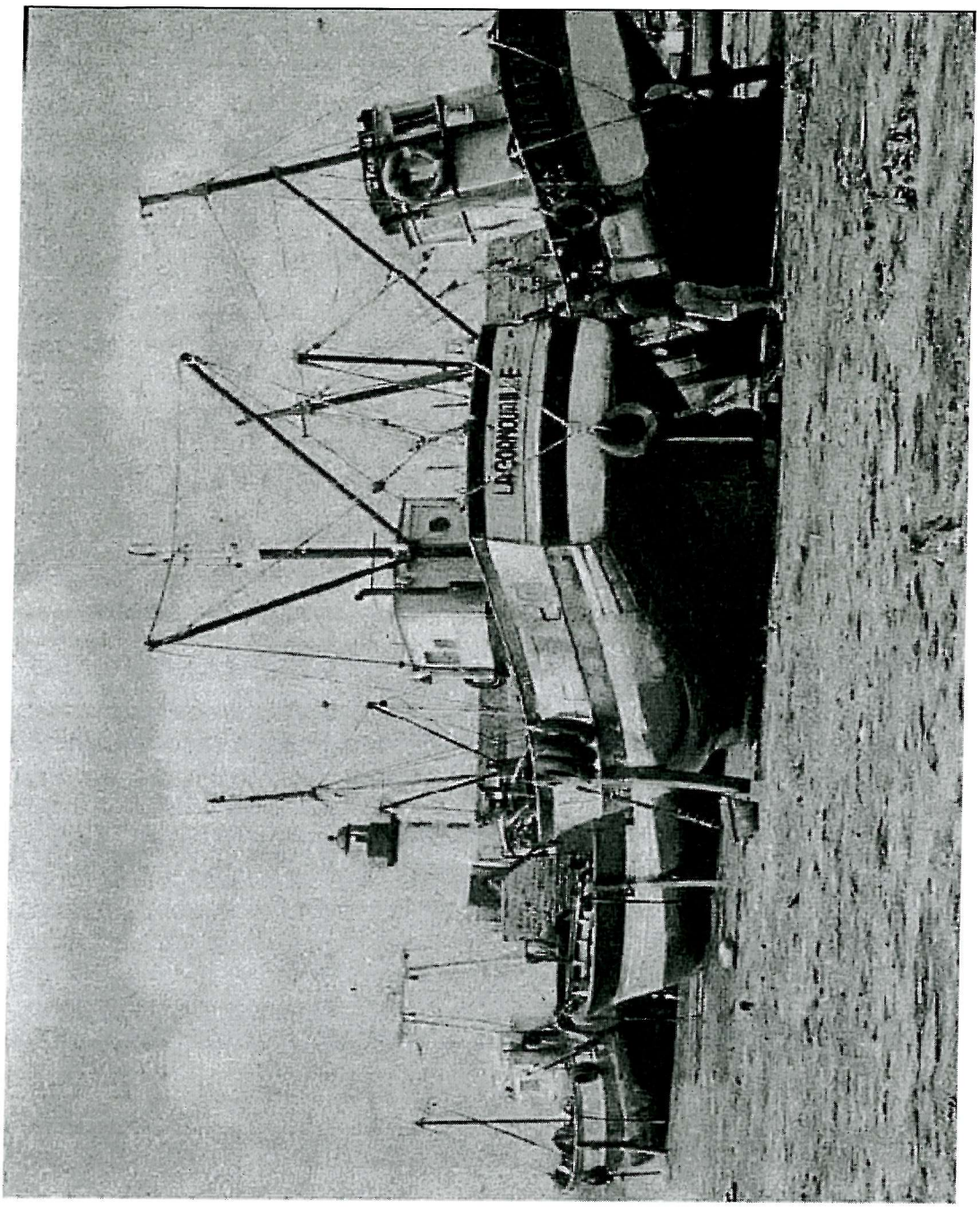


Figure 7-7-b. Image contaminated by BF noise after de-noising.

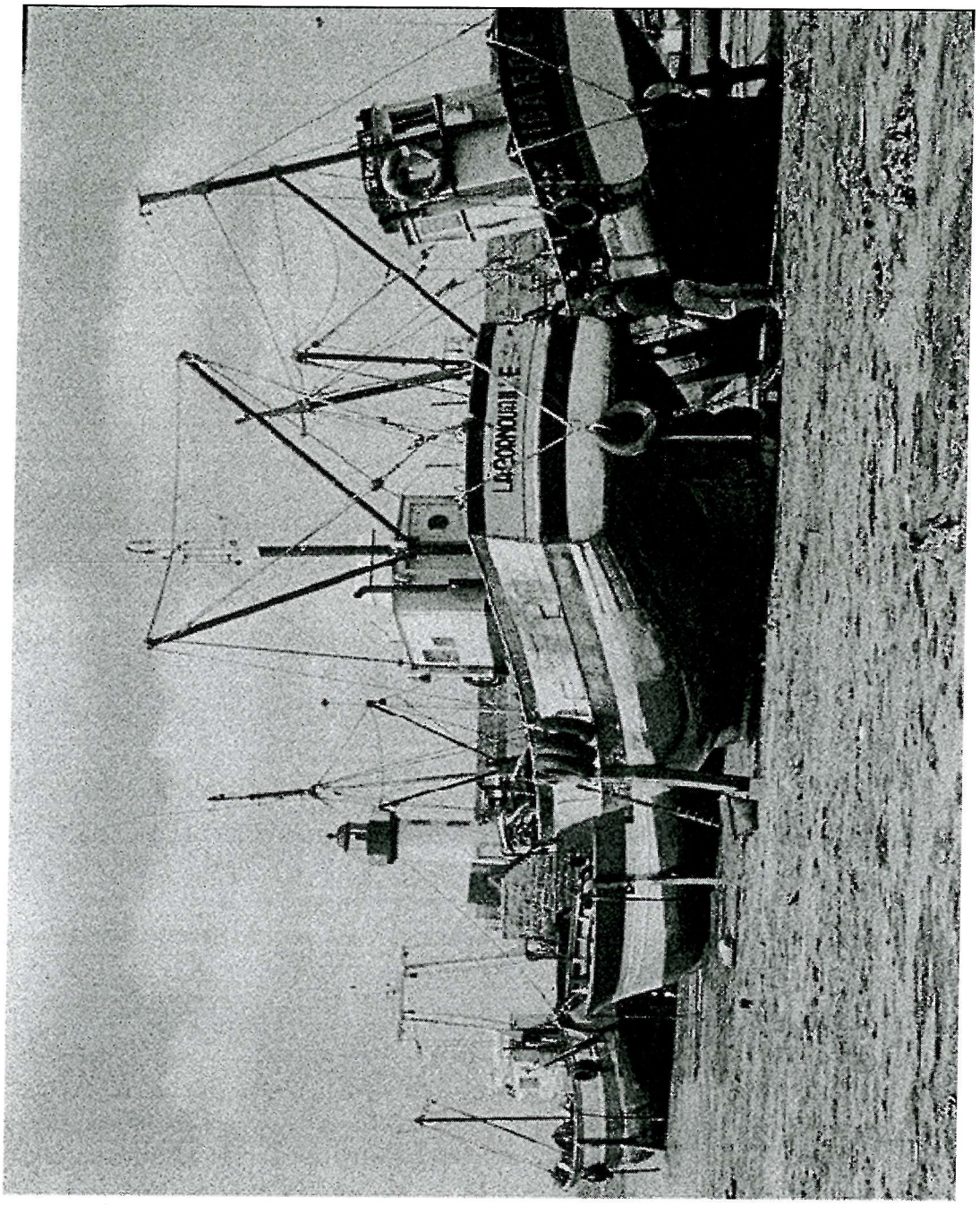


Figure 7-8-a. Image contaminated by HF noise.

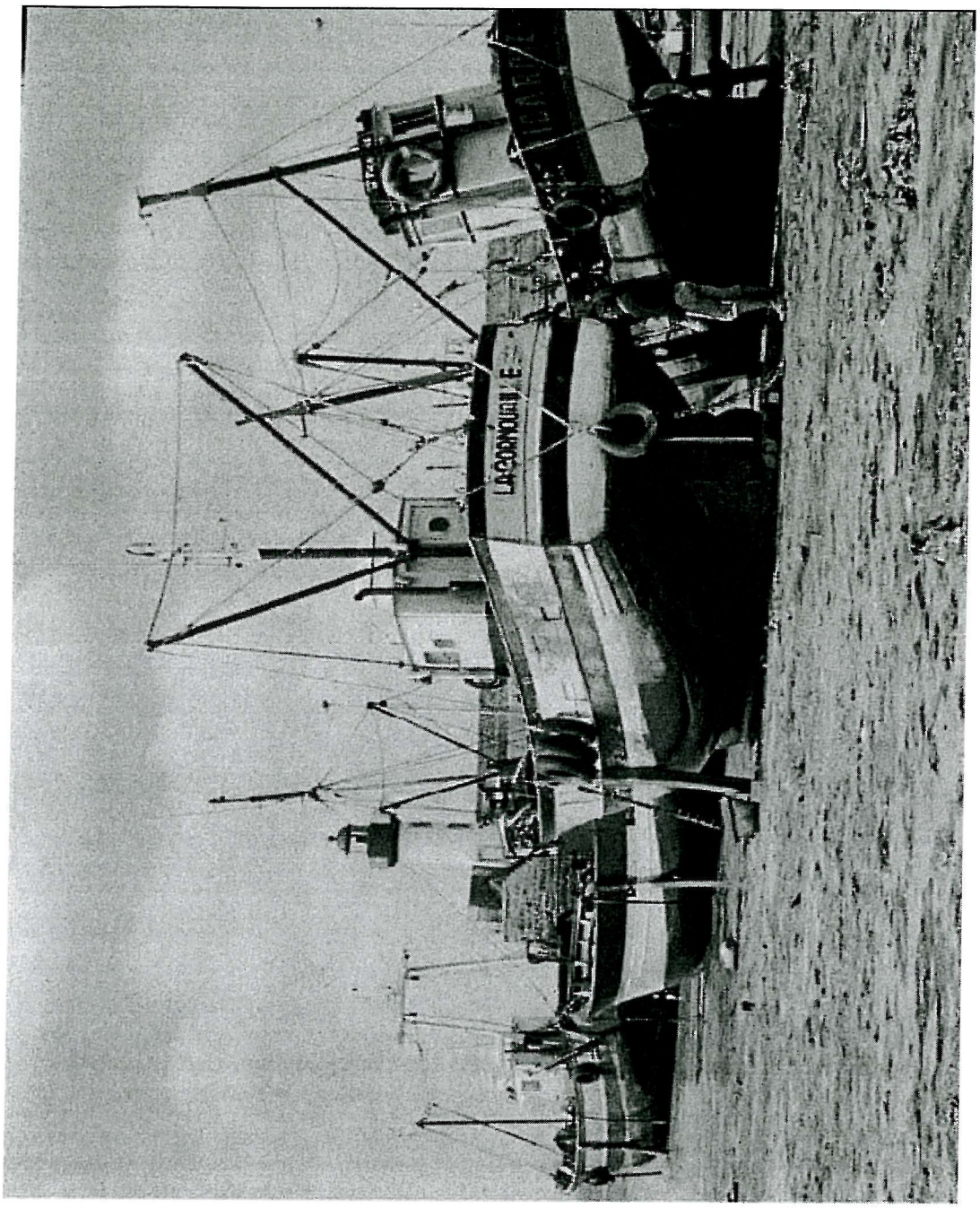


Figure 7-8-b. Image contaminated by HF noise after de-noising.

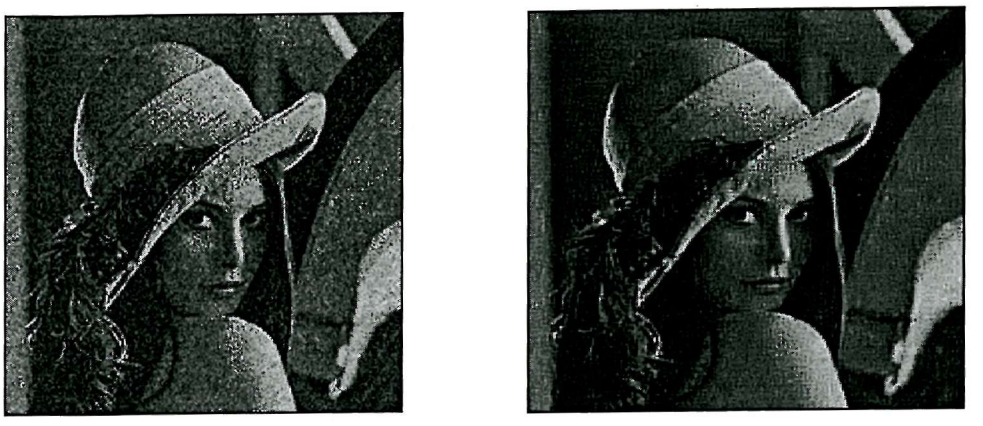


Figure 7-9 Image 'Lenna' contaminated by AWGN and after de-noising.

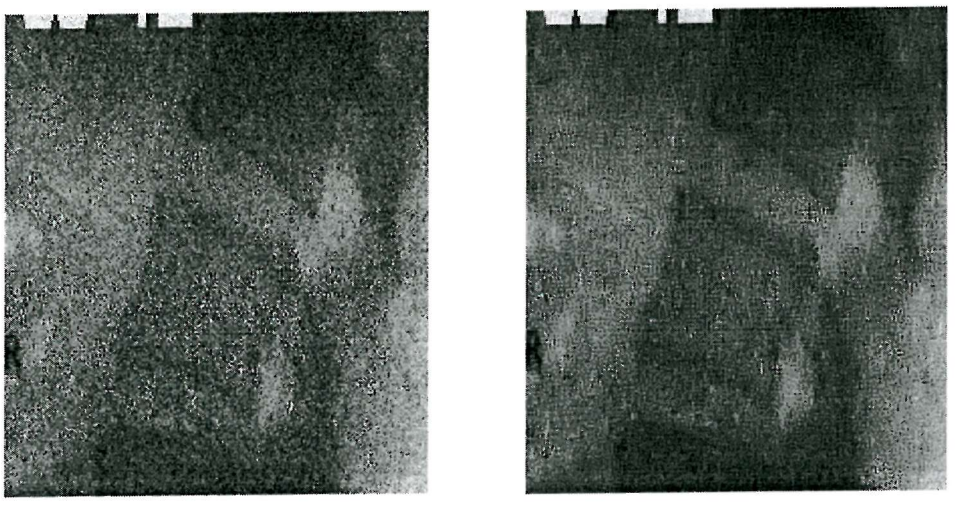


Figure 7-10. Image 'Vertebra' contaminated by AWGN and non-additive noise and after de-noising.

Figure 7-11-a. Image 'girl' original.

Figure 7-11-b. Image 'girl' contaminated by AWGN.

Figure 7-11-c. Image 'girl' after de-noising using MSE optimisation.

Figure 7-11-d. Image 'girl' after de-noising using WMSE optimisation.

Figure 7-12-a. Image 'plane' original (noisy).

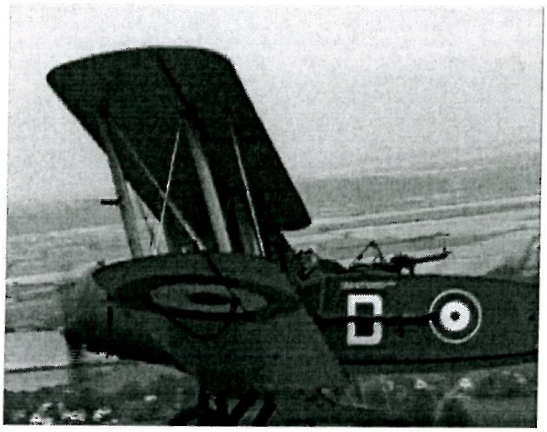


Figure 7-12-b. Image 'plane' after de-noising using MSE optimisation.

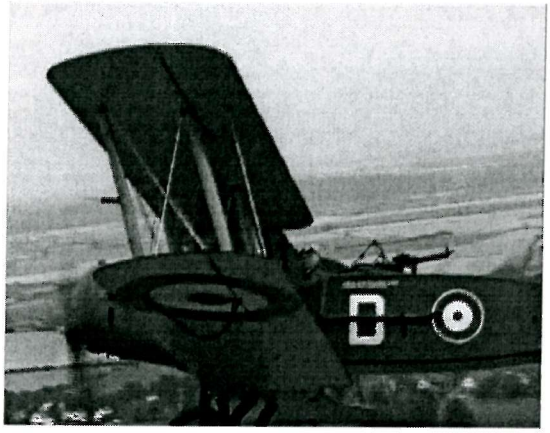


Figure 7-12-c. Image 'plane' after de-noising using WMSE optimisation.

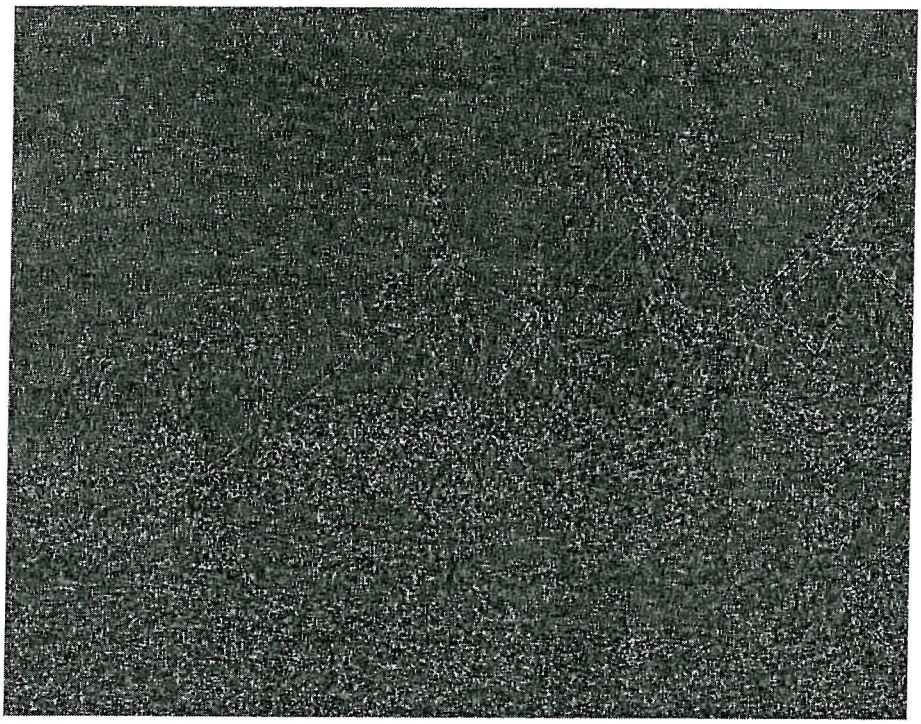


Figure 7-13-a. Difference between noise free image and image contaminated by AWGN with std level 12 and de-noised using WMSE optimisation.

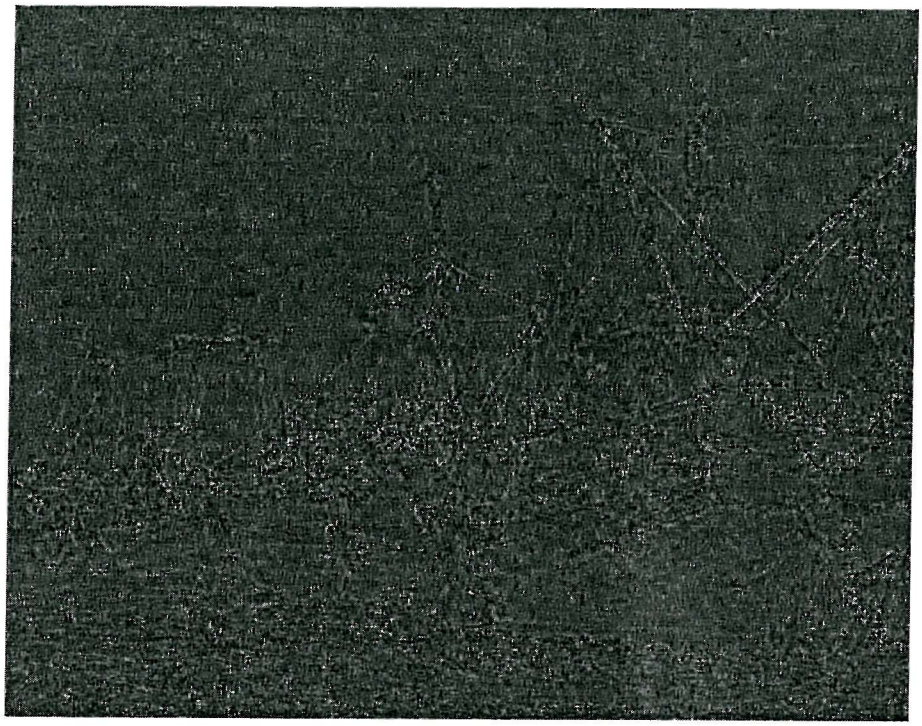


Figure 7-13-b. Difference between noise free image and image contaminated by AWGN with std level 12 and de-noised using MSE optimisation.

Figure 7-14-a. Image de-noised using WMSE optimisation.

Figure 7-14-b. Image de-noised using decimated wavelet decomposition and Visual based thresholding.

Figure 7-14-c. Image de-noised using decimated wavelet decomposition and SURE based thresholding.

Figure 7-14-d. Image de-noised using decimated wavelet decomposition and Hybrid based thresholding.

Figure 7-14-e. Image de-noised using decimated wavelet decomposition and Minimax based thresholding.

Figure 7-14-f. Image de-noised using decimated wavelet decomposition and MAD based thresholding.

Chapter 8

Conclusions and Further work

8.1 Introduction

The subject examined in this thesis is the reduction of spatial noise on video images. The approach used is based on the wavelet decomposition. The basic elements of a wavelet based noise reduction method are the filter bank used to decompose the image and the method used to reduce the noise on the components. The decomposition applied in this work has been described in chapter 3 and the related conclusions are presented in the next section. The method utilised to reduce the noise on the components employs non-linear functions to threshold the pixel values. These functions are characterised by a shape common for all the components and by a number of parameters characterising the function for each component. The new shape utilised in this work has been detailed in chapter 4 and the relevant conclusions are presented in section 8.3. The parameters have been selected using a training procedure that is presented in chapter 5 and the conclusions deduced from this procedure are presented in section 8.4. The parameters of the thresholding functions are dependent on the standard deviation of the noise contaminating the image; consequently a procedure is required to estimate this quantity. The estimator utilised in this work is illustrated in chapter 6 and conclusions about its performance are given in section 8.5. In chapter 7 the performance of the noise reduction approach presented in this work has been evaluated on a set of images and for white and coloured additive Gaussian noises. In chapter 7 the approach has also been compared with other wavelet based noise reduction approaches. The conclusions concerning chapter 7 are presented in section 8.6. Finally, section 8.7 illustrates the potential extensions of the methods and discusses the areas where further investigation is required.

8.2 Wavelet decomposition

In chapter 3, the wavelet analysis theory leading to the design of filter banks for decomposing images was presented and an innovative bank was presented. In this bank the decomposition is performed utilising a tree bank derived from a biorthogonal set of basis functions and produces undecimated components. The absence of decimation simplifies the selection of the filters and allows shift invariance. The price paid is an increase in computation and memory due to components having constant size.

In the absence of the aliasing introduced by the decimation, the filter design procedure to obtain perfect reconstruction is constrained only by a non-distortion condition. This reduces dramatically the complexity of the filters and allows the synthesis step to be a simple summation of components. The absence of decimation also reduces the very unpleasant visual artefacts due to shift-variance. The size of the components to be thresholded is constant and equal to the size of the original image. This computational burden is partly compensated for by the decrease in the number of filtering operations required.

In this work a very simple non-orthogonal set of filters[1] has been utilised and the characteristics of these impose the concurrent training of all the thresholding functions to obtain the noise reduction over the entire image. This choice causes an increase in complexity in the training procedure precluding parallel, independent training of the components.

The number of levels of decomposition used has a strong influence on the complexity of the algorithm. The results of tests performed, and presented in chapter 7, show that the

[1] This set has been derived from a biorthogonal set

number of decomposition levels needed to obtain an efficient noise reduction depends in general on the level of contaminating noise. Moreover it has been demonstrated that there are certain images having spectral distributions that require either a different decomposition scheme or a larger number of decomposition levels in order to be de-noised efficiently.

8.3 Shape of the thresholding functions

Chapter 4 considered the shape of the thresholding functions to apply to the components. Two shapes (soft and hard thresholding) have been widely used in the literature. The chapter analysed the relation between the shape of the thresholding functions and the statistical distribution of the components derived from the wavelet decomposition.

For practical reasons it is convenient to use a common shape for all the components, but the totality of the statistical distributions of the components is difficult to model using a single PDF. The distribution of each component is sensitive to the distributions of both a) the complete image to which it belongs, and b) the band related to that component. Optimal thresholding shapes have been deduced from theoretical statistical distributions using Bayesian estimation. A new shape has been defined using a parameterised family of piecewise linear functions approximating these optimal shapes.

The new shape is optimal with respect to a large number of statistical distributions for the components. This number is larger than in the case of the thresholding schemes used in the literature. Nevertheless the shape requires a superior number of parameters to be characterised compared with soft and hard thresholding schemes. This drawback has been almost completely eliminated using simplifications of the scheme able to reduce the number of parameters required and concurrently maintaining the advantages of the scheme.

8.4 Parameters of the thresholding function

Chapter 5 examines the problem of selecting the parameters characterising the thresholding function for each component. This selection is performed using a training procedure in order to minimise the noise on images contaminated by white noise. The influence of several factors on the training has been examined: the minimisation algorithm used, the number of parameters and the form of the cost function, the level and the realisation of the noise contaminating the image, and the spectral distribution of the training image.

Three minimisation algorithms have been tested and they are shown to produce similar parameters. The results in chapter 7 show that the parameters derived from two different algorithms produced comparable noise reductions. The Nelder (Simplex based) algorithm is preferred because it has been demonstrated to have superior convergence properties compared with the other two algorithms considered.

Simplifications to the shape of the thresholding functions have been presented and they prove to be able to dramatically reduce the number of parameters and consequently to simplify training. The training can also be simplified by reducing the number of parameters through a reduction of the number of levels of decomposition. Nevertheless, as stated in section 8.2, the results in chapter 7 demonstrate that the number of decomposition levels needed to obtain satisfactory performance depends on noise level and spectral distribution of the image. Consequently a general optimal number of levels cannot be selected in general.

Two cost functions have been considered that aim to reflect the visual quality of an image. The mean square error (MSE) is the simplest index but fails to reflect the visual noise reduction. The noise has a more unpleasant visual impact over some frequency bands than

over others because the human eye is frequency sensitive. A weighted version of the mean square error (WMSE) has been considered and it has been shown that this index was reflecting theoretical models related to the human visual system. The thresholding parameters obtained by minimising MSE and WMSE are different and in chapter 7 the quantitative performance using the two cost function values are compared. This qualitative analysis of the images shows that the WMSE produces superior visual performance when compared with the MSE.

The effect of noise level and realisation on the results of training has been examined. It has been shown that, using training from a single image, the majority of the parameters have a near linear dependence on the noise level. Employing this approximation produces a dramatic reduction in the complexity of the noise reduction procedure but serves to emphasise the need for a noise level estimation procedure. The training has been tested on one image contaminated by different noise realisations. The parameters derived show that the influence of the noise realisation is very small and mainly only affects the parameters related to the first level of decomposition (high frequencies).

To understand the relation between the image spectral characteristics and the thresholding parameters is not a trivial task. The training has been applied to six images in order to examine the differences in the parameters derived. The parameters from the training were significantly influenced by the image used, nevertheless in chapter 7 it has been shown that the performance of the method in terms of WMSE is only slightly degraded by using an average set of parameters. This demonstrates that this average produces reasonable performances for all the images included in the set.

8.5 Noise estimation

The parameters derived from training depend on the noise level on each component. The problem of estimating the noise level on each component is treated in chapter 6. The noise contaminating the image is assumed to be white or coloured, additive and Gaussian. In the presence of white noise estimating the noise level on one component is sufficient to infer the level of the noise contaminating all the components. Three noise level estimators and a combined version of these have been proposed. Two performance indices are considered to analyse the estimator performance over sets of 13 images and 6 noise levels. When the noise is coloured information regarding the noise spectral distribution is required. A procedure for the estimation of the noise spectral distribution has been presented.

The performance indices demonstrate that the estimator based on the analysis of the cumulative distribution function obtains, in general, the best performance. Nevertheless the two indices demonstrate that the other two estimators (based on moment matching and linear regression) produced performances superior for several images over a range of noise levels. This indicates that the performance of the noise level estimators depends on image and noise level. The three methods considered seem to cover a large class of images and noise levels, and one solution is to average their estimates. Using the average of the three estimates has been shown to produce a performance superior to any individual method over the set of images and at various noise levels. Further, these results have been shown to extend to images not included in the training phase of these methods.

8.6 Noise reduction results

Chapter 7 analysed the performance of the noise reduction method in terms of MSE, WMSE and a subjective judgement of visual quality. The influence of the decomposition and thresholding function characteristics on the results has been evaluated using a single image and a set of 13 images contaminated by white and coloured noises. Finally the performance of the method has been compared with the performance of other wavelet based schemes.

Analysing the dependence of the performance on the decomposition, thresholding function shape and training, the optimal choices resulted in: 3 levels of undecimated decomposition, shape of the thresholding function with the first simplification (11 parameters), training based on the Nelder algorithm to minimise MSE or WMSE. The performance has been tested for on the set of 13 images contaminated by 6 levels of noise.

The amount of MSE and WMSE reduction increased with the level of contaminating white noise. The performance averaged over 13 images in the range of noise levels between 5 and 15, went from 30.4 % to 61.4 % and from 16.6 % to 36.2 % respectively for MSE and WMSE. The performance of the method depends on the image being analysed. In particular two images had a noise reduction significantly lower than that seen for other images. This has been interpreted as being a result of the high spectral content of these images in the high frequency bands and it is a limitation of the method. Fortunately this type of frequency distribution is uncommon in most video images. Omitting these two images from the set, the average reduction percentage rises to 66.3 % and 42.3 % for MSE and WMSE respectively for noise level 15.

The subjective judgement of the quality of the images obtained by optimising the MSE and the WMSE showed a preference for the WMSE method over the MSE. The analysis of the visual quality of images de-noised using WMSE and MSE optimisation is hard, nevertheless using images derived from the differences between the noise free and the de-noised images it was shown that the WMSE optimisation preserves more detail than the MSE and introduces a small amount of artefacts. This confirmed that the MSE fails to incorporate any perceptual model.

The performance has been evaluated when the images are contaminated by three types of coloured, additive and Gaussian noises. The performance is compared with the performance in the presence of white, additive and Gaussian noise. Different levels of noise contamination in the components results in the use of coloured noise and a procedure of spectral noise estimation has been applied. In order to compare the performance the four types of noise (three coloured and one white) have been normalised so as to produce the same total noise power. The results showed that MSE and WMSE behave differently. The MSE reduction in the presence of all the types of noise is roughly comparable with that obtained for white noise, whilst the WMSE reduction in the presence of coloured noise is superior to the reduction obtained in presence of white noise. The justification of this characteristic of the WMSE case is as follows. The coloured noises considered here (see figure 6-4) have more power in the frequency band that the WMSE concentrates on (see figures 5-5 and 5-6). Consequently a superior level of reduction is achieved. It is noticeable that this effect depends on the filter design.

Wavelet based approaches characterised by decimated or undecimated decomposition and by five thresholding techniques have been tested and compared qualitatively and quantitatively with the weighted and unweighted methods presented in this work. The

weighted and unweighted methods are shown to perform, in general, better than previously proposed methods. To ensure an appropriate comparison, the previously proposed methods were implemented using the undecimated decomposition described in this thesis. A decimated decomposition was also tested using a filter bank based on a 2-D orthogonal basis (see section 7.4.1). Soft and hard thresholding schemes were considered and level dependent or fixed criteria for selecting the thresholding parameters implemented. The performance was analysed in terms of MSE and WMSE reduction using a set of 13 images and 6 levels of noise. As far as the decimated decomposition was concerned the results showed that, on average, weighted and unweighted methods produced the greatest MSE and WMSE reductions. These results demonstrate that the proposed combination of the undecimated decomposition and the thresholding technique contributed to the efficiency of the methods. As far as the undecimated decomposition was concerned the results show that, on average, the methods presented here performed best. However in this case the differences in performance were significantly smaller, indicating that the use of undecimated decomposition is the single most important factor generating the performance enhancement observed.

8.7 Further work

A large number of areas could be investigated in order to develop this research.

Other types of filters could be tested in order to assess the influence of the decomposition on the noise reduction. Using orthogonal filters increases the computational complexity of the filtering and consequently cost of implementation, but facilitates the training permitting the use of parallel optimisation of all the components. This could be very useful for extending the method to other classes of images.

The tests showed also that for two of the images considered the decomposition is not efficient in separating the noise as was the case for the other 11 images. This could be due to the number of decomposition levels applied or to the geometry of the filter bank used. It could be interesting to evaluate the efficiency of existing criteria based on the analysis of the entropy or the energy of the image [103, 104, 105, 106]. These criteria may be able to define when to increase or to decrease the number of decomposition levels applied and how to modify the geometry of the filter bank utilised in the decomposition.

The optimal shapes for the thresholding function can be deduced by assuming specific statistical distributions for the components of noise and image. In this work a limited number of distributions has been considered. The estimation of other optimal shapes from other statistical distributions could provide precious information. It could be also attractive to deduce an expression for the statistical distribution of the image components from the functional form of the thresholding functions proposed in this thesis.

Three algorithms have been tested for minimising two cost functions. Alternative algorithms and cost functions could be examined. Other algorithms could speed up this procedure. The cost function used in the training aims to represent the visual quality of the

image and other cost functions could be used that better reflect this index. The analysis of the influence on the visual quality of the weighting filter utilised in the applied cost function could also be interesting but unfortunately the judgement of the visual quality is not a trivial task.

The presented noise level estimator when the image is contaminated by white noise analyses the high frequency band in order to evaluate the noise level on all the components. In coloured noise environments the noise level on each band has to be estimated independently. It would be useful to have a noise level estimator able to operate directly on all the components or at least on the majority of them. A solution could be to extend to all the image components the estimators that use training on sets of images and noise levels by introducing parameters able to characterise the spectral distribution of the noiseless image on each component.

Finally, the results described are very promising and demonstrate the efficiency of the procedure in many circumstances. Other images and noise distributions could be considered to understand the limitations of the method. In particular it would be interesting to consider images contaminated by Gaussian multiplicative noise and to see if using the logarithmic transform the method produces performance comparable with the case of additive noise.

Appendix A Training Algorithms

A.1 BFGS Quasi Newton algorithm

The goal of the quasi Newton or variable metric methods is to accumulate information from successive line minimisations so that N of these lead to the minimum of a quadratic expression. There exist many algorithms included in the category; in this section the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method is described.

The information is stored and updated in a matrix of size N x N. The method starts with the hypothesis that the function to be minimised, $f(x)$, can be locally approximated around an initial point **P** using the Taylor series by the quadratic form:

$$
\begin{aligned}
&f(\mathbf{x}) = f(\mathbf{P}) + \sum_i \frac{\partial f}{\partial x_i} x_i + \frac{1}{2}\sum_{i,j} \frac{\partial^2 f}{\partial x_i x_j} x_i x_j + \ldots \approx c - \mathbf{b}\cdot\mathbf{x} + \frac{1}{2}\mathbf{x}\cdot\mathbf{A}\cdot\mathbf{x} \\
&c = f(\mathbf{P}) \\
&\mathbf{b} = -\nabla f\big|_{\mathbf{P}} \\
&[\mathbf{A}]_{ij} = \frac{\partial^2 f}{\partial x_i x_j} x_i x_j \bigg|_{\mathbf{P}}
\end{aligned}
\qquad \text{(A.1)}
$$

P is the point that has been taken as the origin of the system with coordinates x. The parameters **A** (Hessian matrix of the cost function) and **b** (gradient of the cost function) are unknown. The method iteratively builds an approximation of the matrix $\mathbf{A}^{-1}$ by constructing a sequence of matrices $\mathbf{H}_i$ so that

$$
\mathbf{H}_i \xrightarrow{i \to \infty} \mathbf{A}^{-1} \qquad \text{(A.2)}
$$

From (A.1) it results

$$\nabla f = \mathbf{A} \cdot \mathbf{x} - \mathbf{b} \qquad \text{(A.3)}$$

The minimum point $\mathbf{x}_m$ then satisfies (if the gradient vanishes)

$$\mathbf{A} \cdot \mathbf{x}_m = \mathbf{b} \qquad \text{(A.4)}$$

At the same time, for each point $\mathbf{x}_i$ we have

$$\mathbf{A} \cdot \mathbf{x}_i = \nabla f(\mathbf{x}_i) + \mathbf{b} \qquad \text{(A.5)}$$

Subtracting these two equations and multiplying by $\mathbf{A}^{-1}$, one obtains

$$\mathbf{x}_m - \mathbf{x}_i = \mathbf{A}^{-1}\left(-\nabla f(\mathbf{x}_i)\right) \qquad \text{(A.6)}$$

In this expression, the left side is the finite step to be taken to arrive to the minimum, the right side is known once a sequence of $\mathbf{H}$ has been determined. One step of the procedure can be described as follows. Subtracting (A.6) from $\mathbf{x}_i$ gives a new search point $\mathbf{x}_{i+1}$ and leads to

$$\mathbf{x}_{i+1} - \mathbf{x}_i = \mathbf{A}^{-1}\left(\nabla f(\mathbf{x}_{i+1}) - \nabla f(\mathbf{x}_i)\right) \qquad \text{(A.7)}$$

This is the step from $\mathbf{x}_i$ to $\mathbf{x}_{i+1}$. Assuming that the new approximation $\mathbf{H}_{i+1}$ is such that $\mathbf{H}_{i+1} = \mathbf{A}^{-1}$, then one has

$$\mathbf{x}_{i+1} - \mathbf{x}_i = \mathbf{H}_{i+1}\left(\nabla f(\mathbf{x}_{i+1}) - \nabla f(\mathbf{x}_i)\right) \qquad \text{(A.8)}$$

In the BFGS algorithm the updating formula for $\mathbf{H}$ has the form

$$\mathbf{H}_{i+1}= \mathbf{H}_i + \frac{(\mathbf{x}_{i+1} - \mathbf{x}_i) \otimes (\mathbf{x}_{i+1} - \mathbf{x}_i)}{(\mathbf{x}_{i+1} - \mathbf{x}_i)(\nabla f(\mathbf{x}_{i+1}) - \nabla f(\mathbf{x}_i))} -$$

$$- \frac{[\mathbf{H}_i (\nabla f(\mathbf{x}_{i+1}) - \nabla f(\mathbf{x}_i))] \otimes [\mathbf{H}_i (\nabla f(\mathbf{x}_{i+1}) - \nabla f(\mathbf{x}_i))]}{(\nabla f(\mathbf{x}_{i+1}) - \nabla f(\mathbf{x}_i)) \mathbf{H}_i (\nabla f(\mathbf{x}_{i+1}) - \nabla f(\mathbf{x}_i))} + \quad \text{(A.9)}$$

$$+ [(\nabla f(\mathbf{x}_{i+1}) - \nabla f(\mathbf{x}_i)) \mathbf{H}_i (\nabla f(\mathbf{x}_{i+1}) - \nabla f(\mathbf{x}_i))] \mathbf{u} \otimes \mathbf{u}$$

In this expression $\otimes$ indicates the outer product, and

$$\mathbf{u} = \frac{\mathbf{x}_{i+1} - \mathbf{x}_i}{(\mathbf{x}_{i+1} - \mathbf{x}_i)(\nabla f(\mathbf{x}_{i+1}) - \nabla f(\mathbf{x}_i))} - \frac{\mathbf{H}_i (\nabla f(\mathbf{x}_{i+1}) - \nabla f(\mathbf{x}_i))}{(\nabla f(\mathbf{x}_{i+1}) - \nabla f(\mathbf{x}_i)) \mathbf{H}_i (\nabla f(\mathbf{x}_{i+1}) - \nabla f(\mathbf{x}_i))} \quad \text{(A.10)}$$

In the work presented in the thesis, the line minimisation criterion used is a mixed quadratic and cubic procedure. This algorithm requires the computation of the first derivative at each step. A limitation of the algorithm is the heavy computational burden. Another limitation is the assumption of quadratic shapes for the error surfaces related to each parameter. In the image noise reduction application tests demonstrated that, for the error surfaces related to some of the parameters, the quadratic approximation is not suitable. Moreover the convergence speed of this algorithm is in general inferior compared with the other algorithms.

A.2 Downhill simplex (Nelder) algorithm

This method for multidimensional minimisation requires, like the BFGS algorithm, only function evaluations and not derivatives. N+1 points are selected to identify a geometrical figure in N dimensions, called a *simplex*. The procedure tends to the optimum using three operations: reflection, contraction and expansion. The method requires three parameters: an acceleration factor α ($\alpha >$ 1), a contraction factor β ($0 < \beta < 1$), and a reflection factor γ. If $\mathbf{P}_0$ is the starting point, the other N points P_i which define the initial simplex are given by:

$$\mathbf{P}_i = \mathbf{P}_0 + \lambda \mathbf{e}_i \quad i = 1,...N \qquad \text{(A.11)}$$

where $\mathbf{e}_i$ are the N unit vectors and λ (potentially different in every direction) is a constant depending on the scale of the problem. The logic of the procedure is determined by the evaluation of the function in each corner of the simplex. Call $\mathbf{P}_h$ the point where the function has the maximum value y_h, and $\mathbf{P}_l$ the point where the function has the minimum value y_l. At each step the point $\mathbf{P}_h$ is moved through the opposite face of the simplex (reflection). Three circumstances can occur.

1) The function value in the reflected point lies between y_h and y_l, then this point replaces the old to create a new simplex.

2) The function value in the reflected point is smaller than y_l, then a new minimum has been produced and an expansion of this point is performed.

2-a) If the expansion produces a new minimum, the reflected and expanded point replaces the old point.

2-b) If the expansion does not produce a new minimum, the reflected and not the expanded point replaces the old point.

3) The function value in the reflected point is bigger than the function values in the points $\mathbf{P}_k$ for all the $k \neq h$, then the procedure defines a new maximum to be either the old $\mathbf{P}_h$ or the reflected point, whichever has the lower function value, and a contraction is performed. If the reflected and contracted points have a function value bigger than the smaller of the reflected point and the old maximum, all the $\mathbf{P}_k$ are replaced by $(\mathbf{P}_k+\mathbf{P}_l)/2$ and a new simplex is generated. Otherwise the reflected and contracted point replaces the old point.

The termination criteria can be when:

a) the vector distance moved in a step is a quantity smaller than a specified tolerance, or when

b) the decrease in the function value in a step is smaller than a second tolerance, or when

c) the number of steps reaches a prefixed limit.

The definition of the initial simplex has significant influence on the convergence speed. Therefore the criteria used for the selection of the N points around the initial point is fundamental in the implementation here. The value 0.1 (10 %) was used as a percentage increment if the starting point was close to the origin, and the value 0.2 (20 %) was used if the algorithm started away from the origin. The factors of acceleration, contraction, and reflection are set to 1, 0.5, and 2 respectively. Tests demonstrated that the MSE and WMSE error surfaces, after a certain number of steps, are more efficiently explored by the simplex generation than proceeding with further steps of the algorithm. It is then advisable to apply

this algorithm for a limited number of steps, using *c)* as termination criteria, and restart it from that solution point. 500 iterations and 4 restarts of the algorithm were employed.

A.3 Differential evolution algorithm

The previous method is a member of the category of direct search minimisation algorithms. Another member of this category is the differential evolution algorithm. This algorithm has, like the Nelder algorithm, a self-organising scheme and uses the vector population to alter the search space. The difference between two vectors randomly selected in the population is utilised to perturb an existing parameter. The perturbation for every population vector can be performed independently or in parallel.

The methods utilises N vectors as an initial population for each generation G. A preliminary population (and not a single initial point) is required to cover the entire parameter space. The basic strategy of this algorithm can be described by several operations.

- Mutation. For each target vector $\mathbf{x}_{i,G}$ (i=1...N) in a generation G a mutant vector is generated

$$\mathbf{v}_{i,G+1} = \mathbf{x}_{r1,G} + F(\mathbf{x}_{r2,G} - \mathbf{x}_{r3,G}) \qquad \text{(A.12)}$$

The indices *r1*, *r2* and *r3* are random integers $\in \{1...N\}$ mutually independent and different from the running index i. The parameter F is real, constant, $\in$ [0 2], and controls the amplification.

- Crossover. To increase the diversity of the perturbed vector. The trial vector is formed

$$\mathbf{u}_{i,G+1} = (u_{1i,G+1}, u_{2i,G+1}, u_{3i,G+1} \ldots u_{Mi,G+1})$$

with

$$u_{ji,G+1} = \begin{cases} v_{ji,G+1} & \text{if } (\mathbf{randb}(\text{j}) \leq \text{CR}) \text{ or } \text{j} = \mathbf{rnbr}(\text{i}) \\ x_{ji,G} & \text{if } (\mathbf{randb}(j) > CR) \text{ and } j = \mathbf{rnbr}(i) \end{cases} \qquad \text{(A.13)}$$

$$j = 1 \ldots M$$

In this expression **randb**(j) is the j[th] evaluation of a uniform number generator with outcome $\in$ [0 1], CR is a constant $\in$ [0 1], and **rnbr**(i) is a randomly chosen index $\in \{1 \ldots N\}$.

- Selection. The criteria to decide whether or not to include the newly generated vector in the population is based on a comparison between the cost function values in $\mathbf{u}_{i,G+1}$ and $\mathbf{x}_{i,G}$.

This algorithm is an attractive solution when the training procedure can be carried out using parallel processors. This was not the case of the training procedure described in this thesis where this algorithm exhibits slower convergence than the Nelder algorithm.

Appendix B Optimal Thresholding Schemes

B.1 Laplacian image model

In this case

$$p(\hat{u}) = \frac{\sqrt{2}}{\sigma_u} e^{-\frac{\sqrt{2}}{\sigma_u}|\hat{u}|} \qquad \text{(B.1)}$$

and

$$p(v/\hat{u}) = \frac{1}{\sigma_\eta \sqrt{2\pi}} e^{-\frac{(v-\hat{u})^2}{2\sigma_\eta^2}} \qquad \text{(B.2)}$$

The MAP optimal estimator is evaluated by solving

$$\begin{aligned} &\frac{\partial \log\{p(\hat{u}/v)\}}{\partial \hat{u}} = 0 \\ &\frac{\partial}{\partial \hat{u}} \log\left\{\frac{p(v/\hat{u})p(\hat{u})}{p(v)}\right\} = 0 \\ &\frac{\partial}{\partial \hat{u}} \log\{p(v/\hat{u})p(\hat{u})\} = 0 \end{aligned} \qquad \text{(B.3)}$$

Substituting the condition becomes

$$\frac{\partial}{\partial u}\left\{-\frac{(v-\hat{u})^2}{2\sigma_\eta^2} - \frac{\sqrt{2}}{\sigma_u}|\hat{u}|\right\} = 0$$

$$\frac{v-\hat{u}}{\sigma_\eta^2} - \frac{\sqrt{2}}{\sigma_u} sign(\hat{u}) = 0 \qquad \text{(B.4-a)}$$

$$\hat{u} = v - \frac{\sigma_\eta^2 \sqrt{2}}{\sigma_u} sign(\hat{u}) = \begin{cases} v - \dfrac{\sigma_\eta^2 \sqrt{2}}{\sigma_u} & \hat{u} > 0 \\ v & \hat{u} = 0 \\ v + \dfrac{\sigma_\eta^2 \sqrt{2}}{\sigma_u} & \hat{u} < 0 \end{cases}$$

which can also be expressed as:

$$\hat{u}_{\substack{MAP \\ Laplacian}} = \begin{cases} v - \dfrac{\sigma_\eta^2 \sqrt{2}}{\sigma_u} & v > 0 \\ 0 & v = 0 \\ v + \dfrac{\sigma_\eta^2 \sqrt{2}}{\sigma_u} & v < 0 \end{cases} \quad \text{(B.4-b)}$$

The LMSE optimal estimator is evaluated by solving

$$\hat{u}_{LMSE} = \int_{-\infty}^{+\infty} up(u/v)du = \int_{-\infty}^{+\infty} u \frac{p(v/u)p(u)}{p(v)} du = \frac{\int_{-\infty}^{+\infty} up(v/u)p(u)du}{p(v)} \qquad \text{(B.5)}$$

Solve for $p(v)$

$$p(v) = \int_{-\infty}^{+\infty} \frac{1}{\sigma_\eta \sqrt{2\pi}} e^{\frac{-(v-u)^2}{2\sigma_\eta^2}} \frac{\sqrt{2}}{\sigma_u} e^{\frac{-\sqrt{2}|u|}{\sigma_u}} du = \frac{\sqrt{2}}{\sigma_u \sigma_\eta \sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{\frac{-(v-u)^2}{2\sigma_\eta^2}} e^{\frac{-\sqrt{2}|u|}{\sigma_u}} du =$$

$$= \frac{1}{\sigma_u \sigma_\eta \sqrt{\pi}} \left[\int_{-\infty}^{0} e^{\frac{-(v-u)^2}{2\sigma_\eta^2}} e^{\frac{+\sqrt{2}u}{\sigma_u}} du + \int_{0}^{+\infty} e^{\frac{-(v-u)^2}{2\sigma_\eta^2}} e^{\frac{-\sqrt{2}u}{\sigma_u}} du \right] \qquad \text{(B.6)}$$

Using

$$\begin{aligned} a &= \frac{1}{2\sigma_\eta^2} \\ b_1 &= \frac{\sqrt{2}}{\sigma_u} + \frac{v}{\sigma_\eta^2} \\ b_2 &= \frac{\sqrt{2}}{\sigma_u} - \frac{v}{\sigma_\eta^2} \\ c &= \frac{v^2}{2\sigma_\eta^2} \end{aligned} \qquad \text{(B.7)}$$

the result is

$$p(v)=\frac{1}{\sigma_u\sigma_\eta\sqrt{\pi}}\left[\int_0^{+\infty}e^{-\left[au^2+b_1u+c\right]}du+\int_0^{+\infty}e^{-\left[au^2+b_2u+c\right]}du\right] \quad \text{(B.8)}$$

and because

$$\int_0^{+\infty}e^{-\left(au^2+bu+c\right)}du=\frac{\sqrt{\pi}}{2\sqrt{a}}e^{\frac{b^2}{4a}-c}erfc\left(\frac{b}{2\sqrt{a}}\right) \quad \text{(B.9)}$$

one can write

$$p(v)=\frac{1}{\sigma_u\sigma_\eta\sqrt{\pi}}\left[\frac{\sqrt{\pi}}{2\sqrt{a}}e^{\frac{\left[b_1^2\right]}{4a}-c}erfc\left(\frac{b_1}{2\sqrt{a}}\right)+\frac{\sqrt{\pi}}{2\sqrt{a}}e^{\frac{\left[b_2^2\right]}{4a}-c}erfc\left(\frac{b_2}{2\sqrt{a}}\right)\right] \quad \text{(B.10)}$$

But

$$\frac{b_1^2}{4a}-c=\frac{\frac{2}{\sigma_u^2}+\frac{v^2}{\sigma_\eta^4}+\frac{2v\sqrt{2}}{\sigma_u\sigma_\eta^2}}{\frac{2}{\sigma_\eta^2}}-\frac{v^2}{2\sigma_\eta^2}=\frac{\sigma_\eta^2}{\sigma_u^2}+\frac{v^2}{2\sigma_\eta^2}+\frac{v\sqrt{2}}{\sigma_u}-\frac{v^2}{2\sigma_\eta^2}=\frac{\sigma_\eta^2}{\sigma_u^2}+\frac{v\sqrt{2}}{\sigma_u}$$

$$\frac{b_2^2}{4a}-c=\frac{\frac{2}{\sigma_u^2}+\frac{v^2}{\sigma_\eta^4}-\frac{2v\sqrt{2}}{\sigma_u\sigma_\eta^2}}{\frac{2}{\sigma_\eta^2}}-\frac{v^2}{2\sigma_\eta^2}=\frac{\sigma_\eta^2}{\sigma_u^2}+\frac{v^2}{2\sigma_\eta^2}-\frac{v\sqrt{2}}{\sigma_u}-\frac{v^2}{2\sigma_\eta^2}=\frac{\sigma_\eta^2}{\sigma_u^2}-\frac{v\sqrt{2}}{\sigma_u}$$

$$\frac{b_1}{2\sqrt{a}}=\frac{\frac{\sqrt{2}}{\sigma_u}+\frac{v}{\sigma_\eta^2}}{\frac{\sqrt{2}}{\sigma_\eta}}=\frac{\sigma_\eta}{\sigma_u}+\frac{v}{\sigma_\eta\sqrt{2}}$$

$$\frac{b_2}{2\sqrt{a}}=\frac{\frac{\sqrt{2}}{\sigma_u}-\frac{v}{\sigma_\eta^2}}{\frac{\sqrt{2}}{\sigma_\eta}}=\frac{\sigma_\eta}{\sigma_u}-\frac{v}{\sigma_\eta\sqrt{2}} \quad \text{(B.11)}$$

Then finally

$$p(v)=\frac{e^{\frac{\sigma_\eta^2}{\sigma_u^2}}}{\sqrt{2}\sigma_u}\left[e^{\frac{v\sqrt{2}}{\sigma_u}}erfc\left(\frac{\sigma_\eta}{\sigma_u}+\frac{v}{\sigma_\eta\sqrt{2}}\right)+e^{\frac{-v\sqrt{2}}{\sigma_u}}erfc\left(\frac{\sigma_\eta}{\sigma_u}-\frac{v}{\sigma_\eta\sqrt{2}}\right)\right] \quad \text{(B.12)}$$

Solving now for the numerator

$$\int_{-\infty}^{+\infty}up(v/u)p(u)du=\frac{1}{\sigma_u\sigma_\eta\sqrt{\pi}}\left[\int_{-\infty}^{0}ue^{\frac{-(v-u)^2}{2\sigma_\eta^2}}e^{\frac{+\sqrt{2}u}{\sigma_u}}du+\int_{0}^{+\infty}ue^{\frac{-(v-u)^2}{2\sigma_\eta^2}}e^{\frac{-\sqrt{2}u}{\sigma_u}}du\right] \quad \text{(B.13)}$$

Using again (B.7) results in

$$\begin{aligned}&\int_{-\infty}^{+\infty}up(v/u)p(u)du=\frac{1}{\sigma_u\sigma_\eta\sqrt{\pi}}\left[\int_{0}^{+\infty}(-u)e^{-\left[au^2+b_1u+c\right]}du+\int_{0}^{+\infty}ue^{-\left[au^2+b_2u+c\right]}du\right]=\\&=\frac{1}{\sigma_u\sigma_\eta\sqrt{\pi}}\frac{1}{2a}\left[\left[e^{-\left[au^2+b_1u+c\right]}\right]_0^{+\infty}+b_1\int_{0}^{+\infty}e^{-\left[au^2+b_1u+c\right]}du-\left[e^{-\left[au^2+b_2u+c\right]}\right]_0^{+\infty}-b_2\int_{0}^{+\infty}e^{-\left[au^2+b_2u+c\right]}du\right]=\\&=\frac{1}{\sigma_u\sigma_\eta\sqrt{\pi}}\frac{1}{2a}\left[b_1\int_{0}^{+\infty}e^{-\left[au^2+b_1u+c\right]}du-b_2\int_{0}^{+\infty}e^{-\left[au^2+b_2u+c\right]}du\right]\end{aligned} \quad \text{(B.14)}$$

Solving the integral as before

$$\begin{aligned}\int_{-\infty}^{+\infty}up(v/u)p(u)du=&\frac{\sigma_\eta e^{\frac{\sigma_\eta^2}{\sigma_u^2}}}{\sigma_u}\left[\left(\frac{\sqrt{2}}{\sigma_u}+\frac{v}{\sigma_\eta^2}\right)e^{\frac{v\sqrt{2}}{\sigma_u}}erfc\left(\frac{\sigma_\eta}{\sigma_u}+\frac{v}{\sigma_\eta\sqrt{2}}\right)+\right.\\&\left.-\left(\frac{\sqrt{2}}{\sigma_u}-\frac{v}{\sigma_\eta^2}\right)e^{\frac{-v\sqrt{2}}{\sigma_u}}erfc\left(\frac{\sigma_\eta}{\sigma_u}-\frac{v}{\sigma_\eta\sqrt{2}}\right)\right]\end{aligned} \quad \text{(B.15)}$$

The resulting expression is

$$\hat{u}_{\substack{LMSE \\ Laplacian}} = \frac{\sqrt{2}\sigma_\eta \left[\left(\frac{\sqrt{2}}{\sigma_u} + \frac{v}{\sigma_\eta^2} \right) e^{\frac{v\sqrt{2}}{\sigma_u}} erfc\left(\frac{\sigma_\eta}{\sigma_u} + \frac{v}{\sigma_\eta \sqrt{2}} \right) - \left(\frac{\sqrt{2}}{\sigma_u} - \frac{v}{\sigma_\eta^2} \right) e^{\frac{-v\sqrt{2}}{\sigma_u}} erfc\left(\frac{\sigma_\eta}{\sigma_u} - \frac{v}{\sigma_\eta \sqrt{2}} \right) \right]}{\left[e^{\frac{v\sqrt{2}}{\sigma_u}} erfc\left(\frac{\sigma_\eta}{\sigma_u} + \frac{v}{\sigma_\eta \sqrt{2}} \right) + e^{\frac{-v\sqrt{2}}{\sigma_u}} erfc\left(\frac{\sigma_\eta}{\sigma_u} - \frac{v}{\sigma_\eta \sqrt{2}} \right) \right]}$$

(B.16)

B.2 Generalised Gaussian image model

In this case

$$p(\hat{u},\sigma_u,\beta_u)=\frac{\omega(\beta_u)}{\sigma_u}e^{-c(\beta_u)\left|\frac{\hat{u}}{\sigma_u}\right|^{\frac{2}{1+\beta_u}}} \qquad -\infty<\hat{u}<\infty$$

$$\omega(\beta_u)=\frac{\left\{\Gamma\left[\frac{3}{2}(1+\beta_u)\right]\right\}^{\frac{1}{2}}}{(1+\beta_u)\left\{\Gamma\left[\frac{1}{2}(1+\beta_u)\right]\right\}^{\frac{3}{2}}} \qquad c(\beta_u)=\left\{\frac{\Gamma\left[\frac{3}{2}(1+\beta_u)\right]}{\Gamma\left[\frac{1}{2}(1+\beta_u)\right]}\right\}^{\frac{1}{1+\beta_u}} \qquad -1<\beta_u<1$$

(B.17)

and

$$p(v/\hat{u})=\frac{1}{\sigma_\eta\sqrt{2\pi}}e^{-\frac{(v-\hat{u})^2}{2\sigma_\eta^2}} \qquad \text{(B.18)}$$

The MAP optimal estimator is evaluated by solving

$$\begin{aligned}&\frac{\partial\log\{p(\hat{u}/v)\}}{\partial\hat{u}}=0\\&\frac{\partial}{\partial\hat{u}}\log\left\{\frac{p(v/\hat{u})p(\hat{u})}{p(v)}\right\}=0\\&\frac{\partial}{\partial\hat{u}}\log\{p(v/\hat{u})p(\hat{u})\}=0\end{aligned} \qquad \text{(B.19)}$$

Substituting the condition becomes

$$\frac{\partial}{\partial \hat{u}}\left\{-\frac{(v-\hat{u})^2}{2\sigma_\eta^2}-c(\beta_u)\left|\frac{\hat{u}}{\sigma_u}\right|^{\frac{2}{1+\beta_u}}\right\}=0$$
$$\frac{v-\hat{u}}{\sigma_\eta^2}-|\hat{u}|^{\frac{1-\beta_u}{1+\beta_u}}\left(\frac{2c(\beta_u)}{(1+\beta_u)(\sigma_u)^{\frac{2}{1+\beta_u}}}\right)sign(\hat{u})=0 \qquad \text{(B.20)}$$

In the particular case when β_u=0

$$C(0)=\frac{1}{2}$$
$$\frac{v-\hat{u}}{\sigma_\eta^2}-\frac{\hat{u}}{\sigma_u^2}=0 \qquad \text{(B.21)}$$
$$\hat{u}_{\substack{MAP\\Gaussian}}=\frac{\sigma_u^2}{\sigma_u^2+\sigma_\eta^2}$$

The LMSE optimal estimator is evaluated by solving

$$\hat{u}_{LMSE}=\int_{-\infty}^{+\infty}up(u/v)du=\int_{-\infty}^{+\infty}u\frac{p(v/u)p(u)}{p(v)}du=\frac{\int_{-\infty}^{+\infty}up(v/u)p(u)du}{p(v)} \qquad \text{(B.22)}$$

In this case the solution has no closed analytical form and it is

$$\hat{u}(\beta_u)_{\substack{LMSE\\Gener.Gaussian}}=\frac{\int_{-\infty}^{\infty}ue^{\frac{-(v-u)^2}{2\sigma_\eta^2}}e^{-c(\beta_u)\left|\frac{u}{\sigma_u}\right|^{\frac{2}{1+\beta_u}}}du}{\int_{-\infty}^{\infty}e^{\frac{-(v-u)^2}{2\sigma_\eta^2}}e^{-c(\beta_u)\left|\frac{u}{\sigma_u}\right|^{\frac{2}{1+\beta_u}}}du} \qquad \text{(B.23)}$$

In the particular case when β_u=0

$$C(0)=\frac{1}{2}$$

$$\widehat{u}_{\substack{LMSE\\Gaussian}}=\frac{\int_{-\infty}^{\infty}ue^{\frac{-(v-u)^2}{2\sigma_\eta^2}}e^{-\frac{u^2}{2\sigma_u^2}}du}{\int_{-\infty}^{\infty}e^{\frac{-(v-u)^2}{2\sigma_\eta^2}}e^{-\frac{u^2}{2\sigma_u^2}}du} \qquad \text{(B.24)}$$

If

$$a=\frac{1}{2\sigma_\eta^2}+\frac{1}{2\sigma_u^2}$$
$$b=-\frac{1}{\sigma_\eta^2} \qquad \text{(B.25)}$$
$$c=\frac{v^2}{2\sigma_\eta^2}$$

It results in

$$\widehat{u}_{\substack{LMSE\\Gaussian}}=\frac{\int_{-\infty}^{\infty}ue^{-(au^2+bu+c)}du}{\int_{-\infty}^{\infty}e^{-(au^2+bu+c)}du}=\frac{\int_{-\infty}^{0}ue^{-(au^2+bu+c)}du+\int_{0}^{\infty}ue^{-(au^2+bu+c)}du}{\int_{-\infty}^{0}e^{-(au^2+bu+c)}du+\int_{0}^{\infty}e^{-(au^2+bu+c)}du}=$$

$$=\frac{\int_{0}^{+\infty}ue^{-(au^2+(-b)u+c)}du+\int_{0}^{\infty}ue^{-(au^2+bu+c)}du}{\int_{0}^{+\infty}e^{-(au^2+(-b)u+c)}du+\int_{0}^{\infty}e^{-(au^2+bu+c)}du}=$$

$$=\frac{\frac{1}{2a}\left[\left[e^{-\left[au^2+(-b)u+c\right]}\right]_0^{+\infty}+(-b)\int_0^{+\infty}e^{-\left[au^2+(-b)u+c\right]}du-\left[e^{-\left[au^2+b_2u+c\right]}\right]_0^{+\infty}-b\int_0^{+\infty}e^{-\left[au^2+b_2u+c\right]}du\right]}{\int_{0}^{+\infty}e^{-(au^2+(-b)u+c)}du+\int_{0}^{\infty}e^{-(au^2+bu+c)}du}=$$

$$=\frac{-\frac{1}{2a}\left[b\int_0^{+\infty}e^{-\left[au^2+(-b)u+c\right]}du+b\int_0^{+\infty}e^{-\left[au^2+bu+c\right]}du\right]}{\int_{0}^{+\infty}e^{-(au^2+(-b)u+c)}du+\int_{0}^{\infty}e^{-(au^2+bu+c)}du}=-\frac{b}{2a}=\frac{\frac{1}{\sigma_\eta^2}}{\frac{\sigma_\eta^2+\sigma_u^2}{\sigma_\eta^2\sigma_u^2}}=\frac{\sigma_u^2}{\sigma_\eta^2+\sigma_u^2}$$

(B.26)

Then finally

$$\underset{Gaussian}{\hat{u}_{LMSE}} = \frac{\sigma_u^2}{\sigma_\eta^2 + \sigma_u^2} = \underset{Gaussian}{\hat{u}_{MAP}} \qquad \text{(B.27)}$$

Appendix C

List of Image Examples

C.1 Documents

This Appendix presents the documents including the results of noise reduction tests performed on images in order to assess the performance of the method. These documents are:

1. **Images.ppt**. This is a PowerPoint presentation including 108 images. The de-noising is performed on images contaminated by white and coloured noises.
2. **Seq1.avi** and **Seq2.avi**. These are two videos in AVI (video for windows) non-compressed format, describing the static sequence of images contaminated by white noise where the de-noising technique has been applied.

References

[1] T. Chen; The past, the present and the future of image and multidimensional signal processing; IEEE Signal Processing Magazine; Vol. 15, No. 2 pp. 21-58; March 1998.

[2] S. T. Baron and W. R. Wilson; MPEG Overview; SMPTE Journal, pp.391-392; June 1994.

[3] MPEG what it is and what it isn't; IEE Colloquium; 24 January 1995.

[4] J. Sengstack; MPEG sets the standard, CD-Rom World, pp.56-59; June 1994.

[5] F. Russo and G. Ramponi; A fuzzy operator for the enhancement of blurred and Noisy images; IEEE Transactions on Image Processing, Vol. 4, No. 8, pp. 1169-1171; August 1995.

[6] S. Koch, H. Kaufman, and J. Biemond; Restoration of spatially varying blurred images using multiple model-based extended Kalman filters; IEEE Transactions on Image Processing, Vol. 4, No. 4, pp.520-523; April 1995.

[7] H. C. Andrews and B. R. Hunt; Digital image restoration; Prentice-Hall Signal Processing Series; 1995.

[8] J. Skalansky, P. V. Sankar and R. J. Walter JR; Biomedical image analysis; Academic Press; 1986.

[9] Z. H. Cho, J. P. Jones and M. Singh; Foundation of medical imaging; John Wiley & Sons; 1993.

[10] A. K. Jain; Fundamentals of digital image processing; Prentice-Hall Inc.; 1990.

[11] R. J. Schalkoff; Digital image processing and computer vision; John Wiley & Sons; 1989.

[12] K. R. Castelman; Digital image processing; Prentice Hall International Editions; 1996.

[13] R. C. Gonzalez and R. E. Woods; Digital image processing; Addison-Wesley; 1992.

[14] A. Chambolle, R. A. De Vore, N. Lee and B. J. Lucier; Nonlinear wavelet image processing: variational problems, compression and noise removal through wavelet; IEEE Transactions on Image Processing, Vol. 7, No. 3, pp. 319-335; March 1998.

[15] C. K. Chui; An introduction to wavelets; Boston Academic Press; 1992.

[16] L. L. Schumaker and G. Webb; Recent advances in wavelet analysis; Boston Academic Press; 1994.

[17] H. L. Resnikoff; Wavelet analysis: the scalable structure information; New York Springer; 1998.

[18] F. Murtagh and J. L. Starck; Image processing through multiscale analysis and measurement noise modeling; Statistics and Computing, Vol.10, No.2, pp.95-103; 2000.

[19] Wavelet applications in signal and image processing. Conference; SPIE Proceedings Series, Vol.3169; 1997.

[20] M. Wolkenstein, T. Kolber, S. Nikolov and H. Hutter; Detection of edges in analytical images using wavelet maxima; Journal of Trace and Microprobe Techniques Vol.18, No.1, pp.1-14; 2000.

[21] R. D. Nowak; Wavelet-base rician removal for magnetic resonance imaging; IEEE Transactions on Image Processing, Vol. 8, No. 10, pp. 1408-1419; October 1999.

[22] R. Oktem, L. Yarovslavsky and K. Egiazarian; Signal and image denoising in transform domain and wavelet shrinkage: A comparative study; University of Athens, Greece; Computer Technology Institute, Patras, Greece; European Association for Signal Processing, Lausanne, Switzerland; 1998.

[23] M. Florkowski; Wavelet based partial discharge image de-noising; IEE Conference Publication, No.467; 1999.

[24] A. Pizurica, W. Philips, I. Lemahieu and M. Acheroy; Image de-noising in the wavelet domain using prior spatial constraints; IEE Conference Publication, No.465, pp.216-219; 1999.

[25] S. K. Nath, R. M. Vasu and M. Pandit; Wavelet based compression and denoising of optical tomography data; Optics Communications, Vol.167, No.1-6, pp.37-46; 1999.

[26] E. L. Miller; Efficient computational methods for wavelet domain signal restoration problems; IEEE Transactions on Signal Processing, Vol.47, No.4, pp.1184-1188; 1999.

[27] G. Uytterhoeven and D. Roose, A. Bultheel; A wavelet toolbox for large scale image processing; Lecture Notes in Computer Science, Vol.1557, pp.337-346; 1999.

[28] M. Jansen, G. Uytterhoeven and A. Bultheel; Image de-noising by integer wavelet transforms and generalized cross validation; Medical Physics, Vol.26, No.4, pp.622-630; 1999.

[29] N. Kompella, R. Machiraju, A. Choudhary and A. Skjellum; Wavelet based compression and denoising of convolution operators; SPIE Proceedings Series, Vol.3458, pp.160-171; 1998.

[30] S. F. Lin, G. Rohde and A. Aldroubi; Wavelet denoising of epicardial fluorescence imaging; SPIE Proceedings Series, 1998, Vol.3458, pp.267-273; 1998.

[31] B. Savale, O. Delage and H. H. Arsenault; Iterative denoising and simultaneously reconstructing algorithm: a new approach for images degraded by multiplicative signal dependent noise; SPIE Proceedings Series, Vol.3460, pp.811-820; 1998.

[32] B. Vidakovic and C. B. Lozoya; On time-dependent wavelet denoising; IEEE Transactions on Signal Processing, 1998, Vol.46, No.9, pp.2549-2554; 1998.

[33] A. A. Basuhail and S. P. Kozaitis; Wavelet-based noise reduction in multispectral imagery; SPIE Proceedings Series, Vol.3372, pp.234-240; 1998.

[34] M. Jansen and A. Bultheel; Experiments with wavelet based image de-noising using generalized cross validation; SPIE Proceedings Series, Vol.3034, pp.206-214; 1997.

[35] Z. Xiong, M. T. Orchard and K. Ramchandran; Inverse halftoning using wavelet; IEEE Transactions on Image Processing, Vol. 8, No. 10, pp. 1479-1483; October 1999.

[36] M. E. Alexander, R. Baumgartner, A. R. Summers, C. Windischberger, M. Klarhoefer, E. Moser and R. L. Somorjai; A wavelet-based method for improving signal-to-noise ratio and contrast in MR images; Magnetic Resonance Imaging, Vol.18, No.2, pp.169-180; 2000.

[37] G. Jammal and A. Bijaoui; Multiscale image restoration for photon imaging systems; SPIE Proceedings Series, Vol.3661, No.p.1, pp.1180-1189; 1999.

[38] T. Daubos, Z. Geradts, J. L. Starck, J. Campbell and F. Murtagh; Improving video image quality using automated wavelet-based image addition; SPIE Proceedings Series, Vol.3813, pp.795-801; 1999.

[39] V. Sandor and S. K. Park; Wavelet domain model-based restoration; SPIE

Proceedings Series, Vol.3387, pp.87-98; 1998.

[40] A. V. Bronnikov and G. Duifhuis; Wavelet-based image enhancement in x-ray imaging and tomography; Applied Optics, Vol.37, No.20, pp.4437-4448; 1998.

[41] V. Sandor and S. K. Park; Model-based restoration in the wavelet domain; SPIE Proceedings Series, Vol.3169, pp.26-35; 1997.

[42] N. Gelman and M. L. Wood; Wavelet encoding for improved SNR and retrospective slice thickness adjustment; Magnetic Resonance in Medicine, Vol.39, No.3, pp.383-391; 1998.

[43] M. R. Banham and A. K. Katsaggelos; Spatially adaptive Wavelet-based multiscale image restoration; IEEE Transactions on Image Processing, Vol. 5, No. 4, pp. 619-634; April 1996.

[44] M. R. Banham, N. P. Galatsanos, H. L. Gonzales, and A. K. Katsaggelos; Multichannel Restoration of single channel images using a wavelet-based subband decomposition; IEEE Transactions on Image Processing, Vol. 3, No. 6, pp. 821-832; November 1994.

[45] X. H. Hao, S. K. Gao and X. R. Gao; A novel multiscale nonlinear thresholding method for ultrasonic speckle suppressing; IEEE Transactions on Medical Imaging, Vol.18, No.9, pp.787-794; 1999.

[46] M. John, S. N. Sundaresan and P. V. Ramakrishna; Wavelet based image denoising: VQ-Bayesian technique; Electronics Letters, Vol.35, No.19, pp.1625-1626; 1999.

[47] M. D. Harpen; A computer simulation of wavelet noise reduction in computed tomography; Medical Physics, Vol.26, No.8, pp.1600-1606; 1999.

[48] R. W. Komm, Y. Gu, F. Hill, P. B. Stark and I. K. Fodor; Multitaper spectral analysis and wavelet denoising applied to helioseismic data; Astrophysical Journal, Vol.519, No.1 Pt1, pp.407-421; 1999.

[49] A. V. Bronnikov and G. Duifhuis; Wavelet-based image enhancement in x-ray imaging and tomography; Applied Optics, Vol.37, No.20, pp.4437-4448; 1998.

[50] L. L. Chen, C. W. Chen and K. J. Parker; Adaptive feature enhancement for mammographic images with wavelet multiresolution analysis; Journal of Electronic Imaging, Vol.6, No.4, pp.467-478; 1997.

[51] Y. Dong, B. C. Forster, A. K. Milne and G. A. Morgan; Speckle suppression using recursive wavelet transforms; International Journal of Remote Sensing, Vol.19, No.2, p.317-330; 1998.

[52] S. Zaroubi and G. Goelman; Complex denoising of MR data via wavelet analysis: Application for functional MRI; Magnetic Resonance Imaging, Vol.18, No.1, pp.59-68; 2000.

[53] M. G. Strintzis; Optimal pyramidal and subband decomposition for hierarchical coding of noisy and quantized images; IEEE Transactions on Image Processing, Vol. 7, No. 2, pp. 155-166; February 1998.

[54] M. K. Schneider, P. W. Fieguth, W. C. Karl and A. S. Willsky; Multiscale methods for the segmentation of signal and images; IEEE Transactions on Image Processing, Vol. 9, No. 3, pp. 456-467; March 2000.

[55] J. L. Starck and F. Murtagh; Automatic Noise Estimation from the Multiresolution Support; Publications of the Astronomical Society of the Pacific, Vol.110, No.744, pp.193-199; 1998.

[56] J. Zhang, D. Wang and Q. N. Tran; A wavelet based multiresolution statistical model for texture; IEEE Transactions on Image Processing, Vol. 7, No. 11, pp. 1621-1627; November 1998.

[57] K. M. Iftekharuddin; Orthogonal wavelets in nonlinear speckle reduction for

improved target recognition; Optical Engineering, Vol.39, No.5, pp.1211-1217; 2000.

[58] V. Strela and A. T. Walden; Orthogonal and biorthogonal multiwavelets for signal denoising and image compression; SPIE Proceedings Series, Vol.3391, pp.96-107; 1998.

[59] V. Strela, P. N. Heller, G. Strang, T. Topiwala and C. Heil; The application of multiwavelet filterbanks to image processing; IEEE Transactions on Image Processing, Vol. 8, No. 4, pp. 548-563; April 1999.

[60] D. Wei, J. Tian, R. O. Wells JR and C. S. Burrus; A new class of biorthogonal wavelet system for image transform coding; IEEE Transactions on Image Processing, Vol. 7, No. 7, pp. 1000-1013; July 1998.

[61] S. Malasiottis and M. G. Strintzis; Optimal biorthogonal wavelet decomposition of wire-frame meshes using box splines, and its application to the hierarchical coding; IEEE Transactions on Image Processing, Vol. 8, No. 1, pp. 41-57; January 1999.

[62] J. Liang and T. W. Parks; Image coding using translation invariant wavelet transform with symmetric extension; IEEE Transactions on Image Processing, Vol. 7, No. 5, pp. 762-769; May 1998.

[63] G. Strang, and T. Nguyen; Wavelet and filter banks; Wellesley Cambridge Press; 1996.

[64] S. Mallat; A wavelet tour of signal processing; Academic Press; 1998.

[65] R. R. Coifmann, and D. L. Donoho; Translation-invariant de-noising; Wavelets and Statistics, Springer-Verlag notes, Anestis Antoniadis editor; 1995.

[66] D. Wei, and C. S. Burrus; Optimal wavelet thresholding for various coding schemes; IEEE Transactions on Image Processing, Vol. 1, pp. 610-613; 1995.

[67] M. Lang, H. Guo, J. E. Odegard, C. S. Burrus, and R. O. Wells JR.; Nonlinear processing of a shift-invariant DWT for noise reduction; SPIE Proceedings Series, Vol. 2491, pp. 640-651; 1995.

[68] M. Lang, H. Guo, J. E. Odegard, C. S. Burrus, and R. O. Wells JR.; Noise reduction using an undecimated discrete wavelet transform; IEEE Signal Processing Letters Vol. 3, No. 1, pp. 10-12; January 1996.

[69] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger; Shiftable multiscale transform; IEEE Transactions on Information Theory, Vol. 38, No. 2, pp. 587-607; March 1992.

[70] P. M. B. van Roosmalen, S. J. P. Westen, and J. Biemond; Noise Reduction for image sequences using oriented Pyramid thresholding Technique; IEEE International Conference on Image Processing, Vol. 1, pp. 375-378; 1996.

[71] N. Kingsbury; Complex wavelet and shift-invariance; IEE Colloquium; Section 5; pp. 1-10; January 2000.

[72] M. Holschneider, R. Kroland-Martinout, J. Morlet and Ph. Tchamitichian; A real time algorithm for signal analysis with the help of the wavelet transform; Wavelet, Time-Frequency Methods and Phase Space; Berlin Springler IPTI, pp. 286-297; 1989.

[73] P. Duttilleux; An implementation of the algorithm 'a trous' to compute the wavelet transform; Wavelet, Time-Frequency Methods and Phase Space; Berlin Springler IPTI, pp. 298-304; 1989.

[74] M. J. Shensa; The discrete wavelet transform: wedding the a trous and Mallat algorithms; IEEE Transactions on Signal Processing, Vol. 40, No. 10, pp. 2464-2482; October 1992.

[75] P. Carre, H. Leman, C. Fernandez and C. Marque; Denoising of the uterine EHG

by undecimated wavelet transform; IEEE Transactions on Biomedical Engineering, Vol. 45, No. 9 pp. 104-1013; September 1998.

[76] Y. Xu, J. B. Weaver, D. M. Healy JR., and J. Lu; Wavelet transform domain filters: a spatially selective noise filtration technique; IEEE Transactions on Image Processing, Vol. 3, No. 6, pp. 747-757; November 1994.

[77] M. Jansen and A. Butheel; Multiple wavelet threshold estimation by generalised cross validation for images with correlated noise; IEEE Transactions on Image Processing, Vol. 8, No. 7, pp. 947-953; July 1999.

[78] S. Mallat, and W. L. Hwang; Singularity detection and processing with wavelets; IEEE Transactions on Information Theory, Vol. 38, No. 2, pp. 617-643; March 1992.

[79] S. Mallat, and S. Zhong; Characterisation of signals from multiscale edges; IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 14, No. 7, pp. 710-732; July 1992.

[80] M. Malfait and D. Roose; Wavelet based denoising using a Markov random field a priori model; IEEE Transactions on Image Processing, Vol. 6, No. 4, pp. 549-565; April 1997.

[81] A. Bruce, D. Donoho, and H. Gao; Wavelet Analysis; IEEE Spectrum, pp. 26-35; October 1996.

[82] D. L. Donoho, and I. M. Johnstone; Wavelet Shrinkage: Asymptopia?; Journal Royal Statistics Society, B, Vol. 57, No. 2, pp. 301-369; February 1995.

[83] D. L. Donoho, and I. M. Johnstone; Ideal spatial adaptation by wavelet shrinkage; Biometrika, Vol. 81, No. 3, pp. 425-455; March 1994.

[84] R. Landry JR., P. Mouyon and D. Lakaim; Interference mitigation in spread spectrum systems by wavelet coefficients thresholding; Wavelet Journal (special Issue), Vol. 9, No. 2, pp. 191-202; March-April 1998.

[85] G. H. Kaufmann, and E. Galizzi; Speckle noise reduction in television holography fringes using wavelet thresholding; Optical Engineering Vol. 35, No. 1, pp. 9-14; January 1996.

[86] F. Abramovich, and Y. Benjamini; Thresholding of wavelet coefficients a multiple hypotheses testing procedure; Wavelets and Statistics, Springer-Verlag notes, Anestis Antoniadis editor; 1995.

[87] M. H. Neumann, and R. von Sachs; Wavelet thresholding: beyond the Gaussian i.i.d. situation; Wavelets and Statistics, Springer-Verlag Notes, Anestis Antoniadis Editor; 1995.

[88] D. L. Donoho; De-noising by soft-thresholding; IEEE Transactions on Information theory, Vol. 41, No. 3, pp. 613-627; May 1995.

[89] R. D. Nowak and R. G. Baraniuk; Wavelet domain filtering for photon imaging systems; IEEE Transactions on Image Processing, Vol. 8, No. 5, pp. 666-678; May 1999.

[90] M. Jansen and A. Bultheel; Geometrical priors in a Bayesian approach to improve wavelet threshold procedures; SPIE Proceedings Series, Vol.3813, pp.580-590; 1999.

[91] J. M. Lina and B. MacGibbon; Non-linear shrinkage estimation with complex Daubechies wavelets; SPIE Proceedings Series, Vol.3169, pp.67-79; 1997.

[92] M. Fligge and S. K. Solanki; Noise reduction in astronomical spectra using wavelet packets; Astronomy & Astrophysics, Suppl. Ser. 124, pp. 579-587; 1997.

[93] J. A. Nelder and R. Mead; A simplex method for function minimisation; Computer Journal, No. 7, pp. 308-313; 1965.

[94] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vettering; Numerical recipes in C; Cambridge University Press; 1988.

[95] R. Storn and K. Price; Differential evolution: a simple and efficient heuristic for global optimisation over continuous spaces; Journal of Global Optimisation, Vol. 11, pp. 341-357; 1997.

[96] T. A. Hentea, and V. R. Algazi; Perceptual models of filtering of high-contrast achromatic images; IEEE Transactions on Systems, Man, and Cybernetics, Vol. SMC-14, No. 2, pp. 230-246; March 1984.

[97] V. R. Algazi, G. E. Ford, and H. Chen; Linear filtering of images based on properties of vision; IEEE Transactions on Image Processing, Vol. 4, No. 10, pp. 1460-1464; October 1995.

[98] V. R. Algazi, and M. Suk; On the frequency weighted least-square design of finite duration filters; IEEE Transactions on Circuit and Systems, Vol. CAS-22, No. 12, pp. 943-953; December 1975.

[99] B. R. Hunt; Digital image processing; Proceedings of the IEEE, Vol. 63, No. 4, pp. 693-708; April 1975.

[100] G. L. Anderson, and A. N. Netravali; Image restoration based on a subjective criterion; IEEE Transactions on Systems, Man, and Cybernetics, Vol. SMC-6, No. 12, pp. 845-853; December 1976.

[101] W. H. Pun and B. D. Jeffs; Adaptive image restoration using a generalised Gaussian model for unknown noise; IEEE Transactions on Image Processing, Vol. 4, No. 10, pp. 1451-1455; October 1995.

[102] R. Bernstein; Adaptive Nonlinear filters for simultaneous removal of different kind of noise in images; IEEE Transactions on Circuits and Systems, Vol. CAS-34, No. 11, pp. 1275-1291; November 1987.

[103] D. Barchiesi and T. Gharbi; Local spectral information in the near field with wavelet analysis and entropy; Applied Optics, Vol.38, No.31, pp.6587-6596; 1999.

[104] S. P. Kozaitis and T. Olmstead; Spectrally sensitive wavelet analysis of multispectral imagery for object detection; SPIE Proceedings Series, Vol.3071, pp.2-11; 1997.

[105] M. Pilard and Y. Epelboin; Multiresolution analysis for the restoration of noisy X-ray topographs; Journal of Applied Crystallography, Vol.31, No.Pt1, pp.36-46; 1998.

[106] J. L. Starck; Image restoration by the multiscale entropy; SPIE Proceedings Series, Vol.3813, pp.97-107; 1999.

[107] A. Papoulis; Probability, random variables, and stochastic processes; McGraw-Hill International Editions; 1991.

[108] C. Polak; Computational methods in optimisation; New York Academy Press; pp. 56; 1990.

[109] W. B. Collis and P.R. White, Volterra Serie, Internal Report ISVR University of Southampton UK, 1997.

[110] A. De Stefano, P. R. White and W. B. Collis; An Innovative approach for spatial video noise reduction using a wavelet based frequency decomposition; IEEE ICIP 2000, Vancouver, Canada; September 2000.

[111] A. De Stefano, P. R. White and W. B. Collis; Selection of thresholding scheme for video noise reduction on wavelet components using Bayesian estimation; IEE Fifth IMA Conference, Warwick UK; December 2000